CDT Europe's Response to the Council of Europe's Ad hoc Committee on Artificial Intelligence (CAHAI) Multistakeholder Consultation on the Elements of a Legal Framework on AI

Section 1: Definition of AI Systems

7. In view of the elaboration of a legal framework on the design, development and application of AI, based on the standards of the Council of Europe on human rights, democracy and the rule of law, what kind of definition of artificial intelligence (AI) should be considered by the CAHAI? (select 1 option)

- No definition, with a legal instrument focused on the effect of AI systems on human rights, democracy and the rule of law
- A technologically-neutral and simplified definition, such as "a set of sciences, theories and techniques whose purpose is to reproduce by a machine the cognitive abilities of a human being" (See the CAHAI feasibility study, §5)
- A definition focusing on machine learning systems
- A definition focusing on automated decision-making
- Other
- No opinion

8. If other, please explain below:

The chosen definition should take a broad view of AI, focused on the distinguishing features of the system and the role it plays in any process it is a part of. The definition should be encompassing enough to include predictive systems that replace or support decisions traditionally made by humans, or offer input, advice, or influence into a human decision-making process. It should also include predictive systems designed to increase the efficiency of existing processes, even if they do not aim to offer any suggestion or input into the outcome of the process (such as a system that extracts and organizes information from CVs, but does not rank or otherwise organize the candidates) as these systems may also introduce bias or errors that will affect the human decision makers.

9. What are the reasons for your preference?

Whilst 'automated decision making' describes the purpose of many AI systems, it is too narrow as systems that do not make decisions may nonetheless qualify as AI. Conversely, an impacts-based definition risks being too broad and could include human decisions without computer assistance.

CENTER FOR DEMOCRACY & TECHNOLOGY

Section 2.1: Opportunities and Risks arising from AI Systems

10. Please select the areas in which AI systems offer the most promising opportunities for the protection of human rights, democracy and the rule of law: (select max. 3 options)

- Banking, finance and insurance
- Justice
- Law enforcement
- Customs and border control
- Welfare
- Education
- Healthcare
- Environment and climate
- Election monitoring
- National security and counter-terrorism
- Public administration
- Employment
- Social networks/media, internet intermediaries
- Other
- No opinion

11. If other, which areas and why?

While the community should have as a goal AI systems that protect and advance human rights, the first and immediate step must be to understand and avoid the potential for harm these systems are already exhibiting. Through our work, we have unfortunately found repeated examples of where the use of AI can perpetuate and even cause discrimination. CDT has done research on the use of AI in hiring tools and in access to disability benefits and found evidence of discrimination in both cases. Because algorithms learn by identifying patterns and replicating them, algorithm-driven tools can reinforce existing inequalities in our society. Algorithmic bias can also be harder to detect than human bias, because many people think of technology as 'neutral.' So although AI can help with increasing efficiency of certain tasks, in order to ensure that the risk of discrimination is mitigated against, it will be important to ensure humans' ability to understand, question, test, verify, and challenge the output and function of systems and also to recognise that the use of such technologies is not neutral and will need further safeguards in place to protect human rights.

In many of the listed areas, 'promising' uses are at least possible. For example, AI could expand job applicant pools if it captures applications that humans may miss (or be quick to dismiss), but that contain info reflecting applicants' qualifications. 'Good' AI could equitably allocate other resources (i.e., social security benefits) to people who are disproportionately policed, incarcerated, or

otherwise denied. Unfortunately, the reality is that many current applications of AI perpetuate discrimination.

12. Please indicate which of the following AI system applications in your view have the greatest potential to enhance/protect human rights, democracy and the rule of law: (select 5 maximum)

- Facial recognition supporting law enforcement
- Emotional analysis in the workplace to measure employees' level of engagement
- Smart personal assistants (connected devices)
- Scoring of individuals by public and private entities
- Medical applications for faster and more accurate diagnoses
- Automated fraud detection (banking, insurance)
- AI applications to predict the possible evolution of climate change and/or natural disasters
- AI applications for personalised media content (recommender systems)
- Deep fakes and cheap fakes
- Recruiting software/ AI applications used for assessing work performance
- AI applications to prevent the commission of a criminal offence (e.g. anti-money laundry AI applications)
- AI applications aimed at predicting recidivism
- AI applications providing support to the healthcare system (triage, treatment delivery)
- AI applications determining the allocation of educational services
- AI applications determining the allocation of social services
- AI applications in the field of banking and insurance
- AI applications to promote gender equality (e.g. analytical tools)
- AI applications used for analysing the performance of pupils/students in educational institutions such as schools and universities

13. Please briefly explain how such applications would benefit human rights, democracy and the rule of law:

Any application of AI that involves making crucial decisions about people's lives or well-being should be carefully considered. A common theme from our listed choices, is that these are applications trying to predict and prevent adverse impact by systems on individuals, and not the other way around. As outlined below, however, even in these instances, AI is not without its limitations.

Medical applications for faster and more accurate diagnoses: As the OECD has documented, AI has played an important role in helping to detect, predict and prevent outbreaks of Covid-19 in

the context of the pandemic. An important caveat of the application of AI in a medical context is the need to recognise that this tech may embed longstanding biases pertaining to race and gender (e.g., beliefs that Black people have a higher pain tolerance or that women exaggerate their pain). Image recognition software used for medical diagnosis may not work equally well on different skin tones. Like other applications of AI, use in the area of medicine is not automatically free from risk of discrimination and bias and so should be treated accordingly.

AI and the environment: Only this week, environmental conservation experts in Kenya used AI applications to predict increased flooding, and acted early to evacuate endangered giraffes. Whilst AI algorithms can be used to build better climate models and determine more efficient methods for example of reducing CO2 emissions, AI itself often requires substantial computing power and therefore consumes a lot of energy. For example, a study carried out by the University of Massachusetts found that creating a sophisticated AI to interpret human language led to the emissions of around 300,000 kilograms of the equivalent of C02.

AI applications to promote gender equality: AI may be able to help promote gender equality in certain cases: for example, AI tools can help employers check whether their job postings use gender-sensitive language to help support diversity in the workforce. However, even these tools pose risks, because they may cause humans to rely unduly on automated review processes, which cannot capture all forms of discriminatory language in the way a human reviewer might do. A key challenge is to ensure users know the limitations of the program and consider the AI tool as a supplement, instead of a replacement for human judgment.

14. What other applications might contribute significantly to strengthening human rights, democracy and the rule of law?

Systems to identify discriminatory practices/outcomes and their sources, systems to analyse governance trends leading to rights abuses, systems to predict and identify new viruses or other sources of risk, systems to map policy approaches to results, and identifying other large-scale trends (population/migration/etc) for informed decision making.

Section 2.2: Impact on human rights, democracy and the rule of law

15. Please select the areas in which the deployment of AI systems poses the highest risk of violating human rights, democracy and the rule of law: (select 3 maximum)

- Banking, finance and insurance
- Justice
- Law enforcement
- Customs and border control

- Welfare
- Education
- Healthcare
- Environment and climate
- Election monitoring
- National security and counter-terrorism
- Public administration
- Employment
- Social networks/media, internet intermediaries
- No opinion
- Other: (written answer)

16. Please briefly explain how such applications might violate human rights, democracy and the rule of law:

Use of AI systems for law enforcement, national security and counter-terrorism is risky because it can provide the fuel for decisions that result in a deprivation of liberty without due process. They can effectively lengthen a term of imprisonment when used to predict recidivism. They can contribute to over-policing in neighbourhoods that are already over-policed, and result in disparate rates of imprisonment. Those who are affected adversely by AI used in these areas are effectively barred from mounting challenges to such use because the algorithms employed are proprietary, classified, or jealously guarded by law enforcement or the entities that provide them.

Use of AI in sentencing decisions can also have an adverse impact on access to justice for minorities and communities at risk. Evidence has previously shown how automated risk assessment of a defendant to guide a judge's sentence can have very unreliable results and be biased against race. That is because such systems have the potential to incorporate and amplify the aggregate biases of all of the decisions it was trained on.

We also strongly caution against the use of AI for automated analysis of social media content in law enforcement, justice, and counter-terrorism contexts. The tools that law enforcement officials and others use to conduct sentiment analysis, semantic analysis, and other forms of automated evaluation of individuals' social media content are prone to bias and error, often with a disparate impact on racial and ethnic minorities. Collection of social media content by law enforcement and national security officials can involve a substantial invasion of privacy for individuals and yields little useful information.

CENTER FOR DEMOCRACY & TECHNOLOGY

17. Please indicate the types of AI systems that represent the greatest risk to human rights, democracy and the rule of law: (select max. 5 options)

- Facial recognition supporting law enforcement
- Emotional analysis in the workplace to measure employees' level of engagement
- Smart personal assistants (connected devices)
- Scoring / scoring of individuals by public entities
- Medical applications for faster and more accurate diagnoses
- Automated fraud detection (banking, insurance)
- AI applications to predict the possible evolution of climate change and/or natural disasters;
- AI applications for personalised media content (recommender systems)
- Deep fakes and cheap fakes
- Recruiting software/ AI applications used for assessing work performance
- AI applications to prevent the commission of a criminal offence
- AI applications aimed at predicting recidivism
- AI applications providing support to the healthcare system (triage, treatment delivery)
- AI applications determining the allocation of educational services
- AI applications determining the allocation of social services
- AI applications in the field of banking and insurance
- AI applications to promote gender equality (e.g., analytical tools)
- AI applications used for analysing the performance of pupils/students in educational institutions such as schools and universities

18. Please briefly explain how such applications might violate human rights, democracy and the rule of law:

Facial recognition is particularly problematic in the law enforcement arena because it has been shown to be less accurate when being used to identify dark-skinned people and women. Mis-identification in the criminal justice arena can deprive a person of liberty. As mentioned above, because AI learns by identifying patterns and replicating them, algorithm-driven tools can reinforce existing inequalities in our society. Given that racial-profiling is already a concerning trend across society there is a real danger that facial recognition technology can exacerbate or even increase this phenomenon that violates people's rights.

The European Data Protection Supervisor has called for a moratorium on the use of remote biometric identification systems - including facial recognition - in publicly accessible spaces. This arises from the data protection body's concern that a stricter approach is needed to automated recognition in public spaces of human features - such as of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioural signals - whether these are used in a commercial or administrative context, or for law enforcement purposes. A stricter

approach is necessary in the view of the EDPS in light of the extremely high risks of deep and non-democratic intrusion into individuals' private lives. Outside the European Union, in Council of Europe member states there is an even higher risk of use and adverse impact of these technologies given the lack of equivalent data protection rules. For example, it has recently come to light, the extent to which the Russian authorities are using facial recognition to identify and arrest people that attend protests, including those who were simply peacefully protesting. Such use of the technology has a chilling effect on freedom of association and expression. Such developments in non-EU states makes it is even more pertinent that a Council of Europe Convention ensure a higher layer of protection for human rights across the Council of Europe jurisdiction and potentially beyond.

CDT concurs that law enforcement's use of facial recognition can pose a particularly high threat to human rights given the risks of racial profiling and indiscriminate surveillance. It therefore would be desirable, in such cases where there is a high risk of rights violations, to consider a moratorium until such a time that robust safeguards and effective limitations are in place. Governments are also increasingly turning to algorithms to determine whether and to what extent people should receive crucial social security benefits. Billed as a way to increase efficiency and root out fraud, these algorithm-driven decision-making tools are often implemented without much public debate and are incredibly difficult to understand once underway. Reports from people on the ground confirm that the tools are frequently reducing and denying benefits, often with unfair and inhumane results. As research from CDT has confirmed, people with disabilities in particular experience disproportionate and particular harm because of unjust algorithm-driven decision-making. To prevent such harms, thoughtful design, use, and oversight of algorithm-driven decision-making systems will be crucial.

Further, employers turn to algorithm-driven technologies to analyse employees in the workplace. These technologies are purported to measure employees' engagement and productivity. Instead, they enable employers to impose more stringent productivity requirements and prevent workers from unionising. CDT is examining how these tools facilitate worker exploitation by penalising employees for needing breaks or alternative work schedules.

19. What other applications might represent a significant risk to human rights, democracy and the rule of law?

Content moderation also features prominently in discussion of the use of AI, but as with other uses, it is rife with potential risks to human rights and the rule of law. AI/machine learning and other forms of automation are sometimes incorporated by online intermediaries to enable them to manage the massive quantities of user-generated content that people upload onto their systems. These automated tools can be useful for some aspects of sorting and organising user-generated content, but they also have distinct limitations.

Tools or techniques may not be robust; that is, they may perform well in an experimental or training environment but poorly in the real world. Data quality issues can mean that tools are trained on unrepresentative data sets that end up baking bias into the algorithmic processes. Automated tools for analysing user-generated content typically assess a limited degree of context; they may evaluate a given image, for example, but not understand crucial information about the caption, account, or commentary around the image that is essential to its meaning. The operation of automated tools can be difficult to measure, and the creators of these tools may report 'accuracy' rates that fail to meaningfully characterise the tool's impact on different speakers and communities. And the decision processes for some machine learning techniques are difficult to explain in terms that are relevant and useful to human understanding, making interventions and mitigation tactics to protect human rights potentially very difficult. Finally, automation/AI will never be able to achieve consensus decisions or analysis of issues on which humans do not already agree; a machine-learning classifier trained to identify hate speech will nevertheless make determinations with which some people strongly disagree.

In addition to these technical limitations in the use of AI for content moderation, it is important to recall that 'automation' in these circumstances is typically a form of content filtering. Content filtering raises significant threats to human rights, particularly when mandated by law. Filtering is a form of prior restraint on speech, where all statements by anyone using a service must be pre-approved by the filter in order to be posted. Filtering requires a form of total surveillance of people's communications to ensure that whatever is being said abides by the filter's standards. While content filtering can have a very useful role to play in the management of massive quantities of online content (think, for example, of spam filtering), it is crucial for any voluntary use of filters to incorporate opportunities for review of the filter's decisions and operation, and opportunities for appeal of the inevitable errors the filter will make. Filtering, whether it uses simple techniques or sophisticated machine learning, should never be mandated in law.

20. In your opinion, should the development, deployment and use of AI systems that have been proven to violate human rights or undermine democracy or the rule of law be: (select 1 option)

- Banned
- Not banned
- No opinion
- Other: (written answer)

21. In your opinion, should the development, deployment and use of AI systems that pose high risks* with high probability to human rights, democracy and the rule of law be: (select 1 option)**

** High negative impact on human rights, democracy and rule of law*
*** High probability of occurrence of these risks*

- Banned
- Subject to moratorium
- Regulated (binding law)
- Self-regulated (ethics guidelines, voluntary certification)
- None of the above
- No opinion

22. In your opinion, should the development, deployment and use of AI systems that pose low risks* with high probability to human rights, democracy and the rule of law be: (select 1 option)**

** Low negative impact on human rights, democracy and rule of law*
*** High probability of occurrence of these risks*

- Banned
- Subject to moratorium
- Regulated (binding law)
- Self-regulated (ethics guidelines, voluntary certification)
- None of the above
- No opinion

23. In your opinion, should the development, deployment and use of AI systems that pose high risks* with low probability to human rights, democracy and the rule of law be: (select 1 option)**

** High negative impact on human rights, democracy and rule of law*
*** Low probability of occurrence of these risks*

- Banned
- Subject to moratorium
- Regulated (binding law)
- Self-regulated (ethics guidelines, voluntary certification)
- None of the above
- No opinion

24. What are the most important legal principles, rights and interests that need to be addressed and therefore justify regulating the development, deployment and use of AI systems? (select max. 5 options)

- Respect for human dignity
- Political pluralism
- Equality
- Social security
- Freedom of expression, assembly and association
- Non-discrimination
- Privacy and data protection
- Personal integrity
- Legal certainty
- Transparency
- Explainability
- Possibility to challenge a decision made by an AI system and access to an effective remedy

25. In your opinion, in what sectors/areas is a binding legal instrument needed to protect human rights, democracy and the rule of law? (select max. 3 options)

- Banking, finance and insurance
- Justice
- Law enforcement
- Customs and border control
- Welfare
- Education
- Healthcare
- Social networks/media, internet intermediaries
- Environment and climate
- Election monitoring
- Public administration
- No opinion
- Other: (written answer)

Given the broad range of use and application of AI, a sector-specific approach will be required. In some areas, audits and stronger obligations on explainability would be desirable. In other areas, there is already existing legislation and so it is less pressing to regulate. At the same time, we need to carefully monitor changes to existing regulations in case they change in a manner that limits their control over AI. For example, in the U.S. the Dep't of Housing and Urban

Development proposed rule changes that would have impacted people's access to legal redress for discrimination resulting from algorithmic models.

Section 3: Potential Gaps in Existing Binding Legal Instruments Applicable to AI

(In the following section, please indicate to what extent you agree or disagree with the following statements or if you have no opinion on a given issue.)

26. Self-regulation by companies is more efficient than government regulation to prevent and mitigate the risk of violations of human rights, democracy and the rule of law:

- *1=I completely disagree; 2=I rather disagree; 3=Indifferent/no opinion; 4=I rather agree; 5=I fully agree;*

27. Self-regulation by companies is sufficient to prevent and mitigate the risk of violations of human rights, democracy and the rule of law:

- *1=I completely disagree; 2=I rather disagree; 3=Indifferent/no opinion; 4=I rather agree; 5=I fully agree;*

28. Which of the following instruments of self-regulation do you consider to be the most efficient?

- Ethics guidelines
- Voluntary certification
- No opinion
- Other: (written answer)

Self-regulation has proven to be an inadequate approach to mitigate against human rights violations and ensure access to effective remedy to those whose rights are impacted. CDT therefore recommends (see response 39 below) a combination of risk-based assessments and human rights impact assessments, as well as obligations with regard to explainability and AI.

29. Existing international, regional and/or national binding and/or non-binding legal instruments are sufficient to regulate AI systems in order to ensure the protection of human rights, democracy and the rule of law:

- *1=I completely disagree; 2=I rather disagree; 3=Indifferent/no opinion; 4=I rather agree; 5=I fully agree;*

CENTER FOR DEMOCRACY & TECHNOLOGY

30. If you responded disagree/completely disagree to previous question, please indicate why existing international, regional and/or national (binding and/or non-binding) legal instruments are not sufficient to regulate AI systems: (select all you agree with)

- There are too many and they are difficult to interpret and apply in the context of AI
- They provide a basis but fail to provide an effective substantive protection of human rights, democracy and the rule of law against the risks posed by AI systems
- They lack specific principles for the design, development and application of AI systems
- They do not provide enough guidance to the designers, developers and deployers of AI systems
- They do not provide for specific rights (e.g. transparency requirements, redress mechanisms) for persons affected by AI
- They create barriers to the design, development and application of AI systems

31. Please provide examples of existing international, regional and/or national (binding and/or non-binding) instruments that in your view are effective in guiding and regulating the design, development and use of AI systems to ensure compatibility with the standards for human rights, democracy and the rule of law:

The EU's Regulation (EU) 2016/679 General Data Protection Regulation (GDPR) places obligations on EU member states with regard to data governance and explainability. However, as the EU's Fundamental Rights Agency has found, despite the existence of GDPR, many actors do not understand how to carry out a fundamental rights-based approach to data governance in order to prevent algorithmic discrimination, particularly in the private sector. So, whereas GDPR has had a positive impact on privacy and better data governance, further thought is needed on combatting discrimination in particular.

CDT concurs with the opinion of the EU European Data Protection Supervisor that recommender systems should by default not be based on profiling within the meaning of Art. 4(4) of the GDPR. In theory GDPR can be a helpful tool in limiting the AI-driven spread of disinformation by limiting such profiling, however the GDPR is currently not adequately enforced to make this positive potential a reality. Furthermore, GDPR is focussed on individual consent, but in reality today's complex info-ecosystems mean that data-subjects often do not understand the full implications of what they are consenting to. GDPR also empowers data-subjects to delete information that is inaccurate or where they simply wish to withdraw consent, but deleting specific data points from machine-learning is currently very challenging. Overall, whereas GDPR is an essential privacy and data protection law, for some of the reasons outlined above it is not fully equipped to deal with the issues of collective algorithmic harm. See also The Limits of the GDPR in the Personalisation Context.

Given the risks that micro-targeting in the context of elections in particular and profiling pose in a democracy, CDT has further agreed with the EU EDPS that advertising based on pervasive tracking should be phased out.

32. Please indicate other specific legal gaps that in your view need to be addressed at the level of the Council of Europe:

Article 14 of the Council of Europe Convention on Human Rights (ECHR) enshrines the protection against discrimination in the enjoyment of the rights set forth in the Convention. According to the Court's case law, the principle of non-discrimination is of a "fundamental" nature and underlies the Convention together with the rule of law, and the values of tolerance and social peace (S.A.S. v. France [GC], 2014, § 149; Străin and Others v. Romania, 2005, § 59). Furthermore, this protection is completed by Article 1 of Protocol No. 12 to the Convention which prohibits discrimination more generally, in the enjoyment of any right set forth by law.

Vital decisions which impact our lives are being made using automated decision-making (ADM). These systems, now used in job recruitment, decisions on benefits, access to educational opportunities and other settings frequently perpetuate existing prejudice and discrimination. This is particularly urgent considering that those most affected are already marginalised and at-risk. The current problem is a lack of access to effective remedies. Further legal safeguards and obligations around the obligation to meaningfully explain the use of algorithms is needed in order to ensure access to justice and improved oversight of AI. If an individual has suffered discrimination as prohibited under European human rights law they need to have access to effective remedy in practice. Ensuring such access will involve some mandatory transparency over how AI is used and stronger obligations on explaining how decisions were reached.

Furthermore, in cases such as recruitment and access to social benefits, auditing could also be considered. In December 2020, the EU's Fundamental Rights Agency released a report of the results of an interview of 100 public officials and private and civil sector experts. The report found that despite the existence of GDPR, many actors did not understand how to carry out a fundamental rights-based approach to data governance in order to prevent algorithmic discrimination, particularly in the private sector. There is a need to make it imperative for private companies to take action in this area. There is a lack of case studies and case-law in the area of AI and discrimination across Europe to help inform upcoming legislative proposals with evidenced-based policy suggestions.

CENTER FOR DEMOCRACY & TECHNOLOGY

Section 4: Elements of a Legal Framework on AI Systems

(In relation to some AI systems, we can reasonably foresee a significant risk to human rights, democracy and the rule of law. Bearing this in mind, in the following section, please indicate to what extent you agree or disagree with the following statements or if you have no opinion on a given issue.)

33. Please indicate to what extent you agree or disagree with the following statements or if you have no opinion on a given issue:

	I completely disagree	I rather disagree	Indifferent/ no opinion	I rather agree	I fully agree
Individuals should always be informed when they interact with an AI system in any circumstances					
Individuals should always be informed when a decision which affects them personally is made by an AI system					
Individuals should always be informed when an AI system is used in a decision-making process which affects them personally					
Individuals should have a right to a meaningful explanation of algorithmic based decisions, in particular how the algorithm reached its output					
Individuals should always have the right that any decision taken by an AI system in the framework of judicial proceedings are reviewed by a “human” judge					
Individuals should have a right to demand the review of an algorithmic based decision by a human being					
There should always be a person responsible for reviewing algorithmic based decisions in the public sector and private companies					
Public institutions should not use AI systems to					

promote or discredit a particular way of life or opinion (e.g. “social scoring”)					
States should be obliged to design, develop and apply sustainable AI systems that respect applicable environmental protection standards					
The code behind AI systems used in the public and private sectors should always be accessible to the competent public authorities for the purposes of external audit					
There should be higher transparency standards for public entities using AI than for private entities					
There should be higher standards for access to an effective remedy for individuals in relation to decisions informed and made by an AI system in the field of justice than in the field of consumer protection					
Member States should establish public oversight mechanisms for AI systems that may breach legally binding norms in the sphere of human rights, democracy and the rule of law					
Errors and flaws discovered in AI systems which have led or could lead to the violation of human rights, democracy and the rule of law must be reported to the competent authorities					
The use of facial recognition in public spaces should be prohibited					
The information obtained through the use of facial recognition systems should always be reviewed by a human being before being used for purposes that have an impact on individual freedom, such as in relation to a person boarding an airplane, upon police arrest or in the framework of judicial proceedings					
The use of AI systems in democratic processes (e.g., elections) should be strictly regulated					

CENTER FOR DEMOCRACY & TECHNOLOGY

34. Should a future legal framework at Council of Europe level include a specific liability regime in relation to AI applications? (select 1 option)

- Yes
- No
- No opinion

35. If yes, what aspects should be covered?

For systems that support human decisions, hold the deciding humans responsible for any rights violations or illegal outcomes.

For systems that make or act on their own decisions/analysis, hold the humans that selected/deployed the system responsible.

Section 5: Policies and Measures for Development

36. In your opinion, how useful would the following compliance mechanisms be in preventing and mitigating the risks to human rights, democracy and the rule of law arising from the design, development and application of AI?

** Intersectional audits consider intersection of multiple sensitive attributes (race, gender, etc) jointly instead of attributes alone - for an example of such audits with machine learning, see for instance: Morina, Giulio & Oliinyk, Viktoriia & Waton, Julian & Marusic, Ines & Georgatzis, Konstantinos. (2019). Auditing and Achieving Intersectional Fairness in Classification Problems*

	Not useful	Rather not useful	Indifferent/ no opinion	Rather useful	Highly useful
Human rights, democracy and rule of law impact assessments					
Certification and quality labelling					
Audits and intersectional audits*					
Regulatory sandboxes					
Continuous automated monitoring					

CENTER FOR DEMOCRACY & TECHNOLOGY

37. Please indicate what combination of mechanisms should be preferred to efficiently protect human rights, democracy and the rule of law: (select max. 3 options)

- Human rights, democracy and rule of law impact assessments
- Certification and quality labelling
- Audits and intersectional audits
- Regulatory sandboxes
- Continuous automated monitoring
- Other: (written answer)

38. Please select which mechanism(s) should be part of either a binding instrument or a non-binding instrument to best protect human rights, democracy and the rule of law:

	Binding instrument	Non-binding instrument	No opinion
Human rights, democracy and rule of law impact assessments			
Certification and quality labelling			
Audits and intersectional audits*			
Regulatory sandboxes			
Continuous automated monitoring			

39. If any other mechanism(s) should be considered, please list them and mention if they should be part of either a binding or non-binding instrument:

A risk-based approach helps to set the parameters for particularly high-risk applications of AI which should be subject to further regulation. At the same time the analysis of risk should be more nuanced.

Key factors for inclusion in a risk assessment:

(1) the likelihood/probability of the occurrence of a certain use of AI;

(2) the impact of that application; acknowledgement that any application of AI can potentially be high risk depending on the specific purpose for which it is used, i.e., recommender systems in music streaming might be categorised as 'low-risk' but should a streaming-

app use speech recognition to detect emotional state or gender etc. this would be a high-risk application;

(3) user choice, whether an individual has the ability to choose not to be subject to the AI application, i.e., in applying for a job that you need you have little choice but to be subject to a recruitment process that may deploy AI.

Process towards a risk-based assessment:

(1) the State should set the parameters of what constitutes a risk and what processes, processes, procedures and safeguards should apply in each case;

(2) companies may do more than that which a government requires, and adopt additional safeguards;

(3) governments should not take such decisions alone, the categorisation of risks should involve a robust multi-stakeholder process and in particular allocate resources to ensure dialogue and feedback from at-risk or vulnerable groups most likely to suffer the adverse impacts of the application of high-risk AI.

Auditing and impact assessments:

Risk-based approaches are based on predicted outcomes. Given the complexity of and constant evolution of the applications of AI, in addition to such an *ex ante* analysis, *ex post* human rights impact assessments can be a crucial tool to assess the actual impact. These impact assessments should be analysed for trends that can inform future risk assessments.

Auditing applications of AI for discriminatory and other adverse impacts is also an important tool. National authorities/regional laws can and should set the parameters that the audit should entail, as well as which specific harms that audit should seek to uncover. Companies may have overall responsibility that such an audit is carried out, but an independent third party with relevant expertise should conduct the audit. The State should set out clear rules to ensure the independence, competence of such third-party auditors. The obligation and basic procedures to guarantee a multistakeholder consultative process should also be mandated by law. There will be situations where it is more appropriate that a State authority itself has investigatory powers to check certain applications of AI. For example, you could imagine a situation where national equality bodies are mandated to investigate discrimination in the allocation of social security benefits by a Government Department. In addition, the Convention should provide a legal framework that enables privacy-preserving access to research data for third parties such as academic researchers and civil society. This can add an additional layer of oversight.

CENTER FOR DEMOCRACY & TECHNOLOGY

40. In your opinion, how useful would the following follow-up activities be if implemented by the Council of Europe?

	Not useful	Rather not useful	Indifferent/ no opinion	Rather useful	Highly useful
Monitoring of AI legislation and policies in Member States					
Capacity building on Council of Europe instruments, including assistance to facilitate ratification and implementation of relevant Council of Europe instruments					
AI Observatory for sharing good practices and exchanging information on legal, policy and technological developments related to AI systems					
Establishing a centre of expertise on AI and human rights					

41. What other mechanisms, if any, should be considered?

The development of standards and practices in relation to the auditing of AI for discrimination in particular.