



## Cognitive Insights for Artificial Intelligence

Response to WH/OSTP RFI

Submitted by Monica Lopez, PhD, Chief Executive Officer

Organization, Cognitive Insights for Artificial Intelligence (CifAI)

April 3, 2022

On behalf of Cognitive Insights for Artificial Intelligence (CifAI), we write in response to the RFI to *Support the Development of a Federal Scientific Integrity Policy Framework*.

We support the efforts of the Executive Office of the President, Office of Science and Technology Policy (OSTP) in developing a framework for regular assessment and iterative improvement of agency scientific integrity policies and practices, and acknowledge that these efforts build on existing scientific integrity policies and practices as mentioned in the Presidential Memorandum released on January 27, 2021 about the establishment of principles and guidance for protecting scientific integrity.<sup>1</sup>

We at CifAI provide strategic research-based solutions from a human-centered perspective to ensure the safe and ethical design, development, deployment, and management of artificial intelligence (AI)-enabled autonomous systems across various industries. Our values-based approach is founded on accuracy, consistency, and context-dependency, and supports trusted data across every phase of the AI lifecycle to achieve confident and fair decision making.

We underscore the value of scientific integrity and we appreciate the opportunity to provide comments and recommendations on the following RFI focus: *Topic 1. On how scientific integrity policies at Federal agencies and other components of the Executive Branch can be developed or updated to address important and emergent issues of our time, including the subtopic (2) new technologies, such as artificial intelligence, machine learning, and the lack of transparency and potential for bias in computer algorithms and associated data.*

---

<sup>1</sup> The White House - Presidential Actions. Memorandum on Restoring Trust in Government Through Scientific Integrity and Evidence-Based Policymaking. 27 January 2021, <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/memorandum-on-restoring-trust-in-government-through-scientific-integrity-and-evidence-based-policymaking/>

## **Recommendation 1: Implement an ethically informed framework for AI R&D.**

To ensure “that science is conducted, managed, communicated, and used in ways that preserve its accuracy and objectivity,”<sup>2</sup> an ethical framework must guide AI R&D. We are witnessing the propagation of biases and magnification of discrimination as AI-enabled technologies advance, automate a range of low to high-risk decision-making tasks, and transform industries. In healthcare, for example, automated diagnostic tools powered by biased datasets are a detriment to public health.<sup>3</sup> This and other rising downsides of AI highlight that AI-enabled technologies for the benefit of humanity depend on responsibly built AI.

We assert that responsible AI must be firmly grounded in ethical standards of diversity, integrity, reliability, sustainability, and accountability. Moreover, mechanisms are needed to support the practical implementation of robust methodologies to ensure ethical standards are carried out in a dynamically changing landscape that is the field of AI. Today’s AI-enabled tools do not live up to high ethical standards. We propose several recommendations to help achieve high ethical standards.

- **Recommendation 1.1: Precisely define ethical AI.**

‘Ethical AI’ implies a code of conduct that is explainable, inclusive, positive in its purpose, and responsible in its use of data. However, code of conduct must adhere to well-defined ethical principles and values regarding equality, explainability, fundamental human rights, individual civil liberties, non-discrimination, prevention of harm, and respect for human autonomy, among other principles.<sup>4</sup> A standard for an ethical AI R&D code of conduct must be created.

- **Recommendation 1.2: Integrate ethics across the entire AI lifecycle.**

Ethical standards, as precisely determined in Recommendation 1.1, must incontrovertibly inform the design, development, deployment, and management of AI-enabled systems. From the beginning stages of the data acquisition process to the evaluation of impact of the optimized AI model employed during deployment, requirements of equitable, fair, and inclusive treatment of the user and the 100% safety and reliability of the technology should be the gold standards of an ethical approach. Data privacy must also be guaranteed as well as the economical and ecological sustainability of the data and algorithms themselves. Furthermore, given cultural differences across values, context necessitates fair and proper identification and consideration thereof. This requires maintenance of human intervention at all stages and collaboration with as many stakeholders of AI systems and their outcomes as possible.

---

<sup>2</sup> A report by the Scientific Integrity Fast-Track Committee of the National Science and Technology Council. January 2022. Protecting the Integrity of Government Science, [https://www.whitehouse.gov/wp-content/uploads/2022/01/01-22-Protecting\\_the\\_Integrity\\_of\\_Government\\_Science.pdf](https://www.whitehouse.gov/wp-content/uploads/2022/01/01-22-Protecting_the_Integrity_of_Government_Science.pdf)

<sup>3</sup> Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, 383(9), 874-882.

<sup>4</sup> European Commission. December 2018. Ethical guidelines for trustworthy AI, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Regulatory agencies and the development of evaluation methodologies and tools should be created to monitor the dynamic integration of ethics' standards throughout the AI lifecycle.

- **Recommendation 1.3: Eliminate biases, not just mitigate them after-the-fact.**

Assuming the establishment of ethical standards as delineated in Recommendation 1.2, all AI R&D must unequivocally include algorithmic decision explainability and dataset and algorithm bias analysis. This necessitates establishment of step-by-step procedures for AI research. Methodologies need to be designed to detect and eliminate bias to avoid, for example, the pursuit of unregulated and questionable methods to eliminate bias like the creation of synthetic data<sup>5</sup> through the introduction of another bias. Regulatory agencies and the development of evaluation methodologies and tools should be created to monitor bias elimination practices utilized and/or developed.

- **Recommendation 1.4: Be transparent with the type of data used and its purpose(s).**

Algorithms developed need to be acceptable. The success of today's deep learning systems depends on the availability and collection of large and accurate datasets. Insufficient data, poor quality and biased data, and incorrect data labeling produce unexpected and/or erroneous results. In healthcare, for example, the consequences of such can be fatal.<sup>6</sup> Sufficient data, data diversity, data quality, and data integrity need to be prioritized. Regulatory agencies and the development of evaluation methodologies and tools should be created to monitor use-case specifics and their requirements to ensure resulting automated decision-making is explainable and consequent policy decisions are justifiable.

- **Recommendation 1.5: Include a multidisciplinary working group on ethical AI R&D.**

To prevent introduction of discipline biases and thus prioritization of single solutions for the above recommendations, a multidisciplinary team of interdisciplinary advisors must be included in the review of AI R&D. AI is a dynamic field and highly interdisciplinary,<sup>7</sup> intersecting mathematics, computer science, engineering, and robotics with business, the cognitive and brain sciences, environmental science, history, philosophy, and the humanities more broadly. A continuously evolving working group of interdisciplinary experts from outside of the government's network of advisors should be created to provide regular assessment of the determined ethical AI R&D code of conduct.

- **Recommendation 1.6: Empower AI R&D scientists to prioritize ethics.**

Assuming the establishment of Recommendations 1.1, 1.2, 1.3, 1.4, and 1.5, AI R&D Federal scientists should be empowered to both flag concerns pertaining to any specific lack and/or abuse of AI R&D ethical codes of conduct, and identify remediation procedures before

---

<sup>5</sup> Gretel.ai, <https://gretel.ai/>

<sup>6</sup> Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, 383(9), 874-882.

<sup>7</sup> Dignum, V. (2019). AI is multidisciplinary. *AI Matters*, 5(4), 18-21. DOI: 10.1145/3375637.3375644, <https://sigai.acm.org/static/aimatters/5-4/AIMatters-5-4-07-Dignum.pdf>

continuing their scientific duties. Regulatory agencies and the development of evaluation methodologies and tools should be created to enable reporting of positive and negative conduct within ethical AI R&D.

- **Recommendation 1.7: Enforce principles and standards that ensure ethical AI R&D.**

Following from Recommendation 1.6, regulatory agencies and the development of evaluation methodologies and tools should be created to ensure the independent oversight and accountability of ethical AI R&D.

**Recommendation 2: Advocate alignment on the accuracy of Science.**

To ensure that science is protected “from suppression, manipulation, and inappropriate influence—including political interference,”<sup>8</sup> we must consider the current landscape in which modern science happens. Science does not exist in a vacuum. It is informed as much by societal norms and expectations as the platforms that sustain its discovery and dissemination. Since the commercialization of the Internet and the emergence of greater computing power and big data, we have created a new world for how we not only conduct but communicate science.

We assert that alignment between government and the makers of social media platforms—i.e. private sector—is critical to communicating truth about Science and its processes and outcomes. Powered by machine learning (ML), a subset of AI that enables machine systems to make accurate predictions from large quantities of data, modern social networks provide us the means to engage in highly personalized experiences at unprecedented speed with increasingly larger numbers of people. ML powers the features of popular social networks like Facebook, LinkedIn, and Snapchat, among others. Unfortunately, AI algorithms are neither neutral, nor have their own moral compass. AI reflects the biases of its creators, spreading misinformation and disinformation on empirical fact (and much more); recent notable examples include on the COVID-19 pandemic and vaccine efficacy,<sup>9</sup> among many others. Science is being challenged at unprecedented levels. We propose several recommendations to help achieve integrity across scientific communication.

- **Recommendation 2.1: Establish safeguards so AI systems do not promote falsities.**

AI systems must be built in a responsible way for the good of humanity. This includes respect for democracy, culture, nondiscrimination, and the right to safety, health, and prosperity. To minimize the abuse of AI-enabled technology and preserve citizens’ welfare as the primary priority of government, proper safeguards must be developed and put in place. Regulatory agencies must advance a strategy to establish proper safeguards.

---

<sup>8</sup> Protecting the Integrity of Government Science. A report by the Scientific Integrity Fast-Track Committee of the National Science and Technology Council. January 2022, [https://www.whitehouse.gov/wp-content/uploads/2022/01/01-22-Protecting\\_the\\_Integrity\\_of\\_Government\\_Science.pdf](https://www.whitehouse.gov/wp-content/uploads/2022/01/01-22-Protecting_the_Integrity_of_Government_Science.pdf)

<sup>9</sup> How to Address COVID-19 Vaccine Misinformation. Centers for Disease Control and Prevention, <https://www.cdc.gov/vaccines/covid-19/health-departments/addressing-vaccine-misinformation.html>

- **Recommendation 2.2: Collaborate to block the supply of false information.**

Collaborations between government agencies, the companies (and research institutes) leveraging on this AI technology, and a range of citizen stakeholders (e.g. advocates of social and consumer interests) are fundamental to addressing Recommendation 2.1. Regulatory agencies working with industry without immediately setting a prescriptive rule allows for innovation on methods to ensure algorithms protect the accuracy of science as scientific information proliferates across social media platforms. This certainly assumes that clear and detailed official scientific information critical to curtailing the creation and spread of false information is the gold standard at the onset.

- **Recommendation 2.3: Educate on the limits of AI.**

Communicating science is ultimately about educating the public on facts about the natural world. To educate many means acknowledging that everyone has different educational backgrounds, different learning styles, and different levels of access to information. To encourage a collective ability to ably understand, manage, and respond to scientific information through social media platforms, a user-centric approach to the presentation of facts must be taken. Regulatory agencies need to work on communicating science in an accessible way for the general public.