BEFORE THE EUROPEAN COMMISSION

Re: Comments on Proposed Regulation on the Use of AI

The Association of Test Publishers ("ATP") submits its comments to express the serious concerns of the testing industry to the proposed Artificial Intelligence Act ("Regulation" or "Proposed Regulation") published by the European Commission ("EC" or "Commission") for public consultation on April 21, 2021. We welcome this opportunity to express our views and contribute to the EU decision-making process on the adoption of a final regulation by the European Parliament and the EU Council ("Final Regulation"). This feedback is being made by the required date of August 6, 2021. Further, the ATP fully understands that, as required by GDPR Art. 70(4), the EC must make the results of this consultation public, and therefore we acknowledge that our submission will be published in its entirety on the EC website; for clarity, there is no personal information contained in this submission that requires protection by the EC.

The ATP is a not-for-profit international trade association (i.e., business association) for the testing industry, which includes a regional organization representing European organizations, as well as a regional organization for Asia, including China, Japan, and South Korea, an organization representing India, and an ongoing effort to establish a regional organization in the Middle East/Africa. The ATP is comprised of hundreds of publishers, test sponsors (i.e., owners of test content, such as professional certification bodies), test delivery vendors of 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/credentialing, as well as businesses that provide testing services (e.g., test development, proctoring, scoring) or who administer test programs ("Members"). Additionally, many Members are global vendors and service providers of AI or automated decision-making solutions used in testing, who do business in the EU either because they are located there or operate globally, so the Proposed Regulation will have a significant impact on the testing industry.

Since its inception in 1987, the Association has advocated for the use of fair, reliable, and valid assessments, which include ensuring the security of test content and test results. Our activities include providing resources and expertise to the U.S. Congress and state legislatures in the United States on legislative proposals affecting the use of testing in education and employment, as well as representing the industry in federal and state regulatory matters and litigation surrounding the use of testing. In providing industry-specific education on the EU's General Data Protection Regulation ("GDPR") for its Members in the EU and the US, the ATP has published a "Checklist for EU-US Privacy Shield Registration" (2016) and a "Compliance Guide for the EU General Data Protection Regulation" (2017). We also submitted comments about specific concerns of the testing industry to the European Data Protection Board on its proposed 2019 Guidelines for the Use of Video Surveillance under the GDPR.

In general, ATP Members are data-oriented organizations; thus, analysis, predictive analytics, and AI have been vital tools in their research and commercial efforts for many years.

Specific to the uses of AI in the testing industry, the ATP recently published a White Paper to provide information on the background of AI, its historic uses in the industry, and the growing regulatory attention being paid to AI. *See* https://atpu.memberclicks.net/assets/ATP%20White%20Paper_AI%20and%20Testing_A%20Primer_1July2021_Final%20R1%20.pdf . In June, the ATP provided comments to the Organization for Economic Co-operation and Development ("OECD") on its proposed AI Framework. *See* https://atpu.memberclicks.net/atp-comments-on-oecd-framework.[1]

Introduction

The Proposed Regulation contains four key areas: (i) rules for placing on the market, putting into service, and use of AI systems in the EU; (ii) prohibitions of certain AI practices and specific requirements for "high-risk" AI systems; (iii) transparency rules for AI systems; and (iv) compliance rules on marketing, monitoring, and surveillance.

The ATP respects and applauds the goals of the European Commission to ensure that the health, safety, and fundamental rights of individuals are protected when they engage with business process and system that utilize AI. The ATP endorses an AI framework grounded on trustworthiness, responsibility, accountability, transparency, and ethics. The ATP also generally supports a "risk-based" regulatory approach; however, we do not believe that the evaluation and management of risk factors can or should be a "one size fits all" or "all or nothing" analysis so that all uses of current testing are not automatically judged to be "high risk" (*See* Article 6(2)). To that end, the ATP strongly believes that all uses of current testing should not be judged automatically to be "high risk," and that greater differentiation in risk levels ought to be part of the Final Regulation. Moreover, we submit that specific common circumstances exist in the testing industry where the use of AI is both appropriate and necessary to promote innovation and efficiency, where its use is justified when balanced against the rights of individual test takers, and where this technology should be allowed within the requirements of the GDPR.[2] Thus, the

[1] The ATP also is currently discussing with OECD how the testing industry could be involved in the development of an approach to assess the capabilities of AI solutions and compare them with human capabilities. OECD plans to use existing human tests to carry out this assessment, supplemented with AI-specific measures developed by the computer science community. The goal is to provide a set of valid and transparent measures of AI capabilities that give policymakers a meaningful way to understand what current AI can and cannot do. The so-called "Future of Skills" study will apply familiar assessment techniques to the novel problem of assessing the capabilities of a new "population" – the population of AI systems. *See* https://www.oecd.org/education/ceri/future-of-skills.htm.

[2] The ATP notes that investment in educational technology (affecting both learning and assessment) in the U.S. was up 30% in 2020 to $2.2 billion—the highest single-year investment total in US edtech history. *See* "A Record Year amid a Pandemic**:** US Edtech Raises $2.2 Billion in 2020," EdSurge (Jan. 13, 2021) https://www.edsurge.com/news/2021-01-13-a-record-year-amid-a-pandemic-us-edtech-raises-2-2-billion-in-2020. We suspect this surge in technology investment has also occurred in the EU, which focuses even more attention on the role of AI.

ATP requests that the Commission modify its proposals in line with the recommendations made in these comments, and to include the examples presented here, along with explanations clarifying how those examples and proposed edits are consistent with both regulatory goals and the GDPR The ATP believes that the appropriate use of AI going forward has the potential to increase the benefits of testing and assessment for European individuals, organizations, and society; but as currently drafted, the Proposed Regulations presents barriers to those wanting to implement and responsibly use AI.

Summary of the ATP's Concerns

The ATP's comments largely focus on elements of the four key areas in the Proposed Regulations: i) rules for the development and use of AI systems in the EU; (ii) specific requirements for the management of "high-risk" AI systems; (iii) transparency rules for AI systems; and (iv) compliance rules on marketing, monitoring, and surveillance. All of our concerns initially stem from the scope and definitions of the Proposed Regulation, as well as the risk characteristics that underlie it.

As noted in the Introduction, the vast majority of the Proposed Regulation is directed towards AI systems considered to be "high-risk" activities, which are defined very broadly to include those used in employment, education, vocational training, and even healthcare/clinical diagnostics (e.g., evaluating persons on tests that are part of or as a precondition for their employment, training, or education opportunities). *See* Annex 1II. Beyond that risk-based nomenclature, the Proposed Regulation sets out comprehensive compliance requirements for high-risk AI systems – including validating the quality of data used in model development including training activities, maintaining adequate records, providing adequate transparency to users, providing adequate human oversight, and ensuring the accuracy and robustness (e.g., lack of discrimination and bias) of the AI system itself. The ATP notes, however, that the cost of imposing such compliance requirements for every producer or user of AI is going to be significant. In fact, even the Commission's own study (also published on April 21, 2021), estimates AI regulation compliance for high-risk systems is likely to cost upwards of 17% of total AI investment. The ATP strongly believes that the benefits of AI to society in general – and test users and test takers in particular – are significant; therefore, a more relaxed regulatory approach would serve to promote increased R&D to advance these benefits. Consequently, we feel it is imperative that the Proposed Regulation be modified to evaluate AI risk characteristics/ classifications and compliance outcomes with a more granular, balanced perspective, to ensure that the benefits of regulation are truly commensurate with the costs.

Indeed, as urged by 14 EU countries in their position paper addressed to the Commission, entitled, "Innovative and Trustworthy AI: Two Sides of the Same Coin," a "soft law" option would better serve such a uniform, standardized regulatory process, where solutions such as self-

regulation, voluntary labeling,[3] and other similar voluntary practices should be used to supplement current rules (e.g., GDPR), and safety and security standards ("Soft Law Position Paper"). *See* em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf.[4] The ATP strongly identifies with the position that, "Soft law can allow us to learn from the technology and identify potential challenges associated with it, taking into account the fact that we are dealing with a fast-evolving technology." Otherwise, we fear that the acknowledged benefits of AI will be lost amid misunderstood business applications and over-regulation – and in the process, future innovation and the concomitant benefits to Europe will be stifled.

Finally, the ATP urges the Commission to take into account the desirability and value of shaping a global regulatory perspective – a more unified global standard will provide the benefits of "a proportionate, operable, and futureproof regulatory framework" to the benefit of both the EU and every interconnected global commerce network.[5] As such, the EU would become a leader in responsible regulation of AI, just as it is seen as the global leader in establishing individual data privacy protection through the GDPR.

---

[3] From the testing industry perspective, requirements for data traceability and provision of appropriate reasonable explanations of AI to test takers (i.e., allowing for full protection of intellectual property) would be extremely useful components of a regulatory program. Accordingly, adoption of global standards in technology notation, labeling, and disclosure would be very helpful, so long as reporting requirements are reasonable. Such standards would be fully consistent with the Commission's principle of transparency.

[4] The Soft Law Position Paper represents the viewpoint of the following 14 signatory countries: Belgium, the Czech Republic, Denmark, Finland, France, Estonia, Ireland, Latvia, Luxembourg, the Netherlands, Poland, Portugal, Spain and Sweden.

[5] While the Soft Law Position Paper is focused on the benefits that "a single regulatory framework will also enhance possibilities of cooperation between Member States in the public sector," the ATP feels the same principle applies to a global framework. Similar efforts towards strengthening the public/private partnership in the United States have begun. The ATP supports fully funding R&D on AI in Fiscal Year 2022, pursuant to the National Artificial Intelligence Initiative Act passed by the U.S. Congress in 2020 to promote U.S. global leadership in AI and enable the development and use of trustworthy and responsible AI systems. As recognized by supporters of full funding in a letter to the U.S. House of Representatives Appropriations Committee on July 15, 2021, "AI can also pose some risks if improperly created or used, so it is essential that stakeholders collaborate to address and mitigate risks stemming from AI. The U.S. government can meet this challenge through increasing investments in research and development and supporting the development of AI-related voluntary consensus standards."

1. Concerns about Scope and Definitions

The ATP has misgivings about the scope and definition of Artificial Intelligence ("AI") set forth in the Proposed Regulation.[6] The definition includes logic and knowledge-based computerized solutions which can be interpreted so broadly that essentially any basic software used in the testing industry and even discreet logic hardware systems (e.g., a Scantron scoring machine) would be included. In our view, the draft Regulation inappropriately lumps "automated decision-making" (i.e., mere automation of human functions) into the definition of AI. Further, the Proposed Regulation's definition of AI includes any system based on prediction, arrived at by classic statistical means, into the AI definition. Finally, the Proposed Regulation assumes that all computerized technology related to tests used in education, clinical, and employment settings should be automatically characterized as "high risk" activities – a conclusion that the ATP suggests is based: on (1) the mistaken assumption that characterizes knowledge, predictive, analytical, and logic-based practices as AI;[7] (2) the overly-broad definition of AI; and (3) a lack of familiarity with long-standing testing standards and practices, including the well-documented and safe historical uses of data-based methods and computer technology.

Perhaps the most fundamental concern the ATP has with the Proposed Regulation is its apparent lack of appreciation for the science of psychometrics that underlies all of assessment/ measurement. Since the 1950s, rigorous professional standards have governed the development, administration, and scoring of assessments, especially in the areas of education and employment. *See Standards for Educational and Psychological Testing,* American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (eds. 2014).[8] These standards establish how assessments are professionally

---

[6] The Proposed Regulation defines AI as, "any system that generates content, predictions, recommendations or decisions, based on, *inter alia*, machine learning approaches, logic- and knowledge-based approaches, or statistical approaches." Article 3(1).

[7] The ATP acknowledges that these methods are mathematics and computer science practices central to the AI field, but we note they are sufficiently differentiated from more recent learning AI systems. Accordingly, the clustering of all AI practices into a single broad definition results in misclassification of inherent risk in business processes and systems commonly used in testing (and in other industries).

[8] Six versions of the *Standards* have been produced, with the most recent published in 2014. The *Standards* are "joint" in nature in that they have been prepared by a joint committee of testing experts representing the three sponsoring organizations: the American Educational Research Association (AERA); the American Psychological Association (APA); and the National Council on Measurement in Education (NCME) (*see also* fn. 10). The Joint Standards are widely accepted and followed by testing professionals around the world. Despite the title, these standards are widely acknowledged to apply to assessments used in certification/licensure, workforce and professional credentialing, and clinical/diagnostic settings.

evaluated -- based on their accuracy ("validity") and repeatability ("reliability"), as well as their fairness to test takers. The intent is to promote the sound and ethical use of tests and to provide a rigorous professional basis for the quality of testing practices. *See* Eignor, D. R., "Standards for the development and use of tests: The Standards for Educational and Psychological Testing," *European Journal of Psychological Assessment, 17*(3), 157–163 (2001) https://doi.org/10.1027/1015-5759.17.3.157. Accordingly, the ATP requests that the Commission recognize the Joint Standards as an appropriate part of the set of its approved harmonization standards, which would assist in providing a reasonable compliance path for testing organizations to follow.

Validity and reliability are measurement-related principles involving prediction and probability, both of which are long-standing, well-recognized fields of mathematics and statistics. Today, the term "psychometrics" is often applied to the measurement science of the validity and reliability of inferences drawn from answers to test questions. While most AI systems are built on statistical probability, they are NOT the same as psychometrics. Consequently, the ATP views this is as an important distinction in terms of determining the appropriate regulation of AI/machine learning in testing.

For example, classical statistical algorithms in psychometrics (e.g., linear regression, multiple regression analysis) have been commercially deployed for decades with no documented negative impact on test takers' rights. Despite the scary association the term "artificial intelligence" has acquired, in reality AI is only an applied technology – namely, mathematical and statistical functions performed on data. Accordingly, the ATP asserts there is truly no reason for regulating the mathematical and statistical functions themselves. The application of AI practices, as with the previously discussed application of discreet probability, statistical, and predictive practices, requires a transparency of data and algorithmic origin, clarity of application, human oversight, and appropriate review mechanism. These principles – which the ATP sees as the core themes of the Proposed Regulation – equally sit at the heart of the ATP's focus on fairness and transparency using psychometric principles. When developers and users of AI adhere to these principles in practice, we believe that many discrete applications of AI techniques should not be deemed to be high-risk, whether in the testing industry or more broadly.

The ATP recently explored a comprehensive history of AI to assist testing organizations in understanding the growing focus on AI. Key information in the ATP White Paper traces the development of AI, as well as explaining the distinctions between different types of AI systems, and sets forth an appropriate definition of AI for the testing industry. *See* https://atpu.memberclicks.net/assets/ATP%20White%20Paper_AI%20and%20Testing_A%20Primer_1July2021_Final%20R1%20.pdf Background Information/Definition (pp. 4-7). We encourage the Commission to recognize that its proposed definition is not aligned with the one we have used – and with definitions provided by other commenters – and we submit other industries likely have similar concerns. For that reason, the ATP recommends that some industry-specific flexibility surrounding the definition of AI should be built into the Final Regulation, unless more generally agreed definitions are used across the board.

Related to the scope of AI, the Proposed Regulation essentially treats all AI systems as learning AI, when that is not the case.[9] As clearly articulated by Dr. Peter Norvig (UC Berkley), "Regular programming is about writing instructions for the computer to do what you want it to do, when you do know what you want it to do. AI is for when you don't."[10] From its familiarity with testing solutions, the ATP asserts that little commercial AI is truly active "learning AI" – marketing claims to the contrary. Most commercial AI today is strictly symbolic/probabilistic in nature, where AI outcomes are arrived at using big data, with algorithms possibly created via neural network methods.

Equally important, as the above-referenced quote by Dr. Norvig suggests, there is a critical distinction between AI and automated decision-making (ADM), which the Proposed Regulation largely fails to acknowledge or address. The ATP is very concerned that the misperception that automated decision-making is or should be treated the same as machine learning AI will lead to misguided and unfair regulation. Indeed, there is ample evidence to indicate that human decision-making is prone to subjective (even emotional) input, which can actually create bias and result in less accurate decisions than ADM. *See* Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *Journal of Economic Perspectives*, Vol. 5 (1) (1991), pp. 193-206 (discussion of "status quo bias"); Tversky, Amos and Kahneman, Daniel, " Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics*, Vol. 106, No. 4 (Nov., 1991), pp. 1039-106 http://www.jstor.org/stable/2937956 .

The ATP is concerned that the Proposed Regulation includes automated decision-making (ADM) within the definition of AI and seems to hold ADM to a high standard, reflecting an unrealistically high estimate of the degree of transparency (and accuracy) attainable from human decision-makers. Indeed, automated decision-making processes are intentionally designed to function with little or no human interactions during their use precisely because the results are arrived at by rigid application of specific patterns and rules – usually to perform repetitive tasks. Thus, the ironic fallacy surrounding the purported need for human intervention in the Proposed Regulation is that scientific evidence demonstrates that much human decision-making is fraught with transparency problems, can often produce results that are worse than AI in the uniform accuracy of outcomes, and therefore can raise at least some concern that regulatory proposals for explainable AI could end up setting the bar higher than is necessary or indeed helpful. *See* Zerilli, J., Knott, A., Maclaurin, J. *et al.*, "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?," *Philos. Technol.*, Vol 32., pp. 661–683 (2019) https://doi.org/10.1007/s13347-018-0330-6.

[9] A comprehensive explanation of machine learning and its variations is found in a recent MIT article: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained. The ATP is concerned that the Proposed Regulation confuses many examples of automated decision-making with machine learning/AI.

[10] Talati, A. (September 12, 2018). CS 6601 Artificial Intelligence. Retrieved from Subtitles To Transcripts: https://subtitlestotranscript.wordpress.com/2018/09/12/cs-6601-artificial-intelligence/7.

For example, in many instances of automated scoring of tests (*see infra.* at pp. 14-15), the human scoring rubric (i.e., scoring key of right answers written by a human), is merely applied uniformly to all test takers through an automated computerized system, with no chances of human error. By comparison, in situations where human interviews are used to make decisions (e.g., school admissions, employment), a human interviewer who is generally influenced by subjective opinions and extraneous factors, statistically will make the "proper" decision about a person only 50 percent of the time compared to ADM systems. A recent interview conducted by the American Psychological Association with Dr. Fred Oswald, the director of the Organization and Workforce Laboratory at Rice University in Houston, Texas, addressing the utility of AI solutions in Industrial and Organizational Psychology (i.e., employment-related testing), echoes the same concern. *See* https://www.apa.org/research/action/speaking-of-psychology/personality-tests?utm_medium=email&utm_source=rasa_io&PostID=33813611&MessageRunDetailID=5878582355.

Moreover, the ATP believes it is highly significant that ADM is already regulated under the GDPR. Although the GDPR does not specifically reference AI, it does mention automated decision-making (*See* Article 22). Article 22 gives individuals the right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects or similarly significant affects [to] him or her." Accordingly, an organization using automated processing must provide a "clear and separate" notice of the automated process and the individual's right to object. An organization can overcome individual objections to some decisions (except for direct marketing) if they are able to demonstrate that they have a "compelling legitimate ground for the processing which override the interests…of the data subject or for the establishment, exercise of defense of legal claims." Moreover, the GDPR allows personal data processed for various purposes, including statistical purposes, if there is a public interest under Art. 89(1).

Indeed, it is worth noting that Art. 22 regulates more consequential automated decision making, specifically when there are legal or similarly significant effects to the individual and there is no human involvement in the decision making.[11] Thus, when an organization has established a process by which it provides human review or challenge to an action, then the reports or results generated before making any final decisions generally do not violate the GDPR. Given these regulatory requirements already in place, the ATP is concerned that the Proposed Regulation is inconsistent with the GDPR.

[11] The Norwegian DPA ordered the International Baccalaureate Organization ("IBO") to provide information showing its compliance with the GDPR regarding its calculation of individual student grades using an automated decision-making process. The DPA's reasoning was that although the input factors in part may consist of assessments made through human involvement, the calculation of the final grade itself appeared to happen through a wholly automated process where there is no room for meaningful human assessment. The IBO disputed these facts. Norway Data Protection Authority (Datatilsynet) Case Reference 20/03087-2/TJU (July 20, 2020). This investigation was closed on July 19, 2021, because the DPA agreed it had no jurisdiction over the IBO's actions in the UK.

Finally, the Proposed Regulation focuses on the need for autonomy to protect against the misuse of AI. Yet the ATP submits that very little AI is truly autonomous. In fact, just the opposite is true. The concept of "human-in-the-loop" is central to many AI implementations, either in terms of direct development of the AI system, or in terms of its training or in handling or responding to its actual use. Although it may be accurate to state that humans may not pay the necessary close attention to their roles in decision-making that they should when provided input from an AI system, the existence of that problem does not mean the basic regulatory model should be changed. For example, an AI-based GPS system does not drive a car into a swamp or off a road; the driver's continuity bias is responsible for those outcomes – the failure of some drivers to pay appropriate attention to their driving does not mean all GPS systems are suspect or "high risk."

2. Concerns about Risk-Based Classification of AI

Based on its review of the Proposed Regulation, the ATP believes it is too limiting to classify all AI as being either "outcome dependent" (high risk) or having "low impact," especially in areas of assessment that involve high stakes for individuals, such as in education, employment, professional qualifications, career training, or clinical/heath-related diagnostics. This sort of "all or nothing"/ "either/or" determination of risk does not square with the way in which AI and automated decision-making is currently handled in the testing industry (these use cases are explored in depth *infra.* at pp. 10-15). As the ATP recently noted in its comments to the OECD, "risk-based classification of AI is considerably more nuanced and requires using a broader range of possible classification values, not just two." *See* https://atpu.memberclicks.net/atp-comments-on-oecd-framework.[12] The Proposed Regulation adopts a similar risk framework, where essentially every computer-based assessment used in education, employment, and vocational training constitutes a "high risk" activity, a view that is not justified and will not result in a useful regulatory framework.

For example, one common 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 uses an AI solution to flag possible cheating as the test is being administered (e.g., by data analysis). When such a system operates without any human-in-the-loop intervention/review, then we understand the concern that there is a higher risk to individual rights, but if the AI solution is used with human review (either as being used or subsequently on review or challenge), the risk is functionally much lower. Therefore, the ATP believes it would be sensible to link risk to rights and autonomy using more granular (i.e., detailed/differentiated) classifications.

[12] The ATP comments to the OECD primarily sought to encourage expansion of risk classification categories to provide greater range of distinctions as to how risk is determined and calculated. We believe this outcome is needed to ensure that specific activities, especially testing in employment, education, and training, are not automatically assumed or deemed to constitute "high risk."

Similarly if an AI solution is used to score a test, which scores are used as a single factor to help make an education or job-related decision, there is a considerable difference between the test score being automatically used as the sole determinant 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-making process. There is a strong likelihood that if the regulatory risk scale is not more nuanced, then the risk analysis will be seriously flawed and relatively minor risks will end up being classified together with more substantial ones. If that comingling – and misjudgment -- exists, then the Proposed Regulation will not provide regulators with the precise tools needed to apply regulation in a way which encourages the development of new ways of using AI technology in testing, while protecting test taker rights.

In addition to the lack of nuanced risk classifications/criteria, the ATP is also concerned the Proposed Regulation forces assumptions about AI to be classified as "high risk" activities. As we noted above, statements where the only choices are between "yes/no" or "no risk/high risk" create an artificially limited set of classifications/criteria, where virtually all computer-based testing is automatically deemed to be "high risk."

For all of these reasons, then, the ATP firmly believes that automated decision making algorithms and AI, as routinely applied in the testing industry for the purposes of preserving integrity of the process, creating test questions appropriate to individual capabilities, and scoring test results, serves the interests of individuals and the public. The next section presents context and examples to illustrate how these processes typically work and the associated benefits to support our thesis that regulation of AI should be more nuanced.

3. Discussion of Testing Business Processes

To understand the role of AI requires understanding the business process(es) in which the solution is used. There are obvious legal and disclosure implications related to that understanding. AI, or other solutions that may be perceived as AI (e.g., straightforward automated decision-making) have been used within the testing industry for many years to perform a wide variety of business functions. Testing organizations have already embedded AI or computer-based analytical technology into their workflows and decision-making processes, which have resulted in enhancements/improvements for all stakeholders and society as a whole.

Significantly, many of these functions do not rise to the level of "high-risk" activities as assumed by the Commission. Accordingly, the ATP urges that a more granular, nuanced analysis must be undertaken by the Commission to distinguish between instances where automated decision-making or much lower risk AI solutions are employed, often with "human-in-the-loop" protections, and those instances where true machine learning or autonomous AI is involved. Equally significant, the ATP strongly believes that whether any personal data are used – and whether such personal data are retained in violation of the GDPR – are relevant factors that must be taken into account.

The debate over AI regulation is now focusing intense scrutiny on such business functions, so it is critical to examine them closely under what we believe to be the proper definition and scope of regulation. The ATP White Paper discusses the following testing

industry business processes/functionalities where AI, AI-like, or computer automation have existed for decades. https://atpu.memberclicks.net/assets/ATP%20White%20Paper_AI%20and%20Testing_A%20Primer_1July2021_Final%20R1%20.pdf (pages 8-9). In these comments, the ATP reviews the most common of those functions and provides an analysis of how we contend they should be treated under the Final Regulation.

A. Content Analysis

• Question construction. Some forms of AI are used in generating/writing items/questions for use in tests (e.g., by taking some instructional text and using language analysis to construct new, different questions based on that analysis).

ATP Analysis: Significantly, every item that ends up being used in a test, whether it is written by a human item writer or is generated by a computer, is subjected to bias study, other psychometric research, and pilot testing, to make sure that items evidencing any form of bias are removed from the pool of items eventually used in constructing tests (whether by humans or computer). Equally significant, such item construction does NOT involve any use of current test taker personal data: if past test usage is used for research purposes, that information has been de-identified and aggregated so no one involved in the process has any access to individual test taker information. Based on this analysis, the ATP believes this functionality does not rise to the level of a "high risk" AI activity and therefore should not automatically be regulated as such.

• Question selection. Some organizations use algorithms to select questions to be included in a particular test or separate forms of the test (e.g., for a regional or national administration). A similar process is used to present a unique set of test questions to individual test-takers (e.g., in fixed format using Linear on the Fly ("LOFT") testing, in variable forms using Computer Adaptive Testing ("CAT"), or in other situations where each test-taker receives a personalized/ customized assessment). AI may be used to make more effective selections.

ATP Analysis: Generic item selection processes for creating many tests or comparable fixed forms of tests are conducted before any test administrations – and are performed without any reference to specific test taker personal information. Selection decisions for a test are based solely on considerations of ensuring appropriate test content/coverage of subject areas and related psychometric principles for each test. Creating different forms of the same test are similarly performed without reference to test taker personal information to ensure that those multiple forms are equitable (i.e., same level of difficulty, same level of content coverage, and same level of validity/reliability -- so that scores on all forms of a test can be compared). These types of algorithms are built by trained psychometricians to achieve those results – and those algorithms are applied exactly the same way a human would apply them if performing the same scientific work by hand. By comparison, if an AI algorithm is employed in tailoring test items for a test administration to individual test takers (i.e., CAT), then the next question asked of each test taker is dependent on his/her previous answers, which provides a more efficient test

administration and a more accurate scoring methodology. While it is appropriate to explain this process to test takers in reasonable terms, nothing about it is prejudicial or discriminatory – all questions ultimately given to each test taker were previously equated with all other items in the pool of possible questions from which items are selected and all have been pre-determined to be valid and reliable and free from bias. Based on this analysis, the ATP believes this functionality does not rise to the level of a "high risk" AI activity and therefore should not automatically be regulated as such.

- Data analysis. Data analytics techniques that may include AI are used to analyze assessment data sets and make predictions or evaluative analyses (e.g., predicting job competence or identification of learning deficiencies, evaluation of compliance risks).

ATP Analysis: As mentioned earlier, while prediction and statistical analysis are mathematical components of AI, in these applications they are hallmarks of the psychometric process and managed through the history of psychometric governance. Whether used in an educational or employment setting, these types of data analytics (as opposed to profiling of a person for targeted marketing purposes) enable an organization to evaluate uniformly every candidate against a pre-determined set of common criteria (be they job-related or education-based). Based on this analysis, the ATP believes this functionality does not rise to the level of a "high risk" AI activity and therefore should not be regulated as such.

- AI in learning. Edtech companies and other testing organizations focused on various functions (e.g., reading, training, language learning) are using AI to aid in helping individuals to learn, whether that is through traditional instruction or e-learning/e-assessment, at every level of education, including social-emotional learning, life-long learning, and employment training. By definition, personalized learning is intended to adjust the program to the specialized needs of the individual, using systematic, step-by-step methodologies by which the person is able to advance towards identified educational goals.

ATP Analysis: Machine learning and data analytics enable a testing organization to create more effective personalized learning instructional content (e.g., courses, curriculum), and to assess a person's competence/skills or to assist in making career choices, whether that is to identify education weaknesses or positive pathways. Nevertheless, some aspects of personalized learning may also involve ADM to address how the learning program operates. Any program structure is tied to the psychometric principles to demonstrate validity, reliability, and fairness. Critically, when the testing organization gives notice to the individual about how the personalized learning works, the use of personal information is directly related to the profiling used to create the personalized plan is exactly what the individual expects/has agreed to; in other words, the AI solution is co-extensive with the outcomes sought by the individual. In these use cases, the ATP agrees that relevant test taker protections and privacy considerations need to be used in determining how to regulate this functionality, using a more granular risk analysis to evaluate where on the scale of risk any specific AI system falls.

B. Test administration integrity/security[13]:

• Analyzing photographic images. Another useful application of AI supports a testing organization verifying the identity of a test taker. Here, AI is used to compare a form of identification/photographic image provided by a test taker at the time of registration with the identification provided at the time of testing. A match ensures that the proper person is taking the test, and not a surrogate/imposter.

ATP Analysis: This "one-to-one" match function is merely an electronic image evaluation of whether the identification provided by a test taker at two different times match one another, so that only the person who registered (or is eligible) to take a test actually takes it. This type of AI function does not actually constitute (practically or legally) biometrics/facial recognition – and the individual was given notice about and consented to provide the testing organization (or its vendor) with personal identification. Further, the testing organization (or its vendor) provides notice to the test taker about the image matching process, and the individual is asked to provide consent prior to the testing organization (or its vendor) collecting the individual's identification/photographic image at the time of registration. Then the previously provided identification is matched with the identification presented by the individual before the test administration begins. Even if a digital match is performed, it almost always occurs under human supervision, allowing for a digital match to be overruled. This match serves the same exact function as using one's own biometrics to open a mobile device or laptop – to ensure that only the right person is able to get access. Indeed, this image matching process ensures the integrity of the testing event so that all persons involved can be assured that a surrogate/imposter is not cheating the system. No other use of the identification/photographic image is made; the identification/photographic image is not used as part of the test, to change the test administration or scoring, or conduct any profiling of the test taker. Based on this analysis, the ATP believes this functionality does not rise to the level of a "high risk" AI activity and therefore should not be regulated as such.

[13] Standardized test administration is required to assure that everyone who takes a test has the same opportunity to be measured on a test given under the same conditions to achieve fair results. See *Standards for Educational and Psychological Testing* (2014), a set of professional test standards first developed jointly in the 1950s (*see supra.* at p. 5 and fn 7). The Joint Standards have been recognized in most countries around the world. *See also*, ISO 10667 -- Parts 1 and 2 (2011) Section 5.4 (Note), which requires that "… when administering an assessment to one or more individuals, assessment administrators follow the standardized procedures for the delivery of the assessment and document any deviations from those procedures." Standard administration requires observing the test administration, to identify any irregularities that may occur (e.g., use of cheating devices, instance of a power failure, medical emergency, disruption of test takers), as well as to protect the test content from being copied and illegally distributed (e.g., infringing the owner's copyright).

By comparison, other online proctoring systems can also use AI to perform digital facial recognition or other analysis of biometrics to help in identification of test takers. Some of these situations involve a "one-to-many" analysis, where in fact personal profiling of individual test takers occurs. When profiling occurs, the ATP agrees that it is important that the AI system provides accurate profiles for test takers of varying demographics.[14] In these use cases, the ATP agrees that relevant test taker protections and privacy considerations need to be used in determining how to regulate this functionality, using a more granular risk analysis to evaluate where on the scale of risk any specific AI system falls.

- Analyzing video/audio. Using AI enables the analysis of a testing event in real time (either during in person or online administration) with test proctoring/monitoring by one or more humans to determine if any test taker has cheated on the test, or has stolen test content.

ATP Analysis: Such "hybrid" proctoring systems use algorithms to analyze video and/or audio recordings, often along with other data (e.g., observation by a human proctor in either real time or subsequently), to identify test taker actions that could raise questions about the integrity of the test administration (e.g., using a mobile phone, talking to someone through an earpiece, persistent looking away from the screen, seeing a second person in the room who could assist in taking the test). Such issues are flagged, typically for direct review by human proctors or reviewers, to determine if any genuine integrity violations have occurred. Moreover, testing organizations are careful to provide procedures for any test taker to challenge a ruling/score where analysis of video has occurred. Based on this analysis, the ATP believes this functionality does not rise to the level of a "high risk" AI activity and therefore should not automatically be regulated as such.

- Fraud detection. Machine learning and other AI solutions can also be used to look for and analyze patterns in data collected during the test administration to identify anomalies that could represent cheating or other test integrity issues (e.g., forensic data analytics, keystroke analysis).

ATP Analysis. In some cases, the AI is capable of identifying a statistical rationale for a potential anomaly, which establishes the person has not cheated, while in other cases, machine learning or other AI systems are capable of identifying a potential issue that has no apparent rationale or explanation. As implemented, these AI systems generally produce information that is escalated for review – either in real time or subsequently – by a human being to resolve whether a particular action was an attempt to cheat, including a procedure for challenge or appeal of the decision. Consequently, the ATP agrees that relevant test taker protections and privacy considerations need to be used in determining how to regulate this functionality, using a more granular risk analysis to evaluate where on the scale of risk any specific AI system falls.

[14] Significantly, even this use of biometrics in testing is not equivalent to public surveillance (e.g., for law enforcement) inasmuch as test takers have registered for the testing event and have been notified that using an imposter is a violation and that profiling will occur as part of the process.

C. Test Scoring

• Scoring answer sheets. Automation, in the conversion and computerization of data on paper-based assessments, using optical readers to read "fill-in-the-bubble" answer sheets and convert them to digital information, has been used since the 1950s. Identical automated scoring occurs on computer-based assessments, by converting on-screen responses to digital information for scoring against the scoring key. Such scoring is usually associated with multiple choice test items.

ATP Analysis: Critically, such computerized functionality is ADM, not AI; moreover, no test taker's personal information is involved in the process, inasmuch as the scoring completely relies on a human-developed scoring key ('rubric"), comprised of correct/desirable responses based on scientific research. As discussed above, *supra.* at p. 7, the ability of the optical/computerized scoring system to provide more accurate results in a more efficient manner and timeframe benefits all stakeholders. Based on this analysis, the ATP believes this functionality does not rise to the level of a "high risk" AI activity and therefore should not be regulated as such.

• Scoring written test answers. One of the most established uses of AI in the testing industry is to automatically score certain types of questions (e.g., fill-in the blank, short answer, essays), whether those answers are handwritten or electronically captured in a digital format by a computer, using software designed to identify key words or phrases in a test taker's written response, digitize that information, and then provide scores. Computer-based systems for this purpose have been used by testing organizations since the late 1990s.

ATP Analysis: Similar to scoring multiple-choice test items, scoring other written test answers (whether handwritten or computer-entered) results in ADM relying on a human-developed rubric comprised of key word or phrases. Here again, the scoring rubric uses no personal information from test takers but the ADM merely "reads" the test takers' written answers. As with scoring answer sheets, this computer-based scoring performs the function faster and more accurately than human scoring. Accordingly, testing organizations are able to provide test scores on many tests taken on a computer at the end of the testing event, or within a much shorter "turn-around" time. The speed of scoring using this form of ADM is now common-place, demanded by test takers who expect scores quickly, often to enable reporting those scores to an entity (e.g., educational institution, employer, certificate/credential issuer) that uses the scores to make a decision those test takers want or have paid for. All such decisions are made by the third-party entity, not the testing organization providing the test scores. Finally, virtually every testing organization provides test takers with the right to challenge/appeal a score, so human intervention is anticipated. Based on this analysis, the ATP believes this functionality does not rise to the level of a "high risk" AI activity and therefore should not automatically be regulated as such.

• Scoring audio responses. AI systems have been developed that are capable of recognizing speech to enable the scoring of verbal responses (e.g., in spoken English and other language

proficiency exams). For example, a test-taker is asked a question, s/he speaks the answer, and the AI analyzes the response, evaluates it, and determines a score or grade, based on an analysis of the response.

ATP Analysis: Test taker engagement/speech analytics platforms that leverage AI and machine learning operate to capture, transcribe, and evaluate outcomes from those verbal interactions – those outcomes may range from native language speaking proficiency, to foreign language proficiency, to evaluating personal traits/characteristics based on speech patterns. Some of these AI solutions utilize the speaker's personal information to profile or predict the test taker's abilities, while other solutions redact sensitive biometric data and focus exclusively on the words that are spoken. Consequently, the ATP agrees that relevant test taker protections and privacy considerations need to be used in determining how to regulate this functionality, using a more granular risk analysis to evaluate where on the scale of risk any specific AI system falls.

• Scoring video responses. AI systems also score video recordings (e.g., a job applicant asked to respond to a series of recorded questions), where AI is used to evaluate and score the responses, and in some cases to screen out applicants who do not meet set job qualifications necessary for the job, or who fail to demonstrate sufficient skills necessary for a particular job (i.e., communications skills).

ATP Analysis: Although some AI systems are used to assess test takers' job-related skills and attributes, they may also predict how individuals will perform in a specific job. To some extent, such analyses are fully consistent with psychometric principles; in other instances they go beyond the scientific bases for assessment.[15] Other AI solutions are also being made available directly to job candidates, to assist them in preparing for interviews by evaluating them against typical attributes used by employers. Especially in these latter instances, AI producers are striving to deliver computer-based assessments powered by AI without infringing on people's privacy or security through the use of privacy-by-design, anonymization of all data to protect the sensitive information, and avoidance of any facial recognition profiling function. Consequently, the ATP agrees that relevant test taker protections and privacy considerations need to be used in determining how to regulate this functionality, using a more granular risk analysis to evaluate where on the risk scale any specific AI system falls.

4. Compliance Issues with the Proposed Regulation

The ATP is concerned because the Proposed Regulation imposes methods of confirmation of AI compliance through third party evaluation of data sets, models, and implementations. In our view, this approach leads to a number of issues/problems. Significantly, in our view, all of these issues could be remedied if the Commission were to adopt the voluntary labeling "soft law" approach advocated in the 14-country position paper (*see* Soft Law Position Paper*; see also*, https://www.euractiv.com/section/digital/news/eu-nations-call-for-soft-law-solutions-in-future-

[15] These uses of AI should not be confused with those performing video surveillance of testing events for the purpose of evaluating if test takers are attempting to cheat on the test or to identify other irregularities in the test administration (*see, supra.* at p. 13).

artificial-intelligence-regulation/; https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf .

Related to any "soft law" approach, the ATP is concerned about who exactly must comply, because "producers" and "users" do not seem to be the most appropriate terms for describing the various roles that exist in developing and marketing AI systems.[16] It seems to us that a "producer" is the developer of the AI system, but some producers merely license their products for marketing/distribution ("deployment") by others – we question whether the distributor is a producer or a user. On the "user" side of the spectrum, some "users" (i.e., those who are clearly not producers or distributors) actually implement or use an AI system in their products or services, while others never actually use an implemented AI system in their operating businesses but act as "middlemen" or service providers in the process of marketing AI systems. Finally, of course, some users are in fact the developers and deployers of the AI system. The ATP urges the Commission to clarify these definitions to ensure that each entity is clear as to its responsibilities.

A major focus of this definitional uncertainty involves the legal issues surrounding the intellectual property rights (IPR) of an AI system, which will be owned by one entity, yet deployed (used) by literally dozens/hundreds of individual organizations. Only the owner of the IPR has legal access to all of the supporting documentation about the AI technology.[17] Every AI system producer required to make a compliance filing will have to take into account patent filings, extensions, and prior art, which is likely to result (as noted in other feedback on the compliance annex) in a filing that could be thousands of pages. On the other hand, for organizations that have no role in the production of the AI system, those entities do not have

[16] Under the Proposed Regulation as written, "Provider" means "a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge". "User" means "any natural or legal person, public authority, agency or other body using an AI system", except when used during a personal non-professional activity." Moreover, "Users" are "any person or entity that employs an AI system located within the EU or one located outside the EU if the system output is used within the EU." As we suggest, these definitions do not adequately describe the variety of roles that exist in the testing industry, and we suspect in other industries.

[17] Moreover, international legal standards for "trade secrets" (i.e., the way in which an AI system uses technology or operates, which would qualify as IPR separate from any patents or software copyrights, require that the owner take all "reasonable steps" to ensure that the secret information is fully protected from any release; failure to secure trade secrets is likely to be determined by courts to constitute the legal loss of the right of IP protection. *See* Agreement on Trade-Related Aspects of Intellectual Property Rights (the "TRIPs Agreement"), part of the World Trade Organization agreements, which requires each member country to adopt laws covering both the substantive requirements (Article 39) and procedural requirements (Article 42) for protection of trade secrets. Today, these requirements have been adopted by more than 100 countries.

control over or access to such information. By the same token, a producer will have limited, if any, access to information under the control of an entity that is only a user. Finally, a license/distributor that sits in the middle between the producer and user is likely to have a difficult time obtaining access either producers' or users' information.

In this vein, a number of practical/operational difficulties exist. As the ATP reads the Proposed Regulation, organizations engaged in the development, manufacturing, importation, distribution, servicing, or use of AI for "high risk" testing activities, must address a series of regulatory requirements. Under the Proposed Regulation, a "provider" of high-risk AI systems is compelled, among other requirements, to:

- have a quality management system in place;
- perform a conformity assessment to demonstrate that the AI system is compliant, including that the AI system is "error-free and complete";
- report serious incidents of any malfunctioning of the high-risk AI system to the competent authority immediately and no later than 15 days after becoming aware of any such problem;
- establish and document a risk management system, a quality management system, and a post-market monitoring system;
- develop detailed technical documentation and maintain automatically generated logs; and
- register the AI system in the EU Database (maintained by the Commission).

In addition to the above requirements, AI "providers" must meet the transparency requirements of the Proposed Regulation by ensuring that all AI systems intended to interact with individual test takers are designed and developed in such a way as to ensure individuals are informed that they are interacting with an AI system (unless this is obvious from the circumstances and the context of use, which could still be challenged by individuals). The ATP has pointed out that, especially in the testing industry, as with many others we suspect, the "provider" of the AI solution is often not the user of that solution – so there is an immediate disconnect between the requirements and the information that a provider would need to have available to meet the Proposed Regulation, but which is not within its control – the use of the AI system is by another entity which may not have any direct legal relationship to the producer.

While some of the above requirements are sensible and do not create practical obstacles, the ATP feels that some requirements create unworkable, practical problems, which should be changed. In particular, the ATP suggests the following modifications, as well as making them apply to producers and deployers:

1. Article 83(2) provides that AI systems already on the market are exempt from compliance with the Proposed Regulation, but would undergo conformity assessment only when "those systems are subject to significant changes in their design or intended purpose." The ATP shares the concern of other commenters that the definition of the words "significant change" is open to major interpretation and confusion, and should be clarified so that producers/deloyers of grandfathered high-risk AI systems are able to easily understand when their AI systems would be

required to undergo conformity assessment. The ATP agrees with the proposal by Google to modify the words "significant changes" to "substantial modifications," as used in Article 3(23), to align with existing product regulation as outlined in Recital 66.

2. The ATP is unclear whether the release of open-source software (OSS), as used in the testing industry, constitutes "placing it on the market" or "putting into service" or developing an AI system "with a view to" it being placed on the market from the point of view of the Proposed Regulation. These related issues are problematic because the use of OSS is important to AI innovation, thus, if the Proposed Regulation imposes general, routine conformity assessment requirements on OSS, it would have a chilling effect on open collaboration in the AI ecosystem. The ATP therefore recommends that the Proposed Regulation should be clarified to state that compliance requirements only apply when an AI system becomes operational – thus, compliance shifts to the provider or deployer (as redefined) who has opted to include OSS in its operational AI system.

3. The ATP recommends that language should be added to the Final Regulation to clarify that providers and deployers of high-risk AI systems should "take reasonable measures to address risks, consistent with industry best practices." This language would recognize that there are limits to what is possible with the current state of technology.

4. Additionally, the ATP makes the following recommendations on other specific proposals:

a) Article 10(3) requires that "training, validation and testing data sets shall be relevant, representative, free of errors and complete." Real-world data sets are rarely, if ever, "free of errors," particularly the large data sets, which often contain millions (or even billions) of individual data points, used in the most advanced AI applications available today. Furthermore, what constitutes "relevant" and "representative" is a matter of interpretation – there are few standards and metrics to measure them or frameworks to consistently apply them. "Completeness" is also a complex concept to apply universally to datasets - there will always be additional datapoints that would improve a dataset, but at some point a decision must be made that it is good enough. We agree with other commenters that this Article should be modified to require that developers and deployers of AI systems "take appropriate measures to ensure that validation and testing data sets are sets are sufficiently accurate and complete to meet the intended purpose" – the Commission should delete the requirement to provide access to training data because testing data sets should be sufficient, especially since testing data sets cover more sources of bias than only those caused by training data.

b) Other requirements of Article 10 for data set and source code disclosure should be removed or modified. These requirements are overly broad and create unnecessary legal obstacles. The data governance requirements in Article 10 should provide reasonable protection, but giving market surveillance authorities access to data sets themselves would in many cases be

unworkable. Sharing source code is also unwarranted as alternative approaches are available that would be more effective and not undermine trade secrets or IP security.

c) Article 14(4)(a) requires that individuals that exercise human oversight of AI systems "fully understand the capacities and limitations of the high-risk AI system." For many AI systems, whether highly complex models with millions or billions of parameters or relatively simple hand-coded models, "fully understanding" the system is impossible. Rather individuals should be required to "adequately understand" the system to exercise effective oversight. Specific related recommendations include:
    - Article 14 (1) should be modified to ensure that human oversight is guaranteed where necessary to reduce risks "as far as possible and achieve accurate performance of an AI system"; and
    - Article 14 (4) should require that reasonable information about the operation of the AI system should be made available so the user sufficiently understands the AI system to ensure to the extent possible that it functions as intended by the producer.

The Final Regulation should clarify that human oversight of AI (or "human-in-the-loop") can occur on a continuous, intermittent, or retrospective basis (as it does in the testing industry). High-throughput computing allows systems to monitor millions (or even billions) of data points simultaneously. The system can act much faster than a typical human response time. Any AI system that ultimately relies solely or primarily on human attention and oversight cannot possibly keep up with the volume and velocity of algorithmic decision-making and is likely to be outmatched by the scale of the problem, causing potential harm to test takers.

5. Finally, the Proposed Regulation anticipates maximum administrative fines of up to €30m or 6% of total worldwide annual turnover in the event of non-compliance (meaning fines are higher than those under the GDPR). In this context, the ATP notes that enforcement guidance will be provided by a newly formed European Artificial Intelligence Board, which presumably will be similar in construction and form to the EDPB. However, unlike under the GDPR, actual enforcement under the Final Regulation will be the responsibility of national authorities competent in AI matters – there would be no single enforcement mechanism, and no guidance has been provided on what happens in the event of cross-border enforcement – resultant different rulings between countries. The ATP strongly recommends that this oversight be corrected in the Final Regulation.

Conclusion

The ATP appreciates the Commission's attention to its feedback on the application of the Proposed AI Regulation to the testing industry. First and foremost, the testing industry needs to have an appropriate regulatory definition of AI that helps the public understand where and what to be concerned about when AI is used, rather than treating every test that has some kind of mathematical formula attached to it as AI as a "high risk" activity. When measured against a proper, granular risk scale, uses of AI should be regulated in a fair, consistent, uniform, and reasonable manner.

The testing industry also needs – and supports – a simple unified global "soft law" regulatory standard; the ATP hopes that the Final Regulation can serve as such a standard, providing reasonable regulations without onerous reporting requirements. The testing industry is not afraid of regulation, provided the regulations make sense, are applied in a consistent manner across the board, and the cost of regulation bears a rational relationship to the benefits. When organizations are truly relying on AI/machine learning and/or autonomous systems, the ATP supports requiring transparency, assuming that compliance reporting systems are straightforward, reasonable, and easy to use.

The ATP is available to answer any questions the Commission regulators may have in response to this feedback. We would suggest that the Commission schedule public hearings with organizations that have submitted comments as part of its presentation to the European Parliament and Council; the ATP would welcome the opportunity to participate in such a hearing.

Sincerely,

ASSOCIATION OF TEST PUBLISHERS

William G. Harris, Ph.D.
CEO

John Kleeman
2021 Chair of the Board of Directors (representing Europe-ATP)
Past Co-Chair, ATP International Privacy Subcommittee, Test Security Committee

Alan J. Thiemann
General Counsel