AMERICAN PSYCHOLOGICAL ASSOCIATION SERVICES, INC.

February 2, 2024

Rachel Trello
National Institute of Standards and Technology
100 Bureau Drive
Stop 8900
Gaithersburg, MD 20899–8900

Response to Request for Information Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence (Sections 4.1, 4.5, and 11)

Submitted Electronically

To Whom It May Concern:

The American Psychological Association Services, Inc. (APA Services) respectfully submits the following response to your Request for Information to aid the implementation of new policies concerning Artificial Intelligence (AI). APA Services is the companion organization of the American Psychological Association, which is the nation's largest scientific and professional nonprofit organization representing the discipline and profession of psychology, as well as over 157,000 members and affiliates who are clinicians, researchers, educators, consultants, and students in psychological science.

Helping ensure that technology shapes the future for the better requires understanding the psychology of human-technology interaction. Just as new technological tools are emerging, the psychological science on human interaction with new products and services is also developing. A burgeoning area of psychological science involves the development, use, and impact of artificial intelligence.

Psychologists who study human intelligence are unlocking ways to enhance artificial intelligence and to ensure its ethical and appropriate use. Applied psychologists are studying how important characteristics of AI technologies impact identity and psychological wellbeing. Clinicians are navigating an increasingly complex environment where AI tools can assist in evaluation and diagnosis. Meanwhile, research points out that the deployment of AI is not a one-size-fits-all solution and is perceived differently by different groups of people. For example, even when factors such as age, gender, education, technology skills, and country-level GDP per capita are held constant, those living in societies with greater inequality are more likely to perceive AI as a threat to their role in the workplace.[1]

1. Developing Guidelines, Standards, and Best Practices for AI Safety and Security

[1] Shoss, M. K., & Ciarlante, K. (2022). Are Robots/AI Viewed as More of a Workforce Threat in Unequal Societies? Evidence From the Eurobarometer Survey. *Technology, Mind, and Behavior*, *3*(2: Summer 2022). https://doi.org/10.1037/tmb0000078

It is essential that NIST and the broader federal government take steps to create guidelines and standards for best practices in AI safety and security. These new and evolving technologies present important questions across our society that must be carefully considered. Executive Order 14110 takes important steps towards directing agencies to create these guidelines and we encourage NIST to consider the implications in the following areas when determining how best to provide guidance for other government actors. We also encourage NIST and other government agencies to continue to solicit comments and to engage with the expertise psychologists bring to the important questions surrounding these new technologies.

Health Care

We are already seeing the deployment of AI-based tools to assist in the diagnosis and treatment of mental and behavioral health conditions. APA strongly supports the development of evidence-based guidelines and regulations regarding the use of AI technology in a healthcare setting, particularly in behavioral health care. AI technology must undergo rigorous design and development quality control processes to mitigate the biases to which it is susceptible. Ongoing research is needed to determine the efficacy of the varied AI-based tools that currently are and will become available. Psychologists are uniquely positioned to contribute to these research efforts as their expertise in representative sampling may help to mitigate biased datasets; psychologists also are well equipped to study the impact of interactions between people and human-like AI-tools on both the individual users and society more broadly. Additionally, a lack of regulatory and reimbursement pathways may hinder the ability for these new technologies to scale; therefore, APA encourages continued collaboration between the FDA and CMS to develop regulatory and reimbursement pathways for the appropriate use of AI technology in mental health care.

Supporting Medical Innovation

While we need to continue to support and expand the mental and behavioral health care workforce, artificial intelligence and innovative technologies have the potential to positively impact the health care system and the future of service delivery. For example, AI technologies may improve patient care by enhancing diagnostic precision, allowing for more individualized treatment, and improving engagement. AI technologies may assist psychologists and other health professionals in their clinical work by automating administrative tasks and supporting clinical decision making. 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.

Technology aimed at identifying and treating social isolation, mental and physical health issues, and providing new tools from transportation to caregiving show promise in improving health.[2] Some of these tools provide a means for helping predict health risks or recommend

[2] 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*

personalized treatment plans, and others interface directly with patients in the form of therapeutic chatbots.[3]

Yet, the use of AI technology within mental health care raises several ethical considerations. First, AI can be susceptible to bias; if the datasets used to train AI technologies are incomplete or reflect systemic biases, then the resulting AI technologies also will reflect those biases. Ultimately, this could exacerbate existing disparities and inequities. Ethics and diversity training for individuals involved in the development of new AI tools may help mitigate this effect. Secondly, given the sensitivity of mental health data, there needs to be considerable attention to data safeguarding to prevent public harm. Users of apps, chatbots, and other AI-based technology must be comprehensively informed about the privacy implications, including what personal data is being collected, how it is stored, who has access to it, and what happens to their data after they stop using the tool. Thirdly, 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, to establish efficacy, and to minimize unintended consequences.

Ultimately, finding a balance between continued health care innovation and public safety is imperative. Psychologists should be at the forefront of AI development and employment to ensure that human-technology interactions are effective and ethical.

Medical Ethics and Protecting Patients

APA wholeheartedly agrees that new AI tools should centralize transparency and effectiveness in their development and deployment. Data being employed in the training of AI models raises serious privacy and transparency concerns. The prevalence of AI tools is often unknown by the end user, fueling misunderstandings and misconceptions about the 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 experience in real world and online environments. The use and exploitation of data will have unknown consequences on the perception and experience of AI tools.[4] AI, like humans, are varied; yet, people often see AI as homogeneous. This may result in AI failures being treated as global rather than specific, affecting perception and trust in AI.[5] Trust between humans is earned. The human, the AI, and the context in which AI is employed determine trust in the technology, and the limits of this trust must be closely considered.[6]

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

[3] 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/

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

[5] 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

[6] 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

Transparency requirements should be paired with public education efforts centered around technological literacy. Users should be informed about the way new AI technologies work and use user-provided data, how they might impact their lives, and how best to interact with these platforms to minimize negative impacts.

Deployment of AI tools is often done without proper consideration of ethics and efficacy. AI must be properly tested prior to widespread deployment and the use of these technologies must be limited to applications where it is safe and effective, minimizing unknown, unintended, and inequitable impacts. These new AI technologies can be greatly improved by psychological science.[7] User experience researchers and designers with strong backgrounds in psychological science can play an important role in ensuring new tools are more human, more effective, and more ethical.

The data collection being fueled by the increased deployment of AI tools is leading to increased risks for individual data privacy. Companies must be required to redouble their efforts to protect data and ensure privacy around data collected by AI technologies, especially in the case of sensitive information possessed by health care providers and when copywritten material is present. Companies must also be more transparent related to user data practices and data used for AI training.

Education

General Policy

As with other areas of our daily lives, AI technologies can massively impact the way students at all education levels learn and consume information. Technologies aimed at personalized learning programs, adaptable lesson planning, and creating new levels of accessibility within education all have the potential to 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 Office of Educational Technology, both challenges and opportunities for AI in the classroom are explored.[8] APA encourages NIST to review this comprehensive report to ensure that new policies made around AI and education do not inherently limit the positive impacts of new technologies.

As with other concerns expressed throughout this response, there are inherent risks that must be anticipated to avert any negative consequences associated with the deployment of these new tools. Young people often consume, process, and retain information much differently than adults, and tools aimed at young 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 is observable (i.e., approximately beginning at 10 years of age), and lasting at least until dramatic changes in youths' social environment (e.g., peer, family, and school context) and neurological changes have completed (i.e., until approximately 25 years

[7] Gopnik, A. (2017). Making AI More Human. *Scientific American, June 2017*. alisongopnik.com/Papers_Alison/scientificamerican%20Gopnik0617-60.pdf
[8] Artificial Intelligence and the Future of Teaching and Learning Insights and Recommendations. (2023). https://www2.ed.gov/documents/ai-report/ai-report.pdf

of age).[9] Age-appropriate use of artificial intelligence should be based on each adolescent's level of maturity (e.g., self-regulation skills, intellectual development, comprehension of risks) and home environment.[10] Because adolescents mature at different rates, and because there are no data available to indicate that children become unaffected by the potential risks and opportunities posed by artificial intelligence usage at a specific age, research is in development to specify a single time or age point when it might be appropriate to begin introducing a young person to artificial intelligence-backed tools. In general, potential 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.[11]

Tools that are used more broadly across an adult population should not be directly repurposed for the youth population. Ethical considerations must be centralized to avoid unintended consequences. We encourage additional study of the opportunities and risks associated with AI technologies in the education space 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.

Fostering Students' Understanding of AI

Adolescents' use of artificial intelligence tools 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 social media use. Emerging science offers preliminary support for the efficacy of Digital Citizenship and Digital Literacy to increase the frequency of positive interactions with technology; however, more research is needed in this area.[12] Additional competencies could also 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.

Higher Education Admissions

The higher education admissions process is one with potential challenging issues related to the integration of AI-based tools. This presents several fundamental research opportunities the government must investigate prior to the deployment of tools in this area. AI Ethics and Psychology is an evolving discipline essential to the study of how AI learns from society and

---

[9] 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

[10] 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

[11] 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

[12] 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

humans and how AI makes consequential decisions in critical settings.[13] Studies have demonstrated that AI automatically learns implicit biases from language corpora and accordingly perceives the world in a biased manner.[14] These implicit biases that have been documented in social psychology for decades include racial, gender, sexuality, ability, and age attitudes.[15] There is even evidence that AI has been responsible for keeping Black patients from receiving adequate mental health care.[16] 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 the context of higher education admissions process and if not studied, regulated, and carefully deployed, have the potential to perpetuate racial, class, and structural barriers to entry for individuals seeking admission.

Labor

Practical Uses for AI in the Workplace

APA acknowledges that new AI technologies can have profound positive impacts on the workplace and worker productivity when developed, tested, and deployed responsibly. Already, psychologists are working on understanding perceptions of AI in the workplace and are involved in developing AI for more effective use in the workplace. For example, psychologists recently used job descriptions, in tandem with natural-language processing algorithms, to estimate the knowledge and skills necessary for existing and even future jobs.[17]

In a recent survey fielded by APA and The Harris Poll, workers were asked about their perceptions of AI in the workplace.[18] The survey shows 46% of workers 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, high school graduates worry about AI at a rate 10 percentage points higher than those that completed a four-year degree. Similar disparities were noted in workers of color and younger workers.[19] This upheaval will impact the job satisfaction of workers as well as their security. These disparities are likely to grow as the prevalence of AI technologies impact more parts of the workforce. It is essential to understand and create resources to serve the workers most vulnerable to worry and displacement by AI. This work underscores the importance of including psychological scientists in the development, regulatory, and evaluation of these new technologies.

---

[13] 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.

[14] 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.

[15] Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4; Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464.

[16] 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.

[17] 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

[18] 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

[19] *Id.*

Researchers are beginning to understand how AI technologies might displace individual 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."[20] The federal government has the ability to set resources aside to prepare for these negative impacts and create an environment that helps to minimize them.

2. Reducing the Risk of Synthetic Content

The World Health Organization has declared we are experiencing an "infodemic that continues to undermine the global response and jeopardizes measures to control the pandemic."[21] This infodemic is exemplified by an overabundance of both online and offline information and co-evolved with the COVID-19 epidemic; to improve communication, equal consideration should be given to the provision of accurate and concise health information and general information.[22] Misinformation is dangerous; without being able to believe the information being shared, people may opt not to use all available tools to improve their safety and the safety of the larger community. This risk, and the availability and prevalence of inaccurate information is likely to increase through the proliferation of new AI tools.

Psychological research suggests several methods for countering misinformation that should be included in regulations and developmental recommendations for AI technologies.[23] One method is to debunk incorrect information once it has been shared; however, a more effective method of "prebunking" to inoculate people against the misinformation "virus" in the first place. It involves distributing truthful, science-based information from reputable sources that counter known misinformation. Another method for reducing people's willingness to believe misinformation is to help them deliberate and think about the information.[24] Data suggest that people share information on social media without giving much thought to its accuracy, but, when prompted, in a lab setting, to assess the validity of an unrelated headline, participants were then more discerning in what information they would share. Research has also found that people are

---

[20] 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

[21] World Health Organization, (2020) *Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation*, *https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation*

[22] Gallotti, R., Valle, F., Castaldo, N. *et al.* Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nat Hum Behav* 4, 1285–1293 (2020). https://doi.org/10.1038/s41562-020-00994-6

[23] Greene CM, Murphy G. Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation. J Exp Psychol Appl. 2021 Dec;27(4):773-784. doi: 10.1037/xap0000371. Epub 2021 Jun 10. PMID: 34110860.

[24] Bao, H., Cao, B., Xiong, Y., Tang, W., 2020. Digital Media's Role in the COVID-19 Pandemic. JMIR mHealth and uHealth 8, e20156.. doi:10.2196/20156.

more susceptible to misinformation when they fail to carefully consider the information, regardless of whether it aligns with their political beliefs, further supporting the theory that helping people to slow down and deliberate about the information they are consuming may reduce the spread of misinformation.[25] Additionally, morality plays a significant role in people's acceptance of corrective information (pointing out what is untrue) and tailoring messages correcting COVID-19 misconceptions to connect with the morality of the recipients helps them receive the information.[26] Narrative methods for disseminating information may be more effective in influencing actual health attitudes and behaviors.[27] This is supported by another research group who found that using people-centered, first-person narratives with emotional language may help communicate pro-vaccine messages, by using narrative stories to personalize the issue.[28]

Further research is needed to understand the complex interactions between demographic factors such as age and misinformation, especially as it relates to interaction with new AI tools. To connect with the broadest possible audience, public health authorities must work to build relationships with trusted, influential stakeholders and media companies to reach culturally and linguistically diverse groups.[29] NIST must take steps to facilitate and centralize this research to ensure that new standards incorporate the impacts on diverse populations and design recommendations take into account the risks of misinformation.

3. Advance Responsible Global Technical Standards for AI Development

Unified standards for AI development and deployment should center the principles below to ensure ethical, efficient, and effective results from these new technologies. Policymakers and employers alike should be aware of the risks and benefits inherent to developing and using AI. To maximize the efficiency and effectiveness of AI, and minimize damages to labor markets, workplaces, and employees, APA believes that artificial intelligence standards should include:

1. **Human-in-the-loop requirements –** People must remain a central part of technologies involving artificial intelligence, including in the development, deployment, and evaluation stages.
2. **Center Ethics & Diversity –** Diversity must be present in the creation and deployment of new technologies. Companies must endeavor to embed diversity and reduce unconscious and conscious discrimination in both the development of AI tools and their deployment.

---

[25] Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

[26] Trevors, G., & Duffy, M. C. (2020). Correcting COVID-19 Misconceptions Requires Caution. Educational Researcher, 49(7), 538–542. https://doi.org/10.3102/0013189X20953825

[27] Gesser-Edelsburg A. Using Narrative Evidence to Convey Health Information on Social Media: The Case of COVID-19. J Med Internet Res. 2021 Mar 15;23(3):e24948. doi: 10.2196/24948. PMID: 33674257; PMCID: PMC7962859.

[28] Germani, F., Biller-Andorno, N., 2021. The anti-vaccination infodemic on social media: A behavioral analysis. PLOS ONE 16, e0247642.. doi:10.1371/journal.pone.0247642.

[29] Pickles K, Cvejic E, Nickel B, Copp T, Bonner C, Leask J, Ayre J, Batcup C, Cornell S, Dakin T, Dodd R, Isautier J, McCaffery K COVID-19 Misinformation Trends in Australia: Prospective Longitudinal National Survey J Med Internet Res 2021;23(1):e23805, URL: https://www.jmir.org/2021/1/e23805, DOI: 10.2196/23805.

3. **More research** – Psychological science focused on the development and deployment of artificial intelligence is a relatively new area of study and much more needs to be done to ensure effective development of these technologies and to understand the impact of these technologies on human behavior.
4. **Equitable and fair deployment–** Deployment of AI tools is often done without proper consideration of ethics and efficacy. AI must be properly tested prior to widespread deployment and the use of these technologies must be limited to applications where it is safe and effective, minimizing unknown, unintended, and inequitable impacts.
5. **Privacy –** The data collection being fueled by the increased deployment of AI tools is leading to increased risks for individual data privacy. Companies must be required to redouble their efforts to protect data and ensure privacy around data collected by AI technologies. Companies must also be more transparent related to user data practices and data used for AI training.
6. **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.
7. **Transparency** – The prevalence of AI tools is often unknown by the end user, fueling misunderstandings and misconceptions about the 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 experience in real world and online environments.
8. **Digital literacy training –** 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 is well situated to create cross-cutting government policies related to AI. We also believe that there is a role for the National Science 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.

Data privacy training should be embedded into processes for the development, deployment, and evaluation of AI tools. AI technologies collect and store large 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.

Non-Discrimination in AI Development and Deployment

Three AI challenges facing policymakers 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.[30] 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.[31] These workers are also exposed to high level of graphic images, which is an activity associated with a number of other mental health symptoms.[32] Because of a lack of transparency requirements currently being imposed on companies developing new AI technologies, there is not more information about these workers and their condition. Regulations can and should do more to prevent these working conditions and potential exploitation that can lead to mental health harms.

The European Union, in their recently proposed EU AI Act, have taken an important step to recognize the potential for unfairness and discrimination that new AI tools present.[33] 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 by Union or national law," 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 an important step forward in avoiding discrimination and furthering fairness. A similar framework has been proposed by psychological scientists that can also be considered.[34]

The 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, especially when the AI is perceived to be biased or unfair.

---

[30] Perrigo, B. (2023, January 18). Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer. Time. https://time.com/6247678/openai-chatgpt-kenya-workers/; Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18. https://doi.org/10.1145/3173574.3174023; Fairwork Cloudwork Ratings 2023 | Fairwork Cloudwork Ratings 2023 WORK IN THE PLANETARY LABOUR MARKET. (n.d.). https://fair.work/wp-content/uploads/sites/17/2023/07/Fairwork-Cloudwork-Ratings-2023-Red.pdf.

[31] 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

[32] 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

[33] 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

[34] 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

As with other areas of technology, there remains an imperative for diversity, ethics, and inclusion to be a central component of any oversight model. Requiring diversity training, especially in teams developing AI tech can be an important first step to ensuring teams developing these new products are well versed in the components of DEI, which in turn can help to reduce the possibility of a biased product.[35]

There are, additionally, ethical implications for what AI learns, how AI learns, and AI's subsequent decision-making. 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.[36] Alternatively, these AI methods could examine and analyze current and historical social biases and human cognition.[37] Research programs focused on this area would allow for understanding how AI is co-evolving with humanity, as AI is shaping society and impacting individuals' lives in an accelerated manner and at an unprecedented scale.

Given evidence that AI can reproduce discrimination and bias against individuals and groups, it is imperative the government leverage psychological science and examine people's expectations about and reactions to the fairness and potential discrimination of AI versus human agents. An emerging line of research suggests that people expect AI to be less biased than humans in some cases while being less outraged when they learn of bias from an AI versus human actors.[38] Algorithms appear less discriminatory than humans, perhaps incorrectly engendering trust and comfort from human users.

Without the inclusion of psychological science, policies risk harming already disadvantaged populations and creating systems that perpetuate harmful stereotypes and bias. AI systems are often trained using large data sets of human attributes or demographics 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. Behavioral 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.

Thank you again for the opportunity to weigh in on these questions. To the extent that our psychologists can serve as a resource to your office, please feel free to reach out to Corbin Evans (Senior Director Congressional & Federal Relations, Science Advocacy) at CEvans@apa.org.

---

[35] 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

[36] 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.

[37] 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.

[38] Jago, A. S., & Laurin, K. (2021). Assumptions About Algorithms' Capacity for Discrimination. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/01461672211016187.

Sincerely,

Katherine B. McGuire
Chief Advocacy Officer
American Psychological Association Services, Inc.