Background, Mission and Activities of The IEEE Global Initiative

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.

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

Our Activities
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*; and the identification and recommendation of ideas for Standards Projects focused on prioritizing ethical considerations in A/IS.

Ethically Aligned Design
Version 1 of *Ethically Aligned Design* (EAD) 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.

Version one 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 A/IS.

- You can download *Ethically Aligned Design, V1* at this link.
- You can also see an overview of the document here.
- You can learn about all of The Initiative's Committees and Members via this link.

Ethically Aligned Design, Version 2 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.

- You can download *Ethically Aligned Design, v2* at this link.
- You can also see an overview of the document here.

Recommending Standards Projects
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 descriptions for each of these approved IEEE Standards Projects. More information is available below.

- **IEEE P7000**™ - Model Process for Addressing Ethical Concerns During System Design 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 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

- **IEEE P7002**™ - Data Privacy Process 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 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 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 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.

- **IEEE P7006**™ - Standard on Personal Data AI Agent Working Group 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](#) 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](#) 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](#) 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](#) will establish 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 well-being.