

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

Strasbourg, 24 November 2021

CAHAI(2021)12
Restricted

AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI)

**Compilation of comments received on
Possible elements of a legal framework on artificial intelligence,
based on the Council of Europe's standards on human rights,
democracy and the rule of law**

Document prepared by the CAHAI Secretariat

www.coe.int/cahai

MEMBER STATES	3
<i>Austria.....</i>	<i>3</i>
<i>Estonia.....</i>	<i>5</i>
<i>France.....</i>	<i>5</i>
<i>Germany.....</i>	<i>6</i>
<i>The Netherlands.....</i>	<i>9</i>
<i>Russian Federation.....</i>	<i>15</i>
<i>Slovenia.....</i>	<i>20</i>
<i>Turkey</i>	<i>21</i>
OBSERVER STATES	23
<i>Japan.....</i>	<i>23</i>
COMMITTEES AND OTHER COUNCIL OF EUROPE BODIES	25
<i>Congress of Local and Regional Authorities</i>	<i>25</i>
<i>European Commission Against Racism and Intolerance (ECRI).....</i>	<i>26</i>
COUNCIL OF EUROPE PARTNER INTERNET COMPANIES	26
<i>EuroISPA.....</i>	<i>26</i>
CIVIL SOCIETY ORGANISATIONS, OTHER PRIVATE SECTOR AND ACADEMIC ACTORS RELEVANT TO THE WORK OF THE CAHAI	27
<i>Access Now.....</i>	<i>27</i>
<i>AlgorithmWatch</i>	<i>31</i>
<i>Center For AI and Digital Policy</i>	<i>38</i>
<i>Council of Bars and Law Societies of Europe.....</i>	<i>41</i>
<i>Global Partners Digital.....</i>	<i>42</i>
<i>International Bar Association</i>	<i>44</i>

<i>CAHAI Secretariat</i>	<i>45</i>

MEMBER STATES

Austria

Below you will find the **Austrian preliminary comments** on the draft document subject to additional remarks:

- AT appreciates the elaboration of new legal instruments in the field of AI – a combination of a transversal legally binding horizontal legal instrument and sectoral binding or non-binding instruments.
- Share concern voiced in para 15 (risk of duplication and fragmentation) and support focus on complementing (i.e. filling gaps) and “translating” existing human rights norms to the field of design, development, use and deployment of AI. Need to provide “**translation**” and guidance on the application of existing human rights norms to the design, development and use of AI.
- In principle, a legal instrument can be supported, but care must be taken to ensure **coherence with legal instruments** – already existing ones and those currently being developed, as outlined in para 8 of the Draft Elements. This applies to the full array of **existing international and regional human rights norms, which are already fully applicable to the whole life-cycle of artificial intelligence**, and in particular to **EU norms currently under negotiation (especially the regulation on AI as well as the EU’s Digital Services Act)**. Since these EU norms have not been finalised, our assessment can only be of a preliminary nature, and we strongly urge for constant monitoring of developments, regular cooperation and exchange. It will be essential that definitions, risk classification/assessment and scope do not contradict each other.
- In order to promote innovation in the AI sector, further support measures for companies (especially for SMEs and start-ups) should be encouraged in addition to the establishment of “regulatory sandboxes”.
- In addition, a specific comment on Section IV “Elements relating to fundamental principles of protection of human dignity”: from a data protection perspective, more information on this section would be of interest, in particular regarding the approach not to duplicate or fragment fundamental rights.
- While the risk of such fragmentation is acknowledged in principle, it must follow that the relevant legal acts in which the essential fundamental rights are enshrined must be cited at least in a preamble or the like. It must be made clear that the new legal act to be created does not interfere with the guaranteed rights.
- In this context, it should also be noted in particular that Convention CETS No. 108 and its additional protocol CETS No. 223 are not only relevant in the context of the establishment of real laboratories for the development of artificial intelligence, but also apply in general to the processing of personal data by artificial intelligence systems. This should be addressed accordingly in the legal act.

- At the EU level, we are expecting the presentation of an inter-institutional declaration soon, which will provide a framework for Digital Rights and Principles. It might be useful to include a reference to this. It should be noted that special regulations on AI must remain within the framework of the GDPR and the DSRL PJ. This applies in particular to the use of AI in the area of decision-making (Art. 22 GDPR) as well as the envisaged provisions on real laboratories.
- We encourage to consistently reference **human rights** throughout the document – this includes e.g. rephrasing/replacing terms like “promote human prosperity” and “social wellbeing” with human rights terminology (e.g. “promote the full realization of human rights for all individuals”).
- We also encourage the coherent use of terminology when talking about design, development and use of AI (or whole life-cycle of AI), in order to introduce some clarity into a very complex thematic area. Currently, the document uses a variety of terms.
- We would like to request clarification on para 11 and 12 (scope of the instrument), stating that the instrument shall be applicable to the development, design and application of AI systems, *“irrespective of whether these activities are undertaken by public or private actors”*. In our view, this formulation might create confusion regarding the direct applicability of the instrument to private actors, and we would therefore suggest rephrasing. We suggest to include a reference on States obligation to protect human rights against abuses by private actors in this regard.
- On para 17 (combination of formulation of individual rights and of state obligations) – even though it seems unclear, how this combination *can “ensure a more uniform application of the binding instrument among Parties”*, the proposal can be supported, and should be mirrored accordingly in the elaboration of an effective monitoring mechanism.
- On para 22 (minimum safeguards applicable to all AI systems), we would suggest to include – next to the provisions regarding transparency of AI systems – references to the element of human control (see para 31) and ensuring accountability. There seems to be overlap with paras 30, 31 – suggest to streamline.
- On para 28: suggest to refer to **persons in vulnerable situations**.
- **On safeguards:** there seems to be an overlap between para 34 and Section IX; suggest to streamline with a view to ensure consistency. Also, we would appreciate clarification on the reasoning to include a specific section on AI in the public sector – it seems like the elements included in this section should also be applied to AI used outside of the public sector. Additionally, we would suggest to rephrase the Section “Safeguards” into “Right to an effective remedy” – all elements mentioned in this section (para 40) are prerequisites for the exercise of the right of an effective remedy, this would be in line with a human rights based approach.

Estonia

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 57

The CAHAI ~~recommends that sets out~~ the following ~~elements be considered for inclusion in an additional instrument as recommendations according guidance relating~~ to the different stages in the ~~process design, development and use of the adoption~~ of an AI system by a public entity:

France

III Éléments concernant l'objet et le but, le champ d'application et les définitions

Paragraphe 13

En ce qui concerne les *définitions*, le CAHAI considère que, au minimum, un instrument juridiquement contraignant transversal devrait définir les expressions et termes suivants : « *système d'intelligence artificielle* », « *cycle de vie* », « *fournisseur d'IA* », « *utilisateur d'IA* », « *sujet d'IA* » et « *préjudice illégal* ». Le CAHAI recommande que toutes les définitions utilisées soient, dans la mesure du possible, compatibles avec les définitions analogues utilisées dans d'autres instruments pertinents portant sur l'IA. De plus, les définitions devraient être rédigées avec soin pour garantir, d'une part, qu'elles respectent les exigences de précision juridique et, d'autre part, qu'elles sont suffisamment abstraites pour rester valides malgré les évolutions technologiques futures des systèmes d'IA.

V Éléments concernant la classification des risques liés aux systèmes d'IA et les applications d'IA interdites

Paragraphe 21

En ce qui concerne les *applications interdites de l'IA* (ce que l'on appelle les « *lignes rouges* » ou « *risque inacceptable* »), le CAHAI considère qu'un instrument juridiquement contraignant transversal devrait prévoir la possibilité d'imposer un moratoire ou une interdiction visant l'application des systèmes d'IA qui, en vertu de la classification des risques susmentionnée, sont considérés comme présentant un risque inacceptable d'entrave à la jouissance des droits de l'homme, au bon fonctionnement de la démocratie et au respect de l'État de droit. Une telle possibilité devrait également être envisagée pour la recherche et le développement de certains systèmes d'IA qui présentent un risque inacceptable. Le CAHAI souhaite notamment attirer l'attention, par exemple, sur certains systèmes d'IA utilisant la biométrie pour identifier, catégoriser ou déduire les caractéristiques ou les émotions des individus pouvant conduire à une surveillance généralisée, et sur les systèmes d'IA utilisés pour la notation sociale afin de déterminer l'accès aux services essentiels, en tant qu'applications pouvant nécessiter une attention particulière. Un moratoire ou une interdiction ne devraient toutefois être envisagés que lorsqu'il apparaît qu'un risque inacceptable pour les droits de l'homme, la démocratie et l'État de droit a été identifié sur une base

Commented [LR1]: The proposed revisions are in order to bring the text in line with the wording of points 4 and 10 (see highlighted text) and also with the intent of Part III, i.e. to recommend to the Council of Ministers elements which could be part of possible additional legal instruments. This wording also mirrors the wording of point 50, which also outlines elements which could be part of possible additional legal instruments. We also note that the term „adoption“ has not been used previously in this document, and is not therefore clearly defined. We recommend using the word „use“ or the trio of „design, development and use“, which has been used throughout CAHAI's work. Note that below you have the additional stages of “procurement” and the differing term “deployment”. Are “deployment” and “use” synonyms?

Alternatively, we suggest using “use” in point 61, and adding “deployment” here in point 57.

Commented [LR2]: Cette notion doit être explicitée sans quoi elle laisserait entendre que des préjudices sont légaux, voire acceptables. Nous serions en faveur d'une reformulation de cette notion.

Commented [LR3]: Cette formulation d'exemple est beaucoup trop générale et absolue pour être incluse dans la catégorie des systèmes d'IA présentant un risque inacceptable. Elle pourrait interdire de développer par exemple des interfaces humains-machines à base d'IA dotés de compétences de communications adaptées à l'interlocuteur, y compris en matière de technologies assistives.

Il existe une part émotionnelle dans la communication et un dialogueur doit pouvoir adapter ses réponses à l'état émotionnel du locuteur.

Il convient donc de proscrire les applications de surveillance généralisée comme la notation sociale pour déterminer l'accès aux services essentiels sans pour autant interdire a priori le développement de technologies comme l'informatique dite “affective” qui peut avoir des applications positives traitement automatique de la parole, en éducation, en santé, en prise en charge des maladies cognitives pour rassurer les personnes, etc.

Commented [LR4]: Autre option: préciser le champ de cet exemple

objective et qu'après un examen minutieux, il n'existe pas d'autres mesures applicables et tout aussi efficaces pour atténuer ce risque. Des procédures de révision devraient être mises en place pour permettre de lever une interdiction ou un moratoire si les risques sont suffisamment réduits ou si des mesures d'atténuation appropriées deviennent disponibles, sur une base objective, pour ne plus présenter de risque inacceptable.

Paragraphe 24

En outre, le CAHAI recommande de prévoir une disposition encourageant les Parties à créer des « *bacs à sable réglementaires* » afin de stimuler l'innovation responsable en matière de systèmes d'IA en permettant l'essai de ces systèmes ~~dans un environnement contrôlé~~, sous la supervision de l'autorité de réglementation nationale compétente, tout en assurant le respect des normes énoncées dans la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (STE n° 108) et son Protocole additionnel, ainsi qu'avec les normes énoncées dans cet instrument juridiquement contraignant transversal contraignant sur la conception, le développement et l'application de l'IA.

Commented [LR5]: L'innovation ne peut pas intervenir uniquement en environnement contrôlé et lorsqu'une technologie est nouvelle et qu'il n'existe pas de normes adaptées, il doit rester loisible aux Etats Parties de créer des bacs à sable réglementaires et pour une durée limitée en situation réelle de marché, avec un assouplissement de règles proportionné au cas d'usage. Cette possibilité doit absolument être préservée et correspond à la réalité des développements numériques (pivots frequent des solutions et cas d'usage, importance du retour de marché itératif, importance de l'effet réseau et de la prime au premier entrant).

IX Éléments concernant les garanties

Paragraphe 40

Ces garanties devraient, au minimum, comprendre : le droit à *un recours effectif devant une autorité nationale* contre de telles décisions, le droit d'*être informé de l'application d'un système d'IA dans le processus décisionnel*, et le droit de *choisir d'interagir avec un humain plutôt qu'avec un système d'IA, dans la mesure du possible/si la situation le permet*. D'autres garanties peuvent être opportunes en fonction des spécificités des systèmes d'IA ~~utilisés~~.

Commented [LR6]: Une nuance nous semble devoir être apportée à ces droits, qui ne pourront pas forcément être offerts de manière inconditionnelle, notamment pour les contraintes liées aux enquêtes/détection d'infractions.

XI Éléments concernant les autorités de contrôle, la conformité et la coopération

Paragraphe 43

Le CAHAI considère qu'un instrument juridiquement contraignant transversal devrait comprendre des dispositions faisant obligation aux Parties de prendre toutes les mesures nécessaires et appropriées pour *garantir la conformité effective* avec l'instrument, en particulier par la *mise en place de mécanismes et de normes de conformité*. En outre, des dispositions relatives à la mise en place ou la désignation d'*autorités de contrôle nationales*, définissant leurs pouvoirs, leurs tâches et leur fonctionnement et garantissant leur indépendance et leur impartialité, ainsi que l'allocation de ressources financières et humaines, devraient être considérées pour inclusion. ~~De plus, l'instrument juridiquement contraignant transversal devrait pourrait~~ comprendre des dispositions régissant la *coopération entre les Parties, afin de faciliter la mise en conformité avec l'instrument juridiquement contraignant transversal* ainsi que *l'entraide judiciaire et autre, notamment l'échange de données et d'autres formes d'informations, en s'assurant de la cohérence avec les autres instruments en vigueur du conseil de l'Europe permettant déjà l'entraide judiciaire.*

Commented [LR7]: exemple : le droit d'être informé de l'application d'un système de reconnaissance des émotions ou d'un système de catégorisation biométrique dans certains cas

Commented [LR8]: Nous ne comprenons pas en quoi des dispositions sur la coopération permettraient la mise en conformité avec l'instrument juridiquement contraignant. La mise en conformité des législations nationales avec l'instrument contraignant ne passe pas vraiment par la coopération entre les Parties. Quant à l'opportunité que cet instrument contienne des dispositions en matière d'entraide et notamment en matière d'échange de données ou d'autres information, il nous semble qu'à minima les dispositions en cette matière devraient s'articuler avec les textes déjà existants du conseil de l'Europe en matière d'entraide.

Germany

II General remarks

Paragraph 6

In accordance with [Chapter I](#) Article 1 d of the Statute of the Council of Europe, matters relating to national defence [fall outside the scope of a legal framework of the Council of Europe and are therefore should](#) not be covered by the scope of a legally binding (or non-legally binding) instrument of the Council of Europe. ~~The CAHAI is of the opinion that the issue of whether that scope should cover “dual use” should be further considered in the context of developing a Council of Europe legal framework on AI.~~

V Elements relating to risk classification of artificial intelligence systems and prohibited applications of artificial intelligence

Paragraph 19

In particular, the CAHAI considers that the risk classification should include a number of categories (e.g., “low risk”, “high risk”, “unacceptable risk”, based on a risk assessment in relation to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law. The risk classification will be based on an initial review to determine if a full HUDERIA ([“Human Rights, Democracy, and Rule of law Impact Assessment”](#)) ~~impact assessment~~ is required (cf. chapter XII) just as the impact assessment itself may have an impact on whether to uphold or change the initial risk classification of the AI system in question. This impact assessment is considered as an element of the overall legal framework on AI systems proposed by the CAHAI. However, the specific HUDERIA model need not necessarily form a constituent part of a possible legally binding instrument.

VI Elements relating to the development, design, and application of artificial intelligence systems in general

Paragraph 27

The CAHAI further proposes to include a provision on respect of *equal treatment and non-discrimination* of individuals in relation to the development, design, and application of AI systems to avoid unjustified bias being built into AI systems [and the use of AI systems leading to discriminatory effects](#).

VII Elements relating to the development, design, and application of artificial intelligence systems in the public sector

Paragraph 34

The CAHAI finds that a legally binding transversal instrument when addressing the development, design, and application of AI systems in the public sector should, as a minimum, include provisions on *access to effective remedy*, a mandatory *right to human review* of decisions taken or informed by an AI system except where competing legitimate overriding grounds exclude this, and an *obligation for public*

authorities to *implement adequate human review for decision processes which are informed or supported by AI systems and to provide relevant individuals or legal persons with meaningful information* concerning the role of AI systems in taking or informing decisions relating to them, except where the law prescribes that competing legitimate overriding grounds exclude *or limit* such disclosure. Furthermore, Parties should be obliged to ensure that *adequate and effective guarantees against arbitrary and abusive practices* due to the application of an AI system in the public sector are afforded by their domestic law.

IX Elements relating to safeguards

Paragraph 40

These safeguards should, at least, include the following: the right to *an effective remedy before a national authority (including judicial authorities)* against such decisions; the right to *be informed about the application of an AI system in the decision-making process*; and the right to *choose interaction with a human in addition to or, if feasible, instead of an AI system*. Other safeguards may be relevant depending on the specificities of the AI systems being used.

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 54

As set out in Chapter VII, the development, design and application of AI systems in the public sector *which can interfere with human rights, democracy or the rule of law* should be addressed in a legally binding transversal instrument, covering the most important transversal rights and obligations that should be respected in this domain. Additionally, the CAHAI is of the opinion that, given the context specificity of the risks posed by AI in the public sector in light of its specific role in society, such transversal framework should be supplemented by additional legally binding or non-legally binding instruments at sectoral level.

Paragraph 61

Risk management and mitigation frameworks set up in previous phases should be evaluated, adapted and maintained during the *deployment* phase. Taking into account the nature of the risk, human involvement may need to be guaranteed in order to ensure appropriate oversight over the system. Where appropriate, the AI system should be initially and regularly audited by an independent actor, and the results rendered publicly available to foster public trust. To this end, member States should establish public registers listing AI systems used in the public sector, as well as essential information about the system such as, its purpose, actors involved in its development and deployment, basic information about the model, and performance metrics, where appropriate, and the result of a HUDERIA. In addition, a feedback mechanism should be put in place in order to collect input on how to improve the system directly from its users and those potentially affected thereby. The AI system should be subjected to regular evaluation and update, including by taking into account the aforementioned feedback. The evaluation process should hence be a periodic one. Transparency and communication towards users and citizens

should always be ensured, and they should have access to accountability and redress mechanisms. Last but not least, the public should always-as a general rule have the right to be informed about the fact that they are interacting with an AI system rather than a human being. In such case, they should also be reserved the right to interact with a human being rather than *only* an AI system, and in particular when their rights and interests can be adversely impacted.

The Netherlands

II General remarks

Paragraph 5

The legally binding transversal instrument should focus on mitigating risks emanating from applications of AI systems with the potential to interfere with the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, all the while promoting socially beneficial AI applications. It should be complementary to (and coherent with) relevant legal frameworks like Convention 108 (+). The usefulness and necessity of new legal norms should be demonstrated. The instrument should be underpinned by a risk-based approach: the legal requirements to the design, development and use of AI systems should be proportionate to the nature-of-the-risk they-these systems pose to human rights, democracy and the rule of law given relevant factors like the opaqueness of the algorithm, the purpose of the system and the context in which it is deployed. Exceptions on rights and obligations maybe necessary in the interest of national security, law enforcement or another legitimate public interest, while being in accordance with the law and necessary in a democratic society. Basic principles that enable the determination of such risk (e.g. transparency requirements) should be applicable to all AI systems.

Paragraph 7

The various legal issues raised by the application of AI systems are not specific to the member States of the Council of Europe, but are, due to the many global actors involved and the global effects they engender, transnational in nature. The CAHAI therefore recommends that a legally binding transversal instrument of the Council of Europe, though obviously based on Council of Europe standards, be drafted in such a way that it facilitates accession by States outside of the region without lowering that share the aforementioned standards. Not only will this significantly increase the impact and efficiency of the proposed instrument, but in addition it will provide a much-needed level playing field for relevant actors, including industry and AI researchers which often operate across national borders and regions of the world. The standards of the Council of Europe on human rights, democracy and the rule of law are sufficiently universal in nature to make this a realistic option. There are several precedents of Council of Europe treaties being applied beyond the European region, cf. notably the Budapest Convention (Cybercrime) and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108), which currently have 66 and 55 Parties respectively, many of which are not member States of the Council of Europe.

Paragraph 8

It is further recommended that, to ensure both global and regional legal consistency, a legally binding transversal instrument of the Council of Europe should take into account existing and upcoming legal and regulatory frameworks of other international and regional fora, in particular [the European Union](#), the United Nations, ~~the European Union~~, and the Organisation for Economic Co-operation and Development – all of which are currently involved in developing various forms of standards related to AI systems.

Paragraph 9

The CAHAI notes that the purpose of an international legal framework should not be to lay down any detailed technical parameters for the design, development and application of AI systems, but to establish certain basic principles and norms governing the development, design and application of AI systems. ~~Such a framework should and~~ regulate, in a consistent and deliberate manner, if and on what conditions AI systems potentially posing risks to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law may be developed, designed and applied by all types of organisations, including public and private actors alike.

Commented [LR9]: Moved forward because the EU is also working on legally binding norms and for us as a EU member the compatibility of legally binding frameworks is very important.

III Elements relating to object and purpose, scope, and definitions

Paragraph 13

In so far as definitions are concerned, the CAHAI considers that, as a minimum, the following definitions should be included in a legally binding transversal instrument: “Artificial intelligence system”; “lifecycle”; “AI provider”; “AI user”; “AI subject”; “unlawful harm”. The CAHAI recommends that all definitions used should, in so far as possible, be compatible with similar definitions used in other relevant instruments on AI. Furthermore, definitions should be carefully drafted to ensure, on the one hand, legal precision, while, on the other hand, being sufficiently abstract to remain valid despite future technological developments concerning AI systems. Whereas a proportional and risk based approach with respect to the design and development of AI-systems could benefit from a more narrow definition of AI (i.e. a definition that addresses the specific risks of opaque data driven models), other proposed elements might benefit from a broader definition of AI (e.g. the right to know one interacts with an AI-system) or no definition at all (for instance one could consider banning ‘social scoring’, regardless of what specific technology is used).

V Elements relating to risk classification of artificial intelligence systems and prohibited applications of artificial intelligence

Paragraph 21

Regarding prohibited applications of AI (the so-called “red lines” or “unacceptable risk”), the CAHAI considers that a legally binding transversal instrument should provide for the possibility of putting a moratorium or a ban on the application of AI systems, which in accordance with the aforesaid risk classification are deemed to present an unacceptable risk of interfering with the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. Such possibility should also

Commented [LR10]: To us it is unclear if a prohibition can follow from the impact assessment and thus pertains to a *specific* system (but who will forbid this system then?), or that a prohibition will be part of the instrument itself and will thus be applicable to all AI-systems that the norm addresses.

We have the same question with respect to high risk AI-systems. Will the instrument regulate what is high risk (like the proposed AI-Act), or will the label high risk follow from the risk assessment?

be considered for the research and development of certain AI systems that present an unacceptable risk. Notably, the CAHAI wishes to draw the attention to, for instance, ~~some AI systems using biometrics to identify, categorise or infer characteristics or emotions of individuals, and~~ AI systems used for social scoring to determine access to essential services, as an applications that may require particular attention. A moratorium or ban should, however, only be considered, where on an objective basis an unacceptable risk to human rights, democracy or the rule of law has been identified and, after careful examination, there are no other feasible and equally efficient measures available for mitigating that risk. Review procedures should be put in place to enable reversal of a ban or moratorium if risks are sufficiently reduced or appropriate mitigation measures become available, on an objective basis, to no longer pose an unacceptable risk. Exceptions on a ban or moratorium may be justified in exceptional cases under very strict conditions and safeguards, for instance when regulating the use of systems for biometric identification for the purpose of law enforcement.

Commented [LR11]: We think exceptions on a ban/moratorium may be justified in exceptional cases under very strict conditions and safeguards, for instance when regulating the use of systems for biometric identification for the purpose of law enforcement.

VI Elements relating to the development, design, and application of artificial intelligence systems in general

Paragraph 24

Furthermore, the CAHAI recommends the inclusion of a provision encouraging Parties to establish “regulatory sandboxes” to stimulate responsible innovation in AI systems by allowing for the testing of AI systems in a controlled environment under the supervision of the competent national regulator, all the while ensuring compliance with the standards set out in the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108) and its amending Protocol (CETS No 223), as well as with the standards set out in this legally binding transversal instrument on the design, development and application of AI and any other relevant standards.

Paragraph 27

The CAHAI further proposes to include a provision on respect of equal treatment and non-discrimination of individuals in relation to the development, design, and application of AI systems to avoid prevent unjustified bias being built into AI systems.

Paragraph 30

Finally, the CAHAI recommends the introduction of provisions on *robustness, safety and cybersecurity, transparency, explainability, auditability and accountability* throughout their lifecycles. It should be noted that these se concepts of “transparency”, “explainability” and “accountability” are considered by the CAHAI to be of paramount importance for the protection of the rights of individuals in the context of AI systems and should thus be clearly defined. In addition, the CAHAI recommends that the issue of sustainability in relation to AI systems throughout their lifecycles be considered in a suitable manner.

VII Elements relating to the development, design, and application of artificial intelligence systems in the public sector

Paragraph 32

The development, design, and application of AI systems in the public sector give rise to some concerns about how to ensure the respect for human rights, democracy, and the rule of law when AI systems are used to (predominantly) take or inform decisions that impact the rights and obligations of individuals and legal persons. That said, the CAHAI underlines that not all public sector AI applications pose risks to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law. It is accordingly important to carefully examine the potential for risk posed by a given AI system on a case-by-case basis. Accordingly, a distinction should be made between, on the one hand, AI systems which can interfere with human rights, democracy or the rule of law, and on the other hand, AI systems which though operated by the same public authorities do not present any such risks.

Paragraph 33

Based on the assumption that a legally binding transversal instrument should be general in nature, the CAHAI recommends that such instrument should focus on addressing the potential risks emanating from the development, design, and application of AI systems for the purposes of law enforcement, the administration of justice, and public administration. Concerning “public administration”, in particular, the CAHAI notes that a legally binding transversal instrument should not address the plethora of specific administrative activities undertaken by public authorities, such as health care, education, social benefits etc, but be limited to general prescriptions about the responsible use of AI systems in public administration. Issues related to the various sectors of public administration may, as necessary, be addressed in appropriate sectoral instruments. Any instrument should, where relevant, take into account the special position and task of law enforcement, and should not unnecessarily hinder or burden the execution of this task.

Paragraph 34

The CAHAI finds that a legally binding transversal instrument when addressing the development, design, and application of AI systems in the public sector should, as a minimum, include provisions on access to effective remedy, a mandatory *right to human review* of decisions that impact the rights and obligations of individuals and legal persons and are (predominantly) taken or informed by an AI system (except where competing legitimate overriding grounds exclude this), and an *obligation for public authorities to provide relevant individuals or legal persons with meaningful information* concerning the role of AI systems in taking or informing decisions relating to them, except where the law prescribes that competing legitimate overriding grounds exclude such disclosure. Furthermore, Parties should be obliged to ensure that *adequate and effective guarantees against arbitrary and abusive practices* due to the application of an AI system in the public sector are afforded by their domestic law.

IX Elements relating to safeguards**Paragraph 39**

The CAHAI recommends that a legally binding transversal instrument should include a series of provisions on legal safeguards to be applied to all applications of decisions (predominantly) taken by AI

Commented [LR12]: “or informed by” already implies human review.

We suggest: “decisions (predominantly) taken by an AI system” as this makes clear that what worries us is not that AI-systems are used *persé* but that there might be a lack of meaningful human intervention with respect to key elements in the decision process.

An alternative could be: “decisions taken by an AI system automatically or without meaningful human intervention”

Commented [LR13]: Para 33 states possible rights and obligations. Are these the elements proposed for the fields of for the law enforcement, the administration of justice, and public administration? Or can we expect more elements, possibly specific for one of the aforementioned fields?

Commented [LR14]: Are we talking about a remedy for non-compliance (a fine for instance) or about a remedy for an incorrect decision (new decision, rectification)?

Commented [LR15]: This repeats what already was stated before in the text to make clear about what kind of decisions we are talking.

Commented [LR16]: See earlier remark at #32.

Commented [LR17]: In our view it is not feasible to make a law for each AI-system regulating its transparency.

Commented [LR18]: See our earlier remarks.

systems ~~used for the purpose of deciding or informing decisions~~ and impacting the legal rights and other significant interests of individuals and legal persons.

Paragraph 40

These safeguards should, at least, include the following: the right to an effective remedy before a national authority against such decisions; the right to be informed about the application of an AI system in the decision-making process; and the right to choose interaction with a human instead of an AI system. Other safeguards may be relevant depending on the specificities of the AI systems being used.

Commented [LR19]: Are we talking about a remedy for non-compliance (a fine for instance) or about a remedy for an incorrect decision (new decision, rectification)?

Commented [LR20]: Public sector: competing legitimate overriding grounds can exclude disclosure.

Commented [LR21]: Needs clarification at some point in time.

XII Human rights, democracy, and rule of law impact assessment

Paragraph 53

Finally, the CAHAI is of the opinion that *stakeholder involvement* in the impact assessment should be assured, where possible. The more severe the impact is deemed to be, or the larger its scale, the more extensive the stakeholder engagement should be. In this regard, particular attention should be paid to involving external stakeholders and members of society (i.e., those who are not covered by the categories of “AI providers” and “AI users”, as listed in Chapter III) who could potentially be adversely affected by the deployment of the AI system.

Commented [LR22]: Exceptions on (full) involvement maybe needed, for instance to protect the interest of law enforcement.

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 56

Additionally, considering that the distinction between public and private sector involvement is often ambiguous, and considering the liability issues relating to the contracting out of public services to private actors any provisions applying to the design, development and application of AI in the public sector should also apply to private actors that act on behalf of the public sector.

Commented [LR23]: Does this include notaries?

Paragraph 58

In the *design phase* of the system, due consideration should be given to the analysis of the problem which the public entity intends to solve, in order to assess whether an AI system is the best fit for the problem and, if so, which characteristics it should have. The data sets to be used for the AI system should be clearly identified, and the protection of such data and their origin respected. The design choices of the system should then be rendered explicit and documented. The intended users of the system, both civil servants and the public, as well as those potentially affected by the system should be involved early on, and their capabilities in using the AI system in question should be considered. An open and transparent design should be favoured. Finally, a human rights, democracy and rule of law impact assessment (in any case: the initial review) should be carried out to anticipate, prevent and mitigate potential risks. This also

requires putting in place risk management and mitigation frameworks, which are relevant throughout all phases.

Paragraph 59

In ~~the possible~~ procurement phase, a thorough review of applicable legislation and policy measures in place should be conducted. Where necessary, public procurement processes should be adapted and public procurement guidelines for AI should be adopted, to ensure that procured AI systems comply with human rights, democracy and rule of law standards. A multidisciplinary and multi-stakeholder approach should be respected in order to involve various perspective and angles, including those of vulnerable groups. Because public entities are responsible for the systems they adopt and apply, careful attention should be paid to the potential impact on public accountability.

Commented [LR24]: Not a necessary step. Development maybe 'in house'.

Commented [LR25]: Why wait until procurement? This policy/legal review is just as important in the design phase.

Paragraph 60

During the particularly sensitive phase of *development* of the system (which may also involve the creation of models by algorithms that learn from data), documentation and logging processes should be meticulously kept to ensure transparency and traceability of the system. Adequate test and validation processes, as well as data governance mechanisms should be put in place. Amongst other risks, the potential risk of unequal access or treatment, various forms of bias and discrimination, as well as the impact on gender equality should be assessed.

Paragraph 61

Risk management and mitigation frameworks set up in previous phases should be evaluated, adapted and maintained during the *deployment* phase. Taking into account the nature of the risk, human involvement may need to be guaranteed in order to ensure appropriate oversight over the system. Where appropriate, the AI system should be initially and regularly audited by an independent actor, and the results (or a summary thereof) rendered publicly available to the extent possible to foster public trust. To this end, member States should establish public registers listing AI systems used in the public sector, as well as essential information about the system such as, its purpose, actors involved in its development and deployment, basic information about the model, and performance metrics, where appropriate, and (a summary of) the result of a HUDERIA. In addition, a feedback mechanism should be put in place in order to collect input on how to improve the system directly from its users and those potentially affected thereby. The AI system should be subjected to regular evaluation and update, including by taking into account the aforementioned feedback. The evaluation process should hence be a periodic one. Transparency and communication towards users and citizens should always be ensured, and they should have access to accountability and redress mechanisms. Last but not least, while exceptions for law enforcement may be needed, the public should always-in principle have the right to be informed about the fact that they are interacting with an AI system rather than a human being. In such case, they should also be reserved the right to interact with a human being rather than *only* an AI system, and in particular when their rights and interests can be adversely impacted.

Commented [LR26]: This should already be considered in the design phase.

Commented [LR27]: If you want to do the HUDERIA correctly, it is very likely that there will be information in the impact assessment that is not fit for publishing (in case of law enforcement, but also intellectual property rights etc).

Russian Federation

II General remarks

Paragraph 6

In accordance with Article 1 d of the Statute of the Council of Europe, matters relating to national defence should not be covered by the scope of a legally binding (or non-legally binding) instrument of the Council of Europe. The CAHAI is of the opinion that the issue of whether that scope ~~sh~~ could cover “dual use” should be further considered in the context of developing a Council of Europe legal framework on AI, taking into account possible difficulties in this respect.

Paragraph 9

The CAHAI notes that the purpose of an international legal framework should not be to lay down any detailed technical parameters for the design, development and application of AI systems, but to establish certain basic principles and norms governing the development, design and application of AI systems and regulate, in a consistent and deliberate manner, if and on what conditions AI systems potentially posing risks to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law may be developed, designed and applied by all types of organisations, including public and private actors alike.

IV Elements relating to fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law

Paragraph 14

The CAHAI considers it necessary that a legally binding transversal instrument contains certain *fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law*, which should apply to all development, design, and application of AI systems, irrespective of whether the actor is public or private, ~~and irrespective of the actual level of risk to human rights, democracy and the rule of law posed by these systems.~~

V Elements relating to risk classification of artificial intelligence systems and prohibited applications of artificial intelligence

Paragraph 18

The CAHAI recommends that a legally binding transversal instrument should provide for the establishment of a methodology for *risk classification* of AI systems with an emphasis on human rights, democracy, and the rule of law. The criteria used for assessing the impact of application of AI systems in this regard should be concrete, clear, and with an objective basis and the assessment itself done in a balanced manner, thus providing for both legal certainty and nuance.

Commented [LR28]: This should be understood to include NGOs.

Commented [LR29]: Does this mean that the development of various automatic and autonomous devices which are broadly used in the industry and never counteract with human beings should also be subject to these fundamental principles? This requirement might be excessive.

Commented [LR30]: AI risks are application-specific, rather than system-specific.

Paragraph 21

Regarding *prohibited applications of AI (the so-called “red lines” or “unacceptable risk”)*, the CAHAI considers that a legally binding transversal instrument ~~sh~~ould provide for the possibility of putting a moratorium or a ban on the application of AI systems, which in accordance with the aforesaid risk classification are deemed to present an unacceptable risk of interfering with the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. Such possibility ~~sh~~ould ~~might~~ also be considered by Parties to the instrument for the research and development of certain AI systems that present an unacceptable risk. ~~Notably, the CAHAI wishes to draw the attention to, for instance, some AI systems using biometrics or other data to identify, categorise or infer characteristics or emotions of individuals, and AI systems used for social scoring to determine access to essential services, as applications that may require particular attention.~~ A moratorium or ban should, however, only be considered, where on an objective basis an unacceptable risk to human rights, democracy or the rule of law has been identified and, after careful examination, there are no other feasible and equally efficient measures available for mitigating that risk and given the specific sphere of application. Review procedures should be put in place to enable reversal of a ban or moratorium if risks are sufficiently reduced or appropriate mitigation measures become available, on an objective basis, to no longer pose an unacceptable risk.

Commented [LR31]: In order not to stifle innovation, it is suggested to leave this option at the discretion of Parties to the future instrument.

Commented [LR32]: There is no sufficient basis for listing these examples as potential « red lines », especially considering the sweeping nature of the statement. Something like this could create legal uncertainty and have a chilling effect on innovation.

Commented [LR33]: Automatic behaviour analysis may be useful for identification of terrorists and other criminals in crowded public spaces. Therefore, specific application of an AI systems should be also considered.

VI Elements relating to the development, design, and application of artificial intelligence systems in general

Paragraph 30

Finally, the CAHAI recommends the introduction of provisions on *robustness, safety and cybersecurity, transparency, explainability, auditability and accountability* throughout their lifecycles. It should be noted that the concepts of “transparency”, “explainability” and “accountability” are considered by the CAHAI to be of paramount importance for the protection of the rights of individuals in the context of AI systems. In addition, the CAHAI recommends that the issue of sustainability in relation to AI systems throughout their lifecycles be considered in a suitable manner.

Commented [LR34]: The Russian Federation reserves its position on this sentence.

XI Elements relating to supervisory authorities, compliance, and cooperation

Paragraph 43

The CAHAI considers that a legally binding transversal instrument should include provisions obliging Parties to take all necessary and appropriate measures to *ensure effective compliance* with the instrument, in particular through the *establishment of compliance mechanisms and standards*. Furthermore, provisions on the establishment or designation of *national supervisory authorities*, defining their powers, tasks and functioning as well as ensuring their functional (at a minimum) independence and impartiality and the allocation of sufficient resources and staff, should be considered for inclusion. In addition, the legally binding transversal instrument should contain provisions regulating the *cooperation between Parties* to facilitate compliance with, and *mutual legal and other assistance, including exchange of data and other forms of information* under, the legally binding transversal instrument.

XII Human rights, democracy, and rule of law impact assessment

Paragraph 46

A well-conducted human rights, democracy, and rule of law impact assessment can advance the assessment of how the deployment of AI systems can affect the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. ~~It should though be noted that this type of impact assessment is not designed to balance negative and positive impacts, something which may depend on the specificities of the legal system in the jurisdiction in which the AI system is intended to be applied.~~ In a subsequent stage, it can then be examined if and how risks identified through the HUDERIA can be mitigated, and if and how a legitimate interest can legitimize the system's use despite interference with human rights, democracy and rule of law standards, when such limitations are prescribed by law, proportionate, and necessary in a democratic society.

Commented [LR35]: Positive impact of AI systems should be taken into account. See Feasibility Study, as well as Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems

Paragraph 47

Indeed, a HUDERIA should *not stand alone*, but be supplemented, at the level of domestic or international law, by other compliance mechanisms, such as certification and quality labelling, audits, regulatory sandboxes and continuous monitoring ~~(where appropriate and feasible) as pointed out in the Feasibility Study.~~ It is important that the impact assessment is aligned with such other compliance mechanisms, as it would be unjustifiably costly and burdensome to require human rights, democracy, and rule of law impact assessments that diverge from public supervisory or regulatory approaches laid down under domestic law. In addition to compliance mechanisms, it must also be ensured that effective remedies remain available for those who may be adversely and unlawfully impacted by the deployment of AI systems.

Paragraph 48

Given the time and resources necessary to undertake such an assessment, and in order to safeguard the proportionality of a risk-based approach, the CAHAI believes that, as a rule, a formalised extensive human rights, democracy, and rule of law impact assessment should only be mandated if there are clear and objective indications of relevant risks emanating from the application of an AI system. This requires that all AI systems undergo an initial review in order to determine whether or not they should be subjected to such a formalised assessment. ~~It is recommended that indications as to the necessity for a more extensive assessment be further developed by the CAHAI.~~ It should also be considered that using an AI system in a new or different context or for a new or different purpose or otherwise relevant changes could require a reassessment.

Paragraph 49

The CAHAI underlines that adopting a risk-based approach entails that any relevant impacts by the application of an AI system on the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law should be duly assessed and reviewed on a systematic and regular basis with a view to identifying mitigating measures tailored to the risks at hand, and taking additional steps if such ~~mitigating~~ measures are not ~~deemed~~ sufficient, ~~applying prohibitive measures, as necessary.~~ Furthermore, given the need for an iterative assessment process, such assessment should in any case be carried out again whenever a given AI system undergoes substantial changes.

Commented [LR36]: This needs to be brought in line with the approach to « red lines ».

Paragraph 50

The CAHAI recommends that, at least, the following *main steps* be included in a human rights, democracy, and rule of law impact assessment, subject to an initial review having been conducted, and including stakeholder involvement, where relevant:

- I *Risk Identification*: Identification of relevant risks for human rights, democracy and the rule of law, **taking into account the likelihood and the severity of the effects**;
- II *Impact Assessment*: Assessment of the impact on those rights and principles, **including possible positive effect** ;
- III *Governance Assessment*: Assessment of the roles and responsibilities of duty-bearers, right holders and stakeholders in implementing and governing the mechanisms to mitigate the impact;
- IV *Mitigation and Evaluation*: Identification of suitable mitigation measures and ensuring a continuous evaluation.

Paragraph 51

As regards the *Impact Assessment* step, the CAHAI further recommends that the assessment of an AI system **application**, ~~at least~~ could include, **when the identified risks so warrant**, the following *elements*: assessment of the *context and purpose* of the AI system, *level of autonomy* of the AI system, *underlying technology* of the AI system, *usage* of the AI system (both intended and potentially unintended use), *complexity* of the AI system (part of multiple deep neural networks/building on other AI systems/**dual use**), *transparency* and *explainability* of the system and the way it is used, *human oversight and control mechanisms* for the AI provider and AI user, *data quality, system robustness/security*, involvement of *vulnerable persons or groups*, the *scale* on which the system is used, its *geographical and temporal scope*, assessment of *likelihood* and *extent* of potential harm, the potential *reversibility* of such harm, and, **where applicable**, whether it concerns a “red line” application as **and if** established by domestic or international law. **The assessment should also take into account such other factors as possible positive effects of AI application for human rights, democracy and the rule of law; the existence of competing legitimate overriding grounds; technical feasibility.**

Commented [LR37]: It should only be necessary to perform such deep assessment if risks are determined to exist that warrant it. Otherwise, the impact assessment procedure may become excessively (and needlessly) burdensome.

Commented [LR38]: This issue should not be governed in this part of the document.

Paragraph 52

Moreover, the CAHAI notes that whereas the impact assessment of AI systems is relatively straightforward in relation to human rights, due to the existence of clearly defined and universal **international** obligations in this area, the impact assessment of AI systems on democracy and the rule of law may prove more difficult in some cases. Nevertheless, given the strong interlinkage between human rights on the one hand and democracy and the rule of law on the other hand, in some situations a negative impact on the former can also provide an indication of a negative impact on the latter. For instance, when the right to freedom of assembly and association or the right to free elections is hampered, it hampers the functioning of democracy. In the same vein, an interference with the right to a fair trial negatively impacts the rule of law. Furthermore, other elements can also be considered, such as the purpose and function of the system within a democratic society, its application domain (with particular attention to the use of AI systems in the public sector or the public sphere), and the way it can hamper certain

democratic- and rule of law-principles (such as the principle of legality, the prevention of misuse of power, or judicial impartiality and independence).

Paragraph 53

Finally, the CAHAI is of the opinion that *stakeholder involvement* in the impact assessment should be assured **whenever appropriate and feasible**. The more severe the impact is deemed to be, or the larger its scale, the more extensive the stakeholder engagement should be. In this regard, particular attention should be paid to involving external stakeholders and members of society (i.e., those who are not covered by the categories of “AI providers” and “AI users”, **such as “AI subjects”** as listed in Chapter III) who could potentially be adversely affected by the deployment of the AI system.

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 54

As set out in Chapter VII, the development, design and application of AI systems in the public sector should be addressed in a legally binding transversal instrument, covering the most important transversal rights and obligations that should be respected in this domain. Additionally, the CAHAI is of the opinion that, given the context specificity of the risks posed by AI in the public sector in light of its specific role in society, such transversal framework **may should** be supplemented by additional legally binding or non-legally binding instruments at sectoral level.

Paragraph 59

In the *procurement* phase, a thorough review of applicable legislation and policy measures in place should be conducted. Where necessary, public procurement processes should be adapted and public procurement guidelines for AI should be adopted, to ensure that procured AI systems comply with human rights, democracy and rule of law standards. A multidisciplinary and multi-stakeholder approach should be respected in order to involve various perspective and angles, including those of **persons in vulnerable situations** ~~vulnerable groups~~. Because public entities are responsible for the systems they adopt and apply, careful attention should be paid to the potential impact on public accountability.

Paragraph 61

Risk management and mitigation frameworks set up in previous phases should be evaluated, adapted and maintained during the *deployment* phase. Taking into account the nature of the risk, human involvement may need to be guaranteed in order to ensure appropriate oversight over the system. Where appropriate, the AI system should be initially and regularly audited by an ~~authorised independent~~ actor **which should be functionally independent**, and the results rendered publicly available to foster public trust. To this end, member States should establish public registers listing AI systems used in the public sector, as well as essential information about the system such as, its purpose, actors involved in its development and deployment, basic information about the model, and performance metrics, where appropriate, and the result of a HUDERIA. In addition, a feedback mechanism should be put in place in order to collect input on how to improve the system directly from its users and those potentially affected

thereby. The AI system should be subjected to regular evaluation and update, including by taking into account the aforementioned feedback. The evaluation process should hence be a periodic one. Transparency and communication towards users and citizens should ~~always~~ be ensured whenever possible, and they should have access to accountability and redress mechanisms. Last but not least, the public should ~~always have the right to normally~~ be informed about the fact that they are interacting with an AI system rather than a human being and provided an opportunity. ~~In such case, they should also be reserved the right~~ to interact with a human being rather than *only* an AI system, and in particular when their rights and interests can be adversely impacted (unless there exist competing legitimate overriding grounds to the opposite, or it is technically infeasible).

Commented [LR39]: The contents of this paragraph often clash with other parts of the document. It contains numerous radical requirements and sweeping statements, while not properly taking into account possible exceptions, special circumstances, or indeed the very nature of AI. A set of minimal amendments are suggested to mitigate this.

Slovenia

III Elements relating to object and purpose, scope, and definitions

Paragraph 13

In so far as *definitions* are concerned, the CAHAI considers that, as a minimum, the following definitions should be included in a legally binding transversal instrument: “*Artificial intelligence system*”; “*lifecycle*”; “*AI provider*”; “*AI user*”; “*AI subject*”; “*unlawful harm*”. The CAHAI recommends that all definitions used should, in so far as possible, be compatible with similar definitions used in other relevant instruments on AI. Furthermore, definitions should be carefully drafted to ensure, on the one hand, legal precision, while, on the other hand, being sufficiently abstract to remain valid despite future technological developments concerning AI systems.

Commented [LR40]: It might be useful to also distinguish a separate category of stakeholders, namely an “AI deployer” and include them among definitions. They can sometimes be distinguished from AI providers.

IX Elements relating to safeguards

Paragraph 39

The CAHAI recommends that a legally binding transversal instrument should include a series of provisions on legal safeguards to be applied to all applications of AI systems used for the purpose of deciding or informing decisions impacting the legal rights and other significant interests of individuals and legal persons.

Commented [LR41]: “We would require some further clarification on paragraph 39 as there seems that the reference to risk based classification is missing.”

XI Elements relating to supervisory authorities, compliance, and cooperation

Paragraph 43

The CAHAI considers that a legally binding transversal instrument should include provisions obliging Parties to take all necessary and appropriate measures to *ensure effective compliance* with the instrument, in particular through the *establishment of compliance mechanisms and standards*. Furthermore, provisions on the establishment or designation of *national supervisory authorities*, defining their powers, tasks and functioning as well as ensuring their independence and impartiality and the allocation of sufficient resources and staff, should be considered for inclusion. In addition, the legally binding transversal instrument should contain provisions regulating the *cooperation between Parties* to

Commented [LR42]: It is necessary that the national supervisory authorities possess sufficient expertise in the AI domain.

facilitate compliance with, and *mutual legal and other assistance, including exchange of data and other forms of information* under, the legally binding transversal instrument.

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 55

These instruments could for instance elaborate further principles and requirements, specifically for the public services, regarding *transparency, fairness, responsibility, accountability* and *redress* to ensure the responsible use of AI. The CAHAI recommends that member States subject the use and design, procurement, development and deployment of AI systems in the public sector to adequate *oversight* mechanisms in order to safeguard *compliance* with human rights, democratic principles and the rule of law, and foster public *trust* in this context by rendering AI systems *traceable* and *auditable*.

Commented [LR43]: "Explainability" should be included among the required principles.

Turkey

II General remarks

We strongly agree that the scope of Council of Europe legal framework on AI should cover "dual use."

III Elements relating to object and purpose, scope, and definitions

We agree that certain definitions should be a component of a legally binding instrument and welcome the inclusion of "AI subject," so as to enable persons affected by/subjected to AI to invoke their rights. Furthermore, we would like to reiterate that in particular, AI "lifecycle" would need to be elaborated. The phrase "development, design and application of artificial intelligence" is used throughout the "Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law" (hereinafter "Possible Elements"). However, we propose that the lifecycle should also involve the "research" and "post-application" phases - the former being addressed under Possible Elements" -, as the potential negative effects of AI would still be imminent in these phases.

IV Elements relating to fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law

We strongly support "both the establishment of certain direct, concrete and positive rights of individuals in relation to the development, design and application of AI systems, as well as the establishment of certain obligations upon Parties."

We suggest including the notion of "digital self-determination" as an evolving normative concept to describe the possibility and realization of human flourishing as it relates to the use of digital technologies

and their affordances. This principle will highlight the multifaceted questions of control over personal data, and also emphasize important contexts of self-determination in the digital age such self-expression and participation in civic life and in the digital economy, as well as relationship-building and well-being, to name just a few application areas with policy implications.

V Elements relating to risk classification of artificial intelligence systems and prohibited applications of artificial intelligence

We very strongly propose and reiterate that HUDERIAS should be mandated in the legal instrument and disagree that they “need not necessarily form a constituent part of a possible legally binding instrument.”

We very strongly propose and reiterate that para. 21 should be restored to the previous version to include “AI systems that profile and target children” and “AI systems used for indiscriminate biometric recognition of individuals in publicly accessible places” as systems that pose unacceptable risk and should therefore be considered as “red lines.”

VI Elements relating to the development, design, and application of artificial intelligence systems in general

We strongly support the specific reference to children and vulnerable groups.

We welcome the inclusion of minimum safeguards, such as but not limited to transparency, to all AI systems as opposed to limiting these principles to “relevant AI systems.”

We welcome the introduction of sustainability under para. 30. We contend that the protection of the environment in relation to AI systems must be an integral component of the legally binding instrument and that the wording of para. 30 must be modified to “environmental sustainability.”

We submit that when referring to “necessary level of *human oversight*” the legally binding instrument should set out the scope and minimum requirements for such role.

VIII Elements relating to democracy and democratic governance

We propose the inclusion of “deep fakes” in the legally binding instrument as opposed to dealing with them in more sectoral instruments. This is significant because the negative implications of such AI technologies is not limited to influencing democratic processes. As recent events and studies have shown, women are disproportionately and severely affected by deep fakes. In accordance with the legal instrument’s focus on gender equality, we argue for consideration of deep fakes in general, with specific regard to vulnerable groups and democratic processes.

XI Elements relating to supervisory authorities, compliance, and cooperation

We strongly support that provisions “on the establishment or designation of *national supervisory authorities*, defining their powers, tasks and functioning as well as ensuring their independence and impartiality and the allocation of sufficient resources and staff” be included in the legally binding instrument.

OBSERVER STATES

Japan

II General remarks

Paragraph 5

The legally binding transversal instrument should focus on mitigating risks emanating from applications of AI systems with the potential to interfere with the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, all the while promoting socially beneficial AI applications. It should be underpinned by a risk-based approach: the legal requirements to the design, development and use of AI systems should be proportionate to the nature of the risk they pose to human rights, democracy and the rule of law. Basic principles that enable the determination of such risk (e.g. transparency requirements) should be applicable to all AI systems.

- ⇒ The legally binding instrument should be drafted based on thorough impact assessment of risk and opportunity of AI applications and such assessment should be made before drafting the instrument. Basic principles that enable the determination of risk should not be applied to AI systems that was assessed low risk a priori or at an initial stage of risk assessment. AI-powered Optical Character Recognition (OCR) may be a good example.

Paragraph 8

It is further recommended that, to ensure both global and regional legal consistency, a legally binding transversal instrument of the Council of Europe should take into account existing and upcoming legal and regulatory frameworks of other international and regional fora, in particular the United Nations, the European Union, and the Organisation for Economic Co-operation and Development – all of which are currently involved in developing various forms of standards related to AI systems.

- ⇒ ISO/IEC JTC1 SC42 is discussing standards on AI including AI Governance and AI Management. The document shows only three organizations (i.e. UN, EU and OECD). Although the government of Japan does not think that other international organizations are excluded, ISO/IEC JTC1 SC42 is relevant and should be explicitly mentioned in the document. Risk Identification, Impact Assessment, Governance Assessment and Mitigation and Evaluation are important elements in the work of ISO/IEC JTC1 SC42, which are mentioned later in Paragraph 50.
- ⇒ In addition to ISO/IEC JTC1 SC42, the document should mention that the legal instrument should be consistent not only at global and regional levels but also a national level. National level instruments

are pooled in the OECD AI Observatory. For example, the government of Japan published AI R&D Guidelines, AI Utilization Guidelines and AI Governance Guidelines.

https://www.soumu.go.jp/main_content/000507517.pdf

https://www.soumu.go.jp/main_content/000658284.pdf

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_9.pdf

Actually, ISO/IEC DIS 38507 mentioned not only Guidelines for Trustworthy Artificial Intelligence by EU and the OECD Principles on AI but also Social Principles of Human-Centric AI by Japan and Model Artificial Intelligence Governance Framework, Second Edition by Singapore.

IV Elements relating to fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law

Paragraph 14

The CAHAI considers it necessary that a legally binding transversal instrument contains certain *fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law*, which should apply to all development, design, and application of AI systems, irrespective of whether the actor is public or private, and irrespective of the actual level of risk to human rights, democracy and the rule of law posed by these systems.

⇒ Some AI applications such as OCR are clearly irrelevant to human rights and democracy. Some AI applications should be excluded a priori or after initial impact assessment of risk and opportunity.

VI Elements relating to the development, design, and application of artificial intelligence systems in general

Paragraph 22

The CAHAI recommends that a legally binding transversal instrument should include a number of provisions applicable to all development, design and application of AI systems, so as to enable their appropriate classification in terms of potential risk to the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law, and to ensure their compliance therewith by setting out minimum safeguards. These can include, for instance, provisions regarding the transparency of AI systems. In line with the risk-based approach mentioned above, further provisions should be rendered applicable to AI systems based on and in proportion with their risk classification, in order to ensure that the risks they pose to human rights, democracy and the rule of law are duly mitigated.

⇒ Some AI applications should be evaluated as low risk a priori such as OCR. The government of Japan suggests that the CAHAI should carefully examine a level of risks of a wide variety of AI applications.

Paragraph 29

The CAHAI also considers it prudent to include a provision on *data governance* for AI systems, in accordance with and building on the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No 108) and its amending Protocol (CETS No 223).

This can include the requirement to establish data governance mechanisms to assess and ensure the data accuracy, integrity, security, **accessibility** and representativeness in a manner that is suitable for the intended purpose of the system and proportionate.

- ⇒ Requirements of data accuracy, integrity, security and representativeness should be carefully defined to ensure the requirements are practical in respect of state of the art.
- ⇒ Requirements of data accessibility should be included to address the issue of data localization.

Paragraph 30

Finally, the CAHAI recommends the introduction of provisions on *robustness, safety and cybersecurity, transparency, explainability, auditability and accountability* throughout their lifecycles. It should be noted that the concepts of “transparency”, “explainability” and “accountability” are considered by the CAHAI to be of paramount importance for the protection of the rights of individuals in the context of AI systems. In addition, the CAHAI recommends that the issue of sustainability in relation to AI systems throughout their lifecycles be considered in a suitable manner.

- ⇒ The government of Japan would like to draw your attention to tradeoff between intellectual property such as source code and trade secret and transparency, explainability, auditability and accountability. Intellectual property should be respected even while the instrument pursues improvement of transparency, explainability, auditability and accountability.

COMMITTEES AND OTHER COUNCIL OF EUROPE BODIES

Congress of Local and Regional Authorities

II General remarks

Paragraph 4

To effectively mitigate these risks, the CAHAI considers that an international legal framework on AI based on the Council of Europe’s standards on human rights, democracy and the rule of law, should take the form of a legally binding transversal instrument. The CAHAI notes that – in addition to the proposed legally binding transversal instrument that sets out general principles and specific legal norms – additional legally binding and/or non-legally binding instruments may be needed at sectoral level, **but also to cover the local and regional dimensions (for municipalities and regions), in order** to provide more detailed guidance on ensuring that the design, development and application of AI occurs in line with human rights, democracy and the rule of law in specific domains **and at all levels of governance**.

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 54

As set out in Chapter VII, the development, design and application of AI systems in the public sector should be addressed in a legally binding transversal instrument, covering the most important transversal rights and obligations that should be respected in this domain. Additionally, the CAHAI is of the opinion that, given the context specificity of the risks posed by AI in the public sector in light of its specific role in society, such transversal framework should be supplemented by additional legally binding or non-legally binding instruments at sectoral level, and to cover the local and regional dimensions (for municipalities and regions).

European Commission Against Racism and Intolerance (ECRI)

The proposal for a binding transversal legal instrument that would include a specific provision on equality and non-discrimination (para. 27) is welcome.

The non-binding nature of additional instruments in the private sector (para. 45) needs to be reviewed. Like the public sector (para. 54), the private should also be governed by binding complementary rules (even more so as discrimination occurs quite often in the private sector). This does not exclude soft law instruments at lower level (e.g. where additional guidance on modalities or procedures are necessary).

COUNCIL OF EUROPE PARTNER INTERNET COMPANIES

EuroISPA

In our view, any future discussion leading to the creation of a new instrument should strive to avoid the creation of new conflicts of Laws with other sector specific legislations at all costs. Furthermore, a specific focus needs to be put on properly addressing the needs of SMEs, without creating additional burdens.

CIVIL SOCIETY ORGANISATIONS, OTHER PRIVATE SECTOR AND ACADEMIC ACTORS RELEVANT TO THE WORK OF THE CAHAI

Access Now

IV Elements relating to fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law

Paragraph 17

The CAHAI further notes that some of the provisions related to these particular elements may be formulated as positive direct rights of individuals, or alternatively as obligations on Parties to ensure the introduction in their domestic law of measures aimed at protecting the rights of individuals in relation to AI systems. Based on its deliberations the CAHAI would, where feasible and necessary, tend to favour a combination of both the establishment of certain direct, concrete and *positive rights of individuals in relation to the development, design and application of AI systems*, as well as the *establishment of certain obligations upon Parties*, to ensure a more uniform application of the legally binding transversal instrument among Parties.

Commented [LR44]: Access Now strongly supports the inclusion of positive direct rights for individuals

Commented [LR45]: Access Now supports this combination, and notes that the initial draft of the EU's AI Act lacks such positive rights and that this lack has been the source of a great deal of criticism as it falls short of properly protecting people impacted by AI systems

V Elements relating to risk classification of artificial intelligence systems and prohibited applications of artificial intelligence

Paragraph 18

The CAHAI recommends that a legally binding transversal instrument should provide for the establishment of a methodology for *risk classification* of AI systems with an emphasis on human rights, democracy, and the rule of law. The criteria used for assessing the impact of AI systems in this regard should be concrete, clear, and with an objective basis and the assessment itself done in a balanced manner, thus providing for both legal certainty and nuance. Clear, non-cumulative criteria must also be provided to determine why a given system is to be assigned to a certain risk level.

Commented [LR46]: Access Now supports this statement, and recommends additionally that clear criteria be given to determine why a given system should be assigned to a certain risk level. Such criteria were developed during the work of the LFG, and we encourage their inclusion in further outputs.

Paragraph 19

In particular, the CAHAI considers that the risk classification should include a number of categories (e.g., "low risk", "high risk", "unacceptable risk", based on a risk assessment in relation to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law. The risk classification will be based on an initial review to determine if a full HUDERIA impact assessment is required (cf. chapter XII) just as the impact assessment itself may have an impact on whether to uphold or change the initial risk classification of the AI system in question. This impact assessment is considered as an element of the overall legal framework on AI systems proposed by the CAHAI. However, the specific HUDERIA model need not necessarily form a constituent part of a possible legally binding instrument.

Commented [LR47]: Access Now strongly supports the inclusion of a category of unacceptable risk.

Paragraph 21

Regarding *prohibited applications of AI* (the so-called “red lines” or “unacceptable risk”), the CAHAI considers that a legally binding transversal instrument should provide for the possibility of putting a moratorium or a ban on the application of AI systems, which in accordance with the aforesaid risk classification are deemed to present an unacceptable risk of interfering with the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. Such possibility should also be considered for the research and development of certain AI systems that present an unacceptable risk. Notably, the CAHAI wishes to draw the attention to, for instance, some AI systems using biometrics to identify, categorise or infer characteristics or emotions of individuals, the use of predictive policing systems and AI systems used for social scoring to determine access to essential services, as applications that may require particular attention. A moratorium or ban should, ~~however, only~~ be considered, where on an objective basis an unacceptable risk to human rights, democracy or the rule of law has been identified and, ~~after careful examination~~, there are no other feasible and equally efficient measures available for mitigating that risk. Review procedures should be put in place to enable reversal of a ban or moratorium if risks ~~have been proven to be~~ sufficiently reduced or appropriate mitigation measures become available ~~and have been tested~~, on an objective basis, such that the system can be demonstrated to not ~~no~~ longer pose an unacceptable risk.

Commented [LR48]: Access Now strongly supports this statement.

VII Elements relating to the development, design, and application of artificial intelligence systems in the public sector

Paragraph 33

Based on the assumption that a legally binding transversal instrument should be general in nature, the CAHAI recommends that such instrument should focus on the potential risks emanating from the development, design, and application of AI systems for the purposes of *law enforcement, the administration of justice, and public administration*. ~~Concerning “public administration”, in particular, the CAHAI notes that a legally binding transversal instrument should not address the plethora of specific administrative activities undertaken by public authorities, such as health care, education, social benefits etc, but be limited to general prescriptions about the responsible use of AI systems in public administration. Issues related to the various sectors of public administration may, as necessary, be addressed in appropriate sectoral instruments.~~

Commented [LR49]: We recommend deleting this section as the use of AI in healthcare, education, and provision of social benefits poses a high risk to human rights.

Paragraph 34

The CAHAI finds that a legally binding transversal instrument when addressing the development, design, and application of AI systems in the public sector should, as a minimum, include provisions on *access to effective remedy*, a mandatory *right to human review* of decisions taken or informed by an AI system except where competing legitimate overriding grounds exclude this, and an *obligation for public authorities to provide relevant individuals or legal persons with meaningful information* concerning the role of AI systems in taking or informing decisions relating to them, ~~except where the law prescribes that competing legitimate overriding grounds exclude such disclosure~~. Furthermore, Parties should be obliged to ensure that *adequate and effective guarantees against arbitrary and abusive practices* due to the application of an AI system in the public sector are afforded by their domestic law.

Commented [LR50]: Access Now enthusiastically support this recommendation. The inclusion of such rights is vital.

XI Elements relating to supervisory authorities, compliance, and cooperation

Paragraph 43

The CAHAI considers that a legally binding transversal instrument should include provisions obliging Parties to take all necessary and appropriate measures to *ensure effective compliance* with the instrument, in particular through the *establishment of compliance mechanisms and standards*. Furthermore, provisions on the establishment or designation of *national supervisory authorities*, defining their powers, tasks and functioning as well as ensuring their independence and impartiality and the allocation of sufficient resources and staff, should be considered for inclusion. In addition, the legally binding transversal instrument should contain provisions regulating the *cooperation between Parties* to facilitate compliance with, and *mutual legal and other assistance, including exchange of data and other forms of information* under, the legally binding transversal instrument.

Commented [LR51]: Access Now believes it is absolutely essential that any such authority must be independent, impartial and well funded and resourced. Any legislation, no matter how good, will be undermined by ineffective enforcement, and independence and impartiality are essential to that.

XII Human rights, democracy, and rule of law impact assessment

Paragraph 46

A well-conducted human rights, democracy, and rule of law impact assessment can advance the assessment of how the deployment of AI systems can affect the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. It should though be noted that this type of impact assessment is not designed to balance negative and positive impacts, something which may depend on the specificities of the legal system in the jurisdiction in which the AI system is intended to be applied. In a subsequent stage, it can then be examined if and how risks identified through the HUDERIA can be mitigated, and if and how a legitimate interest can legitimize the system's use despite interference with human rights, democracy and rule of law standards, when such limitations are prescribed by law, proportionate, and necessary in a democratic society.

Commented [LR52]: Access Now strongly supports this statement. Under no circumstances should such a balancing act be part of an impact assessment, and the role of an impact assessment is also not to enumerate positive benefits; this is the role of marketing materials and public communications.

Paragraph 51

As regards the *Impact Assessment* step, the CAHAI further recommends that the assessment of an AI system, at least, could include the following *elements*: assessment of the *context and purpose* of the AI system, *level of autonomy* of the AI system, *underlying technology* of the AI system, *usage* of the AI system (both intended and potentially unintended use), *complexity* of the AI system (part of multiple deep neural networks/building on other AI systems/dual use), *transparency* and *explainability* of the system and the way it is used, *human oversight and control mechanisms* for the AI provider and AI user, *data quality*, *system robustness/security*, involvement of *vulnerable persons or groups*, the *scale* on which the system is used, its *geographical* and *temporal* scope, assessment of *likelihood* and *extent* of potential harm, the potential *reversibility* of such harm, and whether it concerns a "red line" application as established by domestic or international law.

Commented [LR53]: Access Now supports the extensiveness of this proposal.

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 56

Additionally, considering that the distinction between public and private sector involvement is often ambiguous, and considering the liability issues relating to the contracting out of public services to private actors, any provisions applying to the design, development and application of AI in the public sector should also apply to private actors that act on behalf of the public sector.

Commented [LR54]: Access Now enthusiastically supports this statement, and notes that it accords with the recommendations made in the Commissioner for Human Rights, Dunja Mijatović's recommendation. 'Unboxing artificial intelligence: 10 steps to protect human rights'

Paragraph 58

In the *design phase* of the system, due consideration should be given to the analysis of the problem which the public entity intends to solve, in order to assess whether an AI system is the best fit for the problem and, if so, which characteristics it should have. The data sets to be used for the AI system should be clearly identified, and the protection of such data and their origin respected. The design choices of the system should then be rendered explicit and documented. The intended users of the system, both civil servants and the public, as well as those potentially affected by the system should be involved early on, and their capabilities in using the AI system in question should be considered. An open and transparent design should be favoured. Finally, a human rights, democracy and rule of law impact assessment should be carried out to anticipate, prevent and mitigate potential risks. This also requires putting in place risk management and mitigation frameworks, which are relevant throughout all phases.

Commented [LR55]: We thoroughly support this statement. It is an essential first step to ensure AI is used in an effective manner.

Paragraph 61

Risk management and mitigation frameworks set up in previous phases should be evaluated, adapted and maintained during the *deployment* phase. Taking into account the nature of the risk, human involvement may need to be guaranteed in order to ensure appropriate oversight over the system. Where appropriate, the AI system should be initially and regularly audited by an independent actor, and the results rendered publicly available to foster public trust. To this end, member States should establish public registers listing AI systems used in the public sector, as well as essential information about the system such as, its purpose, actors involved in its development and deployment, basic information about the model, and performance metrics, where appropriate, and the result of a HUDERIA. In addition, a feedback mechanism should be put in place in order to collect input on how to improve the system directly from its users and those potentially affected thereby. The AI system should be subjected to regular evaluation and update, including by taking into account the aforementioned feedback. The evaluation process should hence be a periodic one. Transparency and communication towards users and citizens should always be ensured, and they should have access to accountability and redress mechanisms. Last but not least, the public should always have the right to be informed about the fact that they are interacting with an AI system rather than a human being. In such case, they should also be reserved the right to interact with a human being rather than *only* an AI system, and in particular when their rights and interests can be adversely impacted.

Commented [LR56]: Access Now believes that this is one of the most important recommendations in the document. A public register should contain information about what AI systems are on the market in a given jurisdiction, but also information about individual deployments. This provides an essential, baseline transparency that enables the protection of human rights.

AlgorithmWatch

II General remarks

Paragraph 5

The legally binding transversal instrument should focus on mitigating risks emanating from applications of AI systems with the potential to interfere with the enjoyment of human rights, the functioning of democracy and the observance of the rule of law, ~~all the~~ while promoting socially beneficial AI applications. It should be underpinned by a risk-based approach: the legal requirements to the design, development and use of AI systems should be proportionate to the nature of the risk they pose to human rights, democracy and the rule of law. Basic principles that enable the determination of such risk (e.g. transparency requirements) should be applicable to all AI systems.

Paragraph 6

In accordance with Article 1 d of the Statute of the Council of Europe, matters relating to national defence should not be covered by the scope of a legally binding (or non-legally binding) instrument of the Council of Europe. However, ~~purely~~ civilian applications of AI systems, which were originally developed for military use, ~~could~~ must fall within the scope of Council of Europe instruments. It is thus the actual application of the AI system – not the formal categorisation of that AI system as for military or civilian use – which should be the determining factor for the assessment of whether or not it would be covered by a Council of Europe instrument. The CAHAI is of the opinion that the issue of whether that scope should cover “dual use” should be further considered in the context of developing a Council of Europe legal framework on AI.

Paragraph 7

The various legal issues raised by the application of AI systems are not specific to the ~~M~~member States of the Council of Europe, but are, due to the many global actors involved and the global effects they engender, transnational in nature. The CAHAI therefore recommends that a legally binding transversal instrument of the Council of Europe, though obviously based on Council of Europe standards, be drafted in such a way that it facilitates accession by States outside of the region without lowering the aforementioned standards. Not only will this significantly increase the impact and efficiency of the proposed instrument, but in addition it will provide a much-needed level playing field for relevant actors, including industry and AI researchers which often operate across national borders and regions of the world. The standards of the Council of Europe on human rights, democracy and the rule of law are sufficiently universal in nature to make this a realistic option. There are several precedents of Council of Europe treaties being applied beyond the European region, cf. notably the Budapest Convention (Cybercrime) and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108), which currently have 66 and 55 Parties respectively, many of which are not member States of the Council of Europe.

Paragraph 8

It is further recommended that, to ensure both global and regional legal consistency, a legally binding transversal instrument of the Council of Europe should take into account ~~and complement~~ existing and upcoming legal and regulatory frameworks of other international and regional fora, in particular the

United Nations, the European Union, and the Organisation for Economic Co-operation and Development – all of which are currently involved in developing various forms of standards related to AI systems.

IV Elements relating to fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law

Paragraph 16

Concerning the concept of “human dignity”, the CAHAI notes that the dignity of the human person is universally agreed to constitute the real basis of human rights, cf. also the prominence given to the concept in the preamble of the 1948 Universal Declaration of Human Rights. In the view of the CAHAI, it makes particularly good sense to use this concept in a legally binding transversal instrument on the potential adverse impacts on fundamental human rights of individuals caused by the development, design, and application of AI systems.

V Elements relating to risk classification of artificial intelligence systems and prohibited applications of artificial intelligence

Paragraph 18

The CAHAI recommends that a legally binding transversal instrument should provide for the establishment of a methodology for *risk classification of applications of AI systems* with an emphasis on human rights, democracy, and the rule of law. The criteria used for assessing the impact of *applications of AI systems* in this regard should be concrete, clear, and with an objective basis and the assessment itself done in a balanced manner, thus providing for both legal certainty and nuance. The risk classification should not be regarded as a flexible categorization but as a momentary classification.

Paragraph 19

In particular, the CAHAI considers that the ~~risk classification should include a number of categories (e.g., “low risk”, “high risk”, “unacceptable risk”, based on a risk assessment in relation to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law. The risk classification~~ will be based on an initial review (triage) to determine if a full HUDERIA impact assessment is required (cf. chapter XII) just as the impact assessment itself may have an impact on whether to uphold or change the initial risk classification of the application of the AI system in question. This impact assessment is considered as an element of the overall legal framework on AI systems proposed by the CAHAI. However, the specific HUDERIA model need not necessarily form a constituent part of a possible legally binding instrument.

Paragraph 21

Regarding *prohibited applications of AI (the so-called “red lines” or “unacceptable risk”)*, the CAHAI considers that a legally binding transversal instrument should provide for the possibility of putting

a moratorium or a ban on the application of AI systems, which in accordance with the aforesaid risk classification are deemed to present an unacceptable risk of interfering with the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. Such possibility should also be considered for the research and development of certain AI systems that present an unacceptable risk, inter alia with the aim of preventing an extraterritorial protection vacuum. ~~Notably, the CAHAI wishes to draw the attention to, for instance, some~~ Examples of potential red lines would include, for instance, certain AI systems using biometrics to identify, categorise or infer characteristics or emotions of individuals – especially when those enable mass surveillance –, and AI systems used for social scoring to determine access to essential services, ~~as applications that may require particular attention~~. A moratorium or ban should, however, only be considered, where on an objective basis an unacceptable risk to human rights, democracy or the rule of law has been identified and, after careful examination, there are no other feasible and equally efficient measures available for mitigating that risk. Review procedures should be put in place to enable reversal of a ban or moratorium if risks are sufficiently reduced or appropriate mitigation measures become available, on an objective basis, to no longer pose an unacceptable risk.

VI Elements relating to the development, design, and application of artificial intelligence systems in general

Paragraph 22

The CAHAI recommends that a legally binding transversal instrument should include a number of provisions applicable to all development, design and application of AI systems, so as to enable their appropriate classification in terms of potential risk to the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law, and to ensure their compliance therewith by setting out minimum safeguards. These can include, for instance, provisions regarding the transparency of AI systems. In line with the risk-based approach mentioned above, further provisions should be rendered applicable to applications of AI systems based on and in proportion with their risk classification, in order to ensure that the risks they pose to human rights, democracy and the rule of law are duly mitigated.

Paragraph 23

A legally binding transversal instrument should, ~~as a general rule~~, state that, subject to certain limitations if conforming to the requirements of this instrument, *the development and design of, as well as the research in, AI systems should be carried out freely, with due consideration for safety and security, and in full compliance with the Council of Europe standards on human rights.*

Paragraph 25

To promote a multi-stakeholder approach, and in order to raise awareness in society about the impact of the development, design and application of AI systems, the CAHAI considers it useful to include a provision calling for Parties to promote evidence-based *public debate* on and inclusive engagement with this topic. Inspiration for the wording of such a provision may be found in Article 28 of the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (CETS No 164).

Paragraph 28

For the same reasons, a legally binding transversal instrument should contain provisions on ensuring that gender equality and rights related to *vulnerable groups*, including *children*, are being upheld throughout the lifecycle of artificial intelligence- AI systems.

Paragraph 30

Finally, the CAHAI recommends the introduction of provisions on *robustness, safety and cybersecurity, transparency, explainability, auditability and accountability* throughout their lifecycles. It should be noted that the concepts of “transparency”, “explainability” and “accountability” are considered by the CAHAI to be of paramount importance for the protection of the rights of individuals in the context of AI systems. Individuals should have (upon request) the right to have access to information on decisions taken by AI systems that have an impact on them and access to legal remedies. In addition, the CAHAI recommends that the issue of sustainability in relation to AI systems throughout their lifecycles be considered in a suitable manner.

VII Elements relating to the development, design, and application of artificial intelligence systems in the public sector

Paragraph 32

The development, design, and application of AI systems in the public sector give rise to some concerns about how to ensure the respect for human rights, democracy, and the rule of law when AI systems are used to take or inform decisions that impact the rights and obligations of individuals and legal persons. That said, the CAHAI underlines that not all public sector AI applications pose risks to the enjoyment of human rights, the functioning of democracy and the observance of the rule of law. It is accordingly important to carefully examine the potential for risk posed by a given application of an AI system on a case-by-case basis. Accordingly, a distinction should be made between, on the one hand, AI systems which can interfere with human rights, democracy or the rule of law, and on the other hand, AI systems which though operated by the same public authorities do not present any such risks.

VIII Elements relating to democracy and democratic governance

Paragraph 37

The role of private entities, for instance, online platforms that help shape the public sphere, should also be considered in this respect, insofar as the growing concentration of economic power and of data could undermine democratic processes. To enable an evidence-based public debate regarding their role and impact on democracy, public interest research conducted by academics and civil society organisations should be legally ensured.

Paragraph 38

In this context, the CAHAI underlines the need for respecting *the full catalogue of human rights, in particular the right to freedom of expression, including the freedom to form and hold opinions and to receive and impart political information and ideas, and the right to freedom of assembly and association*, with the aim of ensuring that all parties and interest groups have access to democratic processes in equal conditions, and that a free space for public debate can be ensured.

XI Elements relating to supervisory authorities, compliance, and cooperation**Paragraph 43**

The CAHAI considers that a legally binding transversal instrument should include provisions obliging Parties to take all necessary and appropriate measures to *ensure effective compliance* with the instrument, in particular through the *establishment of compliance mechanisms and standards*. Furthermore, provisions on the establishment or designation of *national supervisory authorities*, defining their powers, tasks and functioning as well as ensuring their independence and impartiality and the allocation of sufficient resources, *expertise* and staff, should be considered for inclusion. In addition, the legally binding transversal instrument should contain provisions regulating the *cooperation between Parties* to facilitate compliance with, and *mutual legal and other assistance, including exchange of data and other forms of information* under, the legally binding transversal instrument.

XII Human rights, democracy, and rule of law impact assessment**Paragraph 46**

A well-conducted human rights, democracy, and rule of law impact assessment (*HUDERIA*) can advance the assessment of how the deployment of AI systems can affect the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. It should though be noted that this type of impact assessment is not designed to balance negative and positive impacts, something which may depend on the specificities of the legal system in the jurisdiction in which the AI system is intended to be applied. In a subsequent stage, it can then be examined if and how risks identified through the HUDERIA can be mitigated, and if and how a legitimate interest can legitimize the system's use despite interference with human rights, democracy and rule of law standards, when such limitations are prescribed by law, proportionate, and necessary in a democratic society.

Paragraph 48

Given the time and resources necessary to undertake such an assessment, and in order to safeguard the proportionality of a risk-based approach, the CAHAI believes that, as a rule, a formalised extensive human rights, democracy, and rule of law impact assessment should only be mandated if there are clear and objective indications of relevant risks emanating from the application of an AI system. This requires that all AI systems undergo an initial review (*triage*) in order to determine whether or not they should be subjected to such a formalised assessment. It is recommended that indications as to the

necessity for a more extensive assessment be further developed by the CAHAI. It should also be considered that using an AI system in a new or different context or for a new or different purpose or otherwise relevant changes would require a reassessment.

Paragraph 50

The CAHAI recommends that, at least, the following *main steps* be included in a human rights, democracy, and rule of law impact assessment, subject to an initial review having been conducted, and including stakeholder involvement, where relevant:

- (1) *Risk Identification*: Identification of relevant risk signals for human rights, democracy and the rule of law;
- (2) *Impact Assessment*: Assessment of the impact on those rights and principles;
- (3) *Governance Assessment*: Assessment of the roles and responsibilities of duty-bearers, right holders and stakeholders in implementing and governing the mechanisms to mitigate the impact;
- (4) *Mitigation and Evaluation*: Identification of suitable mitigation measures and ensuring a continuous evaluation.

Paragraph 51

As regards the *Impact Assessment* step, the CAHAI further recommends that the assessment of an AI system, at least, could include the following *elements*: assessment of the *context and purpose* of the AI system, *its impact on decisions that affect human beings or democratic societies, the level of autonomy* of the AI system, *underlying technology* of the AI system, *usage* of the AI system (both intended and potentially unintended use), *complexity* of the AI system (part of multiple deep neural networks/building on other AI systems/dual use), *transparency* and *explainability* of the system and the way it is used, *human oversight and control mechanisms* for the AI provider and AI user, *data quality, system robustness/security*, involvement of *vulnerable persons or groups*, the *scale* on which the system is used, its *geographical* and *temporal* scope, assessment of *likelihood* and *extent* of potential harm, the potential *reversibility* and *compensability* of such harm, and whether it concerns a “red line” application as established by domestic or international law.

Paragraph 52

Moreover, the CAHAI notes that whereas the impact assessment of *applications of* AI systems is relatively straightforward in relation to human rights, due to the existence of clearly defined and universal obligations in this area, the impact assessment of *applications of* AI systems on democracy and the rule of law may prove more difficult in some cases. Nevertheless, given the strong interlinkage between human rights on the one hand and democracy and the rule of law on the other hand, in some situations a negative impact on the former can also provide an indication of a negative impact on the latter. For instance, when the right to freedom of assembly and association or the right to free elections is hampered, it hampers the functioning of democracy. In the same vein, an interference with the right to a fair trial negatively impacts the rule of law. Furthermore, other elements can also be considered, such as the purpose and function of the system within a democratic society, its application domain (with particular attention to the use of AI systems in the public sector or the public sphere), and the way it can hamper certain

democratic- and rule of law-principles (such as the principle of legality, the prevention of misuse of power, [procedural guarantees](#), or judicial impartiality and independence).

Paragraph 53

Finally, the CAHAI is of the opinion that *stakeholder involvement* in the impact assessment should be assured. The more severe the impact is deemed to be, or the larger its scale, the more extensive the stakeholder engagement should be. In this regard, particular attention should be paid to involving external stakeholders, [civil society organizations](#), and members of society (i.e., those who are not covered by the categories of “AI providers” and “AI users”, as listed in Chapter III) who could potentially be adversely affected by the deployment of the AI system [as well as their representatives](#).

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 55

These instruments could for instance elaborate further principles and requirements, specifically for the public services, regarding *transparency, fairness, responsibility, accountability and redress* to ensure the responsible use of AI. The CAHAI recommends that ~~M~~member States subject the use and design, procurement, development and deployment of AI systems in the public sector to adequate *transparency, oversight and accountability* mechanisms in order to safeguard *compliance* with human rights, democratic principles and the rule of law, and foster [justified public trust](#) in this context by rendering AI systems [trustworthy, i.e. intelligible, traceable and auditable](#).

Paragraph 61

Risk management and mitigation frameworks set up in previous phases ~~must~~[should](#) be evaluated, adapted and maintained during the *deployment* phase. Taking into account the nature of the risk, human involvement may need to be guaranteed in order to ensure appropriate oversight over the system. Where appropriate, the AI system should be initially and regularly audited by an independent actor, and the results rendered publicly available to foster public trust. To this end, member States should establish public registers listing AI systems used in the public sector, [containing as well as](#) essential information about the system such as, its purpose, actors involved in its development and deployment, basic information about the model, and performance metrics, where appropriate, and the result of a HUDERIA. In addition, a feedback mechanism should be put in place in order to collect input on how to improve the system directly from its users and those potentially affected thereby. The AI system should be subjected to regular evaluation and update, including by taking into account the aforementioned feedback. The evaluation process should hence be a periodic one. Transparency and communication towards users and citizens should always be ensured, and they should have access to accountability and [individual and collective](#) redress mechanisms. Last but not least, the public should always have the right to be informed about the fact that they are interacting with an AI system rather than a human being. In such case, they should also be reserved the right to interact with a human being rather than *only* an AI system, and in particular when their rights and interests can be adversely impacted.

Paragraph 62

Finally, measures should be put in place to increase digital literacy and skills among both civil servants and the general public, notably through investment in capacity building (initial and continuous training and education) of public officials and awareness raising about the benefits, risks, capabilities and limitations of AI systems. Such skills should encompass theoretical as well as practical knowledge on the interplay between the design, development and application of AI systems on the one hand, and human rights, democracy and the rule of law on the other hand. Moreover, it should be of high importance to enable and foster public interest research through ensuring data access to relevant stakeholders in society. Furthermore, attention should also be given to the way in which these systems should be supervised and the risks arising ~~therefrom~~ from those should be managed.

Center For AI and Digital Policy

The Center for AI and Digital Policy (CAIDP) welcomes the opportunity to provide comments on the CAHAI draft of "Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law." We support the effort to strengthen human rights, democracy and the rule of law for the field of artificial intelligence through transparent and open consultation. We specifically support the recommendations set out in the CAHAI draft, including the strong emphasis on transparency, accountability, fairness, and redress, as well as further recommendations described below.

We also want to highlight the importance of creating a legally binding transversal instrument within the framework of the Council of Europe to enable AI based systems to promote a better society where technology promotes broad social inclusion based on fundamental rights, democratic institutions, and the rule of law.

Over the last few years, many international organizations have adopted important frameworks for AI policy. The OECD adopted AI Principles in 2019. The G20 adopted AI Guidelines in 2019. This week UNESCO adopted the AI Ethics Recommendation. The EU and US have proposed a framework for AI policy in the context of the EU-US Trade and Technology Council. Although all of these policy instruments establish important AI policy norms, none are legally binding.

The Council of Europe is uniquely situated to establish a global standard for AI. The Council has previously developed successful treaties in the related fields of data protection and cybercrime. Within the program for Strengthening the Rule of Law, the CAHAI was instructed to complete a feasibility study on "a legal framework for the development, design and application of artificial intelligence"¹. The substantial draft of the CAHAI for the Council of Ministers, setting out the elements of a legal framework, reflects the high level of interest in a formal, legally binding instrument for AI.

Based on your invitation, our team of experts also proposes the following suggestions, which we believe should be considered for the CAHAI report for the attention of the Council of Ministers.

¹ COE Committee of Ministers, *Ad hoc Committee on Artificial Intelligence (CAHAI), Terms of Reference*, 1353rd meeting, 11 September 2019, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016809737a1

● **First, we encourage the addition of “sustainability” as a fourth criteria in the human rights, democracy and rule of law impact assessment.** The goal of sustainability has become an ever more prominent principle in the field of AI policy. The G20 digital ministers recognised the need to tackle digital technologies' significant consumption of energy and resources in their 2021 Declaration². In the recently adopted UNESCO Recommendations on the Ethics of AI the value of “Environment and ecosystem flourishing” was included. The UNESCO Recommendation states that environmental impact should be reduced to “ensure the minimisation of climate change and environmental risk factors . . .”³ We urge you to consider adding this emerging value to the foreseen ones namely human rights, democracy and rule of law.

The addition of sustainability is important because the other three norms are strongly reliant on sustainable development. If we do not develop these systems sustainably, then we will be left with far more divided societies and more unevenly distributed consequences which will strongly hamper the upholding of the other three values set out.

● **Second, CAIDP recommends a specific provision on the rights of the child.** It is encouraging that you mention the importance of protecting gender equality and the rights of vulnerable groups. Children as a group of special exposure in digital environments face the risk of harmful experiences, which is why the necessity of protection is of great importance. The G20 has recently recognized the importance of protecting the rights of children in the AI content. We share the view of responsibility of providers of digital services and products, including governments, companies, parents, guardians, civil society, educators, representative groups and children themselves, to provide and engage with technologies in a safe and responsible manner⁴.

● **Third, we propose a prohibition on AI-enabled weapon systems that implicate human rights.** We noted you do not address matters related to national defense in accordance with Article 1(d) of the Statute of the Council of Europe⁵. As the protection of human rights remains within the Council's mandate, we recommend a prohibition of AI-enabled autonomous weapon systems that implicate human rights. Our recent review of country policies strongly indicates support among democratic nations for limits on these systems⁶.

● **Fourth, in relation to Human Dignity (pars. 14-17), we recommend a ban on biometric categorization of individuals and emotion analysis.** Manual forms of these practices have been used throughout history to justify hierarchies of humankind, practices of slavery, eugenics, and oppression of whole groups. None of these applications have any scientific basis or validity. Currently they use spurious correlations and

² G20 (2021), Declaration of G20 Digital Ministers: Leveraging Digitalisation for a Resilient, Strong, Sustainable and Inclusive Recover, p. 2-3, https://www.g20.org/wpcontent/uploads/2021/08/DECLARATION-OF-G20-DIGITAL-MINISTERS-2021_FINAL.pdf (visited 18 November 2021).

³ UNESCO (2021), Recommendation on the Ethics of AI, p. 8-9, <https://unesdoc.unesco.org/ark:/48223/pf0000377897> (visited 18 November 2021).

⁴ G20 (2021), Declaration of G20 Digital Ministers: Leveraging Digitalisation for a Resilient, Strong, Sustainable and Inclusive Recover, p. 6, https://www.g20.org/wpcontent/uploads/2021/08/DECLARATION-OF-G20-DIGITAL-MINISTERS-2021_FINAL.pdf (visited 18 November 2021).

⁵ CAHAI, *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law*, N 6.

⁶ *Artificial Intelligence and Democratic Values* (CAIDP 2020) (findings), <https://www.caidp.org/aidv-2020>

questionable science to make inferences about personality, character, political and religious beliefs – again to deny a person their humanity, dignity, identity. We further propose a **Ban on scoring of individuals by public and private entities**. Human dignity relates to the recognition of the intrinsic and equal worth of each individual human being or a normative judgement of one group over another about what is good behavior. It requires no system segregate or objectify or categorize especially when the criteria are based on political concepts, physical traits, expressed opinions or spurious correlations. Biased risk or trustworthiness scores, coupled with biased datasets and huge power imbalances will result in cumulative disadvantages⁷ – deepening the structural imbalances we are trying to solve. Humans should be treated as moral subjects, and not as objects to be algorithmically scored or manipulated⁸. We also propose **Expansion of protections and HUDARIA assessment**. People seeking refugee protections, asylum and those who are in prison should also be included in the recommendation.

• **Fifth, in relation to Rule of Law (pars. 45-47), we recommend a ban on prediction of future crime.** Human dignity also relates to aspirations and personal development of each individual without being defined by every single past behavior, network of affiliations, or biased datasets reflective of discriminatory practices. Every person deserves presumption of innocence, right to fair trial, due process, judicial independence, effective remedy and impartiality. This requires also a **Restriction on public authorities use of data collected by private entities**. Law enforcement and welfare management agencies increasingly purchase data from databroker vendors, or use private AI systems to access data that they could not legally collect or collect without authorization.

• **Sixth, in relation to Democratic Values, we recommend a Ban on biometric recognition (facial, voice and gait) systems used for mass surveillance purposes.** Indiscriminate mass surveillance whether implemented by public or private companies is intended precisely to manipulate or coerce social behavior and to control populations. We also want to draw **direct connection with CAHAI’s recommendations for elements of a legal framework and CAIDP’s recommendations to national governments.**

The CAIDP is an independent, non-profit organization established to advise national governments and international organizations on AI and digital policy. In 2020 we published *Artificial Intelligence and Democratic Values*⁹, a comprehensive report of the AI policies and practices in 30 countries. As set forth in this report, we recommend that countries:

- Establish national policies for AI that implement democratic values
- Ensure public participation in AI policymaking and create robust mechanisms for independent oversight of AI systems
- Guarantee fairness, accountability, and transparency in all AI systems
- Commit to these principles in the development, procurement, and implementation of AI systems for public services
- Halt the use of facial recognition for mass surveillance

⁷ Gandy, O.H. (2009). *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage* (1st ed.). Routledge. <https://doi.org/10.4324/9781315572758>

⁸ The Alan Turing Institute (2021). Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A proposal prepared for the Council of Europe’s Ad hoc Committee on Artificial Intelligence. <https://rm.coe.int/huderaf-coe-final-1-2752-6741-5300-v-1/1680a3f688>

⁹ *Artificial Intelligence and Democratic Values*, (CAIDP 2020), <https://www.caidp.org/aidv-2020/>

We recognize that several of these recommendations are reflected in the draft CAHAI legal framework for AI, and express our support for this initiative on that basis.

Council of Bars and Law Societies of Europe

Regarding the general remarks

(§4) The CCBE welcomes the consideration of **sectoral instruments** to ensure human rights, democracy and the rule of law in specific domains. In our views, there should be a specific legally binding instrument targeting specific risks in specific circumstances, such as a risk of an unfair trial if parties in a case are not given the opportunity to assess, discuss and raise objections against the results produced by an AI tool which was used in the judicial decision-making process.

(§7)The CCBE notes that the CAHAI recommends to the Council of minister that **a legally binding transversal instrument of the Council of Europe should be drafted in such a way that it facilitates accession by third member States**, without lowering the aforementioned standards. It is said that the standards of the Council of Europe are sufficiently universal in nature to make this a realistic option. This assertion is also based on precedents, such as the Budapest Convention. However, the CCBE considers that the risk to lower the standards of the foreseen instruments are high. The standards of the Council of Europe are universal, but they are not upheld universally. During the draft of the foreseen instrument, it should be make sure that no compromises are made on any standards, in particular those enforce by the ECHR. The example of the Budapest Convention is instructive: during the draft of the 2nd additional protocol, European standards have been lowered in order to give signatories the possibility to choose between different level of lower/higher standards.

§8: the CAHAI further recommends that the future legal instrument on AI takes into account existing and upcoming legal and regulatory frameworks of other international and regional ear, in particular the UN, the EU and the OECD. Here again, by taking into account such instruments, the CoE should assess them to ensure that the highest standards are/will be upheld.

Regarding the elements for a legally binding transversal instrument

§11: the CCBE notes that the CAHAI recommends to ensure a full consistency with respect for human rights [...] in the development, design and use of AI systems, irrespective of whether these activities are undertaken by private or public actors. This is of paramount importance. The proposed AI Act of the EU does not follow this basic principle. The impact of AI on democracy and the rule of law cannot be underestimated. Restrictions and prohibitions on the development, design and use of AI should apply to all actors.

§15: the CCBE considers that the CAHAI should also insist on the importance for the CoE to seek the highest possible standards. The minimisation of the risks of unwarranted duplication or fragmentation of existing standards should not lower the level of the sought standards.

§21: Bans and moratorium should be clear in their scope and not limited with extended exceptions. The review procedure to enable reversal of a ban should bez clearly detailed in the instruments.

§33: the CCBE welcomes that the CAHAI recommends that the foreseen instrument should focus on the potential risks emanating from the development, design, and application of AI systems for the purposes of law enforcement, the administration of justice, and public administration. According to the CCBE a targeted approach is needed in order to set legal requirements tailored to the needs of specific sectors, uses-cases and circumstances.

§34: If the foreseen legally binding instrument focuses in particular on the use AI systems developed, designed and used for Law enforcement purposes and the administration of justice, it must be mentioned that there shall be a right to a human judge, at all stages of the proceedings. In those matters, there can be no exceptions, as the rights to a fair trial is absolute based on Article 6 of the ECHR.

§43: The CAHAI should mention the participation of concerned stakeholders in the monitoring of AI systems, at national and European levels. The **promotion of a multi-stakeholders approach** is mentioned in §25 and should be reflected in the elements relating to supervisory issues, as it is for the human rights, democracy and rule of law impact assessment under §53. With regard to the administration of justice and law enforcement, all justice actors, as well as citizens and litigants, should be able to participate in the supervision of AI systems. At the European level, such monitoring could take place within the framework of the Council of Europe. It should go beyond a simple feedback mechanism. As mentioned in §61.

Global Partners Digital

II General remarks

Paragraph 3

The CAHAI observes that the application of **certain** artificial intelligence (AI) systems has the potential to promote human prosperity and individual and social well-being by enhancing progress and innovation, yet at the same time certain applications of AI systems give rise to concern, as they potentially pose risks to human rights, democracy and the rule of law.

Paragraph 6

In accordance with Article 1 d of the Statute of the Council of Europe, matters relating to national defence should not be covered by the scope of a legally binding (or non-legally binding) instrument of the Council of Europe. The CAHAI is of the opinion that the issue of whether that scope should cover “dual use” should be further considered in the context of developing a Council of Europe legal framework on AI.

Paragraph 8

It is further recommended that, to ensure both global and regional legal consistency, a legally binding transversal instrument of the Council of Europe should take into account existing and upcoming legal and regulatory frameworks of other international and regional fora, in particular the United Nations, the European Union, and the Organisation for Economic Co-operation and Development – all of which are currently involved in developing various forms of standards related to AI systems, **provided that ensuring**

Commented [LR57]: GPD proposes the addition of the word "certain" to here to make clear that while certain applications of AI have the "the potential to promote human prosperity and individual and social well-being by enhancing progress and innovation", it is not true of all applications. This change would also ensure consistency in the paragraph as the second sentence makes reference to "certain" applications.

Commented [LR58]: GPD proposes that this text be retained in its current form. We recognise that there has been a lot of discussion on this point, and that the issue is contentious, but we support the wording proposed by the Secretariat here.

~~consistency would not result in any diminution in the level of protection for human rights, democracy and the rule of law provided in any instrument developed by the Council of Europe.~~

Commented [LR59]: GPD proposes this additional wording to reflect the fact that other legal and regulatory frameworks developed by international and regional for a may not take a human right-based approach or may contain weaker standards for human rights than a Council of Europe instrument. The additional wording makes clear that while consistency is a worthy aim, it should not come at the cost of diminishing the level of protection of human rights in any Council of Europe instrument.

III Elements relating to object and purpose, scope, and definitions

Paragraph 12

The CAHAI considers that the legally binding transversal instrument should contain a provision defining its *scope*. This provision should clarify that the instrument shall be applicable to the development, design and application of AI systems, irrespective of whether these activities are undertaken by public or private actors, with a particular focus on such systems which are assessed to pose potential risks to the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law. ~~As necessary, potential exceptions to the scope should also be addressed.~~

Commented [LR60]: GPD does not support the inclusion of this sentence. The previous sentences make clear that the scope of the treaty would need to be defined in the instrument, and it is implicit that this definition would contain any limitations or exceptions. Explicitly signposting the potential need for exceptions, however, signals too strongly that CAHAI considers that exceptions are needed. In the absence of any guidance as to what they could be, we do not believe that this sentence is a helpful addition to the document.

IV Elements relating to fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law

Paragraph 15

At the same time the CAHAI, recognising the risks of duplicating or even fragmenting existing general standards of international law, including human rights law, recommends that such fundamental principles be drafted in such a way that the risks of unwarranted duplication or fragmentation are duly minimised. ~~This entails, *inter alia*, further tailoring rights and obligations relating to human rights, democracy and the rule of law for the purpose of this instrument only where and when, after careful examination, the conclusion is reached that existing standards in their current form cannot provide sufficient protection of the rights of individuals in the specific context of the development, design and application of AI systems.~~

Commented [LR61]: GPD supports this language. We recognise that there has been a lot of discussion about the wording here, and so wish to reiterate our support for the text as written.

VII Elements relating to the development, design, and application of artificial intelligence systems in the public sector

Paragraph 33

Based on the assumption that a legally binding transversal instrument should be general in nature, the CAHAI recommends that such instrument should focus on the potential risks emanating from the development, design, and application of AI systems ~~for the purposes of law enforcement, the administration of justice, and public administration. Concerning "public administration", in particular, the CAHAI notes that a legally binding transversal instrument should not address the plethora of specific administrative activities undertaken by public authorities, such as health care, education, social benefits etc, but be limited to general prescriptions about the responsible use of AI systems in public administration. Issues related to the various sectors of public administration may, as necessary, be addressed in appropriate sectoral instruments.~~

Commented [LR62]: GPD does not support this proposed focus of the instrument. To ensure that the instrument is future-proof (and thus does not focus on issues that may not be the most important in future years), we believe that the instrument should be fully general in nature and thus apply to all AI systems, rather than those developed, designed or applied for different purposes. In practice, the risk-based approach proposed by this document would allow for greater attention to be paid to higher-risk applications, regardless of sector.

XI Elements relating to supervisory authorities, compliance, and cooperation

Paragraph 43

The CAHAI considers that a legally binding transversal instrument should include provisions obliging Parties to take all necessary and appropriate measures to *ensure effective compliance* with the instrument, in particular through the *establishment of compliance mechanisms and standards*. Furthermore, provisions on the establishment or designation of *national supervisory authorities*, defining their powers, tasks and functioning as well as ensuring *their independence and impartiality* and the allocation of sufficient resources and staff, should be considered for inclusion. In addition, the legally binding transversal instrument should contain provisions regulating the *cooperation between Parties* to facilitate compliance with, and *mutual legal and other assistance, including exchange of data and other forms of information* under, the legally binding transversal instrument.

Commented [LR63]: GPD supports the inclusion of this language. We recognise that there has been some discussion over whether it should be included or not, and so wish to reiterate our support for the text as written.

International Bar Association

The International Bar Association (IBA) Working Group on AI & Human Rights welcomes the contributions made in this document and appreciates the recognition of the OECD guidelines and the UNGP framework, as well as the link made between due diligence and remedies. However, this proposal can reach further and more is needed in order to achieve effective regulation of AI compatible with human rights, democracy and law.

Mandatory Impact Assessments

Point 48 of the document identifies that *“Given the time and resources necessary to undertake such an assessment, and in order to safeguard the proportionality of a risk-based approach... extensive human rights, democracy, and rule of law impact assessment should only be mandated if there are clear and objective indications of relevant risks emanating from the application of an AI system.”* This is of particular concern as it is our view that human rights, democracy, and rule of law impact assessments (HUDERIA) should not be optional nor restricted in scope. In order to identify the risks that are assessed, a preliminary impact assessment is required for all risks, regardless of their perceived threat. From here, the risks that are identified as high-risk warrant further, and more detailed assessment. This should not be at the exclusion of a risk impact assessment of the lower risks. We recognise the UNGP allowance for prioritisation of high-risk applications for impact assessments, once all the risks have been mapped. However, the notion that impact assessment action only occurs with high-risk applications, and no impact assessment is required of low risk applications, is potentially non-compliant with the UNGPs or the ECHR.

Dynamic Risk

Point 19 suggests a risk categorisation of AI applications. We emphasise that this assessment should be contextual as, with the exception of a few (e.g. Autonomous weapons) most uses of AI cannot be identified as inherently bad or good for human rights. As this proposed framework focuses on high-risk applications, it is unclear how the initially perceived lower risk applications are guaranteed to continue to be fit for purpose. Risk is dynamic and the highly contextual and rapidly developing nature of AI highlights the importance of continuous risk assessment of all levels of risk to minimise the probability and severity

of human rights violations. The process for assessing risk impact does not need to be a highly complex procedure, and the UNGP and OECD principles provide a strong initial framework for such analysis.

Role of Private Sector

As far as the role of the private sector is concerned, particular attention should be paid to the role of human rights due diligence. To improve the due diligence process and the measures to be taken, the company needs to monitor the effectiveness of its due diligence policy and communicate the progress made with its stakeholders. In order to avoid excessive regulatory burdens on the private sector, the existing provisions mandating human rights due diligence should be accounted for and, in particular, at the upcoming EU Legislation on mandatory human rights and environmental due diligence.

CAHAI Secretariat

XIII Complementary elements relating to artificial intelligence in the public sector

Paragraph 54

As set out in Chapter VII, the development, design and application of AI systems in the public sector should be addressed in a legally binding transversal instrument, covering the most important transversal rights and obligations that should be respected in this domain. Additionally, the CAHAI is of the opinion that, given the context specificity of the risks posed by AI in the public sector in light of its specific role in society, such a transversal framework should be supplemented by additional legally binding or non-legally binding instruments at sectoral level.

Paragraph 55

These instruments could for instance elaborate further principles and requirements, specifically for the public services, regarding robustness, safety and cybersecurity, transparency, explainability, auditability and accountability ~~transparency, fairness, responsibility, accountability and redress~~ to ensure the responsible use of AI. The CAHAI recommends that ~~member States subject~~ the use and design, procurement, development and deployment of AI systems in the public sector is subject to adequate oversight mechanisms in order to safeguard compliance with human rights, democratic principles and the rule of law, and foster public trust, ~~in this context by rendering AI systems traceable and auditable.~~

Paragraph 56

Additionally, considering that the distinction between public and private sector involvement is often ambiguous, and considering the liability issues relating to the contracting out of public services to private actors any provisions applying to the design, development, and application of AI in the public sector should also apply to private actors that act on behalf of the public sector, cf. paragraph 35.

Paragraph 57

The CAHAI ~~sets out the following recommendations according to the different stages in the process of the adoption of an AI system by a public entity~~ considers that the following elements relating to the adoption of an AI system by a public entity could, in addition to those elements already described in Chapter VII, be addressed as part of a – legally, or non-legally binding instrument on AI in the public sector:

Paragraph 58

In the *design phase* of the system, ~~the CAHAI is of the opinion that a legally, or non-legally binding instrument could address how~~ due consideration ~~sh~~ould be given to the analysis of the problem which the public entity intends to solve, in order to assess whether an AI system is the best fit for the problem and, if so, which characteristics it should have. ~~A legally, or non-legally, binding instrument could furthermore address the following issues:~~ The data sets to be used for the AI system should be clearly identified, and the protection of such data and their origin respected. The design choices of the system should then be rendered explicit and documented. The intended users of the system, both civil servants and the public, as well as those potentially affected by the system should be involved early on, and their capabilities in using the AI system in question should be considered. An open and transparent design should be favoured. Finally, a human rights, democracy and rule of law impact assessment should be carried out to anticipate, prevent and mitigate potential risks. This also requires putting in place risk management and mitigation frameworks, which are relevant throughout all phases, cf. Chapter XII.

Paragraph 59

In the *procurement phase*, a thorough review of applicable legislation and policy measures in place should be conducted. Where necessary, public procurement processes should be adapted and public procurement guidelines for AI should be adopted, to ensure that procured AI systems comply with human rights, democracy and rule of law standards. A multidisciplinary and multi-stakeholder approach should be ~~respected-ensured~~ in order to involve various perspective and angles, including those of vulnerable groups. Because public entities are responsible for the systems they adopt and apply, careful attention should be paid to the potential impact on public accountability.

Paragraph 60

During the particularly sensitive phase of *development* of the system, documentation and logging processes should be meticulously kept to ensure transparency and traceability of the system. Adequate test and validation processes, as well as data governance mechanisms should be put in place. Amongst other risks, the potential risk of unequal access or treatment, various forms of bias and discrimination, as well as the impact on gender equality should be assessed.

Paragraph 61

Risk management and mitigation frameworks set up in previous phases should be evaluated, adapted and maintained during the *deployment* phase. Taking into account the nature of the risk, human involvement may need to be guaranteed in order to ensure appropriate oversight over the system. Where appropriate, the AI system should be initially and regularly audited by an independent actor, and the results rendered publicly available to foster public trust. To this end, ~~member States should the CAHAI considers that the~~ establishment of public registers listing AI systems used in the public sector, as well as essential information about the system such as, its purpose, actors involved in its development and deployment, basic information about the model, and performance metrics, where appropriate, and the result of a HUDERIA, should be addressed in the context of a legally binding, or non-legally binding, instrument on AI in the public sector. In addition, the aforesaid instrument could address the establishment of a feedback mechanism ~~should be put in place~~ in order to collect input on how to improve the system directly from its users and those potentially affected thereby. Further, the instrument could address the need for t~~he AI system should to~~ be subjected to regular evaluation and update, including by taking into account the ~~aforementioned~~ feedback. The evaluation process ~~sh~~ould hence be a periodic one. Transparency and communication towards users and citizens should likewise be addressed~~always be ensured, and they~~as should ~~the possibility of have~~ access to accountability and redress mechanisms. Last but not least, the ~~public should always have~~instrument should address the right of the public to be informed about the fact that they are interacting with an AI system rather than a human being, as well as ~~in such case, they should also be reserved~~ the right to interact with a human being rather than *only* an AI system ~~(and in particular when their~~ rights and interests of individuals or legal persons can be adversely impacted).

Paragraph 62

Finally, a legally binding, or non-legally binding, instrument on AI in the public sector could address measures ~~should be put in place~~ to increase digital literacy and skills among both civil servants and the general public, notably through investment in capacity building (initial and continuous training and education) of public officials and awareness raising about the benefits, risks, capabilities and limitations of AI systems. Such skills should encompass theoretical as well as practical knowledge on the interplay between the design, development and application of AI systems on the one hand, and human rights, democracy and the rule of law on the other hand. Furthermore, the aforesaid instrument could also address attention should also be given to the way in which these systems should be supervised and the risks arising therefrom should be managed.