

July 27, 2021

Consumer Technology Association
Comments on
OECD Framework for Classifying AI Systems

The Consumer Technology Association (“CTA”)®¹ respectfully submits these comments in response to the Organisation for Economic Co-operation and Development (“OECD”) Framework to Classify AI Systems (the “Framework”) offering a potential classification of artificial intelligence (“AI”) systems based on their potential impact on public policy, consistent with the OECD Principles on Trustworthy AI (the “Principles”). CTA applauds OECD’s and the OECD.AI Network of Experts’ thoughtful work on these issues and the effort to establish a global governance framework for assessing the opportunities and risks presented by different AI systems.

CTA supports the development of values-based principles related to AI to further the beneficial development and use of AI systems. CTA firmly believes *AI-related principles and frameworks must be forward-looking*: setting the groundwork for a stable global policy environment that considers the interests of both AI developers and users (including the public). Given the rapid development of AI technology, frameworks that are comprehensive, nimble, and based on overarching principles for inclusive and sustainable growth and development of AI systems will best adapt to and encourage development of new technologies and uses.

¹ CTA® is the tech sector. Our members are the world’s leading innovators—from startups to global brands—helping support millions of jobs. CTA owns and produces CES®—the most influential tech event in the world.

Principles and frameworks, such as those developed by the OECD, will have lasting impact on the development of AI systems to the extent that stakeholders “buy-in.” *Uniform adoption of and respect for global AI-related principles and frameworks* is most likely if these principles and frameworks encourage the development of accurate, ethical, inclusive, and trustworthy AI, while promoting research, flexible oversight and compliance solutions, and economic incentives to innovate.

AI technology offers tremendous opportunities for human and societal development, and CTA agrees that *AI must be trustworthy*. AI can promote inclusive growth, improve the welfare and well-being of people, and enhance global innovation and productivity. But doing so requires trust in systems: from policymakers, regulators, users, and the public. The Principles and Frameworks can help build this trust. Because trust is inescapably linked to perception of risk, the Framework must enhance efficiency, effectiveness, and user satisfaction that are tied to perceived technical trustworthiness. Of course, trustworthiness varies with perceived and actual risk factors: we know that AI medical and diagnostic systems present higher risks, but music and media recommendation systems present very low (no) risk. The Framework must recognize these distinctions based on both perceived and actual risks.

Focused principles and frameworks, like those proposed by the OECD, complement existing policy to encourage AI developers to self-regulate, through the adoption of codes of conduct, voluntary standards, and best practices. When based upon clear and targeted frameworks, self-regulation can result in meaningful protection for users and profound innovation while enhancing trustworthiness. OECD has an opportunity with this Framework and the Principles to champion such an approach.

Respectfully, CTA urges the OECD to consider opportunities to strengthen the Framework by reviewing its potential application in practice. This is necessary to better understand and evaluate the potential costs, and benefits, of implementing a framework that may lead to new rules and restrictions on the development, use, and sale of AI. In order to meet the OECD’s desire to consider the Framework’s “usability and user-friendliness,” the OECD must

recognize the very significant precedent-setting outcomes of the Framework, including their potential impact on technology development and innovation.²

I. The Framework Offers Significant Benefits For AI Systems' Stakeholders and The Public, But Is Weakened By Complexity

Standardizing a framework to support best practices will facilitate research and development into new AI systems, which in turn offer significant potential benefits to the public. Standardizing international frameworks supports the development of strong, consistent industry standards and best practices. To the extent regulation develops in the future, regulators and policymakers can rely on global, agreed-upon principles and frameworks developed by the OECD to adopt uniform regulatory policies and best practices. However, an overly complex framework may create obstacles to broad adoption by key stakeholders.

A strong Framework and uniform regulatory regimes benefits AI system developers operating internationally as they create better models, understand how AI is utilized, and adopt system limitations or mitigations. A strong Framework also serves the public, as members of the public begin to better understand how their data is used, trust the systems that are developed from such data, and reap the social and economic benefits of the numerous AI systems that increase efficiency, enhance safety and protect against outward threats. Thus, a unified approach to AI, as proposed by the Framework, supports efficient development of AI systems that serve the public's interest.

To realize these benefits of efficient innovation and strengthened trust, the Framework should focus on articulating core criteria to reduce complexity and lower compliance costs. As currently developed, the Framework is complex and can be difficult to understand by both non-technical actors (e.g., compliance officers) and non-legal actors (e.g., AI engineers, product managers, leadership, etc.). It seems to anticipate a multi-level review prior to engaging in AI project. The questions asked by the Framework would require input by many individuals with varied roles and functions, as no one actor or even team within an organization would be able to readily answer the questions. Many questions are subjective, potentially resulting in weeks or months of deliberation and divergent conclusions between (or even within!) organizations.

² See OECD Framework for the Classification of AI Systems – Public Consultation on Preliminary Findings <https://www.oecd.ai/classification>.

The costs—in money and time—to understand, implement, and comply with a complex Framework (and any regulations built in accordance with the Framework) will undermine investment in innovation and public benefit. Resources spent on compliance cannot be used for research and development of new technologies and tools. While compliance costs are a reality of technology development, a simplified Framework would achieve its goals of public benefit while encouraging innovation. The framework appropriately recognizes that both perceived, and likely, risks may require different levels of review and oversight. Where the risk context is low for individuals, such as systems offering recommendations for dining or media, a multi-layered and costly oversight process is not necessary. As risks increase, a more robust review process may be appropriate and necessary to enhance user trust.

II. The Framework Should Leverage Existing Laws

Existing laws and norms at the international, multi-level, national, and sub-national level address data protection and privacy, discrimination, and consumer protection. For example, AI providers in Europe are subject to certain notice, opt-out and transparency obligations under the GDPR for any data processing arising from the use of EU subjects' data. Further, in the U.S., numerous federal and state laws protect against discriminatory behavior involving individuals' access to housing, finance, healthcare and other areas of fundamental rights. These laws are successful in protecting consumers and enabling recourse and set standards and compliance expectations already applicable to companies developing AI systems.

The Framework should leverage these existing and proven laws by (a) streamlining criteria to focus on compliance with existing applicable law, and (b) emphasizing that industry groups, jurisdictions, or other actors should acknowledge that existing law already establishes important ground rules when developing self-regulatory standards or any potential new regulation.

Leveraging existing law not only ensures that the public and businesses understand their rights and responsibilities but also furthers the efficient implementation of the Framework. Businesses can utilize their existing compliance processes to ensure compliance with existing laws, resulting in further efficiency and cost reduction as they ensure their practices are aligned with the Framework and aligned with perceived and actual risk.

III. The Framework's Implementation is Critical to its Success

The Framework must be supported in future regulations or guidance with attention to flexibility. An extensive, double-digit component process before an AI system can be developed or used may not be appropriate across AI applications and instances. Prescriptiveness may stifle innovation and use of this technology for good.

CTA commends efforts to make the Framework user-friendly and urges the OECD to prioritize usability. Experimentation and consideration in the implementation of the Framework should be encouraged; the Framework can serve as a mechanism to consider the operational aspects of AI systems, but should not form the basis for adopting strict new regulatory obligations. Otherwise, mandates that are burdensome and difficult to implement would be costly and inhibit the development and enjoyment of the benefits of AI. This would particularly harm small and medium businesses and enterprises and their customers. The contributions of small and medium businesses to the public good and technology development could be insurmountably hamstrung by overly prescriptive implementations of the Framework.

CTA encourages the OECD to address: (a) how it recommends the Framework be implemented by policymakers in statute, regulation, or other formal policy; and (b) targeted relief or carve-outs for smaller entities, especially those with less identifiable risk or where risks are less likely to emerge unexpectedly.

IV. Balancing the Risks and Benefits of AI Systems Requires Attention to Nuance

CTA suggests the OECD consider the necessary nuance in developing the Framework. For instance, the Framework should recognize the benefits of broad data collection. Collecting additional and varied data as inputs improves AI systems and helps address and mitigate harmful bias over time. Better functioning AI (i.e., AI trained on greater amounts and more diverse data, and deployed with appropriate considerations and mitigations) can provide further benefits to the public, such that the AI systems produce more accurate, trustworthy, and ethical outputs. Without sufficient and varied data inputs, desirable outputs may remain unattainable.

CTA emphasizes that the risks and benefits of an AI system cannot be captured in a formulaic calculation. For example, determinations about whether answer A to criteria 1 creates risk may be subjective and complicated. While a formula, or risk assessment framework, may be

a *starting point* for assessing opportunities and risks, the OECD should emphasize that it is only the start of a process that will require additional data, evidence and risk-benefit analyses by policymakers and regulators. Further iterations of the Framework and any resulting implementations should recognize these complexities.

V. The Framework Should Consider the Importance of Consumer Buy-In as Risk Mitigation

CTA encourages the OECD to consider whether the Framework sufficiently accounts for the ability of consumers to buy-in to processing of their data for AI as a means of mitigating risks. Risk mitigation based on consumer opt-in could include principles of transparency (i.e., disclosure of the use of AI) and explainability to the consumer, where appropriate. Thus, where AI systems developers can produce systems that are transparent to the affected individual, or where the system includes explainability components that allow the affected individual to understand the basis for a recommendation, decision or action taken by the system, then consumer buy-in and acceptance of this technology is likely to increase. That, in turn, can mitigate potential risks and support broader adoption of this technology.

CTA notes that some existing laws require user agreement to collection or processing of certain of their personal information (e.g., biometrics, sensitive information). Where those requirements are in place and complied with, the Framework should consider such compliance as a mitigated risk which would not necessitate further risk mitigation requirements by regulators.

VI. Specific Recommendations

CTA also recommends the following specific revisions to the Framework:

- a. **Address performance metrics.** The Framework lacks performance metrics which would enable developers to ensure that their systems are compliant with and addressing key ethical and trustworthiness principles. CTA suggests the inclusion of performance metrics and indicators within the framework.
- b. **Include additional examples.** While the [Qlector.com](https://www.qlector.com) LEAP system addressed in the Framework represents a set of algorithms working in conjunction in a pipeline (a *composite system*, per the terminology in the document), there is no concrete individual example for a computer vision model similar to what is specified for

GPT-3. CTA recommends inclusion of an example for the computer vision domain. It would also be beneficial to see this type of classification applied to multimodal AI algorithms (e.g., AI models capable of processing image and text, such as OpenAI’s DALL·E), either as a separate example or simply described as a use case in the text.

- c. **Include additional transparency and explainability questions.** CTA respectfully suggests the addition of two further questions to Box 3 on page 24: “Possible questions to help determine AI system transparency and explainability.”
 - 1. “Does your model have an uncertainty score, report confidence intervals, or provide heat maps to understand the reliability of the AI system?” and
 - 2. “Are users able to weight components/parts of the AI system?”
- d. **Clarify definitions.** CTA notes that several definitions could benefit from additional clarity or definition.
 - i. “Reasoning with knowledge structures” – A clearer definition of this would be helpful by characterizing the exact properties of this type of causal reasoning system and why this may differ from a traditional AI application. Furthermore, the Framework does not mention methods leveraging uncertainty-based measurements.
 - ii. “Robustness,” “fairness,” and “privacy” – Although these are specified as principles, are there quantitative measurements for each principle, such that developers can evaluate compliance with these principles?
- e. **Expand on mitigation strategies.** The Framework insufficiently describes strategies to mitigate risk. Some mitigations are offered for data quality, but there is no comprehensive mitigation assessment for AI algorithms. The Framework should assist AI developers in understanding not only what risks might be present, and whether they are high-risk or low-risk, but also how to address them in the context of the specific deployment environment.
- f. **Emphasize integration of AI algorithms in an internet of things ecosystem.** The sections on automated sensor data collection and [Qlector.com](https://www.qlector.com) LEAP system

does a good job of characterizing the integration of AI algorithms in IoT. However, the Framework should address whether the principle or characterization of AI algorithm integration differs in local vs. remote setups (e.g., edge vs. the cloud in an IoT setup) and if so, how (e.g., is localization on fairness, explainability, or similar relevant)?

VII. Conclusion

The development and deployment of AI systems present multiple challenges as well as potentially immense and transformative benefits for society. AI systems are expected to become increasingly interwoven with, and beneficial to, all aspects of our everyday professional and personal lives. In order to realize the potential to improve our lives, developers and users of technology must enhance user trust for AI systems to be accepted and widely used, and regulators must recognize the various risk environments that will control how AI systems can safely and effectively operate. “If the AI system has a high level of technical trustworthiness, and the values of the trustworthiness characteristics are perceived to be good enough for the context of use, and especially the risk inherent in that context, then the likelihood of AI user trust increases. It is this trust, based on user perceptions, that will be necessary of any human-AI collaboration.”³ As we move from research and development to actual use, the Framework must allow for maximal data collection and encourage innovation that will lead to deployment and use of trustworthy AI.

Respectfully submitted,

/s/ Douglas K. Johnson
Douglas K. Johnson
Vice President, Technology Policy

/s/ Michael Petricone
Michael Petricone
Sr. Vice President, Government and Regulatory Affairs

³ <https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8332-draft.pdf>