IEEE Standards Association Statement of Intention
Our Role in Addressing Ethical Considerations of Autonomous and Intelligent Systems (A/IS)

Autonomous and Intelligent Systems (A/IS) are rapidly spreading beyond transportation, into health and social care, enterprise productivity and advanced cyber-defense, to name a few. They hold great promise to benefit society, but they also bring forth social, legal and ethical challenges, as well as the related issues of major systemic risk, diminishing trust, privacy challenges and issues of data transparency, ownership and agency. There is a need for awareness and use of consensus-based global best practices, recommendations and standards for ethically aligned design that recognize and align end-users' and citizen's values when building and deploying A/IS.

A/IS - once activated - are capable of operation without human intervention and they can do so because they have the capacity for adaptation and learning. In short, they adapt, learn and can make decisions. A/IS are expected to play a central role in technology development and use, and this includes the need for standardization of these technologies. Standards are vital to the modern world, they are no less a part of the infrastructure of human civilization in the 21st century than roads or airports. In the present case, standards would formalize ethical principles into actionable elements of a system, and these standards could also be used to establish or evaluate levels of compliance with those principles. In addition to functional rules and interoperability, ethics would therefore also underpin such standards, potentially creating a market differentiator. Standards developed in a voluntary, bottom up standardization system can be important tools to promote public-policy objectives. Standards can reduce the need for regulation through public policies that encourage use of systems that are certified as compliant with certain types of standards, or parts of more complex standards can be enablers to demonstrate or clarify potential means for compliance with regulatory requirements. This has been used successfully in the past in order to promote use of products that comply with environmental criteria (for instance, IEEE 1680 series of standards).[1].

As a global, consensus building standards development body, the IEEE Standards Association (IEEE-SA) is committed to increasing trust in the specific sector of A/IS and their underlying and related technologies. This will happen through appropriate standards and other consensus-built products that contribute to transparency, education at all levels of expertise, technical community building and partnerships across regions and nations, thus serving humanity. We offer in particular a process that provides an opportunity for all stakeholders and other interested persons to participate in open and transparent dialogue and consensus building on the key issues and to work together developing tangible and actionable resources and programs.

There are several frameworks and activities around intelligent, autonomous technical systems and their applications in IEEE, including the IEEE Global Initiative on the Ethics of Autonomous and Intelligent Systems (Global Initiative). The Global Initiative is an Industry Connections program of IEEE-SA, formally established in April 2016 and supervised by the IEEE-SA Standards Board (SASB). In its short life, it has produced a very significant body of work, including the *Ethically Aligned Design (EAD): A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* and the creation of a new series of IEEE standardization projects. The extraordinary quantity and quality of the produced material and

[1] Winfield, A., Written evidence submitted to the UK Parliamentary Select Committee on Science and Technology Inquiry on Robotics and Artificial Intelligence, Discussion Paper, Science and Technology Committee (Commons), Website, 2016. http://eprints.uwe.ac.uk/29428/

ideas position IEEE as a pioneer in the field, and connect it with a host of very influential networks and actors. IEEE-SA supports both morally and materially the Global Initiative in its mission to ensure that every stakeholder involved in the design and development of A/IS is educated, trained and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity.

Through this statement, IEEE-SA declares its intention to make use of material produced by the Global Initiative as well as of its vast network of experts, in order to engage in this field with other actors and entities, world-wide, in particular with the private and public sector, including through delivering input and feedback on demand.

The target groups are policy makers, regulators, technical standards developers, pertinent international and inter-governmental organizations and technology developers and users around the globe. The engagements would be aimed at:

- Addressing the fundamental ethical and societal issues and implications of A/IS in regulations, policies, standards and treaties. This would contribute in IEEE taking a leadership role in addressing ethically aligned design in A/IS technology development, standardization and use with the aim of building confidence and trust in A/IS. By taking into account human rights, human and social well-being, accountability and transparency in A/IS technology development, standardization and use, the societal and economic benefits of A/IS would be realized more readily.
- Establishing and implementing practices and instruments that will allow innovative and impactful A/IS, while ensuring that these technologies are developed and used responsibly and with accountability.
- Being proactive in proposing adequate policies for the successful and safe development and implementation of A/IS.
- Identifying opportunities to bring the IEEE Global Initiative's body of work -- *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* and the associated IEEE 7000™ series of standards -- into practice.
- Collaborating with global partners to set precedents through standards and best practices.

This statement was developed by the IEEE-SA and represents the considered judgment of a group of IEEE-SA members with expertise in the subject field. IEEE-SA, a globally recognized standards-setting body within IEEE, develops consensus standards through an open process that engages industry and brings together a broad stakeholder community. IEEE standards set specifications and best practices based on current scientific and technological knowledge. The IEEE-SA has a portfolio of over 1,200 active standards and over 650 standards under development. The positions taken by IEEE-SA do not necessarily reflect the views of IEEE, or its other organizational units.

Background

IEEE's collective experience, expertise, and programs that are (per IEEE's tagline) "Advancing Technology for Humanity" are outlined below. IEEE is not only large in numbers and influence but is also capable of understanding key societal and ethical challenges with A/IS technologies and transforming this understanding – through IEEE SA – into actionable, consensus-based deliverables, such as guidelines, policy recommendations and standardization projects

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (A/IS) was launched in April of 2016 to move beyond the paranoia and the uncritical admiration regarding autonomous and intelligent technologies and to illustrate that aligning technology development and use with ethical values will help advance innovation while diminishing fear in the process.

The goal of The IEEE Global Initiative is to incorporate ethical aspects of human well-being that may not automatically be considered in the current design and manufacture of A/IS technologies and to reframe the notion of success so human progress can include the intentional prioritization of individual, community and societal ethical values.

The Mission of The IEEE Global Initiative is to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity.

The IEEE Global Initiative has two primary outputs – the creation and iteration of a body of work known as *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*[2]; and the identification and recommendation of ideas for Standards Projects focused on prioritizing ethical considerations in A/IS.

Version 1 of *Ethically Aligned Design* (EADv1) was released in December of 2016 as a Creative Commons document so any organization could utilize it as an immediate and pragmatic resource. Launched as a Request for Input (RFI) to solicit response from the public in a globally consensus-building manner, the document received over two hundred pages of feedback at the time of the RFI's deadline.

This version of EAD was created by over 100 Global AI/Ethics experts, in a bottom up, globally open and transparent process, featuring eight sections focused on key areas like Law, Personal Data, Autonomous Weapons and Methodologies for Ethical Design. It contains over eighty key Issues and Candidate Recommendations and is designed as the "go-to" resource to help technologists and policy makers prioritize ethical considerations in AI/AS.

- An overview of the document can be viewed here.[3]
- Information about the Initiative's Committees and Members can be viewed here.[4]

Ethically Aligned Design, Version 2 (EADv2) features five new sections in addition to updated iterations of the original eight sections of EADv1. The IEEE Global Initiative has now increased from 100 AI/Ethics experts to more than 250 individuals including new members from China, Japan, South Korea, India and Brazil, and EADv2 now contains over 120 key Issues and Candidate Recommendations.

- Download *Ethically Aligned Design version 2* at this link.[5]

Along with creating and evolving *Ethically Aligned Design*, members of The IEEE Global Initiative are encouraged to recommend Standards Projects to IEEE based on their work. Below are titles and

[2] http://standards.ieee.org/news/2016/ethically_aligned_design.html
[3] http://standards.ieee.org/develop/indconn/ec/ead_brochure.pdf
[4] http://standards.ieee.org/develop/indconn/ec/ec_bios.pdf
[5] http://standards.ieee.org/develop/indconn/ec/auto_sys_form.html

descriptions for each of these approved IEEE Standards Projects. Information on each is available via the links included:

IEEE P7000™ - Model Process for Addressing Ethical Concerns During System Design[6] outlines an approach for identifying and analyzing potential ethical issues in a system or software program from the onset of the effort. The values-based system design methods addresses ethical considerations at each stage of development to help avoid negative unintended consequences while increasing innovation.

IEEE P7001™ - Transparency of Autonomous Systems[7] provides a Standard for developing autonomous technologies that can assess their own actions and help users understand why a technology makes certain decisions in different situations. The project also offers ways to provide transparency and accountability for a system to help guide and improve it, such as incorporating an event data recorder in a self-driving car or accessing data from a device's sensors.

IEEE P7002™ - Data Privacy Process[8] specifies how to manage privacy issues for systems or software that collect personal data. It will do so by defining requirements that cover corporate data collection policies and quality assurance. It also includes a use case and data model for organizations developing applications involving personal information. The standard will help designers by providing ways to identify and measure privacy controls in their systems utilizing privacy impact assessments.

IEEE P7003™ - Algorithmic Bias Considerations[9] provides developers of algorithms for autonomous or intelligent systems with protocols to avoid negative bias in their code. Bias could include the use of subjective or incorrect interpretations of data like mistaking correlation with causation. The project offers specific steps to take for eliminating issues of negative bias in the creation of algorithms. The standard will also include benchmarking procedures and criteria for selecting validation data sets, establishing and communicating the application boundaries for which the algorithm has been designed and guarding against unintended consequences.

IEEE P7004™ - Standard on Child and Student Data Governance[10] provides processes and certifications for transparency and accountability for educational institutions that handle data meant to ensure the safety of students. The standard defines how to access, collect, share and remove data related to children and students in any educational or institutional setting where their information will be access, stored or shared.

IEEE P7005™ - Standard on Employer Data Governance[11] provides guidelines and certifications on storing, protecting and using employee data in an ethical and transparent way. The project recommends tools and services that help employees make informed decisions with their personal information. The standard will help provide clarity and recommendations both for how employees can share their information in a safe and trusted environment as well as how employers can align with employees in this process while still utilizing information needed for regular work flows.

[6] https://standards.ieee.org/develop/project/7000.html
[7] https://standards.ieee.org/develop/project/7001.html
[8] https://standards.ieee.org/develop/project/7002.html
[9] https://standards.ieee.org/develop/project/7003.html
[10] https://standards.ieee.org/develop/project/7004.html
[11] https://standards.ieee.org/develop/project/7005.html

IEEE P7006™ - Standard on Personal Data AI Agent Working Group[12] addresses concerns raised about machines making decisions without human input. This standard hopes to educate government and industry on why it is best to put mechanisms into place to enable the design of systems that will mitigate the ethical concerns when AI systems can organize and share personal information on their own. Designed as a tool to allow any individual to essentially create their own personal "terms and conditions" for their data, the AI Agent will provide a technological tool for individuals to manage and control their identity in the digital and virtual world.

IEEE P7007™ - Ontological Standard for Ethically driven Robotics and Automation Systems[13] establishes a set of ontologies with different abstraction levels that contain concepts, definitions and axioms that are necessary to establish ethically driven methodologies for the design of Robots and Automation Systems.

IEEE P7008™ - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems[14] establishes a delineation of typical nudges (currently in use or that could be created) that 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. "Nudges" as exhibited by robotic, intelligent or autonomous systems are defined as overt or hidden suggestions or manipulations designed to influence the behavior or emotions of a user.

IEEE P7009™ - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems[15] 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. The standard serves as the basis for developers, as well as users and regulators, to design fail-safe mechanisms in a robust, transparent and accountable manner.

IEEE P7010™ - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems[16] establishes wellbeing metrics relating to human factors directly affected by intelligent and autonomous systems and establish a baseline for the types of objective and subjective data these systems should analyze and include (in their programming and functioning) to proactively increase human wellbeing.

IEEE P7011™ - Standard for the Process of Identifying and Rating the Trustworthiness of News Sources[17] 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. This process will be used to produce truthfulness scorecards through multi-faceted and multi-sourced approaches. The standard defines an algorithm using open source software and a score card rating system as methodology for rating trustworthiness as a core tenant in an effort to establish trust and acceptance.

---

[12] https://standards.ieee.org/develop/project/7006.html
[13] https://standards.ieee.org/develop/project/7007.html
[14] https://standards.ieee.org/develop/project/7008.html
[15] https://standards.ieee.org/develop/project/7009.html
[16] https://standards.ieee.org/develop/project/7010.html
[17] https://development.standards.ieee.org/get-file/P7011.pdf%3Ft=95323300003

IEEE P7012™ - Standard for Machine Readable Personal Privacy Terms[18] provides individuals with means to proffer their own terms respecting personal privacy, in ways that can be read, acknowledged and agreed to by machines operated by others in the networked world. In a more formal sense, the purpose of the standard is to enable individuals to operate as first parties in agreements with others--mostly companies--operating as second parties. Note that the purpose of this standard is not to address privacy policies, since these are one-sided and need no agreement. (Terms require agreement; privacy policies do not.)

About IEEE

IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. IEEE and its members inspire a global community to innovate for a better tomorrow through its highly-cited publications, conferences, technology standards, and professional and educational activities. IEEE is the trusted "voice" for engineering, computing, and technology information around the globe.

For more information, please contact Karen McCabe at k.mccabe@ieee.org

[18] https://development.standards.ieee.org/get-file/P7012.pdf%3Ft=95323600003