10 April 2024

U.S. Agency for International Development **(**USAID)

IPI/ITR/T, Rm. 2.12–213, RRB
1300 Pennsylvania Avenue NW
Washington, DC 20004

In re: *USAID "Global AI Research Agenda"*

IEEE is pleased to provide input to USAID's Request for Comments on the agency's proposed high-level structure for the Global AI Research Agenda.

IEEE is a globally recognized standards-setting body within IEEE, the largest organization of technology professionals in the world. We develop consensus standards through an open process that engages industry and brings together a broad stakeholder community and adhere to the WTO Principles for International Standardization.

IEEE Standards Association (SA), the standards development arm of the IEEE, has taken a keen interest in the ethical implications of technology, especially as AI and autonomous systems have grown in prominence. IEEE SA has developed a robust portfolio of resources and standards globally recognized in applied ethics and systems engineering and offers standards, training and education, certification programs, and more, to empower stakeholders designing, developing, and using Autonomous Intelligent Systems (AIS). IEEE, through its global community, continues to develop accessible and sustainable approaches and solutions for pragmatic application of Artificial Intelligence Systems (AIS) principles and frameworks.

Emerging and new technology comes with unknown obstacles and unintended risks requiring accountable design and lifecycle planning to ensure ethical and responsible innovation. Forecasting the impact that AI may have should not be understated when considering the importance of advances in AI research and development.

IEEE SA released a Communique, 9 June 2023 on Generative Artificial Intelligence Applications, which notes that the deployment of large language models and other generative artificial intelligence (AI) applications has precipitated a worldwide conversation about the benefits and possible harms of AI systems, as well as the need for appropriate standards and measures that should be taken to ensure scientific integrity and the safety of the public.

In addition, IEEE SA would like to provide the following in response to the questions raised by USAID regarding the Global AI Research Agenda:

Research Best Practices

IEEE-USA | 2001 L Street, N.W., Suite 700, Washington, D.C. 20036-4928 USA
Office: +1 202 785 0017 | Fax: +1 202 785 0835 | E-mail: ieeeusa@ieee.org | Web: https://ieeeusa.org

Since 2016, IEEE, through the IEEE SA, has been developing a wide range of standards and methods to address safety, biases, transparency, privacy, and corporate governance. These tools could be used also to address many AI governance and implementation issues.

To address AI decision-making bias, research best practices should be created to integrate formal commitment towards AI Ethics Principles, especially how, in practice, they anticipate AI impacts and which mechanisms are implemented to ensure accountability and transparency.

IEEE has published a report on AI Ethically Aligned Design (EAD), which explains the cultural and behavioral change needed within organizations, and what it takes to enable AI ethics as a core organizational competency. This report provides an 'AI readiness framework' with steps to help facilitate a change in the organizational culture, as a proactive measure to build AI ethics into the governance structure of the organization. This could be a useful guide to support and build responsible research best practices.

- **EAD, First Edition**, sets forth scientific analysis and resources, high-level principles, and actionable recommendations. It offers specific guidance for standards, certification, regulation or legislation for design, manufacture, and use of A/IS that provably aligns with and improves holistic societal well-being.

- **Defining A/IS Ethics – Glossary**, provides interdisciplinary teams a shared resource for reference to terms which may have meanings that are discipline specific. Within this document there are six definitions given for most terms. In those cases where six full definitions are not given for each term a suitable definition within the discipline examined could not be found. The six categories of disciplines from which definition were drawn include: ordinary language; computational disciplines (e.g., mathematics and statistics); economics and social sciences; engineering disciplines; philosophy and ethics; and international law and policy.

Review of IEEE Standard 7001™ will provide guidance on where and when appropriate levels of transparency in AI should be considered and potentially mandated. Use of a 'System Transparency Assessment' or STA, a method for assessing the transparency of an existing autonomous system, or use of a 'System Transparency Specification' or STS for specifying the transparency requirements of a system prior to its development or procurement may be considered, and when and where these are appropriate to use.

Ethical guidelines and standards such as **IEEE P7003™ Algorithmic Bias Considerations** can serve as a roadmap for fair AI development, emphasizing principles like transparency, fairness, and accountability.

IEEE offers recognized AIS standards that are developed to ensure consistency, transparency, and accountability in AI system development as well as the mechanism to update standards in line with technological advancements. The full list of applicable standards can be found at: https://standards.ieee.org/initiatives/autonomous-intelligence-systems/

The value of adhering to a standards-based approach in developing and deploying AI is that product development and compliance to standards can help mitigate potential risks as well as to enable

interoperability between products and within systems as applications are implemented and deployed.

International Engagement

IEEE collaborates with various international organizations such as ISO, IEC, and OECD and the UN, and research institutions to conduct research projects on safe and trustworthy AI. These projects bring together researchers from different countries and disciplines to work on common research questions and share knowledge and expertise in public and private sectors.

In addition, IEEE has organized various conferences and workshops on safe, secure, and trustworthy AI, bringing together researchers, industry experts, and policymakers from around the world. Such as the annual IEEE Conference on Artificial Intelligence (IEEE CAI).

The most challenging aspect of international engagement are the differences in cultural and ethical perspectives and practices.

Foundation Models

Developers of foundation models may focus on research related to model architecture, training, and evaluation, while downstream users may focus on research related to specific applications, use cases, and contexts, and as such, may need to conduct more comprehensive risk assessments and implement more robust safety and security measures to address potential risks associated with the models. Downstream users may need to conduct more targeted risk assessments and implement context-specific safety and security measures to address risks associated with specific use cases. They may also need to provide more detailed documentation and transparency related to model architecture, training data, and evaluation metrics. Downstream users may need to provide more detailed documentation and transparency related to specific use cases, data inputs, and output interpretations.

The reliability, accuracy, diversity, and applicability of datasets also merit care, as the models that learn from them are dependent on good data to function correctly. Changes to datasets may be reflected in downstream differences in how models derived from them perceive certain stimuli. This is a challenge in a environment where AI models are not just trained once, they are continually updating, or even learning on user-targeted information. This enables various forms of exploits, such as worms which can cause one AI system to hijack others. These quickly emerging issues present enormous security risks, even without willful intention on the part of humans. Unit tests on a range of standard inputs should be conducted carefully before deployment to ensure that typical model responses are not meaningfully altered.

IEEE P3426™ Standard for Defining and Measuring the Capabilities of AI Foundation Models defines and provides criteria to measure the capabilities of foundation models. The standard focuses on measurable and objective capabilities such as perception-oriented capabilities, cognitive capabilities, and learning capabilities. The standard, currently in development, provides guidelines for evaluating these capabilities, drawing inspiration from human intelligence quotient (IQ) definition and measurement, and includes:

- Definition of key capabilities pertinent to foundation models.
- Standardized methods and metrics for evaluating perception-oriented, cognitive, and learning capabilities.
- Framework for comparing different models based on their cognitive capabilities.
- Ethical considerations and potential biases in the evaluation process.

IEEE P7014.1™ Recommended Practice for Ethical Considerations of Emulated Empathy in Partner-based General-Purpose Artificial Intelligence Systems is a recommended practice in development that defines ethical considerations and good practices regarding the use of emulated empathy in general-purpose artificial intelligence systems for human-AI partnerships. This includes general-purpose artificial intelligence products marketed as empathic partners, personal AI, co-pilots, assistants, and related phrasing for human-AI partnering.

Human Impacts

Benefits of AI systems may include the automation of routine tasks, improvement in decision-making, and enhancement of human lifestyles and longevity. Risks are many and some are broadly well-known, such as possible infringement on human safety and security (e.g., deep fakes, hostile adversarial machine behavior, etc.). IEEE suggests consideration of other dimensions of engaging with AI, such as, but not limited to, impacts involving human job displacement and labor market changes (different skills needed) and bias, discrimination, and privacy issues associated with human activity, and notes that care should be taken with regard to unfair algorithmic management processes.

Considerations for safe and ethical research into human impacts of AI systems can vary significantly in different global contexts and IEEE often considers these issues through conferences, working groups, and other organization-sanctioned activities. It is important to take a culturally sensitive and contextually appropriate approach to research, and to engage with local stakeholders to understand their perspectives, priorities, and practices.

An AI system, in its entire life cycle, brings potential technical and organizational risks affecting it that must be identified, documented, and accessible.

A growing risk issue is one of algorithmic management in the workplace. Such systems may make unfair misattributions of employee's intentions and character, ones which may be difficult to challenge. Consideration should be afforded as to what degree such systems can operate, especially for remote workers who may labor in domestic environments. Transparency should also be provided as to any decisions or predictions made by such systems, including the underlying predicates.

IEEE's Trustworthy AI Portfolio has been developed and is available for public access to help organizations differentiate and build competencies and methodologies to achieve the outcome of a 'trustworthy organization.'

A 'risk-based approach' should be used, in conjunction with an 'impact-based approach', to assess potential and actual impacts on individuals, communities and society in accordance with qualitative and quantitative wellness criteria, in alignment with IEEE 7010 – 2020™, Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being. Impacts of artificial intelligence or autonomous and intelligent systems (A/IS) on humans are measured by this standard. The positive outcome of A/IS on human well-being is the overall intent of this standard. Scientifically valid well-being indices currently in use and based on a stakeholder engagement process ground this standard. Product development guidance, identification of areas for improvement, risk management, performance assessment, and the

identification of intended and unintended users, uses and impacts on human well-being of A/IS are the intents of this standard.

Enabling Infrastructure

IEEE posits that the adoption of strategies for access to computing resources, data and other prerequisites for AI research that include diversity, equity, and inclusion frameworks that help ensure fair, safe, and secure access to appropriate resources would be beneficial.

Global Equity Considerations

Members of emerging economies might not have the same societal makeup, skillset and infrastructure or financial access as more developed economies. Legislation and regulation regarding AI and data use also differ by jurisdiction. Such differences should be considered for any AI research such as bias is removed from any data collected and used. In addition, laws defining a citizen's rights, data protections, privacy, authorized access to data, and enforcement regulations should be considered.

If you have any questions, please do not hesitate to contact Erica Wissolik at e.wissolik@ieee.org or Karen Mulberry at k.mulberry@ieee.org.

Addendum
List of IEEE Artificial Intelligence (AI) Standards and Related Activity

The IEEE Standards Association (SA) portfolio of Artificial Intelligence (AI) and Autonomous and Intelligent Systems (AIS) standards, includes the following:

For a complete listing of IEEE AI-related standards please see:
https://standards.ieee.org/initiatives/autonomous-intelligence-systems/

IEEE 7000™, Standard Model Process for Addressing Ethical Concerns during System Design incorporates a set of processes by which organizations can include consideration of ethical values throughout the stages of concept exploration and development Processes incorporated in the standard provide for traceability of ethical values in the concept of operations, ethical requirements, and ethical risk-based design are described in the standard.

IEEE 7001-2021™ Standard for Transparency of Autonomous Systems establishes measurable, testable levels of transparency, so that autonomous systems can be objectively assessed, and levels of compliance determined.

IEEE 7002-2022™ Standard for Data Privacy Process contains requirements for a systems/software engineering process for privacy-oriented considerations regarding products, services, and systems utilizing employee, customer, or other external user's personal data.

IEEE P7003™ Algorithmic Bias Considerations describes specific methodologies to help users certify how they worked to address and eliminate issues of negative bias in the creation of their algorithms, where "negative bias" infers the usage of overly subjective or uniformed data sets or information known to be inconsistent with legislation concerning certain protected characteristics (such as race, gender, sexuality, etc); or with instances of bias against groups not necessarily protected explicitly by legislation, but otherwise diminishing stakeholder or user well-being and for which there are good reasons to be considered inappropriate.

IEEE P7004™ Standard for Child and Student Data Governance provides stakeholders with certifiable and responsible child and student data governance methodologies.

IEEE P7004.1™ Recommended Practices for Virtual Classroom Security, Privacy and Data Governance provides best practices for meeting the requirements of IEEE P7004: Standard for Child and Student Data Governance when designing, provisioning, configuring, operating, and maintaining an online virtual classroom experience for synchronous online learning, education,

IEEE 7005-2021™ Standard for Transparent Employer Data Governance contains specific methodologies to help employers in accessing, collecting, storing, utilizing, sharing, and destroying employee data, including specific metrics and conformance criteria regarding the types of uses from trusted global partners and how third parties and employers can meet them.

IEEE 7007-2021™ Ontological Standard for Ethically Driven Robotics and Automation Systems

contains a set of ontologies with different abstraction levels that contain concepts, definitions, axioms, and use cases that assist in the development of ethically driven methodologies for the design of robots and automation systems.

IEEE P7008™ Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems establishes a delineation of typical nudges (currently in use or that could be created). It contains concepts, functions and benefits necessary to establish and ensure ethically driven methodologies for the design of the robotic, intelligent and autonomous systems that incorporate them.

IEEE P7009™ Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems establishes a practical, technical baseline of specific methodologies and tools for the development, implementation, and use of effective fail-safe mechanisms in autonomous and semi-autonomous systems. The standard includes (but is not limited to): clear procedures for measuring, testing, and certifying a system's ability to fail safely on a scale from weak to strong, and instructions for improvement in the case of unsatisfactory performance.

IEEE 7010-2020™ Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being provides specific and contextual well-being metrics that facilitate the use of a Well-Being Impact Assessment (WIA) process in order to proactively increase and help safeguard human well-being throughout the lifecycle of autonomous and intelligent systems (A/IS).

IEEE P7010.1™ Recommended Practice for Environmental Social Governance (ESG) and Social Development Goal (SDG) Action Implementation and Advancing Corporate Social Responsibility provides recommendations for next steps in the application of IEEE Std 7010, applied to meeting Environmental Social Governance (ESG) and Social Development Goal (SDG) initiatives and targets. It provides action steps and map elements to review and address when applying IEEE 7010(™). This recommended practice serves to enhance the quality of the published standard by validating the design outcomes with expanded use. It provides recommendations for multiple users to align processes, collect data, develop policies and practices and measure activities against the impact on corporate goals and resulting stakeholders.

IEEE P7011™ Standard for the Process of Identifying and Rating the Trustworthiness of News Sources provides semi-autonomous processes using standards to create and maintain news purveyor ratings for purposes of public awareness. It standardizes processes to identify and rate the factual accuracy of news stories in order to produce a rating of online news purveyors and the online portion of multimedia news purveyors.

IEEE P7012™ Standard for Machine Readable Personal Privacy Terms identifies/addresses the manner in which personal privacy terms are proffered and how they can be read and agreed to by machines.

IEEE P7014™ Standard for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems defines a model for ethical considerations and practices in the design, creation and use of empathic technology, incorporating systems that have the capacity to identify, quantify, respond to, or simulate affective states, such as emotions and cognitive states. This includes coverage of 'affective computing', 'emotion Artificial Intelligence' and related fields.

IEEE P7015™ Standard for Data and Artificial Intelligence (AI) Literacy, Skills, and Readiness establishes an operational framework and associated capabilities for designing policy interventions, tracking their progress, and empirically evaluating their outcomes. The standard includes a common set of definitions, language, and understanding of data and AI literacy, skills, and readiness.

IEEE P2840™ Standard for Responsible AI Licensing describes specifications for the factors to beconsidered in the development of a Responsible Artificial Intelligence (AI) license. Possible elements in the specification include (but are not limited to): (1) What a 'Responsible AI License' means and what its aims are (2) Standardized definitions for referring to components, features and other such elements of AI software, source code and services (3) Standardized reference to geography specific AI/Technology specific legislation and laws (such as the EU General Data Protection Regulation - GDPR) as well as identification of violation detection, penalties, and legal remedies. (4) The specification lists domain specific considerations that may be applied in developing a responsible AI license.

IEEE P2863™ Recommended Practice for Organizational Governance of Artificial Intelligence specifies governance criteria such as safety, transparency, accountability, responsibility and minimizing bias, and process steps for effective implementation, performance auditing, training and compliance in the development or use of artificial intelligence within organizations.

IEEE P3119™ Standard for the Procurement of Artificial Intelligence and Automated Decision Systems establishes a uniform set of definitions and a process model for the procurement of Artificial Intelligence (AI) and Automated Decision Systems (ADS) by which government entities can address socio-technical and responsible innovation considerations to serve the public interest.

IEEE 2089-2021™ Standard for an Age Appropriate Digital Services Framework (Based on the 5Rights Principles for Children) sets out processes through the life cycle of development, delivery and distribution, that will help organizations ask the right relevant questions of their services, identify risks and opportunities by which to make their services age appropriate and take steps to mitigate risk and embed beneficial systems that support increased age appropriate engagement.

IEEE P2890™ Recommended Practice for Provenance of Indigenous Peoples' Data outlines the core parameters for providing and digitally embedding provenance information for Indigenous Peoples' data. The recommended practice establishes common descriptors and controlled vocabulary for provenance, including recommendations for metadata fields that can be used across industry sectors, including machine learning (ML) and artificial intelligence (AI) contexts, biodiversity and genomic science innovation and other associated databases, and supports proper and appropriate disclosure of originating data information.

Certification

IEEE has an AIcertification program, called IEEE **CertifAIEd(™)**, which offers a risk-based framework supported by a suite of AI ethical criteria that can be contextualized to fit organizations' needs– helping them to deliver a more trustworthy experience for their users. IEEE CertifAIEd Ontological Specifications for Ethical Privacy, Algorithmic Bias, Transparency, and Accountability are an introduction to our AI Ethics criteria.

Reports and Resources

IEEE has several reports related to ethically aligned design (EAD). These include:

- **EAD For Artists,** which identifies and addresses several significant ethical, social, political and economic challenges presented by AI for the creating and technical communities and policymakers and standards-setting organizations.

- **EAD For Business**, which explores the question: what are the metrics of success for Responsible AI? The paper provides direction for business readers so they can utilize these metrics—large enterprises as well as small- and medium-sized businesses (SMBs)—while also informing policy makers of the issues these metrics will create for citizens as well as buyers.

- **EAD, First Edition**, which sets forth scientific analysis and resources, high-level principles, and actionable recommendations. It offers specific guidance for standards, certification, regulation or legislation for design, manufacture, and use of A/IS that provably aligns with and improves holistic societal well-being.

- **Defining A/IS Ethics – Glossary**, which provides interdisciplinary teams a shared resource for reference to terms which may have meanings that are discipline specific. Within this document there are six definitions given for most terms. In those cases where six full definitions are not given for each term a suitable definition within the discipline examined could not be found. The six categories of disciplines from which definition were drawn include: ordinary language; computational disciplines (e.g., mathematics and statistics); economics and social sciences; engineering disciplines; philosophy and ethics; and international law and policy.

AI-Related Programs and Initiatives

The IEEE SA has a portfolio of programs and initiatives in the pre-standardization space, where individuals and groups explore various topics and outline standards roadmaps and other outputs to help inform the standardization ecosystem. In the AI space, programs include:

IC21-007 Ethical Assurance of Data-Driven Technologies for Mental Healthcare

Exploring how ethical assurance—supported by ethically sound and rigorous standards—can promote trust in the use of digital technologies for mental healthcare and support transparent communication between stakeholders.

IC21-006 Research Group on Issues of Autonomy and AI for Defense Systems

Creating a knowledge base of best practices regarding the development, use, and governance of autonomous weapons systems (AWS), initially based on the guidelines first established by AWS-related principles and later updated to reflect evolving technological advancements and international norms and regulations.

IC16-002 The Global Initiative on Ethics of Autonomous and Intelligent Systems

Bringing together experts in fields related to autonomous systems (e.g., Robotics, Artificial Intelligence, Computational Intelligence, Machine Learning, Deep Learning, Cognitive Computing, Affective

Computing) to identify and address the ethical considerations related to the design of autonomous systems and the issues they involved.