Written Testimony of
Arthur C. Evans Jr., PhD
Chief Executive Officer
American Psychological Association
U.S. Senate Artificial Intelligence Insight Forum

November 8, 2023

Leader Schumer, Senators Rounds, Heinrich, and Young, and other members of the Senate, thank you for the opportunity to submit testimony on the increasingly significant role psychological science plays in technology development, testing and implementation. Psychological science plays a role in understanding the impact of artificial intelligence (AI), in helping policymakers think about objectives for regulating AI, and in guiding thinking about what types of AI should be encouraged for betterment of human society. AI is being built by humans and introduced into human systems. Therefore, a psychological understanding of humans should be central to decision-making about AI. I am Arthur C. Evans Jr., PhD, Chief Executive Officer of the American Psychological Association and APA Services, the companion organization of the APA. APA is the nation's largest scientific and professional nonprofit organization representing the discipline and profession of psychology, with over 146,000 members and affiliates who are clinicians, researchers, educators, consultants, and students in psychological science.

Helping ensure that technology responsibly shapes the future requires understanding the psychology of human-technology interaction. Just as new technological tools are emerging, the psychological science of human interaction with new products and services including AI is also developing.

Labor

APA acknowledges that new AI technologies can have profound positive impacts on the workplace and worker productivity when developed, tested, and deployed in a responsible way. Already, psychologists are working on understanding perceptions of AI in the workplace and are also involved in the development of more effective workplace tools. For example, psychologists have developed job descriptions, in tandem with natural-language processing algorithms, to estimate the knowledge and skills necessary for existing and future jobs. [i] This work underscores the importance of including psychological scientists in the development, regulatory, and evaluation of these new technologies.

In a recent survey fielded by APA and The Harris Poll, workers were asked about their perceptions of AI in the workplace. [ii] 51% of respondents who worry about AI said their work has a negative impact on their mental health, compared with 29% of workers not worried about AI. The survey shows that 46% of workers overall worried about AI making part or all of their job duties obsolete. We see that worry about AI is more disproportionately felt among workers with less education, workers of color, and younger workers. [iii] These worries, both real and perceived, impact the security and job satisfaction of millions of workers. It is essential to understand and create resources to support the workers most susceptible to worry and displacement by AI.

Three additional AI challenges facing workers are working conditions, discrimination, and job displacement. There have been recent reports that highlight the poor working conditions of those individuals responsible for training and inputting data into new AI tools. [iv] Often for low wages and centralized to workers in the global south, the work necessary to create tools like ChatGPT comes at a high human cost. We know that workers in these conditions are likely to have a lower level of life satisfaction and higher prevalence of depression and anxiety. [v] These workers are also exposed to disturbing images, which is an activity associated with several other mental health symptoms. [vi] Because of lack of transparency requirements currently being imposed on

companies developing new AI technologies, there is not more information about these workers and their condition. Congress can and should do more to prevent these working conditions and potential exploitation that can lead to mental health harms.

We are beginning to understand how AI might displace workers or entire sections of our economy. Researchers "argue that the conditions for AI to either enhance or threaten workers' sense of identity derived from their work depends on how the technology is functionally deployed (by complementing tasks, replacing tasks, and/or generating new tasks) and how it affects the social fabric of work."[vii] Congress has the ability to and should set resources aside to prepare for these impacts and create an environment that helps to minimize them.

Requiring diversity training, especially in teams developing AI tech is an important first step to ensuring teams developing these new products are equipped to help reduce the possibility of a biased product.[viii] Teams developing and testing these technologies must have built-in processes for feedback from a variety of stakeholders and populations during all parts of the development from design to testing. Moreover, there should be transparent processes in place for the potential harms of the technologies to be evaluated, especially with reference to vulnerable populations.

Health Care

AI-based tools already assist in the diagnosis and treatment of mental health conditions. APA supports developing evidence-based guidelines and regulations on AI technology use in health care settings. AI technology must undergo continual design, development, and monitoring quality control processes to mitigate the biases to which it is susceptible. Social science is essential to determine the efficacy of AI-based tools as they are used in the real world. Psychologists are uniquely positioned to contribute to these research efforts as their expertise in representative sampling helps to mitigate biased datasets; psychologists also study the impact of interactions between people and human-like AI-tools on both individual users and society more broadly. Additionally, without established regulatory and reimbursement pathways, these technologies will fail to reach scalability; therefore, APA encourages continued collaboration between the Food and Drug Administration and Centers for Medicare and Medicaid Services to develop regulatory and reimbursement pathways for the appropriate use of AI technology in mental health care.

Although we need to continue supporting and expanding the mental health care workforce, artificial intelligence and innovative technologies can improve health care systems and service delivery. For example, AI technologies may improve patient care by enhancing diagnostic precision, allowing for more individualized treatment, and improving engagement. AI may assist in clinical work by automating administrative tasks and supporting decision making. Technology underpinned by AI aimed at identifying and treating social isolation, mental and physical health issues, and providing new tools from transportation to caregiving already show promise in improving health.[ix] Some of these tools provide a means for helping predict health risks through intake assessments, and others interface directly with patients in the form of therapeutic chatbots.[x]Additionally, AI could foster the early detection of behavioral health concerns and enable the scaling of interventions to reach a much broader segment of the population than currently can access care. These developments, if realized, appropriately designed, and evaluated, would represent improvements in the lives of individuals, significant benefits to overall population health, and decreases in overall health care costs.

Yet, the use of AI technology within mental health care raises several ethical considerations. First, as we well know by now, AI algorithms can be susceptible to biases based on the data they are trained on, only exacerbating existing disparities. Second, given the sensitivity of health data, there needs to be considerable attention to data safeguarding to prevent harm. Users of apps, chatbots, and other AI-based technology must be informed about the privacy implications, including what data are being collected and stored, who has access to the

data, and what happens to data after the use of an AI tool is discontinued. Third, greater transparency is needed regarding the development and technology underlying AI-based tools (while protecting copyright and intellectual property), as well as rigorous testing of the AI-based tools to foster public trust, establish efficacy, and minimize unintended consequences. Because AI operates within a sociotechnical system, social sciences researchers and practitioners are critical in addressing these three considerations.

Unexpected or unacknowledged collection of data being employed in the training of AI models also raises privacy and transparency concerns. The prevalence of AI tools is often unknown by the end user, fueling misunderstandings, misconceptions, and fears about these technologies; companies should take steps to increase transparency around the presence of these tools to ensure individuals are aware of the role AI might play in impacting their experiences in real world and online environments. Using and exploiting personal or even publicly available data has serious consequences on the perception and experience of AI tools.[xi] AI education is essential in order for people to understand that AI is one thing, but instead many types of technology that affect perception and trust.[xii] In other words, the nature of AI and the context in which AI is deployed determines human trust in the technology, and the limits of this trust must be closely considered.[xiii]

Overall, the use of AI technologies must be limited to applications where it is the most safe and effective, minimizing unknown, unintended, and inequitable impacts. New AI technologies have been shown to be improved by incorporating psychological science, which has a long history of navigating complex human-technical interactions and related legal and ethical issues.[xiv] In moving forward with AI applications in health care, we must always strive to monitor and maintain the right balance between continued innovation and public safety.

Education

AI technologies can impact the way students at all levels learn and consume information. Technologies aimed at personalized learning programs, adaptable lesson planning, and new levels of accessibility within education all can revolutionize the way young people learn. In the *Artificial Intelligence and the Future of Teaching and Learning* report, recently published by the Department of Education, both challenges and opportunities for AI in the classroom are explored.[xv] APA encourages Congress to review this report to ensure that new policies made around AI and education do not limit positive impacts of innovative technologies.

Young people often consume, process, and retain information differently than adults, and tools aimed at younger users must be developed and tested with these factors in mind. Adolescent development is gradual and continuous, beginning with biological and neurological changes occurring before puberty (i.e., approximately beginning at age 10), and lasting at least until approximately 25 years of age.[xvi] Age-appropriate use of AI should be based on each adolescent's level of maturity (e.g., self-regulation skills, intellectual development, comprehension of risks) and home environment.[xvii] Because development differs, and because there are no data available to indicate that children become unaffected by the potential risks and opportunities at a specific age, caution must be exercised. In general, risks are likely to be greater in early adolescence—a period of greater biological, social, and psychological transitions—than in late adolescence and early adulthood.[xviii]

It is therefore essential that tools used more broadly across an adult population are not directly repurposed for the youth population. We encourage additional study of the opportunities and risks associated with AI technologies to be untaken in partnership with the Department of Education, the Department of Health and Human Services, and federal research institutions such as the National Science Foundation.

Any use of AI tools by adolescents should be preceded by training in digital literacy programs to ensure that users have developed psychologically informed competencies and skills that will maximize the chances for balanced, safe, and meaningful technology experience. Emerging science offers support increasing the frequency

of positive interactions with technology.[xix] Other competencies should include an understanding of: the AI technologies underlying the tools, the data being consumed by the AI tool and being used to train the tool, the limits of the outputs of AI tools, the potential bias and limitations of AI tools, the accuracy and representativeness of information generated by an AI tool, and how best to use and convey information gleaned through the use of AI tools in an educational context.

Integration of AI tests and tools into admission processes for educational institutions presents several pitfalls. We recommend the government thoroughly investigate those tests and tools in this area prior to their deployment. The AERA/APA/NCME *Standards for Educational and Psychological Testing* contain professional principles for test development that reflect decades of accumulated wisdom on how tests are developed, deployed, evaluated, monitored, and secured.[xx] This document is not only relevant for AI tools and test development, is even more relevant than ever before.

AI Standards

Policymakers and employers alike should be aware of the risks and benefits inherent to developing and using AI in the workplace. To maximize the efficiency and effectiveness of AI, and minimize damages to labor markets, workplaces, and employees, APA believes that AI should:

– **Be Human-centric** – People must remain a central part of technologies involving all stages of artificial intelligence development, including the development, deployment, and evaluation stages.
– **Elevate Ethics & Diversity** – Diversity must be present in the individuals, backgrounds, perspectives, and thought processes of those responsible for the creation and deployment of AI. Companies must endeavor to embed diversity and reduce discrimination in both the development of AI tools and their deployment.
– **Be Well-Researched** – Psychological science focused on the development and deployment of artificial intelligence is a new area of study, and more research is urgently needed to ensure effective development of these technologies and their impact on human behavior.
– **Be Effective & Risk Averse**– New uses for AI technologies are discovered every day and deployment of these tools is often done without proper consideration of ethics and efficacy. AI must be properly tested, and the use of these technologies must be limited to applications where it is the safest and most effective in achieving the intended purpose.
– **Prioritize Privacy** – The data collection fueled by the increased deployment of AI tools is leading to increased risks for individual data privacy. Companies must redouble their efforts to protect data and ensure privacy around data collected by AI technologies. This should include copyright and trademark protection.
– **Maintain Copyright & Intellectual Property Rights** – Current AI tools and programs exist in violation of copyright and intellectual property standards required of other similarly situated technologies, more must be done to ensure compliance with current standards, and new standards must be created where necessary.
– **Be Transparently Deployed** – The prevalence of AI tools deployed in the creation and distribution of content is often unknown by the end user, fueling misunderstandings and misconceptions about the technologies; companies should take steps to increase transparency to ensure individuals are aware of the role AI might play in impacting their experience in real world and online environments.
– **Further Digital Literacy** – It is essential that new avenues of education be paired with the deployment of new technologies. Users should be informed about the way new AI technologies work, how they might impact their lives, and how best to interact with these platforms to minimize negative impacts.

As Congress is considering areas within the federal government to centralize AI leadership, APA believes the Office of Science and Technology Policy and its National AI Advisory Committee are well situated to create cross-cutting government policies related to AI. We also believe that there is a role for the National Science

APA.ORG
APASERVICES.ORG

Advocating for APA members and psychology

750 First Street, NE
Washington, DC 20002-4242

202.336.5800
202.336.6123 TDD

4

Foundation, the National Institutes of Health, specifically the National Institute of Mental Health, and the National Institute on Minority Health and Health Disparities. It is imperative that the government take concrete steps now to form governing bodies and research efforts to better understand the impact of AI.

Early evidence shows that decisions about AI and how it is implemented reflect the world view and values of the human beings who design them and set policy for how it is used. Given the massive and increasing influence of AI on people's lives, it is critical to better appreciate how people understand and react to such influence. [xxi] AI ethics and psychology is an evolving discipline essential to the study of how AI learns from society and humans and how AI makes decisions consequential for humans in critical settings.[xxii] These implicit and explicit biases that can be captured by AI algorithms have been documented in social psychology for decades include racial, gender, sexuality, ability, and age attitudes.[xxiii] There is even evidence that AI has been responsible for keeping Black patients from receiving adequate mental health care.[xxiv] Moreover, these findings provide insights about how language might be impacting the social cognition of both AI and humans. These findings present potential issues in numerous contexts and if not studied, regulated, and carefully deployed, have the potential to perpetuate racial, class, and structural barriers for individuals.

The European Union, in their recently proposed EU (European Union) AI Act, have taken an important step to recognize the potential for unfairness and discrimination that new AI tools present, stating that "'diversity, non-discrimination and fairness' means that AI systems shall be developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited."[xxv] This regulation seeks to minimize the negative impacts we know are possible with AI. Adoption of a similar proposal at the federal level in the U.S. would be a crucial step forward in avoiding discrimination and furthering fairness. A similar framework has been proposed by psychological scientists that can also be considered.[xxvi]

For example, developing transparency enhancing algorithms for measuring and simulating AI bias and equity would make it possible to analyze the ethical implications of AI in a variety of domains, including natural language and computer vision.[xxvii] Alternatively, these AI methods could examine and analyze current and historical social biases and human cognition.[xxviii] Research focused on this area allows for understanding how AI is co-evolving with humanity.

Data privacy training should be embedded into processes for the development, deployment, and evaluation of AI tools. AI technologies collect and store substantial amounts of data, including personally identifiable data. Adequate steps must be taken to ensure that the data collected by these tools is meeting privacy laws and copyright policies. As with a requirement for data privacy, so should there be requirements for transparency around AI tools. Access for researchers should not be limited to only those that work within the company developing the tools and mechanisms must be built into new technologies that easily allow for researchers to access and analyze anonymized data.

Without incorporating psychological science deeply into the development of AI tools, we risk continuing to harm already disadvantaged populations and creating systems that perpetuate harmful stereotypes and bias. AI systems are often trained using large data sets that have the potential to integrate biases related to gender identity, race, and other characteristics. These systems then spread the biases in their interactions with humans or other technology-informed systems, with implications for equity and fairness. Psychologists' research on the various forms of resulting bias and the detrimental impacts are being used to develop data sets that are less biased and AI systems that can detect and compensate for biases in data. Findings from this research should be incorporated into future deployments of artificial intelligence tools, especially when being funded or used by the federal government.

APA.ORG
APASERVICES.ORG

Advocating for APA members and psychology

750 First Street, NE
Washington, DC 20002-4242

202.336.5800
202.336.6123 TDD

5

[i] Putka, D.J., Oswald, F.L., Landers, R.N. *et al.* (2023). Evaluating a Natural Language Processing Approach to Estimating KSA and Interest Job Analysis Ratings. *Journal of Business and Psychology, 38*, 385–410. https://doi.org/10.1007/s10869-022-09824-0

[ii] American Psychological Association. (2023). *2023 Work in America survey: Artificial intelligence, monitoring technology, and psychological well-being.* https://www.apa.org/pubs/reports/work-in-america/2023-work-america-ai-monitoring

[iii] *Id.*

[iv] Perrigo, B. (2023, January 18). Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer. Time. https://time.com/6247678/openai-chatgpt-kenya-workers/

[v] Howell, R. T., & Howell, C. J. (2008). The relation of economic status to subjective well-being in developing countries: A meta-analysis. Psychological Bulletin, 134(4), 536–560. https://doi.org/10.1037/0033-2909.134.4.536; Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. Science, 370(6522), eaay0214. DOI: 10.1126/science.aay0214

[vi] Holman, E. A., Garfin, D. R., Lubens, P., & Silver, R. C. (2020). Media Exposure to Collective Trauma, Mental Health, and Functioning: Does It Matter What You See? Clinical Psychological Science, 8(1), 111–124. https://doi.org/10.1177/2167702619858300

[vii] Selenko, E., Bankins, S., Shoss, M., Warburton, J., & Restubog, S. L. D. (2022). Artificial intelligence and the future of work: A functional-identity perspective. Current Directions in Psychological Science, 31(3), 272-279. https://doi.org/10.1177/09637214221091823

[viii] Cowgill, Bo and Dell'Acqua, Fabrizio and Deng, Sam and Hsu, Daniel and Verma, Nakul and Chaintreau, Augustin, Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics (June 1, 2020). In Proceedings of the 21st ACM Conference on Economics and Computation (pp. 679-681)., Columbia Business School Research Paper Forthcoming, Available at SSRN: https://ssrn.com/abstract=3615404 or http://dx.doi.org/10.2139/ssrn.3615404

[ix] Matheny, M. E., Whicher, D., & Israni, S. T. (2020). Artificial intelligence in health care: a report from the National Academy of Medicine. *Jama*, *323*(6), 509-510. https://jamanetwork.com/journals/jama/article-abstract/2757958; Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H. C., ... & Jeste, D. V. (2021). Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(9), 856-864. https://doi.org/10.1016/j.bpsc.2021.02.001; Dhelim, S., Chen, L., Ning, H., & Nugent, C. (2023). Artificial intelligence for suicide assessment using audiovisual cues: A review. *Artificial Intelligence Review*, *56*(6), 5591-5618. https://doi.org/10.1007/s10462-022-10290-6

[x] Fiske, A., Henningsen, P., & Buyx, A. (2019). Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *Journal of Medical Internet Research, 21*(5). https://www.jmir.org/2019/5/e13216/

[xi] Sethumadhavan, A. (2019). Trust in Artificial Intelligence. *Ergonomics in Design*, *27*(2), 34–34. https://doi.org/10.1177/1064804618818592

[xii] Longoni, C., Cian, L., & Kyung, E. J. (2023). Algorithmic Transference: People Overgeneralize Failures of AI in the Government. *Journal of Marketing Research*, *60*(1), 170–188. https://doi.org/10.1177/00222437221110139

[xiii] Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in Artificial Intelligence: Meta-Analytic Findings. *Human Factors*, *65*(2), 337–359. https://doi.org/10.1177/00187208211013988

[xiv] Gopnik, A. (2017). Making AI More Human. *Scientific American, June 2017*. alisongopnik.com/Papers_Alison/scientificamerican%20Gopnik0617-60.pdf

[xv] Artificial Intelligence and the Future of Teaching and Learning Insights and Recommendations. (2023). https://www2.ed.gov/documents/ai-report/ai-report.pdf

[xvi] Somerville, L. H., & Casey, B. J. (2010). Developmental neurobiology of cognitive control and motivational systems. Current Opinion in Neurobiology, 20(2), 236–241. https://doi.org/10.1016/j.conb.2010.01.006

[xvii] Orben, A., & Blakemore, S.-J. (2023). How social media affects teen mental health: A missing link. Nature, 614(7948), 410–412. https://doi.org/10.1038/d41586-023-00402-9

[xviii] Magis-Weinberg, L., Ballonoff Suleiman, A., & Dahl, R. E. (2021). Context, development, and digital media: Implications for very young adolescents in LMICs. Frontiers in Psychology, 12, Article 632713. https://doi.org/10.3389/fpsyg.2021.632713; Orben, A., Przybylski, A. K., Blakemore, S.-J., Kievit, R. A. (2022). Windows of developmental sensitivity to social media. Nature Communications, 13(1649). https://doi.org/10.1038/s41467-022-29296-3

[xix] Common Sense Media. (2019, May 10). Digital citizenship | Common Sense Education. https://www.commonsense.org/education/digital-citizenship; Magis-Weinberg, L., Muñoz Lopez, D. E., Gys, C. L., Berger, E. L., & Dahl, R. E. (2022). Short research article: Promoting digital citizenship through a school-based intervention in early adolescence in Perú (a pilot quasi-experimental study). Child and Adolescent Mental Health. Advance online publication. https://doi.org/10.1111/camh.12625

[xx] The Standards for Educational and Psychological Testing. (n.d.). https://Www.apa.org. https://www.apa.org/science/programs/testing/standards

[xxi] Pandey, A., & Caliskan, A., (2021). *Disparate Impact of Artificial Intelligence Bias in Ridehailing Economy's Price Discrimination Algorithms.* In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. 822-833.

[xxii] Caliskan, A., Bryson, J.J., & Narayanan, A., (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183-186. 10.1126/science.aal4230.

[xxiii] Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*(2), 171–192. https://doi.org/10.1037/a0032734.

[xxiv] Ziad Obermeyer et al. ,Dissecting racial bias in an algorithm used to manage the health of populations. Science 366,447-453(2019).DOI:10.1126/science.aax2342.

[xxv] DRAFT Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. (n.d.). https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf

[xxvi] Landers, R. N., & Behrend, T. S. (2023). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. American Psychologist, 78(1), 36–49. https://doi.org/10.1037/amp0000972

[xxvii] Steed, R., & Caliskan, A. (2021). A set of distinct facial traits learned by machines is not predictive of appearance bias in the wild. *AI Ethics 1*, 249–260. https://doi.org/10.1007/s43681-020-00035-y.

[xxviii] Caliskan, A., & Lewis, M. (2020, July 16). Social biases in word embeddings and their relation to human cognition. https://doi.org/10.31234/osf.io/d84kg.