

ARTIFICIAL INTELLIGENCE: ACCELERATING INCLUSIVE INNOVATION BY BUILDING TRUST

*Adopted by the IEEE-USA
Board of Directors (21 July 2020)*

The U.S. is a global leader in artificial intelligence and autonomous systems (AI/AS), which is in no small part due to continued investment by the U.S. Government in these technologies. This success has brought about broad adoption of AI/AS in both the public and the private sector, so much so that very few aspects of our lives, individually and collectively, remain untouched by these technologies.

We now stand at an important juncture that pertains less to what new levels of efficiency AI/AS can enable, and more to whether these technologies can become a force for good in ways that go beyond efficiency. We have a critical opportunity to use AI/AS to help make society more equitable, inclusive, and just; make government operations more transparent and accountable; and encourage public participation and increase the public's trust in government. When used according to these objectives, AI/AS can help reaffirm our democratic values.

If, instead, we miss the opportunity to use these technologies to further human values and ensure trustworthiness, and uphold the status quo, we risk reinforcing disparities in access to goods and services, discouraging public participation in civic life, and eroding the public's trust in government. Put another way: Responsible development and use of AI/AS to further human values and ensure trustworthiness is the only kind that can lead to a sustainable ecosystem of innovation. It is the only kind that our society will tolerate.

SUMMARY OF RECOMMENDATIONS

To establish principles and frameworks for creating and operating AI/AS that further human values and ensure trustworthiness, **IEEE-USA recommends:**

1. Promoting the eight principles for creating and operating AI/AS that further human values and ensure trustworthiness, as outlined in IEEE's Ethically Aligned Design.
2. Developing explicit risk-benefit analysis frameworks for the use of AI/AS by governmental entities.

To achieve inclusive innovation of AI/AS, **IEEE-USA recommends:**

3. Significantly increasing federal investment for research and development on the responsible and ethical development and use of AI/AS, and on assessing the societal impact of these technologies.
4. Building capacity within governments to develop, procure, and manage AI/AS in ways that increase efficiency, while carefully weighing the benefits and the risks of these technologies to a range of stakeholders.
5. Supporting AI/AS education and retraining opportunities to create a diverse AI/AS workforce.
6. Developing public outreach strategies and educational resources to inform members of the public about AI/AS.

To ensure the fairness, safety, security, and reliability of AI/AS, **IEEE-USA recommends:**

7. Developing, updating, and enforcing standards and certifications for AI/AS and their operators, and funding recurring benchmarking exercises and independent studies to ensure their effectiveness, competence, inclusiveness, accountability, and transparency in operation.
8. Prioritizing standards, certifications, benchmarking exercises, and independent studies for high-risk AI/AS, and requiring high-risk AI/AS be independently verified and validated prior to deployment, and to informing decisions in the legal system, law enforcement, governance, and related compliance.
9. Ensuring awareness, access, and research on the existence, fairness, safety, security, privacy, and ethical and societal impacts of AI/AS.

10. Eliminating barriers to parties' access to information needed to ascertain relevant evidence about and from AI/AS in legal disputes.
11. Promoting accountability and transparency in government procurement and contracting for AI/AS.

RECOMMENDATIONS

To establish principles and frameworks for creating and operating AI/AS that further human values and ensure trustworthiness, **IEEE-USA recommends:**

1. **Promoting the eight principles for creating and operating AI/AS that further human values and ensure trustworthiness, as outlined in IEEE's Ethically Aligned Design:** (a) human rights: AI/AS shall be created and operated to respect, promote, and protect internationally recognized human rights; (b) well-being: AI/AS creators shall adopt increased human well-being as a primary success criterion for development; (c) data agency: AI/AS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity; (d) effectiveness: AI/AS creators and operators shall provide evidence of the effectiveness and fitness for the purpose of AI/AS; (e) transparency: the basis of a particular AI/AS decision should always be discoverable; (f) accountability: AI/AS shall be created and operated to provide an unambiguous rationale for all decisions made; (g) awareness of misuse: AI/AS creators shall guard against all potential misuses and risks of AI/AS in operation; and, (h) competence: AI/AS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.
2. **Developing explicit risk-benefit analysis frameworks for governing the use of AI/AS by governmental entities.** Governmental entities, and the U.S. Federal government in particular, should convene working groups to establish explicit risk-benefit analysis frameworks, with direct input and participation from members of the public, industry, and academia.

To achieve inclusive innovation of AI/AS, **IEEE-USA recommends:**

3. **Significantly increasing federal investment for research and development on the responsible and ethical development and use of AI/AS, and on assessing the societal impact of these technologies.** Significant research progress has been made on assessing the inequities that can arise due to the irresponsible use of AI/AS, but additional investment is required to develop responsible and ethical AI/AS frameworks and techniques, and to bring them to market. This includes investment in the creation and dissemination of open data and models, in innovative data

governance infrastructures, and in standards for data and model quality and fitness for purpose.

4. **Building capacity within governments to develop, procure, and manage AI/AS in ways that increase efficiency, while carefully weighing the benefits and the risks of these technologies to a range of stakeholders.** We recommend increasing technical expertise at all levels of government to develop appropriate critical thinking about AI/AS. This can be achieved through multiple means, including: (a) training existing federal employees on AI/AS technologies, and on the associated benefits and risks to stakeholders; (b) establishing new permanent offices and positions with specific focus on AI/AS expertise; (c) establishing and scaling up programs for permanent and temporary placement of technical experts within government (e.g., academic personnel under the Intergovernmental Personnel Act, Presidential Innovation Fellows, and the Science and Technology Policy Fellows program); (d) establishing government liaisons with private-sector developers of AI/AS technologies and means for government funding in support of private-sector R&D in such technologies; and, establishing and increasing funding for Offices that assist Congress' understanding of emerging technologies and their policy implications.
5. **Support AI/AS education and retraining opportunities to create a diverse AI/AS workforce.** When building the future AI/AS workforce, we recommend making an explicit investment in diversity and inclusion, and embedding social, legal, and ethical concerns into training. For example, scholarships and fellowships should focus on supporting underrepresented groups to enter the responsible AI/AS workforce.
6. **Developing public outreach strategies and educational resources to inform members of the public about AI/AS.** The government should invest in the development of public outreach strategies and educational resources to inform the public about AI/AS, to enable broad and inclusive private sector and public participation in the design and regulation of these technologies. The goal should be to move beyond the extremes of unbounded techno-optimism and techno-criticism, and into a nuanced and productive conversation about the role of technology in society.

To ensure the fairness, safety, security, and reliability of AI/AS, **IEEE-USA recommends:**

7. **Developing, updating, and enforcing standards and certifications for AI/AS and their operators, and funding recurring benchmarking exercises and independent studies to ensure their effectiveness, competence, inclusiveness, accountability, and transparency in operation.** These standards, certifications, exercises, and studies should

address: (a) the requirements for informed trust by the general public in AI/AS, and the development of metrics that are immediately and easily accessible by experts and non-experts alike; (b) the existence or absence of reliable and unbiased underlying scientific principles and methods in AI/AS systems; (c) the requirements for recurring testing and auditing of the operation of AI/AS systems, including the operators, field conditions, testing data, environments, methodologies, and performance metrics; (d) the requirements for publicly available documentation by developers and testers of AI/AS, and of the use of AI/AS in individual and aggregate cases and decisions; (e) the requirements for certification or loss of certification of operators and AI/AS, and for their validation for AI/AS already in use; (f) the requirements for individuals to be able to access, review, contest, and correct the data about them, to review and contest the decisions that affect them, and to request human review of such data and decisions; and (g) the requirements for operation in an ethical manner.

8. **Prioritizing standards, certifications, benchmarking exercises, and independent studies for high-risk AI/AS, and requiring high-risk AI/AS be independently verified and validated prior to deployment, or to informing decisions in the legal system, law enforcement, governance, and related compliance.** Before being deployed, high-risk AI/AS ought to be independently verified and validated (IV&V) in accordance with IEEE Standard 1012, IEEE Standard for System, Software, and Hardware Verification and Validation, and be subject to recurring post-deployment audit, including with respect to their operators. Furthermore, governmental entities should make the reports documenting the required IV&V and audits of their high-risk AI/AS public.
9. **Ensuring awareness, access, and research on the existence, fairness, safety, security, privacy, and ethical and societal impacts of AI/AS.** The government should clarify whether and how proprietary AI/AS may be reverse engineered, modified, and evaluated under federal law such as the Computer Fraud and Abuse Act and the anti-circumvention provision of the Digital Millennium Copyright Act. Governmental entities should: (a) publicly identify and disclose the AI/AS used by the government; (b) conduct and publicly disclose a methodological validation study that establishes the value of using the AI/AS in place of existing practices prior to deploying AI/AS; (c) adopt clear procedures relating to the collection, usage, storage, and sharing of personal information in the context of developing, using, and validating a given AI/AS in a privacy-preserving manner; and (d) prevent intellectual property, confidentiality claims, lack of funding, or lack of an designated independent body within government to monitor compliance from impeding duly limited independent validation and verification and publicly disclosed review of the fairness, safety, security, privacy, and ethical and societal impacts of AI/AS. AI/AS are to be submitted voluntarily, to the agency performing validation and verification thereof, and the agency using related

private intellectual property or proprietary data in its evaluation must adopt rules to protect such private rights from misappropriation. Constituents should be allowed to (a) request and receive an explanation of how a government determination using AI/AS was reached; (b) determine whether an AI/AS used in government decision-making disproportionately impacts a protected class; and (c) rectify, challenge, or complete inaccurate or incomplete personal data that is part of an AI/AS system or decision.

10. **Eliminating barriers to parties' access to information needed to ascertain relevant evidence about and from AI/AS in legal disputes.** Intellectual property protections should not be used as a shield to prevent duly limited disclosure of information needed to ascertain whether A/IS meet acceptable standards of effectiveness, fairness, and safety. Specifically, in legal disputes, tribunals should permit disclosure under appropriate protective orders of intellectual property related to AI/AS necessary to obtain evidence in compliance with other judicial requirements, including constitutional requirements, discovery laws, or subpoenas. Furthermore, laws, procedures, and public funding should not make it more difficult for non-government parties in legal disputes to develop, obtain expertise regarding, or gain access to evidence from AI/AS than for government parties to do so.
11. **Promoting accountability and transparency in government procurement and contracting for AI/AS.** The government should not procure AI/AS that (a) require the governmental entity to indemnify vendors for any and all negative outcomes; (b) do not adhere to the eight principles in IEEE's Ethically Aligned Design for creating and operating AI/AS that further human values and ensure trustworthiness (as may be reflected in articulated guidelines, standards, certifications, audits, and other sound documentation); (c) do not comply with federal, state, and local anti-discrimination laws; or, (d) are shielded from independent validation and verification, and public review.

This statement was developed by the IEEE-USA Artificial Intelligence & Autonomous Systems Policy Committee, and represents the considered judgment of a group of U.S. IEEE members with expertise in the subject field. IEEE-USA advances the public good, and promotes the careers and public policy interests of the nearly 180,000 engineering, computing and allied professionals who are U.S. members of the IEEE. The positions taken by IEEE-USA do not necessarily reflect the views of IEEE, or its other organizational units.