

Strasbourg, 31 March 2021

CAHAI-LFG(2021)03  
Restricted

# **AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI)**

## **Legal Frameworks Group (CAHAI-LFG)**

---

### **Outcome from Sub-Working Group 2**

---

Prepared by CAHAI-LFG Sub-Working Group 2

## Table of Contents

Working method .....	3
Steps taken.....	3
Minutes internal SG2 meeting .....	4
Compilation of inputs.....	7
Protection of Human Dignity .....	7
Alternative 1.....	7
Alternative 2.....	8
Human autonomy (Right to opt-in/opt-out) .....	9
Human oversight.....	12
Alternative 1.....	12
Alternative 2.....	13
Right to know .....	14
Alternative 1: Right to be informed .....	14
Alternative 2: Right to know when interacting with an AI system .....	15
Prevention of harm / prejudice .....	15
Alternative 1: Prevention of harm – precautionary principle” .....	15
Alternative 2: Prevention of harm prejudices resulting from AI systems.....	17
Privacy and data protection.....	18
Preliminary Comment .....	18
Alternative 1.....	19
Alternative 2.....	19
Right to human contact.....	20
Right to autonomy and self-determination .....	20
Right to reasoning .....	22
Right to Digital Identity .....	23
Right to digital integrity .....	24
Right to the protection of attention .....	24
General Remarks.....	25

## Working method

### Steps taken

1. SG2 co-leads solicited first ideas/inputs from members; those were collected via Framapad and email in the period 10 – 24 March 2021. The following guidance was given:

#### **\*Preliminary remark**

Our starting point: CAHAI Feasibility Study, Chapter 7.1, points 1-2-3 (paras 95-104; related paras in other chapters where relevant). N.B. SG2 is also expected to cover aspects of privacy, self-determination, digital identity, where not yet covered by SG5 under "responsible data governance".

When adding your contribution, please keep in mind:

- \* stick to CAHAI-LFG's mission (focus on a possibly binding, horizontal instrument), hence, paragraphs/provisions should go in the direction of drafting legal provisions (though not yet amounting to the actual preparation of a convention)

- \* be to the point - input limited to max. 3 pages

- \* suggest text for: 1° concrete legal provisions; 2° a motivation for the provision (why do we propose this provision? Which impact of AI on human rights, democracy, rule of law (as relevant) motivates this provision? If possible, link with existing gaps as listed in Chapter 5); 3° clarification of the link with existing rights and/or case law.

#### **About the editing of text:**

- \* Every author can choose his/her colour code. This will allow us to keep track of contributions and amendments. You can choose this colour yourself by clicking on the dot next to your name.

- \* Since we are still in the phase of collecting everyone's inputs, please do not make significant amendments to contributions posted here, but rather present alternatives or include suggestions for change (under the original text or via the comments function), without immediately amending the original text itself. We will have the opportunity to discuss inconsistencies or important gaps during our LFG-SG2 meeting on Thursday 25th March. Thank you for your understanding!

2. A document containing a compilation of those inputs was circulated via email to all SG2 members the evening of Wednesday 24 March.
3. SG2 held an internal call meeting on Thursday 25 March 2021.
4. A revised version of the document, together with the minutes of the internal meeting, was circulated the evening of Thursday 25 March, inviting written feedback by members by Monday 29 March, noon.
5. One of the co-leads integrated all feedback received into this single document with the view of submitting it to the secretariat and LFG co-chairs on Monday 29 March 2021.
6. **Update: comments received from one SG2 member on Tuesday 30 March 2021 have still been integrated in the document with a view of circulation to LFG members ahead of the meeting on 8-9 April 2021.**

## Minutes internal SG2 meeting

Date: Thursday 25 March 2021, 4 - 5.15 pm, via MS Teams

Attendees (alphabetically by family name): Siiri Aulik (EE); Eleftherios Chelioudakis (Homo Digitalis); Michele Finck (LU); Abigél György (HU); Mario Hernandez (ES); Konstantin Kosorukov (RF); Jürg Lindenmann (CH); Alessandro Mantelero (T-PD); Catelijne Muller (ALLAI); Lesia Stangret (CA); Eva Thelisson (AI Transparency); Peggy Valcke (BE; co-lead SG2). (Excused: Kata Bencze (HU; co-lead SG2); Francesca Fannuci (CINGO))

The following attention points were raised:

- A decision will need to be taken at the LFG meeting in April about the style of formulating our input. On the one hand, CAHAI has not been tasked with the preparation (let alone, the negotiation) of a convention; on the other hand, CAHAI's final deliverable (while not being a legal document *as such*) should give the CM at least the option to consider provisions for a binding instrument. It was noted that we should not merely repeat the feasibility study, but go one step beyond. By way of compromise, it was suggested to formulate our task as follows: we can draft provisions that could be put into a convention, while not being a convention by themselves.
- Such provisions should be drafted carefully, following a legal-technical methodology, in order to be correct, accurate, feasible (especially since the work of CAHAI is closely followed by many parties). This implies a.o.:
  - o The need to be cautious about introducing 'new rights', while we may rather be interpreting existing rights in the AI context;(\*)

*(\*) Comment made by ALLAI: The need to be cautious about introducing 'new rights', while we may rather be interpreting existing rights in the AI context while at the same time dealing with the situation where the existing human rights framework does not adequately protect us from invasive, socially disrupting, harmful or otherwise adverse effects of the use of AI. In other words: **a dual approach**:*

*1. The first element of such a dual approach deals with the interlinkage between AI and existing human rights. Many AI developers, deployers and users (public and private) seem to be unaware of the (potential) impacts of AI on human rights, democracy and the rule of law. As a first step, an iteration or (re)articulation exercise in which existing Human Rights of the ECHR are 'translated' to an AI context;*

*2. The second element of the dual approach deals with the gaps and limitations of existing human rights (including measures for compliance, accountability and redress) and could consist of several steps:*

*a) Adjusting (or tweaking) existing human rights to ensure their protection in light of AI. Think of:*

*\* An extended/clarified right to physical, psychological and moral integrity in light of AI-profiling and biometric recognition*

*\* A strengthened right to privacy to protect against AI-driven mass surveillance*

*\* Adapting the right to data privacy to protect against indiscriminate, society-wide online tracking of individuals, using personal and non-personal data (which often serves as a proxy for personal identification)*

*\* Appropriate measures for compliance, accountability and redress in light of harm or adverse effects by AI*

*b) Proposing new human rights in light of AI where they do not exist but are deemed necessary such as:*

*\* A right to human autonomy, agency and oversight over impactful AI*

*\* A right to transparency/explainability of AI outcomes (where they have a serious impact), including the right to an explanation of how the AI functions, what logic it follows, and how its use affects the interests of the individual concerned, even if the AI-system does not process personal data, in which case there is already a right to such information under GDPR.<sup>1</sup>*

*c) Drawing Red Lines: Due to the invasiveness of some AI-applications or uses, there might be situations in which our current framework of human rights, democracy and the rule of law fails to adequately or timely protect us and where we might need to pause for reflection and find the appropriate answer to what one could consider "question zero": Do we want to allow this particular AI-system or use and if so, under what conditions? Answering this question should force us to look at the AI-system or use from all perspectives, which could result in several 'solutions':*

*\* A particular AI-system or use is put under a moratorium, (temporarily or indefinitely) banned or put under restrictions (for example in terms of scope, term or area of use, under conditions of necessity and proportionality)*

*\* A particular AI-system or use is made subject to a specific democratic oversight-mechanism*

<sup>1</sup> Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence.

*\* Private owners of powerful AI-systems are obliged to align their AI development and governance structures with the interests of those affected by the system and society at large, which could include measures to involve relevant parties (such as workers, consumers, clients, citizens, policy makers)*

*(\*) Comment made by Canada: The rights proposed in this contribution remain a starting point for the discussion on what may be needed in a new international legal instrument along these lines. As the next step, each of them will be examined further within the LFG to determine whether they may be interpreted as falling within the scope of the existing international human rights framework, consisting of human rights instruments within the Council of Europe as well as more broadly applicable human rights instruments, including those in the UN.*

*Where these proposals for AI-specific rights may be linked to existing human rights or existing obligations on states in the Council of Europe or more broadly, consideration will be given to how best to clearly identify this link and how these proposals may be framed within the context of existing rights and obligations. This may include, for example, ensuring appropriate links in the preamble or general principles, as applicable, or ensuring that there is a clear link in any provisions where these contextualizations are being proposed.*

*Where the proposed rights go beyond existing rights or establish a different legal standard or threshold specific to AI, consideration will be given to whether these objectives may be achieved through other means short of the creation of new rights, such as measures that may be imposed by states (whether through legislation, regulation, administrative policy tools or otherwise). Consideration will also be given to the potential implications of establishing different international legal standards specific to AI from an international law perspective.*

*Proposals for new rights set out here should be considered together with any recommendations for new rights provided by the other subgroups of the LFG to enable the LFG to make appropriate links between them and to ensure a consistent approach."*

- The need to be very thorough, in terms of motivating the provisions and grounding them in existing international legal rules and ECtHR case law; while it can be useful to refer to recommendations and declarations, those cannot as such create legal rights or obligations. At the same time, we should make good use of existing materials that have been produced by the CoE, EU, other organisations, and that can provide valuable inspiration.
- The need to choose appropriate and precise wording when formulating rights and obligations; provisions which remain vague are to be avoided. *At the same time, one needs to be cautious of over- or under including certain AI-systems and uses due to too precise wording, leaving no room for interpretation by the Court. A careful balance should be sought. [added by ALLAI]*
- *The need for CAHAI's second deliverable to be revised by Council of Europe professionals before submission to the CAHAI for approval to ensure that the formulation is suitable for a policy document that is recommending the inclusion of potential elements for a potential legal framework [added by EE].*
- The need to stick to legal concepts, rather than philosophical notions.
- *The need to bear in mind that the CAHAI is tasked with working towards a potential legal framework that will ensure that the design, development and use of AI complies with the Council of Europe's standards on human rights, democracy and the rule of law [added by EE].*
- The need(\*) to keep in mind that, while Council of Europe member States will be the ones potentially negotiating a future convention on AI, the text should be mindful of other States who may want to join those discussions (as some have indeed already joined the CAHAI).

*(\*) Comment made by EE: This was said in the context of human dignity. We strongly disagree with this. The role of human dignity as the basis for all human rights and freedoms is central the European understanding of human rights. CAHAI must stay within its mandate which is to work towards a potential legal framework that will ensure that the design, deployment and use of AI will comply with the Council of Europe's standards on human rights, democracy and the rule of law.*

*It should also be noted that human dignity also underlies the UN Declaration on Human Rights and related conventions – so if some countries do not adhere to it, they do not adhere to international human rights standards. There is no justification for making any concessions to them.*

*The Council of Europe (as well as the EU) are striving to provide a European model for AI, which hopefully will be exported and become global. While we agree that the framework should allow for the broadest possible use and acceptance, this cannot be at the expense of our understanding of human rights, democracy and the rule of law.*

- Overlaps with input from other SGs are to be expected and will have to be tackled during the LFG meeting in April.
- We may also have to revise our working method, since drafting in parallel, with several SGs, may not be the most efficient strategy.
- The work of the CAHA in general, and LFG in particular, is very ambitious given the limited timeframe. We may have to prioritise certain aspects in order to deliver a good quality end product.

A number of comments were formulated on specific parts of the compiled input, but it was agreed to give all SG2 members the opportunity to provide detailed feedback in writing in the coming days. It was stressed that the intention of this first phase was merely to collect a variety of ideas, and that there will in any case be an occasion for further discussing, re-ordering, complementing, finetuning the proposed provisions in the next steps. In order to inform those discussions to a maximum extent, it was decided not to condense the current input in order to make it fit within 3 pages (as initially requested), but to keep all proposals, suggested alternatives and modifications in the document, at least for the time being.

It was agreed that Monday 29 March, noon, will be the cut-off date to send feedback on the document (using track changes or comments fields) via email to the SG2 co-leads, who will then compile everything in one document and submit it to the secretariat on 29 March, COB.

# Compilation of inputs

## Protection of Human Dignity

### Alternative 1<sup>2</sup>

1. ~~Member states should~~ Each Party shall provide to<sup>3</sup> ensure that the design, development and use of AI systems respect<sup>4</sup> the dignity of the human beings interacting therewith or impacted thereby. ~~AI systems can<sup>5</sup> only be deployed under the condition that individuals are treated as moral subjects and that their human agency is not weakened or undermined as a result.~~<sup>6,7</sup>

*Alternative<sup>8</sup>: (1) Parties will provide that AI systems [should][must] be designed, developed and used in a manner that is respectful of the dignity of persons interacting with or affected by such systems; OR (2) Parties will seek to ensure that AI Actors design, develop and use AI systems in a manner that is respectful of the dignity of persons interacting with or affected by such systems.*

2. AI applications can be used to foster human dignity and empower individuals, yet their use can also challenge it and (un)intentionally run counter to it.<sup>9,10,11</sup> The allocation of certain tasks may need to be reserved for humans rather than machines given their potential impact on human dignity. More generally, AI

<sup>2</sup> Original text suggested by CINGO.

<sup>3</sup> Amend. (T-PD)

<sup>4</sup> Reformulate – design cannot respect... (EE)

<sup>5</sup> ? who is this addressed to. They must ensure that ... Human agency weakened – at all? Is this too categorical? Unduly? (EE)

<sup>6</sup> The second sentence doesn't seem fitting for a binding legal instrument. It lacks clarity and legal certainty. Proposal to delete. (RF)

<sup>7</sup> Delete. (T-PD)

<sup>8</sup> Suggested by CA. Motivation: the opening words of this proposal appear to meld together two different approaches, resulting in a lack of clarity on what is intended. If it is meant to say “Each Party shall provide to...”, who is each Party providing these things to? If it is meant to say “Each Party shall provide that” or “Each Party shall ensure that”, it should refer to one or the other.

With respect to the standard, is this the type of provision where it is feasible to impose an expectation of result rather than an expectation of intent? Can Parties, in other words, really “ensure” in each and every case that the design, development and use of AI systems will be conducted in a manner that would be consistent with the dignity of the persons interacting with or affected by them? It may be one thing for Parties to adopt measures requiring certain levels of treatment or prohibiting certain types of conduct that they would consider to be inconsistent with human dignity. However, with a legal standard like “ensure”, even the fact that just one AI system is designed, developed or used in a way that would be inconsistent with human dignity would seem to place the Party in non-compliance, notwithstanding the fact that this conduct would be inconsistent with whatever measure the Party sought to impose. This would result in a scenario where a Party could only ever meet its obligation under this proposal by actively policing the design, development and use of AI systems, in every scenario, across both the public and private sectors, which is not feasible in many respects.

As to use of the term “respect” in this context, the entities designing, developing or using the AI systems can be said to respect human dignity. Alternatively, we might say that AI systems should be designed, developed or used “in a manner that is respectful of” or “in a manner consistent with” human dignity.

<sup>9</sup> These are not provisions. The first sentence is a statement. Also – may, should ... (EE)

<sup>10</sup> Same issue with this paragraph. Suggestion to reformulate the article in a clearer, more precise manner, focusing on the key provision (respect of human dignity). The rest, which is in essence an explanation of reasoning, may be added to an explanatory report. (RF)

<sup>11</sup> This sentence, on its own, reads more like the kind of commentary that we would see in a policy discussion. It is also not clear what is meant by the use of the parenthesis in “(un)intentionally”? It would not be consistent with legal drafting practice, whether in a convention or otherwise, to place parenthesis around the prefix of a word. That said, it sounds like what is intended here is “intentionally or unintentionally” or “whether intentionally or not”.

**Recommendation:** (1) Replace “(un)intentionally” with whichever term or phrase this is intended to reflect. (2) To the extent it remains helpful to acknowledge somewhere in this instrument that AI applications may end up being used to both foster human dignity and empower individuals but also to potentially harm them, consider reframing this first statement another way (e.g. “Recognizing that AI applications can be used to foster human dignity and empower individuals ...”) and moving it somewhere else. A statement like this, for example, may be pertinent to the general principles or perhaps even the preamble that are being developed by subgroup 1.

systems should be developed and used in a way that secures<sup>12</sup> and promotes the physical and mental integrity of human beings. [DELETE PARAGRAPH]<sup>13</sup>

*Alternative<sup>14</sup>: Parties may consider whether the use of AI systems to perform certain tasks that could affect the dignity of the persons interacting with or affected by them should be subject to safeguards, which may include: (a) ensuring that certain tasks in these contexts are reserved for persons or subject to human oversight; and (b) ensuring that decisions made through the use of AI systems in these contexts may be subject to review by a person.*

### **Rationale**

Although not explicitly mentioned in the European Convention of Human Rights (ECHR), the European Court of Human Rights declares that “the very essence [of the fundamental objectives of the Convention] is respect for human dignity and human freedom.” [ECtHR, S.W. v United Kingdom, 1995, para 44; ECtHR, Goodwin v. UK, 2002, para 90]. Protocol 13 to the ECHR (Abolition of the Death Penalty) also acknowledges “the full recognition of the inherent dignity of all human beings.”<sup>15</sup> The UN Universal Declaration of Human Rights (Preamble, Articles 1, 22-23) explicitly includes human dignity as a precondition of all other human rights and the Charter of Fundamental Rights of the EU also explicitly recognises the “right to human dignity”<sup>16</sup> (Article ).

Also see **CAHAI (2020)23 Feasibility Study, Para 104**: “In general, AI systems should not be used to subordinate, coerce, deceive, manipulate or condition humans, but rather to complement and augment their capabilities.

### *Alternative 2<sup>17</sup>*

[Article X -] Human Dignity

[Parties / Member States] should ensure that the design, development and use of AI systems respect the dignity of the human beings interacting therewith or impacted thereby [, and should aim for AI systems to be developed and used in a way that does not harm the physical and mental integrity of human beings].

*Explanatory report: In accordance with the Preamble of the Universal Declaration of Human Rights, “recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world”. AI applications can be used to foster human dignity and empower individuals, yet their use can also challenge it and (un)intentionally run counter to it. In general, AI systems should be deployed under the condition that individuals are treated as moral subjects and that their human agency is not weakened or undermined as a result. The allocation of certain tasks may need to be reserved for humans rather than machines given their potential impact on human dignity. In principle, AI systems should be developed and used in a way that secures and promotes the physical and mental integrity of human beings. ...*

---

<sup>12</sup> Secure – is not normally used in such texts. What is meant by “secure”? (EE)

<sup>13</sup> This paragraph raises interesting and ambitious questions, but these should be better framed. For the moment, it seems best to leave it out. (T-PD)

<sup>14</sup> Suggested by CA. Motivation: It is not clear why this proposal uses the term “machines”, which appears to be a colloquial way to refer to systems that use AI. The terminology used should be consistent, and we should refrain from using informal language such as “machines” or “robots” (which we also sometimes see) to refer to AI.

It is also not clear how an AI system would “secure and promote the physical and mental integrity of human beings”. Imposing legal standards that are too general or imprecise would inject uncertainty into this instrument.

Recommendation: Clarify “machine” and what it would mean for AI systems themselves to “secure and promote the physical and mental integrity of human beings” or use alternative.

<sup>15</sup> It should be stressed, however, that the Court never refers to human dignity as a particular “right” under the Convention, rather using it as a tool when determining violation of such rights (see, for instance, <https://www.echrblog.com/2016/10/the-role-of-human-dignity-in-echr-case.html>). (RF)

<sup>16</sup> In fact the EU Charter does not explicitly recognise the “right to human dignity”. The Charter sets out “rights, freedoms and principles”, and human dignity is never identified as a “right” - it is only said in Article 1 that human dignity is “inviolable”, “must be respected and protected”. (RF)

<sup>17</sup> Suggested by the RF.



Human beings should ~~in principle~~<sup>22</sup> ~~be able~~<sup>23</sup> remain free to make decisions for themselves and remain autonomous, which includes the right to choose not to interact with an AI system, and to not be subject to a decision solely<sup>24</sup> informed or made by an AI system whenever this can significantly impact their lives, the lives of others or society,<sup>25</sup> especially when this can violate rights related to their<sup>26</sup> human dignity.<sup>27</sup>

*Alternative:*<sup>28</sup>

*Every individual shall have the right not to be subject to interaction with an AI system, except in cases of legitimate use in line with the law and when a negative impact on human dignity, human rights and fundamental freedoms is excluded. Clear instructions should be provided on how individuals can exercise this right and on which alternative non-AI driven methods are available.*

---

<sup>18</sup> Delete. (ALLAI)

<sup>19</sup> Delete. (T-PD)

<sup>20</sup> For these and other proposals that are framed as “rights”, it often remains unclear what exactly they are seeking to achieve. A lack of clarity on whether the right that is being proposed falls under an existing right, on the one hand, or constitutes a new right specific to AI, on the other, risks introducing more certainty rather than less and undermining the application of existing law. For this reason, we recommend that wherever a proposal appears to be contemplating the creation of new rights specific to AI, we determine first whether such rights may be interpreted in the context of any existing rights or obligations on states, whether through relevant instruments in the Council of Europe or in more broadly applicable instruments, such as some of the ones listed above. If so, we should consider how to clearly link the interpretation or contextualization that is being proposed to the relevant existing right or obligation to avoid confusion on what is intended.

If an assessment of such proposed rights demonstrates, clearly, that rights and obligations in existing instruments, in the Council of Europe or elsewhere, are insufficient or if the proposal seeks to impose a different standard than what presently exists, we should consider first whether the objective may be achieved through other means short of the creation of new AI-specific rights and what the implications would be of establishing different standards for AI from an international law perspective. Alternatives to new rights include reframing the right as an “ability” to do something or as a measure that may be imposed by Parties (whether through legislative, regulatory or other means).

This sort of analysis requires a thorough and systematic approach and an ability to review the proposed rights across the different subgroups collectively to identify links and ensure consistency in how such rights are assessed. (CA)

<sup>21</sup> Original text suggested by CINGO.

<sup>22</sup> Delete ‘in principle’ (EE).

<sup>23</sup> Amend (ALLAI).

<sup>24</sup> Add ‘solely’ (EE).

<sup>25</sup> Add (ALLAI).

<sup>26</sup> Delete (ALLAI).

<sup>27</sup> This needs to be carefully reworded. We disagree that there should be a general right to choose whether to interact with any and all AI systems. It is important to add “SOLELY” before “informed or made” – the right to human review makes better sense. (EE)

<sup>28</sup> Suggested by T-PD.

In rare exceptions<sup>29,30</sup> where “opt-in” is not feasible, AI systems should at the very least allow users and subjects to opt-out of the system.<sup>31</sup> Clear instructions should be provided on how individuals can opt-out from the system’s use and on which alternative non-AI driven methods are available.<sup>32,33</sup>

*Alternative for second paragraph:*<sup>34</sup>

Human beings should be able to make informed autonomous decisions regarding AI systems.<sup>35</sup> They should be given the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree and, where possible, be enabled to reasonably self-assess or challenge the system.<sup>36</sup> AI systems should support individuals in making better, more informed choices in accordance with their goals.<sup>37</sup>

*Add:*<sup>38</sup>

Every individual shall have the right to object at any time, on grounds relating to his or her situation, to the use of an AI system affecting him or her unless there is not a legitimate ground for the use of such systems which override his or her interests or rights and fundamental freedoms. Clear instructions should be provided on how individuals can exercise their right to object.<sup>39</sup>

---

<sup>29</sup> ? (EE).

<sup>30</sup> “In rare exceptions” provides little clarity. It also does not indicate who would be making this determination. Is this a determination that would be left to each Party in implementing its obligations under this instrument or is there an expectation that the Parties will arrive at a common mutual understanding of what constitutes “in rare exceptions”? Either way, this principle on opting-in or opting-out may not be feasible in all cases. Framing the choice this way also seems to lose sight of the possibility that where it may not be possible to opt in or out, the Parties could implement appropriate safeguards, such as providing the means for a person to review a decision. (CA)

<sup>31</sup> This is not always feasible not desirable. This will depend on the type of system. This needs to be reworded. (EE)

<sup>32</sup> The proposed rule seems too wide. The relevant documents (CM Rec(2020)1, T-PD guidelines etc.) only talk about situations of “higher risk for human rights” or “freedom of decision-makers not to rely on AI”, rather than a general right for everybody to refuse interaction with any AI system or to reject any decision that was “informed” by an AI system. A more conservative approach is suggested. Proposal: perhaps a conjunction with *Human Oversight* would be a solution. (RF)

<sup>33</sup> I wonder whether this should be proportionate to the risk. If we assume that the word auto-correct is a form of an AI system then it seems acceptable that there are AI system with no opt-in opt-out. This will of course also depend on how “AI systems” are defined by CAHAI. (LU)

<sup>34</sup> Suggested by ALLAI. Rationale: I would not refer to human autonomy solely as a right to opt-in or opt-out, but rather as a more general right to remain autonomous vis-a-vis AI, which in certain circumstances could translate into a right to opt-in or - out, but in others could translate into a right to challenge an AI driven decision.

AI systems can sometimes be deployed to shape and influence human behaviour through mechanisms that may be difficult to detect, since they may harness sub conscious processes, including various forms of unfair manipulation, deception, herding and conditioning, all of which may threaten individual autonomy. The overall principle of user autonomy must be central to the system’s functionality. Key to this is the right not to be subject to a decision based solely on automated processing when this has significant or legal effects.

Secondly, human autonomy does not only relate to people affected by AI, but also to people that ‘operate’ AI but want to deviate from the results, because of negative effects of it on others or society in general.

<sup>35</sup> This first sentence remains imprecise. What sorts of decisions should individuals be entitled to make “regarding AI systems”? This looks like it’s referring to the decision to interact or not interact with an AI system. If this is the case, this context cannot be implied by something as general as “to make informed autonomous decisions regarding AI systems”. (CA)

<sup>36</sup> This language too is rather vague. What does it mean to “reasonably self-assess or challenge the system”? (CA)

<sup>37</sup> This reference to making choices “in accordance with their goals” may be a reasonable objective in certain respects, but it is framed too generally here. It is not clear how this would be applicable in all contexts involving AI. (CA)

<sup>38</sup> Suggested by T-PD.

<sup>39</sup> As with the draft proposed above, this appears much too broad. While there is a balancing proposed here, it treats “rights and fundamental freedoms” on par with an individual’s “interests”. It also does not make it clear that the balancing that would take place on whether there is a legitimate interest from a public policy or other perspective is one that would be left to the determination of the Party, not a general standard that would be implied for all Parties from this very broad formulation of the text. (CA)

## **Rationale**<sup>40</sup>

The right to human dignity<sup>41</sup>, the right to life (Art. 2 ECHR), and the right to physical and mental integrity, among other human rights listed in the ECHR and UDHR, ICCPR, ICESCR.<sup>42,43</sup>

The right to have the opportunity, when it is not excluded by competing legitimate<sup>44</sup> overriding grounds<sup>45</sup>, to choose to have contact with a human being rather than a robot.<sup>46,47</sup>

Also see:

**Council of Europe, Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impact algorithmic systems, Para A 11** (*"In some cases, the application of an algorithmic system can prompt a particular, higher risk to human rights, for instance because it is used by States for their public service or public policy delivery and the individual does not have a possibility to opt out or suffers negative consequences as a result of the decision to opt out."*<sup>48</sup>) and **Council of Europe, T-PD (2019)01 Guidelines on Artificial Intelligence and Data Protection, Para 4** (*"Overreliance on the solutions provided by AI applications and fears of challenging decisions suggested by AI applications risk altering the autonomy of human intervention in decision-making processes. The role of human intervention in decision-making processes and the freedom of human decision makers not to rely on the result of the recommendations provided using AI should therefore be preserved."*)

**CAHAI (2020)23 Feasibility Study, Para 104** (Key Substantive Rights): *"The right to effectively contest and challenge decisions informed and/or made by an AI system and demand that such decision be reviewed by a person (right to opt out)."*<sup>49,50</sup>

---

<sup>40</sup> As suggested above, the rationale should be clear, as should the proposed text itself, on which rights are being implied from existing rights and which are being proposed as potential new rights. The explanation here identifies relevant rights but it does not actually say what the intent here is. It remains unclear, for example, whether the proposed right to have an opportunity to have contact with a human is an interpretation of human dignity or the rights cited in the ECHR and these UN instruments or whether it is meant to be a novel right created specifically and only in relation to AI. (CA)

<sup>41</sup> The "right to human dignity" is not listed in ECHR, UDHR, ICCPR, or ICESCR. We should be precise in our legal wording. (RF)

<sup>42</sup> Why specifically mention the right to life? Does it have any special connection to human autonomy? (RF)

<sup>43</sup> While these references to UN instruments are welcome in the rationale, we would equally welcome a reference to these instruments in the preamble and, if applicable, in the text itself. (CA)

<sup>44</sup> How is legitimate defined? (EE)

<sup>45</sup> What is meant by this? This is not Council of Europe standards language. (EE)

<sup>46</sup> Depending on the circumstances. These are too general and categorical. Compare the language in the cited CM/Rec (2020)1 – "in SOME CASES". Also in T-PD (2019)01 – "OVERRELIANCE ...". Please note they are speaking of human decision makers, not subjects of decisions. (EE)

<sup>47</sup> "Robot" is being used here in a colloquial sense, just like the reference to "machines" above. While these may be appropriate for informal or policy discussions, there should be greater consistency in terminology when explaining the intentions of a legal text. The proposed provisions do not refer to "robots". They refer to AI and AI systems. We would suggest using similar language in the explanations. (CA)

<sup>48</sup> This does not imply that the individual should have the *right* to opt in/out; rather, that whenever the individual cannot choose not to interact with an AI system, and there is a "particular, higher risk to human rights", appropriate safeguards should be put in place. (RF)

<sup>49</sup> The right to have a human review a decision is *not* the right to opt out. This is a good example of the problems which arise in documents such as the feasibility study which are not revised for accuracy. (EE)

<sup>50</sup> This citation seems more suitable for the section on *Human Oversight*. It does not refer to a general "right to have contact with a human rather than a robot". (RF)

## Human oversight<sup>51</sup>

### Alternative 1

1. ~~Each Party shall provide to~~<sup>52</sup> ~~Member States should~~<sup>53</sup> require AI developers and deployers to establish human oversight mechanisms that safeguard human autonomy, in a manner that is tailored to the specific risks arising from the context in which the AI system is developed and used.<sup>54,55</sup>
2. An adequate level of human involvement should be ensured in the operation of AI systems, based on a contextual risk assessment taking into account the system's impact of AI systems<sup>56</sup> on human rights, democracy and the rule of law.<sup>57,58</sup>
3. The extent and frequency of oversight should be tailored to the specific AI application context and the autonomy of such human interventions<sup>59</sup> should be preserved.<sup>60,61,62</sup>

### Rationale

See **Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impact algorithmic systems<sup>63</sup>**, para B.3 (Para B 5.3 (Expertise and oversight): “States should ensure that all human rights impact assessments related to high-risk algorithmic systems are submitted for independent expert review and inspection. Tiered processes should be identified or created where necessary for independent oversight. Human rights impact assessments conducted by or for States should be publicly accessible, have adequate expert input, and be effectively followed up.”)

Human oversight helps ensuring that an AI system does not undermine human autonomy or causes other adverse effects. Oversight may be achieved through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach. HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system. Moreover, it must be ensured that public enforcers have the ability to exercise oversight in line with their mandate. Oversight mechanisms

<sup>51</sup> Original text suggested by CINGO.

<sup>52</sup> Amend. (T-PD)

<sup>53</sup> ? recommendation (EE)

<sup>54</sup> This is far too vague – impossible to implement and control. (EE)

<sup>55</sup> This proposal is unclear in several respects. Is it meant to say “Each Party shall provide that AI developers ...” or “Each Party shall require AI developers ..”? Who is contemplated by “provide to”? What does it mean, too, to “safeguard human autonomy”? Is this referring to the autonomy of individuals interacting with the AI systems, for example, or the autonomy of the decision-makers seeking to use the AI? From the paragraphs that follow, it appears that the intent here is to capture the latter, but this would not be implied from the phrase “that safeguard human autonomy” on its own.

Recommendation: Clarify the meaning on both fronts. If “safeguard human autonomy” cannot be clarified, it can be removed. The provision would be understandable without this phrase. (CA)

<sup>56</sup> Amend. (T-PD)

<sup>57</sup> Isn't this basically the same idea that is behind para.1? Is there a difference between involvement and oversight? (RF)

<sup>58</sup> While it is helpful to make a link here to risk assessments, “adequate level of human involvement” appears rather vague. As with other text proposed above, it would be helpful to specify whether this is a determination that would be left to the Parties, whether it is something that would be left to the AI Actors, or whether there is an expectation that the Parties will arrive at a common understanding of what constitutes an “adequate level of human involvement” in certain contexts, depending on the risk. (CA)

<sup>59</sup> What is meant by “autonomy of such human interventions”? (EE)

<sup>60</sup> Para.3 seems to be a repetition of the last part of para.1. Perhaps conflating them would be appropriate. (RF)

<sup>61</sup> I personally find this formulation a little bit confusing. Does this mean that human interveners should be able to do the “meaningful intervention” that the EDPB has for instance referred to in respect of Art 22 GDPR? Or does it mean that they should be independent from the developers/deployers of AI in a structural/employment sense? (LU)

<sup>62</sup> Key elements of this proposal remain unclear. What is the phrase “the autonomy of such human interventions should be preserved” intended to mean? (CA)

<sup>63</sup> Has there been any consideration given to using the same wording as is already used in the recommendations? (EE)

can be required in varying degrees to support other safety and control measures, depending on the AI system's application area and potential risk. All other things being equal, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required.<sup>64</sup>

#### Alternative 2<sup>65</sup>

##### [Article X+1 - ]Human Autonomy and Oversight

Whenever there is a [high] risk of AI systems causing a violation of human rights and fundamental freedoms, or unjust harm, [Parties shall / Member States should] require AI [developers and deployers / operators<sup>66</sup>] to establish human oversight mechanisms that safeguard human autonomy, in a manner that is tailored to the specific risks arising from the context in which the AI system is developed and used.<sup>67</sup>

*Explanatory report: This provision is inspired by Council of Europe, Recommendation CM/Rec (2020)1 of the Committee of Ministers to member States on the human rights impact algorithmic systems, in particular its paragraph A.11 ("In some cases, the application of an algorithmic system can prompt a particular, higher risk to human rights, for instance because it is used by States for their public service or public policy delivery and the individual does not have a possibility to opt out or suffers negative consequences as a result of the decision to opt out. A similarly heightened risk ensues as a result of use in the context of decision-making processes, by either public authorities or private parties, in situations that carry particular weight or legal consequence.") and paragraph B.3 ("States should ensure that all human rights impact assessments related to high-risk algorithmic systems are submitted for independent expert review and inspection. Tiered processes should be identified or created where necessary for independent oversight. Human rights impact assessments conducted by or for States should be publicly accessible, have adequate expert input, and be effectively followed up."), as well as the T-PD (2019)01 Guidelines on Artificial Intelligence and Data Protection, Para 4 ("Overreliance on the solutions provided by AI applications and fears of challenging decisions suggested by AI applications risk altering the autonomy of human intervention in decision-making processes. The role of human intervention in decision-making processes and the freedom of human decision makers not to rely on the result of the recommendations provided using AI should therefore be preserved.")*

---

<sup>64</sup> Suggested by ALLAI.

<sup>65</sup> Suggested by RF.

<sup>66</sup> AI operator: Legal or physical person certified as the operator of the AI system by a national AI certification authority.

<sup>67</sup> This alternative would be clearer if it specified that the determination of where there would be such a risk of harm is one that would be left to the Parties. For example: "Where Parties determine that there is a [high] risk of an AI system causing..., they should..."). The phrase "safeguard human autonomy", however, remains vague, as it does for Alternative 1. (CA)

## Right to know

### Alternative 1: Right to be informed<sup>68,69</sup>

~~Everyone~~ Every individual<sup>70</sup> has the right to be informed of the fact that they are interacting with an AI system rather than with a human being, in particular when the risk of confusion arises and can affect human dignity and other human fundamental<sup>71</sup> rights.<sup>72</sup>

Everyone has the right ~~not to be subject to interaction with an AI system except in cases of legitimate use as provided by law and when an adverse~~<sup>73</sup> impact on human dignity, human rights and fundamental freedoms is excluded.<sup>75,76,77</sup>

=> Obligations for Parties / Member States:

Each Party shall adopt such legislative and other measures as may be necessary to ensure that, where tasks risk violating human dignity<sup>78</sup> if carried out by AI systems rather than human beings, these tasks are reserved for humans.<sup>79,80,81,82</sup>

Each Party shall adopt such legislative and other measures as may be necessary to ensure that AI developers [N.B. NEEDS DEFINITION?] inform human beings of the fact that they are interacting with an AI system rather than with a human being whenever confusion may arise.<sup>83</sup>

Link with existing instruments:

Such right has also been recommended by the Council of Europe Guidelines on AI and Data Protection, <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8><sup>84</sup>

---

<sup>68</sup> Original text suggested by BE, as modified by T-PD.

<sup>69</sup> We have the same comments here as we do for the rights cited above in terms of the need for clarity on what is intended (i.e. whether this is derived from an existing right or obligation or intended to be a new right applicable only in the context of AI). (CA)

<sup>70</sup> Amend. (T-PD)

<sup>71</sup> Replace 'fundamental' by 'human' (EE).

<sup>72</sup> Delete. (T-PD)

<sup>73</sup> What is a legitimate use? How is this determined? And do you mean also in the private sector? On amazon.com? Facebook? This sweeping provision ultimately makes little practical sense. Are we talking about public services? Private companies?? (EE)

<sup>74</sup> Not every adverse impact should be excluded – again, this wording is too categorical. (EE)

<sup>75</sup> This is not in the right place, but rather fits under “human autonomy” (BE).

<sup>76</sup> Coming back to my reference to Word autocorrect above I wonder whether there should again be a reference to the risk of the system here. (LU)

<sup>77</sup> Delete. (T-PD)

<sup>78</sup> We should talk about risk to human rights. (RF)

<sup>79</sup> Delete. I miss the point. If a task puts human rights at risk, it is a problem whether it is carried out by humans or by AI systems designed and used by humans. (T-PD)

<sup>80</sup> I agree with Alessandro. This is going too deep and could be difficult to implement in practice. And wouldn't this issue already be resolved under Human Oversight? (RF)

<sup>81</sup> I second this. This reads as if it is alright for a humans to violate human dignity but not AI. (LU)

<sup>82</sup> Agree that this sentence needs reformulation. ... to ensure that ... AI systems do not violate human dignity. Otherwise you are asking for a ban on the use of AI systems for the performance of certain tasks. Based on what? And which tasks in which sectors under which circumstances? (EE)

<sup>83</sup> Delete. (T-PD)

<sup>84</sup> 1) These Guidelines only refer to “data subjects”, not all persons. The difference is data subjects are already endowed with a “right to know” by virtue of Convention 108/108+. This will of course automatically translate to AI interactions involving personal data, and we already have a section on Privacy and Data Protection.

2). The Guidelines do not per se “recommend a right” (i.e. a binding legal rule), but only recommend that the data subjects be informed when interacting with AI.

Opinion: we should look for more substantial basis for this provision.



## Alternative 2: Right to know when interacting with an AI system<sup>85</sup>

~~Member States should~~<sup>86</sup> ~~require~~ **Each Party shall provide that**<sup>87</sup> AI deployers ~~to~~ inform human beings<sup>88</sup> of the fact that they are interacting with an AI system rather than with a human being whenever confusion may arise.<sup>89</sup>

### Rationale

To safeguard human dignity, it is essential that human beings are aware of the fact that they are interacting with an AI system and are not misled in this regard. While not a right per se,<sup>90</sup> this has been recommended by the Council of Europe Guidelines on AI and Data Protection (<https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>)

-----

## Prevention of harm / prejudice<sup>91</sup>

### Alternative 1: Prevention of harm – precautionary principle<sup>92,93,94</sup>

1. **AI systems must not prejudice human rights and fundamental freedoms and negatively affect natural environment.**<sup>95,96</sup>
2. When implementing measures to prevent harm ~~of~~ **caused by**<sup>97</sup> AI systems, ~~member states~~ each Party<sup>98</sup> should<sup>99</sup> pursue a risk-based approach assessing the level of risk to human rights, rule of law and democracy.
3. An inclusive/ participatory (~~emphasis on~~ **including**<sup>100</sup> ~~inclusion of~~ marginalised groups)<sup>101</sup>, transparent, and continuous human rights impact assessment should be conducted to identify any adverse impacts on human rights, democracy and rule of law, and implement measures to prevent or mitigate these risks.<sup>102</sup>

<sup>85</sup> Original text suggested by CINGO.

<sup>86</sup> ? recommendation. (EE)

<sup>87</sup> Amend. (T-PD)

<sup>88</sup> This is a question of style. Do the users have the right or the deployers have an obligation – or both? Are both needed? (EE)

<sup>89</sup> While this corresponds to the proposed right set out above, the standard is different. Here, there is a reference to confusion. The parallel reference to confusion in the context of the proposed right has been removed. This could result in uncertainty over what is expected as the two appear to be intended to work together.

Furthermore, while the broader objective in this proposed obligation may be clear, the other elements of this obligation are not. Who would be the one determining whether “confusion may arise”? Is this something that would be left to be decided by each AI Actor or would we expect the Parties to make determinations or provide guidelines on when this would apply? Alternatively, is there an expectation that there would be a single standard applicable to all Parties on when “confusion may arise”? If so, this may not be feasible. What may be confusing for one individual or in one context may not be confusing for another.

As in other cases then, it would be important to clarify, whenever there are decisions involved, who would be making those decisions. (CA)

<sup>90</sup> But do we want to see this as a right? (EE)

<sup>91</sup> See also SG1. (T-PD)

<sup>92</sup> Original text suggested by CINGO.

<sup>93</sup> There are many different understandings of what this [N.B. precautionary principle] means. This would need to be written out before anyone can determine whether they agree with its inclusion. (EE)

<sup>94</sup> Isn't risk assessment a topic for SG 1 and SG 5? Could create too much overlap if we are also dealing with it in SG2.

Furthermore, “bans and moratoriums” are the subject of SG 6 (“Red lines”), which has not even started working yet. We should not pre-empt its work by already inserting such provisions. (RF)

<sup>95</sup> Add. (T-PD)

<sup>96</sup> Can AI systems themselves prejudice human rights and fundamental freedoms or negatively affect the natural environment or is it the design, development or use of AI systems that could do so? It seems it would be more accurate to say the latter. (CA)

<sup>97</sup> (EE)

<sup>98</sup> Amend. (T-PD)

<sup>99</sup> Recommendation. (EE)

<sup>100</sup> (EE)

<sup>101</sup> Disagree with emphasis on inclusion of marginalised groups. It is very important that AI and human rights is not perceived as a minorities or marginalised groups issue. It is not. Just as data protection and privacy are not. Human rights too often are presented or perceived as the rights of minorities, and many in the majority population do not fully comprehend that human rights are relevant for all people. We have to ensure that AI and human rights is perceived as relevant for all people. Inclusion is fine and necessary but should not be emphasized. (EE)

<sup>102</sup> Does this apply to all AI systems or is it proportionate to risk? (autocorrect....). (LU)

4. ~~Where relevant given the specific circumstances, for instance in<sup>103</sup> case of a high level of uncertainty coupled with a high level of risk, a precautionary approach, including potential prohibitions, should be taken. Member States/Each Party<sup>104</sup> should consider establishing bans or moratoria or systems whose design and deployment shows significant adverse impacts on human rights, rule of law, or democracy.~~<sup>105,106</sup>
5. For AI systems that are not subject to a ban or moratorium<sup>107</sup>, member states each Party<sup>108</sup> should ensure that developers and deployers take adequate measures to minimise any physical or mental harm to individuals,<sup>109</sup> society and the environment, as part of a broad and inclusive human rights due diligence program.

## Rationale

Protection of right to life (Art. 2 ECHR), and the right to physical and mental integrity, among other human rights listed in the ECHR and UDHR, ICCPR, ICESCR.

Also see **Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impact algorithmic systems, Para B 5.2 (Human rights impact assessments):** *“States should ensure that they, as well as any private actors engaged to work with them or on their behalf, regularly and consultatively conduct human rights impact assessments prior to public procurement, during development, at regular milestones, and throughout their context-specific deployment in order to identify the risks of rights-adverse outcomes. Confidentiality considerations or trade secrets should not inhibit the implementation of effective human rights impact assessments.”*;

Para B 5.3 (Expertise and oversight): *“States should ensure that all human rights impact assessments related to high-risk algorithmic systems are submitted for independent expert review and inspection. Tiered processes should be identified or created where necessary for independent oversight. Human rights impact assessments conducted by or for States should be publicly accessible, have adequate expert input, and be effectively followed up.”*;

Para B 5.4 (Follow-up): *“In circumstances where the human rights impact assessment identifies significant human rights risks that cannot be mitigated, the algorithmic system should not be implemented or otherwise used by any public authority. If the risk is identified in relation to an algorithmic system that has already been deployed, implementation should be discontinued at least until adequate measures for risk mitigation have been taken. Identified human rights violations should immediately be addressed and remedied, and measures adopted to prevent further violations.”*

<sup>103</sup> How risk is determined and when the precautionary approach should “kick in” must be clearly defined – “for instance”, “potential” and “should” cannot be used. The issue of prohibition/bans/moratoriums is to be addressed elsewhere -- ? (EE)

<sup>104</sup> Amend. (T-PD)

<sup>105</sup> Delete (EE)

<sup>106</sup> The feasibility study only mentions “ban or moratorium” when “no appropriate mitigation measures exist within existing legal frameworks to adequately mitigate these risks”. One of the tasks of our legal framework is to precisely provide for appropriate mitigation measures. Rec(2020)1 also speaks of “preventing and mitigating the risk of human rights violations and other adverse effects”, not just banning AI systems whenever such a risk exists. (RF)

<sup>107</sup> Delete. We should not proceed from the idea that a priori all AI has to be either banned or subject to a moratorium – which is how this is currently worded. (EE)

<sup>108</sup> Amend. (T-PD)

<sup>109</sup> Is this intended to reflect the prevention or mitigation of risk as set out in paragraph 3? Referring in some way to the mitigation of risk may be clearer than “adequate measures to minimise any physical or mental harm to individuals” on its own. (CA)



## Alternative 2: Prevention of ~~harm~~ prejudices resulting from AI systems<sup>110</sup>

Everyone has the right to be protected from prejudice to one's fundamental rights and the natural environment resulting from AI systems.<sup>111</sup>

AI systems must not<sup>112</sup> prejudice human rights and fundamental freedoms and negatively affect natural environment.<sup>113,114</sup>

=> *Obligations for Parties / Member States:*

Each Party shall adopt adequate measures to ensure that AI development and deployment<sup>115</sup> minimise any prejudice to individuals, society<sup>117</sup> and the environment.<sup>118</sup>

*Further explanation/clarification:*

This may require specific measures to ensure that AI systems are developed and used in a sustainable manner, with full respect for applicable environmental protection standards. It also entails that member states/Parties foster the use of AI systems to avoid and mitigate harm from the actions of human beings and of other technological systems, while safeguarding the standards of human rights, democracy and the rule of law. Member states/Parties could also consider fostering<sup>119</sup> AI solutions that protect and support human integrity, and that can help to solve environmental challenges.<sup>120</sup>

Each Party shall take provide that<sup>121</sup> appropriate measures are taken<sup>122</sup> to ensure that the adoption of<sup>123</sup> adequate requirements for the<sup>124</sup> safety, security and robustness of<sup>125</sup> requirements adopted by AI systems.<sup>126,127</sup>

*Further explanation/clarification:*

---

<sup>110</sup> Original text suggested by BE, as modified by T-PD.

<sup>111</sup> Delete. There is not a right to be protected from harms, rights are the protection provided by the law. (T-PD)

<sup>112</sup> AI systems are not actors. They do not have legal personality. (EE)

<sup>113</sup> Delete. (T-PD)

<sup>114</sup> We very much support the spirit of this but I wonder whether it's the best formulation. Any AI system will extract energy and other natural resources and hence "negatively affect the environment". Should be formulated more as a positive obligation, something along the lines of obliging developers to choose the most environmentally-friendly version of their AI product? (LU)

<sup>115</sup> They are not actors. They cannot minimise. (EE)

<sup>116</sup> Delete 'any'. (EE)

<sup>117</sup> Any prejudice to society could also mean that robots should not take over human jobs. Again, this is too categorical. We cannot speak of ANY prejudice – rather human rights, democracy and the rule of law – we must stay within CAHAI's mandate. (EE)

<sup>118</sup> Delete. (T-PD)

<sup>119</sup> ? could also consider – is this a suggestion? (EE)

<sup>120</sup> Delete. (T-PD)

<sup>121</sup> Amend. (T-PD)

<sup>122</sup> Add. (T-PD)

<sup>123</sup> Amend. (T-PD)

<sup>124</sup> Add. (T-PD)

<sup>125</sup> Amend. (T-PD)

<sup>126</sup> This is incorrect in English. Currently it states that AI systems have to adopt requirements -- ??? or something like that. What is meant here? Do MS have to enact legislation to ensure the safety, security and robustness (which would have to be defined) of AI systems (used by whom? Public actors? Private actors?)... (EE)

<sup>127</sup> Using passive voice risks introducing uncertainty. It looks like the intent here may be to say that Parties should adopt measures requiring AI Actors to adopt safety, security and robustness requirements in the context of the design, development and use of AI systems. As drafted, however, the subject (AI Actors) is missing. Without it, this provision could be interpreted as suggesting that the Parties themselves must adopt these requirements in respect of all AI systems.

It is also not entirely clear from the text itself what is meant by the obligation to ensure the "robustness" of the AI system. While the rationale provides an explanation of some of these requirements, the proposal should be clear enough on its face, particularly if it seeks to impose a binding obligation, as is the case here. (CA)

These requirements should include, inter alia, resilience to attacks, accuracy and reliability, and the necessity to ensure data quality and integrity. Moreover, AI systems should be duly tested and verified prior to their use as well as throughout the entire life cycle of the AI system including by means of periodical reviews to minimise such risks.

## Privacy and data protection<sup>128</sup>

### *Preliminary Comment*<sup>129</sup>

*Data protection is part of the right to privacy, but the right to privacy entails far more elements that are also impacted by AI: (i) a person's (general) privacy, (ii) a person's physical, psychological or moral integrity and (iii) a person's identity and autonomy.*<sup>130</sup>

*AI-driven (mass) surveillance, for example with facial recognition, involves the capture, storage and processing of personal (biometric) data (our faces)<sup>131</sup>, but it also affects our 'general' privacy, identity and autonomy in such a way that it creates a situation where we are (constantly) being watched, followed and identified. As a psychological 'chilling' effect, people might feel inclined to adapt their behaviour to a certain norm, which shifts the balance of power between the state or private organisation using facial recognition and the individual.<sup>132</sup> In legal doctrine and precedent the chilling effect of surveillance can constitute a violation of the private space, which is necessary for personal development and democratic deliberation.<sup>133</sup> Even if our faces are immediately deleted after capturing, the technology still intrudes our psychological integrity.*

*Indiscriminate on- and offline tracking of all aspects of our lives (through our online behaviour, our location data, our IoT data from smart watches, health trackers, smart speakers, thermostats, cars, etc.), could have the same impact on our right to privacy, including our psychological integrity.*

### **Biometric recognition**

*Other forms of AI-driven biometric recognition are often also grouped under the 'right to privacy' but these have an even greater impact on our psychological integrity. Recognition of micro-expressions, gate, (tone of) voice, heart rate, temperature, etc. are currently being used to supposedly assess or even predict our behaviour, mental state and emotions.*

*It should be noted upfront that no sound scientific evidence exists corroborating that a person's inner emotions or mental state can be accurately 'read' from a person's face, gate, heart rate, tone of voice or temperature, let alone that future behaviour could be predicted by it. In a recent meta-study, a group of scientists<sup>134</sup> concluded that AI-driven emotion recognition could, at the most, recognize how a person subjectively interprets a certain biometric feature of another person. An interpretation does not align with how that person actually feels, and AI is just labelling that interpretation which is highly dependent on context and culture. Far-fetched statements, that AI could for example determine whether someone will be successful in a job based on micro-expressions or tone of voice, are simply without scientific basis.*

*More importantly, the widespread use of these kinds of AI techniques, for example in recruitment, law enforcement, schools, impacts a person's physical, psychological or moral integrity and thus elements of that person's private life.*

---

<sup>128</sup> Original text suggested by CINGO.

<sup>129</sup> Suggested by ALLAI.

<sup>130</sup> Guidance to art. 8 ECHR, Council of Europe.

<sup>131</sup> The [jurisprudence](#) of the European Court of Human Rights (ECtHR) makes clear that the capture, storage and processing of such information, even only briefly, impacts art. 8 ECHR.

<sup>132</sup> Examined Lives: Informational Privacy and the Subject as Object, Julie E. Cohen, 2000.

<sup>133</sup> The chilling effect describes the inhibition or discouragement of the legitimate exercise of a right. It has been shown that once people know that they are being surveilled they start to behave and develop differently.

Staben, J. (2016). Der Abschreckungseffekt auf die Grundrechtsausübung: Strukturen eines verfassungsrechtlichen Arguments. Mohr Siebeck.

<sup>134</sup> Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, 20(1), 1–68.

*Biometric recognition should be dealt with separately under a 'red lines' section.*

*Alternative 1*<sup>135</sup>

1. Member States should effectively implement<sup>136</sup> the modernised version of the Convention ("Convention 108+") to better address AI-related issues.<sup>137,138</sup> In particular<sup>139,140</sup>, any AI-enabled manipulation, individualised profiling and predictions involving the processing of personal data must comply with the obligations set out in the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.<sup>141</sup>

**Comment**<sup>142</sup>: *Not only personal but also non-personal data is being used to train AI systems, to profile and score people for various purposes such as predictive policing, insurance acceptance, social benefits allowance, performance prediction in hiring and firing processes. Moreover, massive amounts of 'data points' on how we go about our daily lives are used not only to send us targeted advertising, but also to push/influence/induce/nudge us towards certain information and thus influence our options, affecting our moral integrity. Also, personal data is linked to identification of a person, while adverse effects of using data with AI go beyond the mere identification of a person. The limitation to personal data excludes many of these instances and situations. This element could prove to be a gap in current legislation that need to be filled.*

**Rationale**

See rights enshrined in Convention 108+ and ECHR's right to privacy (relevant applications of art. 8 ECHR case) as well as rights under EU GDPR.

*Alternative 2*<sup>143</sup>

[Article X+2 - ]Privacy and Data Protection

To protect the right to privacy and other applicable human rights and fundamental freedoms whenever AI systems engage in processing of personal data, [Parties shall / Member States should] secure an appropriate level of data protection based on the provisions of the Convention for the Protection of Individuals with regard

<sup>135</sup> Original text suggested by CINGO.

<sup>136</sup> Should = recommendation. Shall ratify and implement?? Can a provision in one convention require the ratification and implementation of another? (EE)

<sup>137</sup> Too vague. (EE)

<sup>138</sup> This is unclear on the intent. Is this meant to be a recommendation to implement "effectively" or to "essentially implement"? Is it intended to go further by incorporating by reference the full Convention, even for Parties who are not party to that Convention? (CA)

<sup>139</sup> For all or only for those who ratify? (EE)

<sup>140</sup> Is this meant to be a precision or a qualification of the sentence before it? The use of different standards ("should" in the first sentence; "must" in the second) introduces further uncertainty as to the intent, particularly since not all countries that may be Party to this instrument will necessarily be Party to the other Convention. (CA)

<sup>141</sup> Should we also provide for an alternative for those States which for some reason do not desire to enter Convention 108+? Perhaps an "equivalent regime" of some kind, which provides, at a minimum, the same level of protection as the Convention? We can draw from the Convention's own provisions on "appropriate level of protection".

Also, this title actually belongs to Convention 108 (the one in force now); to reference the revised Convention 108+, we should add something along the lines of "...as amended by its Protocol CETS No.223".

In fact, referencing both Conventions could also be advantageous (so that we do not exclude States Parties only to the original Convention 108).

The second sentence of the paragraph seems more suited to an explanatory report, as it merely provides examples. (RF)

<sup>142</sup> Suggested by ALLAI.

<sup>143</sup> Suggested by RF.

to Automatic Processing of Personal Data (ETS No. 108), opened for signature in Strasbourg on 28 January 1981, as amended by its Protocol (CETS No. 223).<sup>144</sup>

*Commentary:* This provision is inspired by paragraph 2 of Article 14 of the Convention 108+.

## Right to human contact<sup>145,146</sup>

**Everyone Every individual** is entitled to have the opportunity, when it is not excluded by competing legitimate overriding grounds<sup>147,148</sup> or by the nature of the service, to have contact with a human being rather than an AI system with regard to the task assigned to AI systems.<sup>149</sup>

~~Alternative: Everyone has the right to have the opportunity, in the context of care and assistance provided to elderly people and people with disabilities, to choose to have contact with a human being rather than a robot.<sup>150,151,152</sup>~~

-----

## Right to autonomy and self-determination<sup>153</sup>

**Everyone Every individual** has the right not to be subject to a decision significantly affecting him or her based solely<sup>154</sup> on AI systems automated processing without having his or her views taken into consideration<sup>155</sup> when this produces legal effects on or similarly significantly affects individuals, irrespective of whether this automated processing involves the individuals' personal data or not.

**Everyone Every individual** has the right to effectively an effective remedy to<sup>156</sup> contest and challenge<sup>157</sup> decisions informed and/or<sup>158</sup> made by an AI system and demand that such decision be reviewed by a person (right to opt out).<sup>159</sup>

<sup>144</sup> What is meant by the phrase “based on the provisions of the Convention ...” in this context? Does it mean Parties should adopt measures consistent with the Convention or is the expectation higher (i.e. to adopt the precise measures required by the Convention)? (CA)

<sup>145</sup> Original text suggested by BE, as modified by T-PD.

<sup>146</sup> This formulation is somewhat strange. The right to human contact would be something different than described here. That being said, I cannot see that we could say that anyone has a subjective right to human contact as such. What would this be based on? This is similar to right to opt out – but again, that may not be feasible or desirable. (EE)

<sup>147</sup> Shouldn't it be either ‘competing’ or ‘overriding’? (ALLAI)

<sup>148</sup> There appear to be too many qualifiers here. It looks like the intent is to preserve the ability of the Parties to provide exceptions based on a potential balancing of other legal or public policy considerations. There may be a better way to frame this, however, than “when it is not excluded by competing legitimate overriding grounds”. “Legitimate” also imports a value judgment, but it is not clear whether that judgment is one that is being left to each Party. Where the intent is to provide discretion to the Parties, it would be helpful to say so in more concrete terms. (CA)

<sup>149</sup> ? what is meant here? (EE)

<sup>150</sup> Delete; context-specific repetition. (T-PD)

<sup>151</sup> This is far better – it is context specific and would be possible to apply in practice. (EE)

<sup>152</sup> I am not sure it makes sense to single these out as there are other categories of vulnerable individuals with a particular interest in this information (I can think of children) and everyone, even AI developers themselves, may struggle to know whether or not they are interacting with a human or AI system in some circumstances. (LU)

<sup>153</sup> Original text suggested by BE, as modified by T-PD.

<sup>154</sup> This is correct. AI systems could then be used but the outcomes/decisions would have to be subject to human review if so requested. But this is similar to the above which is based on personal data. Should these be addressed together? (EE)

<sup>155</sup> Delete. (EE)

<sup>156</sup> Amend. (EE)

<sup>157</sup> What is the difference that is intended between “contest” and “challenge”? If there is no significant difference, it would be clearer to use one term or the other. (CA)

<sup>158</sup> “And/or” is not treaty language. Where more than one option is intended, this can generally be conveyed by “or”. (CA)

<sup>159</sup> Wouldn't this be already covered under Human Oversight? (RF)

~~Everyone~~ Every individual has the right ~~not to be subject to and to decide freely to be excluded from to opt out of any~~<sup>160</sup> AI-enabled manipulation, individualised profiling and predictions, also in case of non-personal data processing.<sup>161,162,163,164</sup>

~~/alternative/ Everyone has the right to refuse to be subjected to profiling, to have one's location tracked, to be manipulated or influenced by a "coach".~~<sup>165,166</sup>

=> *Obligations for Parties / Member States:*

Each Party shall<sup>167</sup> ~~adopt~~ provide adequate measures to ensure that AI developers and deployers establish human oversight mechanisms that safeguard human autonomy and self-determination, in a manner that is tailored to the specific risks arising from the context in which the AI system is developed and used, as well as to duly and timely communicate options for redress.<sup>168,169</sup>

*Further explanation/clarification:*

This may require that:

- An adequate level of human involvement should be ensured in the operation of AI systems, based on a contextual risk assessment taking into account the system's impact on human rights, democracy and the rule of law.
- Whenever necessary and possible<sup>170</sup>, based on a thorough risk assessment, a qualified human being should be able to disable any AI system or change its functionality.
- Those developing and operating AI systems should have the adequate competences or qualifications to do so, to ensure appropriate oversight that enables the protection of human rights, democracy and the rule of law.
- To protect the physical and mental integrity of human beings, AI deployers should strive to avoid the use of 'attention economy' models that can limit human autonomy.

*Links with other instruments:*

A right not to be manipulated and a right to human contact have been suggested by PACE in its Recommendation 2102 (2017) on Technological convergence, artificial intelligence and human rights (28 April 2017)

A right not to be subject to a decision solely based on automated processing when this produces legal effects on or similarly significantly affects individuals also exists in Convention 108+, but the protection it affords is less comprehensive (Article 9(a) : 'the right not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration'). For instance, it does

---

<sup>160</sup> Amend. (EE)

<sup>161</sup> Is this meant also for the private sector? Amazon? Facebook? Google? FYI – the EU proposal for a Digital Services Act does want to enact the obligation to provide an option that is not based on individualised profiling. (EE)

<sup>162</sup> This appears to be an excessive requirement. E.g. Convention 108+ does not provide for a "right not to be subject to" data processing, only for the data subject's views to be taken into account. Also, the Convention provides for exceptions when there are "legitimate grounds for the processing which override [the data subject's] interests or rights and fundamental freedoms". (RF)

<sup>163</sup> Would this apply to all AIs? I.e. are we saying that people should have the right to a non-personalised version of Amazon or their newspaper? (LU)

<sup>164</sup> This seems extremely broad. As others have noted, it may not be feasible in all contexts. An alternative would be to reframe this as a measure that may be imposed by the Parties (e.g. "Parties may provide that, in certain circumstances, individuals should have the ability to opt out ...."). (CA)

<sup>165</sup> Delete. (T-PD)

<sup>166</sup> The wording has to be legal. (EE)

<sup>167</sup> This is worded in a manner that would make sense in a convention. (EE)

<sup>168</sup> This appears to be a duplication of the provisions under *Human Oversight*. (RF)

<sup>169</sup> This does appear to duplicate proposals under Human Oversight, above. This language is also unclear for the same reason as similar language along these lines in earlier proposals. It looks like a blend of "Each Party shall ensure" and "Each Party shall provide measures to ensure." It is not clear what "provide to ensure that" means. (CA)

<sup>170</sup> What about the issue of necessary but impossible? (EE)

not apply in situations falling outside the scope of Convention 108+, such as where an individuals' personal data has not been processed, while an AI system can also impact individuals without processing their personal data.

Right to reasoning<sup>171</sup>

Alternative: Right to explanation<sup>172</sup>

1. Every individual has<sup>173</sup> the right of an individual to [ALTERNATIVE 1: access the reasoning of a decision taken by or with the support of an AI system<sup>174</sup>] [ALTERNATIVE 2: acquire access to the reasoning of/~~behind~~<sup>175</sup> information on the parameters that influence<sup>176</sup> an algorithmic decision], which when that decision<sup>177</sup> produces legal effects concerning him or her and/<sup>178</sup>or significantly affects him or her. [ALTERNATIVE 1: Such reasoning information<sup>179</sup> shall include, in a plain and clear language, the factual information and the legal considerations<sup>180</sup> on which an algorithmic decision has been made.<sup>181,182</sup>] [ALTERNATIVE 2: Such reasoning shall be provided in a plain and clear language and shall include the factual information, the logic underpinning the AI system, and the legal considerations upon which the decision has been made.<sup>183</sup>]
2. [ALTERNATIVE 1] When AI systems are used to<sup>184</sup> support public authorities<sup>185</sup> in a decision making procedure by public authorities, which that produces legal effects concerning an individual and/or significantly affects him or her, the individual has the right acquire access to information on the reasoning of the algorithmic recommendation that supports this human decision making process. Such reasoning information shall include, in a plain and clear language, the factual information and the legal considerations<sup>186</sup> on which an algorithmic recommendation has been made.<sup>187</sup>  
[ALTERNATIVE 2] When such decisions are taken in the context of a task carried out by a public authority, each Party shall provide that the authority issuing the decision has a duty to provide the reasoning described in paragraph 1.<sup>188</sup>

=> Obligations for Parties / Member States:<sup>189</sup>

When an algorithmic decision outcome<sup>190</sup> produces legal effects concerning an individual and/or significantly affects him or her, it is important for the public authorities that issued this decision to have the duty to provide the

<sup>171</sup> Original text suggested by Homo Digitalis.

<sup>172</sup> Suggested by ALLAI.

<sup>173</sup> Amend. (T-PD)

<sup>174</sup> Amend. (T-PD)

<sup>175</sup> Amend (ALLAI).

<sup>176</sup> Amend. This is not a set term – or correct here -- to my understanding. Algorithmic decisions do not have reasoning. For one, it is the system or rather the algorithm that has parameters which determine outcomes. This would need to be carefully worded. (this is just a quick suggestion – I do not believe it is correct). (EE)

<sup>177</sup> Amend. (T-PD)

<sup>178</sup> Delete. (T-PD)

<sup>179</sup> Amend. (EE)

<sup>180</sup> This is far too vague – making it difficult to implement and impossible to control. (EE)

<sup>181</sup> Would this requirement properly account for the opacity/"black box" nature of AI? (RF)

<sup>182</sup> I like this but wonder whether there may be a better formulation for "factual information" I am unclear whether this only refers to the input into the system or also the parameters used to make the decision (which seems preferable). (LU)

<sup>183</sup> Amend. (T-PD)

<sup>184</sup> Amend. AI systems are not actors. The wording of this entire text must be accurate in relation to who is the obliged party, who is the actor. (EE)

<sup>185</sup> Why limit this to public authorities only? (LU)

<sup>186</sup> Too vague. What is even meant by "factual information"? Data set? (EE)

<sup>187</sup> Isn't paragraph 2 essentially a repetition of paragraph 1? Should they be conflated? (RF)

<sup>188</sup> Amend. (T-PD)

<sup>189</sup> Delete entirely. (T-PD)

<sup>190</sup> Amend (ALLAI).



reasoning of this decision outcome<sup>191, 192</sup>. The same duty shall exist when public authorities base their decision making process ~~in~~ on<sup>193</sup> algorithmic recommendations.

#### *Further explanation/clarification:*

The aim is to further clarify and expand in non-EU member states the right to explanation as described under Article 13, para 2(f) GDPR. There is no single, neat statutory provision labelled the 'right to explanation' in GDPR. Therefore, there is an ongoing debate as regards whether such a right is actually provided by law.

This is a unique opportunity for CoE<sup>194, 195</sup> to provide legal clarity and encapsulate first in a clear and precise manner such a right. Access to reasoning of a decision made by an algorithm is an important prerequisite for the successful exercise of the right to protection of personal data and the right to an effective remedy and a fair trial. Moreover, it is crucial for the right to good administration (EU Charter right<sup>196</sup>) since according to the CJEU, the context in which individual decisions are made is important in determining the extent of the duty to give reasons (See more at FRA, Getting the Future Right – AI and Fundamental Rights, p.81)

Such a right should be provided not only when there is an algorithmic decision, but also when a human decision is based on algorithmic recommendations. For example, when evaluating the risk of an individual to reoffend, the public bodies might inform their decision making process based on an algorithmic risk assessment score.

-----

#### Right to Digital Identity<sup>197</sup>

1. Everyone has the right to know his or her digital identity.
2. Everyone is allowed to accede, to question and, if need be, to correct the data supporting his or her digital identity.<sup>198</sup>

#### **Rationale**

The right to digital identity is, such as the right to digital integrity (see below), a NEW right. While we try to match and adapt the framework of fundamental rights to the context of digitalization and artificial intelligence, it is also true that we will not fully succeed: There are components in the virtual world that do not have a counterpart in the analog world. Digital identity is one of those elements.

Digital identity is our transposition from the analogue into the digital world. Our identity is the way to express ourselves and to frame who we are (or wish we were) in the digital world. The right to digital identity is a highly personal right and marks the entry zone into the digital world or digital worlds. It is therefore essential for everyone to know about the digital identity and to accede to it.

---

<sup>191</sup> Amend (ALLAI).

<sup>192</sup> Is this meant to be an explanation or is it being put forward as additional text? If it is meant to constitute part of the text, it is unclear what additional value a statement such as "it is important for the public authorities that issued this decision to have the duty ..." brings in this respect. This seems to be a descriptive statement that perhaps belongs more in the rationale. (CA)

<sup>193</sup> Amend (ALLAI).

<sup>194</sup> Note that the EU Commission in its recently tabled Digital Services Act has also begun to try to define what parameters of algorithmic systems must be made transparent to users of digital platforms. The same may be addressed in the upcoming proposal for a regulation on trustworthy AI (April 21). (EE)

<sup>195</sup> Framing this as "a unique opportunity for the CoE to provide legal clarity and encapsulate first in a clear and precise manner" a particular right seems to be a peculiar justification for the potential creation of a new right. The assessment of whether there is a need for new rights should be based on a thorough and meaningful legal analysis of not only existing rights and obligations but also other means of achieving the same objectives, not on the potential advantages from a policy or strategic objective if the Council of Europe were the "first" to provide a "clear and precise" formulation of such a potential new right. (CA)

<sup>196</sup> Is there anything similar in the CoE's standards on democracy and the rule of law? (EE)

<sup>197</sup> Suggested by CH.

<sup>198</sup> What is meant by this reference to "accede"? What is meant too by "the data supporting his or her digital identity"? Is data "supporting" someone's digital identity different than the data that collectively comprises that digital identity? (CA)

## Right to digital integrity<sup>199</sup>

~~Member States should~~ **Each Party should provide to**<sup>200</sup> protect the digital integrity of natural persons<sup>201,202</sup>, as ~~a new fundamental right,~~<sup>203,204</sup> consistent with the right to physical and mental integrity (art. 3, al. 1 Charter of fundamental rights of the EU<sup>205,206</sup>). ~~Member States should~~<sup>207</sup> recognise this right to digital integrity in its Constitution.<sup>208,209,210</sup>

Each party shall take appropriate measures to protect the right to digital integrity as a pre-requisite<sup>211</sup> requirement to an effective implementation of the right to autonomy.

Personal data should be considered as a component of the person<sup>212</sup>, rather than things that can be owned by whoever collects that data.

In accordance with the principle of proportionality, the right to digital integrity must not be exercised in such a way as to prevent the exercise of other human rights and fundamental freedoms. It is a matter of reconciling the different rights and freedoms involved.

In particular, the right to digital integrity should not prevent Member States to attribute cyberattacks and to allocate responsibility for criminal actions.<sup>213</sup> The necessity of the limitation of the right to digital integrity has to be assessed together with the proportionality to the purpose and the impact on the rights of the persons concerned.<sup>214,215</sup>

-----

## Right to the protection of attention<sup>216</sup>

~~The economy of attention aims at maximising the time spent by the users on some AI applications, in order to optimise (by design) the volume of personal data collected and processed. This can lead to users addiction and to a higher cost for States on Public Health. Due to the positive responsibility of Member States to protect the citizens, and to the negative impact of the economy of attention on the users health, Members States shall recognise a right to the protection of the attention.~~<sup>217,218,219</sup>

<sup>199</sup> Original text suggested by AI Transparency Institute.

<sup>200</sup> Amend. Also: If we want to propose a new right, a more detailed description of its nature and scope is needed. (T-PD)

<sup>201</sup> Please provide a detailed explanation of what this means. (EE)

<sup>202</sup> I am not sure it will be understandable to most what this means. (LU)

<sup>203</sup> Delete. (T-PD)

<sup>204</sup> Fundamental right – is not a CoE term. The ECHR has human rights. It also would not be legally (methodologically) correct to just state that there is supposedly a new human right. (EE)

<sup>205</sup> Delete. (T-PD)

<sup>206</sup> ECHR equivalent?? (EE)

<sup>207</sup> ? recommendation. (EE)

<sup>208</sup> Delete. Not compatible with the CoE framework. (T-PD)

<sup>209</sup> Delete. It is highly inappropriate to dictate to a member state how they should include anything in their legal system. This is not how international law works. (EE)

<sup>210</sup> As with other proposed text drafted along these lines, “provide to protect” remains unclear. Is is also not clear what is meant by “the digital integrity of natural persons”. (CA)

<sup>211</sup> Amend. (EE)

<sup>212</sup> ? (EE)

<sup>213</sup> Please provide a detailed explanation. (EE)

<sup>214</sup> This is a general clause on the restriction of human rights – it should not be here. (EE)

<sup>215</sup> Is this meant to be part of the text or is this being provided as the explanation? (CA)

<sup>216</sup> Original text suggested by AI Transparency Institute.

<sup>217</sup> Delete. Disagree with this. Attention is not something that can be defined with sufficient clarity. This would seem to be related to the right to mental health. What is being suggested here? That there be some kind of mental health and safety standards for AI systems? The future potential legal framework must be practical. A detailed explanation would be required for this to be considered further. (EE)

<sup>218</sup> This is a very important issue but I struggle to see what this would mean concretely. How should Member States do this? (LU)

<sup>219</sup> I suggest avoiding the introduction of new rights that do not have a solid basis in human rights framework and are closer to policy suggestions. (T-PD)



## General Remarks<sup>220</sup>

**General comment** – We respectfully disagree with the proposed approach of drafting the concrete texts of legal provisions. We would advise against very specific wording because it may be very difficult to get the acceptance of CAHAI members – specific wording should be developed through international negotiations. We suggest as an alternative a general description of what the provisions might entail, which could then be translated into the language of (binding?) horizontal legal provisions by the relevant professionals, should the Committee of Ministers so decide upon completion of CAHAI’s work. **The drafting of convention (or any other legal) provisions is a very specific skill, and must follow a legal-technical methodology.** CAHAI’s second deliverable must be revised thoroughly for legal and technological accuracy.

**Another general comment:** It is our understanding that issues of liability are not to be addressed in this section.

I. While SG1 is to develop the scope and general principles, it would seem that the issues of human value and dignity are actually very basic, and may also best be addressed in the form of basic principles established in the General Provisions of the instrument. Therefore the SG might propose that the General Provisions could include:

- A provision similar to Article 1 of the Oviedo Convention (Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine), which provides the purpose and object of the Convention: *Parties to this Convention shall protect the dignity and identity of all human beings and guarantee everyone, without discrimination, respect for their integrity and other rights and fundamental freedoms with regard to the application of biology and medicine.*

This would need to be expanded to express that the purpose would not only be to ensure respect for human rights and fundamental freedoms, but also for democracy and the rule of law. This may mean wording that reflects not only the need to protect the value and dignity of all human beings, but also the value of the democratic societies that human beings have formed – through adherence to standards on democracy and the rule of law.

- A provision similar to Article 2 of the Oviedo Convention (Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine), which provides for the primacy of the human being: *The interests and welfare of the human being shall prevail over the sole interest of society or science.*

The provision could provide for the primacy of the human being and of human dignity, which is inherent in all people. It could recognise that the right to human dignity is both a human right in itself as well as the foundation for all other human rights and freedoms.

II. With regard to the category “Human Value, Dignity, Autonomy, Freedoms”, this derives from the structure underlying the report by Catelijne Muller on “The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law”, commissioned by CAHAI.

In her paper, respect for human value related to Liberty and Security, Fair Trial, No Punishment without Law (Art. 5, 6, 7 ECHR) and Private and Family Life, Physical, Psychological and Moral Integrity (Art 8 ECHR).

Freedom of the Individual related to Freedom of Expression (Art 10 ECHR, which includes freedom of information), and Freedom of Assembly and Association (Art 11 ECHR).

**To what degree might or should the input of SG2 also reflect this structure or understanding?**

III. Potential articles/provisions:

---

<sup>220</sup> EE.

- Agree in general with CINGO regarding a provision on the protection of human dignity (if not sufficiently covered as a basic principle), but only with regard to paragraph 1. Paragraph 2 is more an explanation than a provision. Also, the *promotion* of human rights, democracy and the rule of law cannot be required, only respect therefor. It is unclear who is the obligated subject of this paragraph.
- A provision on the right to be informed of interaction with an AI system, in particular where this is not obvious (similar to suggestion above). This might include the right to be notified where any decision is either informed or made with the involvement of an AI system; also where such decision produces legal effects on or similarly significantly affects individuals, the right to demand review of any such decision by a human being.

The right not to be subject to interaction with an AI system should be approached very carefully (comment on proposal above). First, a distinction should be made between use by the private and public sectors. If a person does not want to be subject to an AI system used by a business to better promote their online sales (e.g. a recommender system), then the person can simply choose not to use that service – the legislator should be very careful in interfering with the right of enterprise, including the business models of owners of businesses. This right may have more merit with regard to use in the public sector by public authorities. What is meant above by “legitimate use”? Provided by law may not be a bad idea – but what this means in practice would have to be left to national law, as in the case of the ECHR. An “adverse effect” on rights and freedoms may actually be perfectly legitimate and proportionate. In the Feasibility Study, it was suggested that this may entail a “right to refuse interaction with an AI system whenever this can adversely impact human dignity”. Again, “adversely impact” is too broad and not a correct measure. It has been suggested also in the Feasibility Study that there should be a right to completely opt-out of automated decision-making, but this may not be feasible nor desirable. It has also been suggested in the Feasibility Study that there could be the right to choose to have contact with a human rather than a robot, unless excluded by legitimate overriding grounds. What is meant by this is not clear to us. We must be careful when we start giving people the option to opt-out of things as to whether this is proportionate to the additional burden of maintaining parallel possibilities.

- An Article/provisions on the principle of prevention of harm to human rights, democracy and the rule of law. The wording should reflect that while some AI systems may be used to prevent or mitigate harm, this should not be a requirement. The requirement should be that no (new) harm is caused. This should include reference to the need to ensure the safety and robustness (including relating to cybersecurity) of AI systems.

It is not clear to us whether the rights relating to the environment as expressed in the Feasibility Study and in these proposals are Council of Europe standards. The legal framework is to ensure compliance with the CoE’s standards on human rights, democracy and the rule of law. With regard to the proposal by CINGO – it would seem that the risk-based approach is a matter for the general principles and scope. Also, the inclusion of marginalised groups would perhaps be a matter for SG 3 – Non-discrimination, gender equality, fairness, diversity. The precautionary principle should not be added without further analysis and assessment. It could be suggested as something to consider. Also, convention provisions cannot use wording such as “Member States *should consider* bans or moratoria...”.

- An Article/provisions on the right to personal autonomy (as developed from Article 8 ECHR) and self-determination. This right could entail the right not to be subject to a decision based solely on automated processing (? Or only involving an AI system?) when this produces legal effects on or similarly significantly affects individuals. This could include the right to decide freely whether to be subject to AI-enabled manipulation, such as individualised profiling and recommender systems – so for example a person has control over the information to which the person is exposed.

See above for concerns regarding opt-out. Human review/oversight may be preferable.