



AMERICAN PSYCHOLOGICAL ASSOCIATION

March 4, 2022

AI R&D RFI Response Team
Attn: Faisal D'Souza, NCO
2415 Eisenhower Avenue
Alexandria, VA 22314, USA

Submitted electronically via Regulations.gov

RE: RFI Response: National Artificial Intelligence Research and Development Strategic Plan

To Whom It May Concern –

The American Psychological Association (APA) appreciates the opportunity to comment on the National Artificial Intelligence Research and Development Strategic Plan require for information. This request represents a step in the right direction towards ensuring that stakeholders across disciplines are represented in future efforts to deploy artificial intelligence. In addition to the comments below, APA endorses the comment submitted from the Society for Industrial and Organizational Psychology.

APA is the largest scientific and professional organization representing psychology in the U.S., numbering over 133,000 researchers, educators, clinicians, consultants, and students. For decades, psychologists have played a vital role in the development and deployment of technologies and neurological science. These contributions have been essential to the currently available artificial intelligence enabled technologies and psychological science should continue to be at the heart of strategic planning of AI deployment.

The comments below represent three primary areas where the *Strategic Plan* should ensure the discipline of psychology is included: increased investments in research on artificial intelligence, ethics of artificial intelligence, and artificial intelligence and implicit bias.

Increased Investments in Research on Artificial Intelligence

APA strongly supports the need for additional investments in research related to Artificial Intelligence. From the current technological and research standpoint, it is almost impossible to predict the impact of future AI-informed technologies. There is an imperative that as the technologies grow in their capabilities and prevalence, that research surrounding their impact also increases. Future research funding in this area should ensure that psychological and behavioral science is adequately represented. The impact of AI on mental and behavioral health must continue to be examined to ensure we mitigate any harmful impacts caused by new systems.

750 First Street, NE
Washington, DC 20002-4242
(202) 336-5800
(202) 336-6123 TDD

Web: www.apa.org



AMERICAN PSYCHOLOGICAL ASSOCIATION

Ethics of Artificial Intelligence

There are some fundamental research opportunities the AI research community must investigate. AI Ethics and Psychology is an evolving discipline essential to the study of how AI learns from society and humans and how AI makes consequential decisions in critical settings.¹ Studies have demonstrated that AI automatically learns implicit biases from language corpora and accordingly perceives the world in a biased manner.² These implicit biases that have been documented in social psychology for decades include racial, gender, sexuality, ability, and age attitudes.³ Moreover, these findings provide insights about how language might be impacting the social cognition of both AI and humans.

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.⁴ Alternatively, these AI methods could examine and analyze current and historical social and human cognition.⁵ This research program 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. While beyond the scope of this Request for Information, there remains no comprehensive regulation for auditing how AI impacts equity and fairness in democratic societies.⁶ Consequently, these promising research areas of computer and

¹ 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](https://doi.org/10.1126/science.aal4230).

² 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.

³ 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>.

⁴ 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>

⁵ 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>

⁶ Caliskan, A. (2021, May 10). *Detecting and mitigating bias in natural language processing*. Brookings Institution. <https://www.brookings.edu/research/detecting-and-mitigating-bias-in-natural-language-processing/>



AMERICAN PSYCHOLOGICAL ASSOCIATION

information science contribute to data-driven policy making and law while having implications for psychology, political science, sociology, linguistics, and philosophy.

Artificial Intelligence and Implicit Bias

Given evidence that AI can reproduce discrimination and bias against individuals and groups, it is imperative to 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 and are less outraged when they learn of bias from an AI versus human actors.⁷ Algorithms appear less discriminatory than humans, perhaps incorrectly engendering trust and comfort from human users. 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.

Without the help of psychological science, we 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. 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.

While we remain broadly supportive of the strategic aims set forward by the *National Artificial Intelligence Research and Development Strategic Plan*, it is important that psychological and behavioral science is included in each strategy to ensure comprehensive consideration of the broad impact of AI technologies.

APA again thanks you for the opportunity to comment on this policy. If APA can be of any further assistance, please contact Corbin Evans, Senior Director of Congressional and Federal Relations, at CEvans@APA.org.

⁷ Jago, A. S., & Laurin, K. (2021). Assumptions About Algorithms' Capacity for Discrimination. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/01461672211016187>



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

Katherine Burnett McGuire, MS
Chief Advocacy Officer, APA