[The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems](#) (AI/AS) was launched in April of 2016 to move beyond both the paranoia and the uncritical admiration regarding autonomous and intelligent technologies and to show that aligning technology with ethical values will help advance innovation with these new tools while diminishing fear in the process.

The goal of The IEEE Global Initiative is to incorporate ethical aspects of human wellbeing that may not automatically be considered in the current design and manufacture of AIS 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 technologist is educated, trained, and empowered to prioritize ethical considerations in the design and development of autonomous and intelligent systems.

The General Principles Committee has articulated high-level ethical concerns applying to all types of AI/AS that:
- Embody the highest ideals of human rights.
- Prioritize the maximum benefit to humanity and the natural environment.
- Mitigate risks and negative impacts as AI/AS evolve as socio-technical systems.

It is the Committee's intention that the Principles, Issues, and Candidate Recommendations they have identified will eventually serve to underpin and scaffold future norms and standards within a new framework of ethical governance for AI/AS design.

Sample Issues:
- How can we ensure that AI/AS do not infringe human rights? (Framing the Principle of Human Rights)
- How can we assure that AI/AS are accountable? (Framing the Principle of Responsibility)
- How can we ensure that AI/AS are transparent? (Framing the Principle of Transparency)
- How can we extend the benefits and minimize the risks of AI/AS technology being misused? (Framing the Principle of Education and Awareness)

A key ethical dilemma regarding personal information is *data asymmetry*. To address this asymmetry the Personal Data and Individual Access Control Committee of the Initiative has elucidated issues and candidate recommendations demonstrating the fundamental need for people to *define*, *access*, and *manage* their personal data as curators of their unique identity.

The Committee recognizes there are no perfect solutions, and that any digital tool can be hacked.   Nonetheless they recommend the enablement of a data environment where people control their sense of self and have provided examples of tools   and evolved practices that could eradicate data asymmetry for a positive future.
Sample Issues:
- How can an individual define and organize his/her personal data in the algorithmic era?
- What is the definition and   scope of personally identifiable information?
- What is the definition of control regarding personal data?
- How can we redefine data access to honor the individual?

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 Wellbeing with Artificial Intelligence and Autonomous Systems*; and the identification and recommendation of ideas for Standards Projects focused on prioritizing ethical considerations in AI/AS.

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 general 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 <u>the</u> "go-to" resource to help technologists and policy makers prioritize ethical considerations in AI/AS.

Ethically Aligned Design, Version 2 is set to launch in late 2017 and will feature five new sections in addition to updated iterations of the original eight sections. 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 will contain over 120 key Issues and Candidate Recommendations.

Along with creating and evolving *Ethically Aligned Design*, members of The IEEE Global Initiative are encouraged to recommend Standardization Projects to IEEE-SA based on their work. Below are titles and descriptions for each of these approved Standardization Projects.

- **IEEE P7000:** <u>Model Process for Addressing Ethical Concerns During System Design</u> 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:** <u>Transparency of Autonomous Systems</u> 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:** <u>Data Privacy Process</u> 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:** <u>Algorithmic Bias Considerations</u> 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 when 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's 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.