



## Cognitive Insights for Artificial Intelligence

Response to White House/Office of Science and Technology Policy Request for Information  
Submitted by Monica Lopez, PhD, Chief Executive Officer  
Organization, Cognitive Insights for Artificial Intelligence (CifAI)  
February 26, 2022

On behalf of Cognitive Insights for Artificial Intelligence (CifAI), we write in response to the RFI on the *National Artificial Intelligence Research and Development Strategic Plan*.

We support the efforts of the Office of Science and Technology Policy (OSTP) to update the 2019 National AI Research and Development Strategic Plan, and appreciate the opportunity to comment on the proposal concerning updates to the plan's strategic aims.

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. As AI-enabled technologies proliferate across a range of use cases both nationally and globally and calls for safety and reliability increase, we believe that the design, development, deployment, and management of these systems must be first and foremost intentional and responsible. Moreover, we believe that a diverse workforce who understands and actively pursues interdisciplinary solutions for the socio-technical complexity of AI is paramount. In this regard, we propose the following 10 recommendations for the robust advancement of AI R&D:

1) In the absence of significant mention of environmental sustainability and given the urgency of our current climate crisis, we recommend a new strategy aptly titled *Strategy 9: Ensure the Environmental Sustainability of AI Systems*.

- **Statement 1A:** Research has already shown that AI-enabled systems require exponentially rising compute power. This increase in compute power requires substantial energy consumption and, as a result, generates a large carbon footprint.<sup>1</sup> To begin to address this

---

<sup>1</sup> AI and Compute. OpenAI, 16 May 2018, <https://openai.com/blog/ai-and-compute/>.

concern, researchers have developed a new automated AI system for training and running certain neural networks that cut down on carbon emissions.<sup>2</sup> We underscore the need to implement an additional strategy focusing exclusively on ensuring the environmental sustainability of AI-enabled systems. The addition of environmental sustainability as a core AI R&D strategy supports a more holistic well-being approach to human health, prosperity and overall welfare given that human well-being is dependent on ecological well-being.<sup>3</sup> Furthermore, standards on assessing the impact of AI-enabled technologies on human and ecological well-being have been put forward by IEEE<sup>4</sup> and merit wider adoption. This proposed strategy directly supports the recent mention from the OSTP on February 17, 2022 to include, among six policy divisions, the priority of advancing “critical Administration priorities around groundbreaking clean energy investments.”<sup>5</sup> Moreover, a focus of this kind would not only be in concert with UNESCO’s Recommendation on the Ethics of AI whereby environmental impacts must be continuously assessed alongside human, social, and economic implications,<sup>6</sup> but also with the United Nations’s 2030 Agenda for Sustainable Development.<sup>7</sup>

**Recommendation 1A:** This new strategy sets up the critical need to support AI R&D that (i) addresses the design, development, deployment, and management of more efficient computing systems, (ii) informs the standardization and implementation of carbon accounting metrics and their independent oversight thereof, and (iii) supports the readily available public access to carbon accounting practices. An immediate consequence of such initiative will be the demand for inclusion of other disciplines such as environmental science, ecology, and oceanography, among others.

- 2) To significantly advance trustworthy AI-enabled systems and support the goal to “protect civil liberties, privacy, and American values,”<sup>8</sup> we recommend an even greater consideration of the integration of ethical principles via technical means within *Strategy 3: Understand and Address the Ethical, Legal, and Societal Implications of AI*. Specifically, we highlight focal areas ripe for R&D investment that are needed to sustain the balance we must bear in mind of building optimized systems for the benefit and empowerment of users and building such systems that do not compromise users’ rights.

---

<sup>2</sup> Cai, H., Gan, C., Wang, T., Zhang, Z., & Han, S. (2019). Once-for-all: Train one network and specialize it for efficient deployment. *arXiv preprint arXiv:1908.09791*.

<sup>3</sup> Raworth, K. (2017). *Doughnut economics: seven ways to think like a 21st-century economist*. Chelsea Green Publishing.

<sup>4</sup> IEEE. (7010-2020). IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being, <https://standards.ieee.org/ieee/7010/7718/>.

<sup>5</sup> OSTP Blog. A New Chapter for the White House Office of Science and Technology Policy. 17 February 2022, <https://www.whitehouse.gov/ostp/news-updates/2022/02/17/a-new-chapter-for-the-white-house-office-of-science-and-technology-policy/>.

<sup>6</sup> UNESCO. Recommendation on the ethics of artificial intelligence, <https://en.unesco.org/artificial-intelligence/ethics>.

<sup>7</sup> United Nations. Department of Economic and Social Affairs - Sustainable Development, <https://sdgs.un.org/goals>.

<sup>8</sup> Select Committee on Artificial Intelligence of the National Science & Technology Council. (June 2019). The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update, <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>

- **Statement 2A:** As AI-enabled technologies increasingly reflect society’s biases and open-source bias audit toolkits grow to address such,<sup>9,10</sup> the need to explicitly state the methods of data acquisition and use to validate a specific values-driven approach and the standardization of such becomes essential.

**Recommendation 2A:** An emphasis on the building of trusted datasets that are more representative of the users of the AI-enabled systems, unbiased in their decision-making, and thus trustworthy in their efficiency and effectiveness is needed. One way is to engage more intentionally with multiple stakeholders from across different sectors; independent oversight is necessary to prevent repeated engagement with already known stakeholders. Data crowdsourcing is another way to broaden participation.<sup>11</sup> While a citizen science paradigm offers opportunities for scalability and expansion into different areas for the diversification of datasets, it too raises challenges regarding (i) privacy issues, (ii) computational resources which must factor in Recommendation 1A, and (iii) public awareness and (iv) accessibility that could in fact upend the very goal of diverse community engagement and non-discriminatory practices. This recommendation is also useful towards the building of public datasets—the focus of *Strategy 5: Develop Shared Public Datasets and Environments for AI Training and Testing*—because data crowdsourcing platforms are by definition datasets intentionally created by the public, invite questions of ownership, and face similar security challenges by adversaries.

- **Statement 2B:** The intersectional ethical-legal question of accountability and its subareas of informed consent to use, safety, and liability, among others, in the case, for example, of system failure and/or harm still necessitates significant development regarding clarity of definitions, design frameworks, and public communication strategies, to name a few.

**Recommendation 2B:** Interdisciplinary research is inevitable. There needs to be (i) clarification of issues regarding causality, compensation, and justice, as well as transparent determination of the role of responsibility of the human(s) involved. As these issues depend on system transparency and explainability, there needs to be (ii) determination of when explanations are required and what kinds of explanations are acceptable, along with a mechanism of adaptability to changing societal expectations as a result of technological advances.

- **Statement 2C:** The question of ethical decision-making in such high stakes contexts such as facial recognition, healthcare diagnostic tools, driverless vehicles, and lethal autonomous weapon systems necessitate substantial attention. Alongside ethical issues such as bias and

---

<sup>9</sup> Bellamy R. K., Dey K., Hind M., Hoffman S. C., Houde S., Kannan K., Lohia P., Martino J., Mehta S., Mojsilovic A., & Nagar S. (2018). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv preprint arXiv:1810.01943, <https://aif360.mybluemix.net/>.

<sup>10</sup> Saleiro P., Kuester B., Hinkson L., London J., Stevens A., Anisfeld A., Rodolfa K. T., & Ghani R. (2018). Aequitas: A bias and fairness audit toolkit. arXiv preprint arXiv:1811.05577, <http://www.datasciencepublicpolicy.org/our-work/tools-guides/aequitas/>.

<sup>11</sup> Mozilla Common Voice, <https://commonvoice.mozilla.org/en>.

discrimination as addressed in Recommendation 1A, there is also the issue of moral capability and the level of moral decision making we are willing as a society to assign to an AI-enabled system.

**Recommendation 2C:** A moral quandary such as choosing between life or death is a controversial issue, most clearly demonstrated with the Trolley Problem and one that depends on cultural differences.<sup>12</sup> As a result, there is an imperative to (i) determine whether a red line of moral capability should be established, and (ii) after determination of (i), to set system boundaries dependent on regional/cultural norms.

- 3) In agreement with the emphatic need to build safe and secure AI-enabled systems, and to underscore OSTP's recent announcement on February 17, 2022 to "build a society where everyone can live with equal dignity and hope and opportunity, as well as equal safety and security,"<sup>13</sup> we recommend fundamental system requirements necessary to uphold such a promise. These recommendations expand *Strategy 4: Ensure the Safety and Security of AI Systems* and *Strategy 6: Measure and Evaluate AI Technologies through Standards and Benchmarks*.

- **Statement 3A:** As multiple proposals circulate on the definition of AI, most notably from the European Union's AI Act,<sup>14</sup> the OECD,<sup>15</sup> and a recent call for commentary by the U.S. Chamber of Commerce's AI Commission,<sup>16</sup> there is continued need for agreed-upon definitions of AI and standards for AI-enabled systems' robustness. The establishment of an all-encompassing definition is, in part, due to the dynamic nature of technological advancement and global centuries-old debate on our human-machine relationship.<sup>17,18</sup> Moreover, the development of scientific innovation within modern AI and of governance structures operate on different timelines and with different goals.

**Recommendation 3A:** Consensus-based provision of precise definitions of technical terms such as AI, autonomy, transparency, explainability, and interpretability and consistent

---

<sup>12</sup> Awad, E., et al. (2018). The moral machine experiment. *Nature* 563(7729), 59-64.

<sup>13</sup> OSTP Blog. A New Chapter for the White House Office of Science and Technology Policy. 17 February 2022, <https://www.whitehouse.gov/ostp/news-updates/2022/02/17/a-new-chapter-for-the-white-house-office-of-science-and-technology-policy/>.

<sup>14</sup> Artificial Intelligence Act. (21 April 2021). Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. *EUR-Lex - 52021PC0206*, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELLAR:e0649735-a372-11eb-9585-01aa75ed71a1>.

<sup>15</sup> OECD.AI Policy Observatory, <https://oecd.ai/en/ai-principles>.

<sup>16</sup> U.S. Chamber of Commerce. U.S. Chamber Launches Bipartisan Commission on Artificial Intelligence to Advance U.S. Leadership. 18 January 2022, <https://www.uschamber.com/technology/u-s-chamber-launches-bipartisan-commission-on-artificial-intelligence-to-advance-u-s-leadership>

<sup>17</sup> Al-Jazari IR. (1974). *The Book of Knowledge of Ingenious Mechanical Devices: Kitāb fi ma 'rifat al-ḥiyal al-handasiyya* (translated from Arabic and annotated by Donald R. Hill). Springer, Dordrecht.

<sup>18</sup> Muri, A. (2007). *The Enlightenment Cyborg: A history of Communications and control in the human machine, 1660-1830*. University of Toronto Press.

terminology included within the domain of AI safety and security must be achieved. It is critical to note that different cultures and communities have different values towards automation and its role within human identity and culture.<sup>19</sup> As such, it is paramount to include multiple stakeholders when defining AI and related terms. Following such provision is the imperative for system updating and significant R&D in this area. Such updating must include (i) new data resulting from a data-centric strategy for AI model building to support system integrity, and (ii) new core AI functionalities resulting from rapid AI advances to maintain system resilience to change and/or adversarial attack. Both of these requirements are dynamic and necessitate an established review structure to manage in real time.

- **Statement 3B:** As proposals for ethical principles and standards circulate, claims of trustworthiness and their compliance thereof must be verifiable and certifiable. Increasing trust can be achieved through standardization, independent oversight, and overt certification.

**Recommendation 3B:** The following steps invite a host of much needed AI R&D foci. A first step is to determine metrics of standardization around transparency, explainability, and interpretability. This is paramount to protecting public and individual privacy requirements as well as overall human well-being. A second step after standardization is determined, is to implement mechanisms of independent audit.<sup>20</sup> The audit process—for which its methodologies also require standardization—and resulting evaluations must confirm the system performs as intended to be certified. A third step is to improve users' system confidence. This can be achieved by ensuring that all AI-enabled technologies contain a user-friendly label of not only privacy considerations regarding an organization's collection, use, and sharing of personal information,<sup>21</sup> but also a label of ethical certification. The issue of comprehensibility, and thus the consensus-based need to establish standards, becomes paramount because users vary in their AI competency and such must be taken into account.

- **Statement 3C:** AI-enabled technology is neither confined to a single use case nor to a single market. Given that everyone will be affected by AI at some point, if not already, international cooperation is fundamental for system interoperability.

**Recommendation 3C:** AI standards should be produced and harmonized at the international level to ensure common ground around definitions, security, safety, and system resilience. This determination should be made by different groups with a variety of expertise<sup>22</sup> and, in effect,

---

<sup>19</sup> Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 1-12. DOI: 10.1177/2053951717738104.

<sup>20</sup> ForHumanity. Independent Audit of AI Systems (IAAIS), <https://forhumanity.center/blog/auditing-ai-and-autonomous-systems-building-an-infrastructureoftrust/>.

<sup>21</sup> Kelley, P. G., Bresee, J., Cranor, L. F., & Reeder, R. W. (2009, July). A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (pp. 1-12).

<sup>22</sup> CEN-CENELEC response to the EC white Paper on AI, Version 2020-06, [https://www.cencenelec.eu/media/CEN-CENELEC/Areas%20of%20Work/CEN%20sectors/Digital%20Society/Emerging%20technologies/cen-clc\\_ai\\_fg\\_white-paper-response\\_final-version\\_june-2020.pdf](https://www.cencenelec.eu/media/CEN-CENELEC/Areas%20of%20Work/CEN%20sectors/Digital%20Society/Emerging%20technologies/cen-clc_ai_fg_white-paper-response_final-version_june-2020.pdf).

underscores the need to significantly expand the multilateral dialogue mentioned as an activity of *Strategy 6: Measure and Evaluate AI Technologies through Standards and Benchmarks*.

- 4) To significantly advance efforts to educate and include interdisciplinary and diverse voices across the entire AI space and thus raise the competitive stance of the U.S., we recommend further improvements to education and workforce training activities and proposals. These recommendations expand *Strategy 7: Better Understand the National AI R&D Workforce Needs* and also address the OSTP's recent mention on February 17, 2022 of sustaining "a national strategy for STEM equity."<sup>23</sup>

- **Statement 4A:** Computing is no longer an exclusively technical field. It is a multidisciplinary field with social-technical consequences. To readily identify more efficient and effective interdisciplinary solutions, interdisciplinary thinking is a requisite. R&D teams composed of single subject matter experts brought together to collaborate as a multidisciplinary group are insufficient because today's and tomorrow's challenges demand interdisciplinary competences and skills.

**Recommendation 4A:** R&D teams, whether in academia, industry, or government, must be composed of individuals who have both deep subject matter expertise *and* interdisciplinary knowledge of the research agenda at hand.

- **Statement 4B:** In order to build a workforce whereby every individual has an interdisciplinary background and to successfully follow through with Recommendation 4A, education must not only expose learners to different disciplines' knowledge, but connect and integrate the disciplines within the learning process. As such, increasing education programs, fellowships, and activities focused exclusively to STEM fields and their subareas is insufficient because their questions, methods, and goals are insufficient for the complex nature of AI.

**Recommendation 4B:** Learning across K-12, precollege, undergraduate, graduate, and postgraduate education necessitates reforming and resources to sustain. A first step is the acknowledgment of the limits of silos and the impact that has on the nurturing of interdisciplinary minds.<sup>24</sup> A second step is the creation of learning environments that reach across disciplinary boundaries; such can be achieved by implementing curricula that balance in-depth subject matter with expansive related subject matter coverage, as well as traditional assessments with creative and experiential practice-based learning experiences.<sup>25</sup> A third step is expanding academic-industry partnerships and closely tied to the focus of *Strategy 8: Expand Public-Private Partnerships to Accelerate Advances in AI*. Moving past differing

---

<sup>23</sup> OSTP Blog. A New Chapter for the White House Office of Science and Technology Policy. 17 February 2022, <https://www.whitehouse.gov/ostp/news-updates/2022/02/17/a-new-chapter-for-the-white-house-office-of-science-and-technology-policy/>.

<sup>24</sup> National Academies of Sciences, Engineering, and Medicine. The integration of the humanities and arts with sciences, engineering, and medicine in higher education: Branches from the same tree. National Academies Press, 2018.

<sup>25</sup> López-González, M. (2017). For female leaders of tomorrow: cultivate an interdisciplinary mindset. In *2017 IEEE Women in Engineering (WIE) Forum USA East* (pp. 1-6). IEEE.

expectations of achievement and cultural differences will not only accelerate advances in AI through collaboration and co-production,<sup>26</sup> but enhance learning experiences that feed back into both academia and industry in mutually beneficial ways. Of resounding importance and a critical addition to *Strategy 8* as well, these partnerships cannot be solely between STEM fields. Instead, both academia and industry must move away from disciplinary boundaries and also initiate collaborations with non-STEM industries and academic departments, respectively.

- **Statement 4C:** Representation of diverse voices is still significantly lacking within the domain of AI.<sup>27</sup> Diversity, equity, and inclusion (DEI) are fundamental to ensuring that AI-enabled systems are respectful of, safe, and reliable for all users.

**Recommendation 4C:** From the designers, developers, researchers, instructors, project managers and directors to the learners themselves, the inclusion of diverse voices must be mandatory across all educational and R&D initiatives. Specifically, inclusion entails the respectful consideration for the different experiences of all within the various areas AI cuts across including education, design, development, deployment, and management. This also has a direct effect on ensuring the goals of Recommendation 2A.

In conclusion, we recommend that the OSTP take these 10 recommendations into consideration for the subsequent update to the *National Artificial Intelligence Research and Development Strategic Plan*.

We thank you for your consideration, and we look forward to your action on our recommendations.

---

<sup>26</sup> Sannö, A., Öberg, A. E., Flores-Garcia, E., & Jackson, M. (2019). Increasing the impact of industry–academia collaboration through co-production. *Technology Innovation Management Review*, 9(4).

<sup>27</sup> West, S. M., Whittaker, M., & Crawford, K. (2019). Discriminating systems: Gender, race, and power in AI. *AI Now*.