June 30, 2021

Re: Comments by Association of Test Publishers (ATP)

Dear Sirs:

On behalf of the Association of Test Publishers (ATP) , we attach comments on the OECD framework for the classification of AI Systems as part of the OECD public consultation, due by today, June 30, 2021.

The ATP is the international trade association for the testing industry. The ATP is comprised of hundreds of publishers, test sponsors (i.e., owners of test content, such as certification bodies), and vendors that deliver tests used in various settings, including healthcare, employment (e.g., employee selection and other HR functions), education (e.g., academic admissions), clinical diagnostic assessment, and certification/ licensure (e.g., licensure/ recertification of various professionals), and credentialing, as well as businesses that provide testing services (e.g., test security, scoring) or administering test programs. Since its inception in 1987, the Association has advocated for the use of fair, reliable, and valid assessments. The ATP is very active in the area of privacy having published guidance and bulletins on the GDPR and other privacy laws; we have a keen interest in AI and its potential and current uses within the area of assessment.

The ATP welcomes the OECD framework for classification and believes that it will provide a useful framework which regulators and other stakeholders to use in classifying AI. We make some detailed comments within the PDF, including some suggestions of assessment use cases that we believe should be considered as examples. But our main and most substantive point is that we believe that there is more connection between the concept of autonomy and risk within the classification framework.

We suggest that it is over-simplistic to classify AI as being either "outcome dependent" or "no impact", especially in areas of assessment that carry high stakes for individuals, such as the right to education or the work and job quality. We would suggest that the issues here are considerably more nuanced and would justify using a range of at least 4 possible classification values, not just two.

For example, one use case for AI within assessments is to identify possible test fraud (cheating). One can pose a situation where an organization delivering assessments which contribute to education or recruitment choices and AI is used to flag possible cheating at the test(e.g., by data analysis). When such a system is used without human review, then we understand the conclusion that there is a higher risk to individual rights, but if it is used with human review, the risk is functionally much lower, and therefore it would be sensible to link risk to rights and autonomy more specifically in the classification.

Similarly if AI is used to score an exam, and that exam's scores are used to help make an education or job-related decision, there is a significant difference between that score being automatically used by the educational institution or employer to make the decision and that score being a single data point amongst many other data points in the human decision process. There is a risk that if the scale is not more nuanced, then the risk analysis will be seriously flawed and relatively minor risks will be lumped together with more substantial ones. If that comingling exists, the Framework will not provide regulators the tools to apply regulation in a way which encourages the development of new ways of using technology in testing, while protecting test-taker rights.

In addition to the lack of nuanced classifications/criteria, the ATP is also concerned the Framework is generally oriented to forcing assumptions to "high risk" activities. As we point out in our comments, there are a number of instances where choices between yes/no or low/high create an artificially limited set of classifications/criteria.

We urge the OECD to consider our comments on these points within the consultation document. We would be pleased to engage with the OECD in further discussion on any points relating to this if we can be helpful.

Sincerely,

Wm G. Harris, PhD.
CEO