epic.org

Electronic Privacy Information Center
1519 New Hampshire Avenue NW
Washington, DC 20036, USA

+1 202 483 1140
+1 202 483 1248
@EPICPrivacy
https://epic.org

FEEDBACK OF THE ELECTRONIC PRIVACY INFORMATION CENTER

to the

Organisation for Economic Cooperation and Development

Regarding the

OECD Framework for Classifying AI Systems

June 30th, 2021

EPIC submits the following feedback to the Organisation for Economic Cooperation and Development's Public Consultation on the OECD Framework for Classifying AI Systems (hereinafter, the "Framework").[1]

EPIC is a public interest research center in Washington, D.C. that was established in 1994 to focus public attention on emerging privacy and related human rights issues and to protect privacy, the First Amendment, and constitutional values.[2] EPIC has a long history of promoting transparency and accountability for information technology.[3]

EPIC has a particular interest in promoting algorithmic transparency and has consistently advocated for the adoption of the Universal Guidelines for AI ("UGAI") to promote trustworthy algorithms.[4] EPIC has advocated for transparency and accountability internationally, litigating cases

[1] Public consultation on the OECD Framework for Classifying AI Systems, Organisation for Economic Cooperation and Development (June 2021), https://oecd.ai/classification.
[2] EPIC, *About EPIC* (2019), https://epic.org/epic/about.html.
[3] EPIC, *Algorithmic Transparency* (2018), https://www.epic.org/algorithmic-transparency/; EPIC, *Algorithms in the Criminal Justice System* (2018), https://www.epic.org/algorithmic-transparency/crim-justice/; Comments of EPIC, *Consumer Welfare Implications Associated with the Use of Algorithmic Decision Tools, Artificial Intelligence, and Predictive Analytics*, Federal Trade Commission (Aug. 20, 2018), https://epic.org/apa/comments/EPIC-FTC-Algorithmic-Transparency-Aug-20-2018.pdf; Comments of EPIC, *Developing UNESCO's Internet Universality Indicators: Help UNESCO Assess and Improve the Internet*, United Nations Educational, Scientific and Cultural Organization ("UNESCO") (Mar. 15, 2018), 5-6, https://epic.org/internetuniversality/EPIC_UNESCO_Internet_Universality_Comment%20(3).pdf.
[4] *See* e.g. EPIC v. DOJ (D.C. Cir.) (18-5307), https://epic.org/foia/doj/criminal-justice-algorithms/; Comments of EPIC, *Intellectual Property Protection for Artificial Intelligence Innovation*, U.S. Patent and Trademark Office (Jan. 10, 2020), https://epic.org/apa/comments/EPIC-USPTO-Jan2020.pdf; Comments of EPIC, *HUD's Implementation of the Fair Housing Act's Disparate Impact Standard*, Department of Housing and Urban Development (Oct. 18, 2019), https://epic.org/apa/comments/EPIC-HUD-Oct2019.pdf; Testimony of EPIC, Massachusetts Joint Committee on the Judiciary (Oct. 22, 2019), https://epic.org/testimony/congress/EPIC-FacialRecognitionMoratorium-MA-Oct2019.pdf; Statement of EPIC, *Industries of the Future*, U.S. Senate Committee on Commerce, Science & Transportation (Jan. 15, 2020), https://epic.org/testimony/congress/EPIC-

against the U.S. Department of Justice to compel production of documents regarding "evidence-based risk assessment tools"[5] and against the U.S. Department of Homeland Security to produce documents about a program to assess the probability that an individual commits a crime.[6] In 2018, EPIC and leading scientific societies petitioned the U.S. Office of Science and Technology Policy to solicit public input on U.S. Artificial Intelligence Policy.[7] EPIC submitted comments urging the National Science Foundation to adopt the UGAI and to promote and enforce the UGAI across funding, research, and deployment of U.S. AI systems.[8] EPIC has also recently submitted comments to the National Security Commission on Artificial Intelligence, the U.S. Office of Science and Technology Policy, the European Commission, and the U.S. Office of Management and Budget urging adequate regulation to protect individuals.[9]

In an effort to establish necessary consumer safeguards, EPIC recently filed FTC complaints against HireVue,[10] an employment screening company, and AirBnB,[11] the rental service that claims to assess risk in potential renters based on an opaque algorithm. EPIC has also filed a petition with the FTC for a rulemaking for AI in Commerce.[12] EPIC recently published the *AI Policy Sourcebook*, the first reference book on AI policy.[13]

SCOM-AI-Jan2020.pdf; Comments of EPIC, *Request for Information: Big Data and the Future of Privacy*, Office of Science and Technology Policy (Apr. 4, 2014), https://epic.org/privacy/big-data/EPIC-OSTP-Big-Data.pdf.

[5] EPIC, *EPIC v. DOJ (Criminal Justice Algorithms)*, https://epic.org/foia/doj/criminal-justice-algorithms/.

[6] *See Id.* and EPIC, *EPIC v. DHS (FAST Program)* https://epic.org/foia/dhs/fast/.

[7] EPIC, Petition to OSTP for Request for Information on Artificial Intelligence Policy (July 4, 2018), https://epic.org/privacy/ai/OSTP-AI-Petition.pdf.

[8] EPIC, Request for Information on Update to the 2016 National Artificial Intelligence Research and Development Strategic Plan, National Science Foundation, 83 FR 48655 (Oct. 26, 2018), https://epic.org/apa/comments/EPIC-Comments-NSF-AI-Strategic-Plan-2018.pdf.

[9] Comments of EPIC, *Solicitation of Written Comments by the National Security Commission on Artificial Intelligence*, 85 Fed. Reg. 32,055, National Security Commission on Artificial Intelligence (Sep. 30, 2020), https://epic.org/apa/comments/EPIC-comments-to-NSCAI-093020.pdf; Comments of EPIC, *Request for Comments on a Draft Memorandum to the Heads of Executive Departments and Agencies, "Guidance for Regulation of Artificial Intelligence Applications,"* 85 Fed. Reg. 1825, Office of Management and Budget (Mar. 13, 2020), https://epic.org/apa/comments/EPIC-OMB-AI-MAR2020.pdf; Comments of EPIC, *Request for Feedback in Parallel with the White Paper on Fundamental Rights*, European Commission Fundamental Rights Policy Unit (May 29, 2020), https://epic.org/apa/comments/EPIC-EU-Commission-AI-Comments-May2020.pdf; Comments of EPIC, *Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence*, European Commission (Sep. 10, 2020), https://epic.org/apa/comments/EPIC-EU-Commission-AI-Sep2020.pdf.

[10] Complaint and Request for Investigation, Injunction, and Other Relief, *In re HireVue* (Nov. 6, 2019), https://epic.org/privacy/ftc/hirevue/EPIC_FTC_HireVue_Complaint.pdf.

[11] Complaint and Request for Investigation, Injunction, and Other Relief, *In re Airbnb* (Feb. 27, 2019), https://epic.org/privacy/ftc/airbnb/EPIC_FTC_Airbnb_Complaint_Feb2020.pdf.

[12] *In re: Petition for Rulemaking Concerning Use of Artificial Intelligence in Commerce*, EPIC (Feb. 3, 2020), https://epic.org/privacy/ftc/ai/EPIC-FTC-AI-Petition.pdf.

[13] *EPIC AI Policy Sourcebook 2020* (EPIC 2020), https://epic.org/bookstore/ai2020/.

The OECD AI Principles[14] were adopted in 2019 and endorsed by 42 countries—including several European Countries, the United States, and the G20 nations.[15] The OECD AI Principles establish international standards for AI use, centered on the following principles:

1. Inclusive growth, sustainable development, and well-being;
2. Human-centered values and fairness;
3. Transparency and explainability;
4. Robustness, security, and safety; and
5. Accountability.[16]

EPIC urges the OECD to make three key modifications to this framework. First, the OECD should eliminate the distinction between optional and required criteria, such that all framework criteria are required. Second, the OECD should expand the criteria to include more information and consideration relating to bias and fairness of the system. Third, the OECD should clearly articulate the necessary resources for meaningful use and enforcement around this Framework to maximize its utility in helping governments prioritize protecting individuals.

EPIC responses to Key Questions from the OECD

Commented [AB1]: A lot of visual space under this heading and in between paragraphs in this section, consider adjusting format.

1. **Should there be a distinction between core and non-core criteria? In other words, should there be a core classification framework for less-technical audiences plus additional considerations for more technical and informed users**?

There should not be a distinction between core and non-core criteria for the Framework for two reasons. First, assessors completing multiple assessments, overwhelmed by volume, or pressed for time are unlikely to expend extra effort or resources to complete Framework portions considered "optional." Several of the criteria currently designated within the Framework as non-core are essential to considering and understanding potential disparate impacts or unfair uses of AI systems. If assessors are able to choose not to respond to these without penalty, critical questions will remain unanswered and the Framework's protective value will be diminished.

Second, the current Framework describes the non-core criteria as relating to more complex technical questions about the systems being assessed. We did not find this to be the case; both core and non-core criteria in the Framework relate to technical assessments. As it currently stands, the distinction between core and non-core questions is largely arbitrary.

And, regardless of their core or non-core designation, all questions should be kept simple and straightforward. This is necessary both to avoid confusion on the part of assessors and to ensure that responses are comprehensible to any potential future reviewers of the assessment. The actual format of certain questions, such as Question 9 in the "Data and Input" section, are vague and will make it difficult to interpret the responses. For example, Question 9 in "Data and Input" contains a "low," "medium," and "high" option for users to select that corresponds to statements, such as "data is not

[14] *Recommendation of the Council on Artificial Intelligence*, OECD (May 21, 2019) [hereinafter *OECD AI Principles*], https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
[15] *U.S. Joins with OECD in Adopting Global AI Principles*, NTIA (May 22, 2019), https://www.ntia.doc.gov/blog/2019/us-joins-oecd-adopting-global-ai-principles.
[16] *OECD AI Principles*, *supra* note 12.

often incorrect or distorted." These response options will make it difficult for users of the Framework to understand what is being asked and difficult to confirm that their answer is coherent, both because "high," "medium," and "low" options are typically designations of probability and because the statement itself is a double negative and does not clearly correspond to probability. Adjusting the question to read, for example, "what is the likelihood of incorrect or distorted data," would more clearly connect to the response options available.

The current Framework would require an evaluator to have access to individuals with a very detailed understanding of an AI system. Such access is not likely available to someone attempting to perform an independent evaluation and would make it difficult for anyone not specifically trained in AI to complete the evaluation. An assessor would likely be required to consult either detailed technical literature on the AI system in question or speak with a developer or high-level user of the system in order to accurately and completely answer many of the questions in this Framework.

Finally, our experience shows that the estimated time to complete an assessment under the Framework is not realistic. A meaningful response to the Framework, as currently written, would require significantly more time than the estimated time of 15-20 minutes and would likely require multi-party collaboration. This doesn't change when only using core criteria, and, even if limiting solely to core criteria lowered the amount of time it takes to complete the evaluation exercise, such a limitation would decrease the value of the tool and risk reducing it to a box-checking exercise.

One possible path to cater to an audience looking for a more simplified version (while still allowing the option to see the full report for all) would be to mandate completion of all questions but prioritize the display of some core criteria at the top of published reports.

2. Which characteristics should be core criteria and which 'optional'?

Commented [AB2]: Should be fixed by adjusting spacing above but make sure not to have widowed headings at the bottom of pages

None of the criteria listed in this Framework should be optional. The criteria currently classified as non-core are essential to fully understanding the AI system and properly evaluating risk to users, particularly the criteria including business function; scale of deployment; users of AI system; impacted stakeholders; optionality; business model; benefits and risk to human rights and democratic values; and benefits and risks to well-being.

Treating these criteria as secondary and, in many cases, optional, would improperly signal to companies that these considerations are less important. All the questions included in the Framework are important and companies should not be encouraged to skip questions that may help point individuals, advocates, and regulatory bodies to problematic uses of AI inhibits consumer protection efforts. This signal is strongly inconsistent with the OECD guidelines of transparency, accountability, explainability, and more.[17]

[17] *Id*.

3. Can AI systems be classified with the core criteria both consistently and reliably?

Commented [AB3]: Same comment as above on spacing

AI systems cannot be classified consistently and reliably while still being meaningfully evaluated using only the core criteria listed in the Framework. There are too many questions where the structure would not allow for adequate nuanced responses.

There are also several questions – for example, Question 1 under "Context" - that only allow for a single response even though many systems could fall under multiple sectors or require more complex responses. We recommend allowing for more detailed and complex responses with a guide explaining answer options wherever possible.

4. **Which criteria are useful for a more detailed and technically-oriented framework?**

The criteria currently included by the OECD allow for an adequately technically-oriented result. In order to yield a more useful result, however, OECD should consider expanding the options for potential answers to reflect the reality that many AI tools are not used for a single purpose and it is unlikely that a system's effects will be siloed to one sector.

Further, several questions are ambiguous and could lead to a confusing or unclear assessment that fail to flag potential harm or risk of harm. For example, Question 7 of the Context section of the Framework does not clearly define what is meant by each area or value ("Liberty, safety, and security," "Right to property," etc.), which allows for significant subjectivity and may enable companies to evade accountability or scrutiny.

The Framework should build on the models of existing assessment systems, such as Canada's Algorithmic Impact Assessment Tool,[18] which attempts to get at the desired result (flagging potentially harmful systems) of OECD Context Questions 7 and 11 in a more direct way that is less susceptible to conclusory ethics-washing tactics. A few examples of questions in the Canadian tool include prompts to evaluate the stakes of decisions the system in question makes, vulnerability of subjects, whether it is a predictive risk assessment, and allowing for multiple sectors or categories when describing what functions the system uses.[19] Other aspects of the Canadian assessment require identification of the downstream processes of a system. These include asking (i) will the system only be used to assist a decision-maker; (ii) will the system be replacing a decision that would otherwise be made by a human; (iii) will the system be replacing human judgment; (iv) whether the system is being used by the same entity that developed it; and (v) consideration and explanation about both economic and environmental impacts.[20]

Rather than trying to increase the technical complexity of the Framework, EPIC urges the OECD to improve questions to increase the usefulness of the assessments in determining risk in line with the OECD AI Principles of explainability, accountability, transparency, and fairness.

[18] Canada Digital Services, Algorithmic Impact Assessment (last visited June 9, 2021), https://open.canada.ca/aia-eia-js/?lang=en.
[19] *Id.*
[20] *Id.*

5. Should there be criteria and classifications that are specific to industries or application domains, e.g. depending on context?

Although additional criteria and classifications for specific industries or applications could be helpful, it may be logistically infeasible to meaningfully design questions for specific industries in a way that is both comprehensible to a wider audience and not unduly deferential to those industries. Overspecification runs the risk of overcomplicating the systems of a given industry and evading further regulation. Instead, improvements to the single main Framework can be made to adequately assess a majority of systems.

Additional Comments from EPIC

The OECD is in an excellent position to urge governments to adopt this Framework in a way that empowers regulators, allows individuals to understand what systems are used on them, and forces companies to reckon with mindful development and deployment. The OECD successfully secured 44 national adherents to the AI principles in 2019 and the OECD can now help those governments follow through on their commitment to the OECD principles by actualizing this proposed Framework.[21]

This should include accounting for the subjective nature of many of the most important questions contained within the assessment. The assessment should be structured in a way that ensures key considerations will be honestly and fully addressed, taking into consideration that the assessors will often include individuals who stand to profit from the continued use of the system and who may be incentivized to give incomplete or misleading information. In addition, companies whose systems receive high-risk conclusions on their assessments will likely not wish to devote resources toward compliance without monetary and operational consequences enforced by governments. Meaningful enforcement may be required to incentivize necessary changes.

Guidance on properly using the Framework should also specify which party or parties related to the AI system will be required to fill out this Framework; at what stage of use or development the Framework assessment must be completed (by the original maker or by subsequent clients of the original maker); and when the AI system may require additional assessments (after substantive change to the AI system, when a new use is proposed, etc.). We recommend that assessments be either carried out in collaboration between privacy experts or auditors and individuals with detailed knowledge of the system in question OR that assessments be carried out by the companies developing and using the system but be proactively submitted for review and discussion to auditors or enforcement bodies. Both public and private users of a given AI system should be required to perform their own assessments of the system.

Conclusion

EPIC applauds the OECD for its efforts in developing a Framework that helps standardize evaluation and increase access to information about AI systems. EPIC recommends that the OECD improve this Framework by requiring more detailed answers from companies, eliminating optional

[21] *OECD AI Principles*, *supra* note 12.

criteria, and making it clear to member-countries that the use of this Framework should be part of a broader ecosystem in order to ensure that OECD principles remain meaningful for individuals.

Respectfully Submitted,

/s/ *Calli Schroeder*
Calli Schroeder
EPIC Global Privacy Counsel

/s/*Ben Winters*
Ben Winters
EPIC Equal Justice Works Fellow