Center for AI and Digital Policy

Center for AI & Digital Policy (CAIDP) Statement on Proposed EU AI Regulation

28 July 2021

CAIDP welcomes the opportunity to provide feedback as amendments to the text of draft Artificial Intelligence Act ("Proposal") are considered.[1] This statement follows from CAIDP's Statement to the European Commission, the European Parliament, and the European Council on April 20, 2021.[2]

As we wrote in our initial statement on the Proposal, "this initiative may be the single most important legal framework for the digital economy to ensure the protection of fundamental rights." We also wish to express support for the work of the European Commission to seek public comment on the Proposal. We believe that meaningful public participation in the development of AI strategies is a key indicator of the health of democratic institutions.

The **Center for AI and Digital Policy** is a global research organization. In 2020 we published ***Artificial Intelligence and Democratic Values*****,** a comprehensive review of the AI policies and practices in 30 countries.[3] We also created a methodology to assess AI national strategies. Our aim is to promote a world where technology promotes broad social inclusion based on fundamental rights, democratic institutions, and the rule of law.

In the CAIDP 2020 report ***AI and Democratic Values***, we stated:

1. *Countries must establish national policies for AI that implement democratic values*

2. *Countries must ensure public participation in AI policymaking and also create robust mechanisms for independent oversight of AI systems*

3. *Countries must guarantee fairness, accountability, and transparency in all AI systems*

---

[1] European Commission, *Artificial intelligence – ethical and legal requirements, Give Feedback,* https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements_en. See also CAIDP, Public Voice, *European Commission Seeks Comments on AI Regulation* (posted May 31, 2021), https://www.caidp.org/public-voice/

[2] CAIDP Statement to European Commission, European Parliament, and European Council regarding the Draft EU AI Regulation (Apr. 20, 2021), https://www.caidp.org/app/download/8312964663/CAIDP-EU-AI-20042021.pdf

[3] *Artificial Intelligence and Democratic Values* (CAIDP 2020), https://www.caidp.org/aisci-2020/

4. Countries must commit to these principles in the development, procurement, and implementation of AI systems for public services

5. Countries must halt the use of facial recognition for mass surveillance

Our assessment of the draft EU AI Regulation is favorable. The draft EU AI Regulation will promote AI policies and practices and will bring more transparency to allow for ongoing evaluation and monitoring. We also recognize the important step forward in the evolution of the internal market with the integration of fundamental rights compliance. As AI systems become more critical role for the digital economy, compliance with fundamental rights should be a necessary precondition for market participation.[4]

We further fully support the objective to prohibit certain AI systems. Our review of country AI practices found that the clearest distinction between AI systems in authoritarian countries and AI systems in democratic countries is the use of facial recognition for mass surveillance. Such indiscriminate ongoing surveillance is intended precisely to coerce social behavior and to control populations. This AI technique has been used against political protesters and religious minorities, and will almost certainly be more widely deployed unless a clear prohibition is adopted.[5]

Below are the further recommendations of the CAIDP for the deliberations of the EU AI Regulation.

PROHIBITED CASES

CAIDP recommends that all prohibited use cases for AI systems should apply equally to both private and public entities. If a system is a risk to the fundamental rights of an individual and is detrimental to society, then it does not matter which entity uses it and it should be prohibited in both remits.

AI system that deploys subliminal techniques beyond a person's consciousness; AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability:

- A subliminal technique is by its definition not detectable by the person impacted, informed consent is not possible, nor is it possible for an individual to prove that his/her/their behavior was materially distorted.

[4] CAIDP Statement (Apr. 20, 2021).
[5] *Id.*

- It is also not clear from the wording who makes the determination that a practice is 'subliminal', 'materially distorting', 'likely' to cause harm. This provision should be clarified.

- The wording of the prohibited use case with such narrowed scope to define makes it practically impossible to effectively ban any practice or protect any individual from exploitation of vulnerabilities. Examples should be provided to help make clear which practices will be prohibited.

The UN Convention on the Rights of Persons with Disabilities[6] states that 'Persons with disabilities include those who have long-term physical, mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others.' The UN Convention recommends "Universal design" - design of products, environments, programmes and services to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. Same definition has been adopted by the CJEU in the Z case.[7]

CAIDP recommends that the right to accessibility should apply to everyone, as various barriers may hinder full and effective participation in society on an equal basis with others.

Use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement

- The wording of the prohibited use case does not cover the use of a system by private companies in public spaces; nor does it prohibit non-real time biometric identification systems. Even when not done in real-time, the collection of data to be analyzed at a later stage and "any" biometric identification in public spaces can negatively impact a person's expectation of being anonymous and the fundamental rights to express one's self and freedom of association and freedom of movement.

- Such narrow scope of the definition of prohibited practice makes it useless in practice. We support the intent of the Article. We are concerned that the text does not fulfill the purpose of the Article.

CAIDP recommends a ban on biometric recognition systems used for mass surveillance purposes. CAIDP also made this recommendations in our report ***Artificial Intelligence and***

[6] UN Convention on the Rights of Persons with Disabilities (2007), https://www.ohchr.org/EN/HRBodies/CRPD/Pages/ConventionRightsPersonsWithDisabilities.aspx#2

[7] CJEU Decision on Z Case (2014), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:62012CJ0363

***Democratic Value**s (CAIDP 2020). This ban should not only be limited to facial recognition systems, but also new /emerging forms of biometric recognition analyses such as gait, voice, etc.*

If biometric identification systems are to be used by law enforcement for specific investigation purposes 'after' due process is followed and 'for the specific location and target person(s)', there should also be a time limit of how long these records are retained. The regulation should clearly state that the records cannot be retained infinitely. *There should also be independent oversight of the deployment, management, and termination of these AI-based surveillance techniques.*

Evaluation or classification of the trustworthiness of natural persons over a certain period based on their social behavior or known or predicted personal or personality characteristics, with the social score

- The wording of the prohibited use case does not cover the use of this system by private companies for their own commercial purposes, or as a service provided to other private companies (for example risk scoring of individuals for online behavior, employability scoring based on online behavior) or other risk scoring systems used by public entities, such as for criminal sentencing.

- Paragraph 17 states that 'AI systems **evaluate or classify the trustworthiness of natural persons based on their social behavior** in multiple contexts or known or predicted personal or personality characteristics. The social score obtained from such AI systems may lead to the detrimental or unfavorable treatment of natural persons or whole groups thereof in social contexts, which are **unrelated to the context in which the data was originally generated or collected or to a detrimental treatment that is disproportionate or unjustified to the gravity of their social behavior**. Such AI systems **should be therefore prohibited**." However, despite the above acknowledgement, Proposal also allows credit scoring systems and scoring of personality (in recruitment) by classifying them as 'high-risk' systems. The explanations at the beginning of the Proposal conflict with the articles of the Proposal in this sense. It is a fact that majority of these systems incorporate behavioral and social data in their models from a wide range of data sources usually unrelated to the context of their eventual use.

CAIDP recommends a ban on any score-based profiling of individuals by private companies that is not fully compliant with all legal obligations prior to deployment and for which an algorithmic impact assessment has not been conducted ex ante.

HIGH RISK USE CASES

Biometric identification and categorisation of natural persons:

- AI systems categorizing individuals from biometric data into groups according to race, ethnicity, gender, political or sexual orientation, or similar provides for grounds for discrimination under Article 21 of the Charter.
- Such categorization assumes universal traits and pre-decided categories and transforms social constructs such as race, ethnicity, gender, political or sexual orientation into 'objective' truths. It removes one of the most fundamental of human rights – right to express one's identity.
- Categorization systems do not have any scientific validity as they are built on constructed concepts.

CAIDP similarly recommends a ban online categorization and scoring of individuals using biometric features (facial features, voice, DNA data) by both public and private entities.

Although we appreciate that the Proposal states that the classification of an AI system / use case as high risk does not make a particular AI system lawful, we are concerned that inclusion of certain use cases in high-risk AI systems, such as for criminal sentencing, will permit such AI systems even if they lack a scientific basis or have historical roots in discrimination against people. Inclusion of these systems as high risk, but without a clear prohibition, could legitimize or normalize certain improper practices. Although the Proposal acknowledges 'the exercise of important procedural fundamental rights, such as the right to an effective remedy and to a fair trial as well as the right of defense and the presumption of innocence, could be hampered, in particular, where such AI systems are not sufficiently transparent, explainable and documented', it does not require these systems to have any scientific validity or replicability in their outcomes.

Law enforcement:

- The danger of categorization of natural person is particularly critical for AI systems used by law enforcement authorities for individual **risk assessments**, **polygraphs** and similar tools or to detect the **emotional state of natural person** and to **predict** the occurrence or reoccurrence of an actual or **potential criminal offence based on profiling** of natural persons, or **assessing personality traits and characteristics or past criminal behaviour of natural persons or groups. Considering these use cases as high risk only and not prohibited practices will have** chilling effects on democratic institutions, presumption of innocence, due process.

- The Proposal acknowledges that 'Actions by law enforcement authorities involving certain uses of AI systems are characterised by a significant degree of **power imbalance** and may lead to surveillance, arrest or **deprivation of a natural person's liberty as well as other adverse impacts on fundamental rights** guaranteed in the Charter'. However, still gives law enforcement the ability to use systems to '**predict'** a crime.

- The Proposal (in Paragraph 17) acknowledges that 'AI systems **evaluate or classify the trustworthiness of natural persons based on their social behavior** in multiple contexts or known or predicted personal or personality characteristics. The social score obtained from such AI systems may lead to the detrimental or unfavorable treatment of natural persons or whole groups thereof in social contexts, which are **unrelated to the context in which the data was originally generated or collected or to a detrimental treatment that is disproportionate or unjustified to the gravity of their social behavior**. Such AI systems **should be therefore prohibited**." However, there is again conflict within the body of Proposal since it then classifies as high-risk (and hence allows the practice) when law enforcement uses "AI systems for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data."

CAIDP recommends a ban on the use of pseudoscientific AI systems to detect emotional state of a natural person, as well as use of what is described as "predictive policing."

Migration, asylum and border control management:

- The Proposal improperly groups together border management, broadly including migration and asylum and security. This very approach seems to indicate that asylum seekers and refugees who are already in a vulnerable position and seeking protection and have rights assigned to their status are considered a threat. This approach forms part of an elaborate agenda on the embedment of AI this field, as indicated by the Commission report on *Opportunities and challenges for the use of artificial intelligence in border control, migration and security* published in 2020.[8] These technologies, such as iBorderCtrl, have provoked considerable and are currently under review before the Court of Justice of the European Union.[9]

- The Proposal states that 'AI systems used in migration, asylum and border control management affect people who are often in **particularly vulnerable position and who are dependent on the outcome of the actions** of the competent public authorities. The accuracy, non-discriminatory nature and transparency of the AI

[8] Deloitte, Directorate-General for Migration and Home Affairs (European Commission), *Opportunities and challenges for the use of artificial intelligence in border control, Migration and security . Volume 1, Main report* (May 2020), https://op.europa.eu/en/publication-detail/-/publication/c8823cd1-a152-11ea-9d2d-01aa75ed71a1/language-en

[9] Umberto Bacchi, *EU's lie-detecting virtual border guards face court scrutiny*, Reuters (Feb. 5, 2021), https://www.reuters.com/article/europe-tech-court/eus-lie-detecting-virtual-border-guards-face-court-scrutiny-idUSL8N2KB2GT

systems used in those contexts are therefore particularly important to guarantee the respect of the fundamental rights of the affected persons, notably their rights to free movement, non-discrimination, protection of private life and personal data, international protection and good administration'. However, again by classifying these AI systems as high-risk, the Proposal legitimizes polygraphs and other similar pseudoscientific tools, such as emotion-recognition, to be used on vulnerable populations.

- The Proposal includes 'AI systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State'.

CAIDP is concerned that despite the published protections[10], systems such as Digital Green Certificates could be used to discriminate against travelers from certain countries due to their country of origin and not necessarily their vaccination status.

CAIDP recommends that asylum seekers and refugee rights be protected on an equal basis and that these populations not become test beds or experimentation for emerging technologies, as has been the case with border control systems requiring biometric identifiers, such as fingerprints and facial images, in the recent years.

Administration of justice and democratic processes:

- CAIDP asks further clarification regarding "AI systems intended to apply the law to a concrete set of facts." If the intention is to use any natural language processing system (NLP) to review the facts, such as the deposition of witnesses, the requirement of validity of these systems (which is not a requirement in the Proposal) becomes crucial where NLP cannot understand the nuances of language, context of descriptions or analogies.

SOCIETAL RISKS

The Proposal very briefly mentions larger risks to the society but leaves out the impact of AI systems on collectives. Even mentioning these, the Proposal leaves the commentary as "Codes of conduct may also include voluntary commitments related, for example, to environmental sustainability, accessibility for persons with disability, stakeholders' participation in the design and development of AI systems, and diversity of development teams.".

[10] European Commission "Eu Digital COVID Certificate" (2021), https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/safe-covid-19-vaccines-europeans/eu-digital-covid-certificate_en

CAIDP notes further risks to the society as below and recommends measures to address these concerns in the final version of the Proposal:

- **Environment**: Require AI system providers to document impact of large AI systems (especially training systems) on the environment, emission, and waste. This documentation should be part of transparency requirements. Declaration on A Green and Digital Transformation of the EU[11] requires deploying and investing more green digital technologies to achieve climate neutrality. Transparency requirements would greatly benefit the realization of twin objectives.

- **Group discrimination**: Biased recognition, scoring and categorization systems based on unscientific methods can lead to further marginalization of certain minorities (in terms of ethnic / income / educational backgrounds, as well as gender identities). Proposal focuses on individual rights but leaves out discrimination / stigmatization of groups as a whole. UN Special Rapporteur Alston detailed his concerns about specific targeting of poor and other vulnerable groups by public authorities via AI systems.[12]

- **Misinformation:** Prohibition of use of AI systems to mis/disinform citizens and manipulate their political and social interactions

- **Children's rights:** Protection / exclusion of children from surveillance, recognition, and data collection systems all together.

- **Disability rights:** Requiring accessibility from all AI systems, so as not to treat people with disabilities as errors, outliers or edge cases in the development of these systems. In addition, Proposal's current wording assumes that anyone with a disability is 'vulnerable' – a framing that disrespects human dignity and the diversity of people.

To repeat again our earlier comment, the UN Convention on the Rights of Persons with Disabilities[13] states that 'Persons with disabilities include those who have long-term physical,

[11] European Commission, *Declaration on A Green and Digital Transformation of the EU*: https://digital-strategy.ec.europa.eu/en/news/eu-countries-commit-leading-green-digital-transformation

[12] Report of the UN Special Rapporteur on extreme poverty and human rights, *Digital technology, social protection and human rights* (October 2019). https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx

[13] *UN Convention on the Rights of Persons with Disabilities* (2007), https://www.ohchr.org/EN/HRBodies/CRPD/Pages/ConventionRightsPersonsWithDisabilities.aspx#2

mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others.' The Convention recommends "Universal design" - design of products, environments, programmes and services to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. Universal design applies to everyone. *No one should experience barriers that may hinder their full and effective participation in society on an equal basis with others*.

DATASETS

Training, validation and testing data sets shall be relevant, representative, free of errors and complete.

- The Proposal requires examination in view of possible biases. However, it should be noted that bias does not only manifest itself in data sets but also the model design, selection of performance metrics, and also in implementation by the operators.

CAIDP recommends that the Proposal explicitly include 'examination of bias in model design and selection of performance metrics.'

- It is also impossible for a dataset to be 'free of errors'. The wording in the Proposal requires clarification as to point to the outcomes / decisions of AI systems to be free of errors, and not the dataset itself.

CAIDP recommends that the wording amended to 'AI system decision error rates across protected categories should be transparent, made publicly available, along with a statement from Provider as to why that error rate was acceptable level for the AI system to be put into market.'

CONFORMITY ASSESSMENTS

CAIDP recommends conformity assessments for high-risk systems be conducted by certified independent third parties who shall not have any conflict of interest.

- The Proposal currently allows high-risk AI systems to be risk and conformity assessed by the very people and entities who have a vested interest and investment in the sale of these products. The transparency obligations are a great step towards scrutiny. However, the self-conducted ex-ante conformity assessments are by no means an effective way to protect fundamental rights or the individuals and society from the possible harms of these systems.

- The Proposal states 'common mandatory requirements applicable to the design and development of certain AI systems before they are placed on the market that will be

further operationalised through harmonised technical standards.' In separate documentation,[14] European Commission has already made its leaning on which entities should set these standards. Two of those entities, CEN (European Committee for Standardization) and CENELEC (European Committee for Electrotechnical Standardisation)[15] [16] have already moved forward and established a CEN-CENELEC Joint Technical Committee 21 'Artificial Intelligence'.

CAIDP is concerned that business interests will dominate these standard-setting organizations. and that the protection of fundamental rights; and that the voices and concerns of civil society and affected communities will not be effectively represented. CAIDP therefore recommends that these organizations publish annually reports that describe specifically steps taken to ensure broad-based participation in the development of technical standards as well as the consideration of fundamental rights

.

AI systems already in use by the time of Proposal coming into force

CAIDP recommends that any AI system that requires safety and / or conformity assessment be included in the scope and no exemptions be provided to those AI systems already in use by the time the Proposal comes into force. Exclusion might result in certain systems to be pushed into live environments without the due diligence, proper development or risk assessments in place in an effort to beat the dates. Exclusion of AI systems might also leave many AI systems, which have impact on fundamental rights, without any oversight.

TRANSPARENCY OBLIGATIONS

In 5.2.4. of Proposal states that 'Transparency obligations will apply for systems that (i) interact with humans, (ii) are used to **detect emotions** or **determine association with (social) categories based on biometric data**. When persons interact with an AI system or their emotions or characteristics are recognized through automated means, people must be informed of that circumstance...This allows persons to make informed choices or step back from a given situation' is in effect not true or practical." In reality, this obligation puts the onus of understanding the implications of these systems on each separate individual who might or might not have the knowledge, expertise or capacity to do so. It also does not address the

[14] European Commission, *Rolling Plan for ICT standardization*, https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/artificial-intelligence
15 European Committee for Standardization, *European Standards support the EU ambitions on Artificial Intelligence* (2020), https://www.cen.eu/news/brief-news/Pages/NEWS-2021-012.aspx
16 European Committee for Standardization, *CEN and CENELEC launched a new Joint TC on Artificial Intelligence* (2021), https://www.cencenelec.eu/news/brief_news/pages/tn-2021-013.aspx

power imbalance between the implementer of the system and the individual. For example, a job candidate who is informed of an interaction with an employer is in no position to step back or refuse to interact, or for example a refugee in a camp whose access to resources is dependent on being subject to these systems. cannot step away from a given situation. Finally, even if it is transparent, this approach allows for 'emotion recognition' and 'categorization of individuals using biometric data.'

CAIDP repeats the recommendation for a ban online categorization and scoring of individuals using biometric features (facial features, voice, DNA data) by both public and private entities.

NOTABLE PROVISIONS FROM UNESCO AI ETHICS RECOMMENDATION TO BE CONSIDERED FOR EU AI ACT

As the work of the European Union on AI policy has moved forward, so too has work on AI policy in international organizations. We would like to call your attention to several provisions in the recently finalized draft of UNESCO Recommendation on the Ethics of AI[17] that could be incorporated in the EU Proposal.

- Persons may interact with AI systems throughout their lifecycle and receive assistance from them…Within such interactions, **persons should never be objectified, nor should their dignity be otherwise undermined**, or human rights and fundamental freedoms violated or abused (Rec #15)

- All actors involved in the lifecycle of AI systems must comply with applicable international law and domestic legislation, standards and practices. They should **reduce the environmental impact** of AI systems. (Rec #18)

- Respect, protection and promotion of human dignity and rights is essential throughout the life cycle of AI systems. Human dignity relates to the **recognition of the intrinsic and equal worth of each individual human being, regardless of race, color, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, or disability and any other grounds** (Rec #19)

- Peace, inclusiveness and justice, equity and interconnectedness should be promoted throughout the lifecycle of AI systems, in so far as the processes of the lifecycle of **AI systems should not segregate, objectify or undermine freedom and autonomous decision-making as well as the safety of human beings and communities, divide and**

[17] UNESCO Recommendation on the Ethics of AI (2021), https://unesdoc.unesco.org/ark:/48223/pf0000377897

turn individuals and groups against each other, or threaten the coexistence between humans, other living beings and the natural environment. (Rec #24)

- The choice to use AI systems and which AI method to use should be justified in the following ways: (a)the AI method chosen should be **appropriate and proportional to achieve a given legitimate aim**; (b) the AI method chosen should not infringe upon the foundational values captured in this document, in particular, **its use must not violate or abuse human rights**; and (c)the **AI method should be appropriate to the context** and **should be based on rigorous scientific foundations**. In scenarios where decisions are understood to have an impact that is irreversible or difficult to reverse or may involve life and death decisions, final human determination should apply. In particular, AI systems **should not be used for social scoring or mass surveillance purposes** (Rec #26)

- AI actors should make **all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes** throughout the lifecycle of the AI system to ensure fairness of such systems. **Effective remedy should be available against discrimination and biased algorithmic determination**. (Rec #29)

- People should be fully informed when a decision is informed by or is made on the basis of AI algorithms, including when it affects their safety or human rights, and in those circumstances should have the **opportunity to request explanatory information from the relevant AI actor or public sector institutions**. In addition, individuals should be able to access the reasons for a decision affecting their rights and freedoms and have the option of making submissions to a designated staff member of the private sector company or public sector institution able to review and correct the decision. (Rec #38)

- Appropriate oversight, **impact assessment, audit and due diligence mechanisms, including whistle-blowers' protection, should be developed** to ensure accountability for AI systems and their impact throughout their lifecycle. (Rec #43)

- Governments should adopt a regulatory framework that sets out a procedure, **particularly for public authorities, to carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation and address societal challenges**. The assessment should also **establish appropriate oversight mechanisms, including auditability, traceability and explainability**, which enable the assessment of algorithms, data and design processes, as well as include external review of AI systems. Ethical impact assessments should be transparent and open to the public, where appropriate. Such assessments should also be **multidisciplinary, multi-stakeholder, multicultural, pluralistic and inclusive**. The **public authorities should be required to monitor the AI systems**

implemented and/or deployed by those authorities by introducing appropriate mechanisms and tools (Rec #53)

- Member States that acquire AI systems for human rights-sensitive use cases, such as law enforcement, welfare, employment, media and information providers, health care and the independent judiciary system should provide **mechanisms to monitor the social and economic impact of such systems by appropriate oversight authorities, including independent data protection authorities, sectoral oversight and public bodies responsible for oversight**. (Rec #62)

Thank you for your consideration of our views. We would welcome the opportunity to speak with you further about these recommendations.

Merve Hickok
CAIDP Research Director

Marc Rotenberg
CAIDP President