The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

Overview – June, 2017

IEEE STANDARDS ASSOCIATION

Contact | FAQs

standards.ieee.org only ▼ | GO

Find Standards | **Develop Standards** | **Get Involved** | **News & Events** | **About Us** | **Buy Standards** | **eTools**

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

An incubation space for new standards and solutions, certifications and codes of conduct, and consensus building for ethical implementation of intelligent technologies

INDUSTRY CONNECTIONS

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

ICAID ⬙

Download the IEEE Global Initiative *Ethically Aligned Design* document ⬙

Download the IEEE Global Initiative brochure ⬙

NEWS AND EVENTS

ABOUT

The purpose of this 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.

- View specifics regarding the Mission and deliverables for the Initiative.
- See a list of The Initiative's Executive and other Committees.
- Learn more from Frequently Asked Questions.

Ethically Aligned Design, Version 1 - Request For Input

Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems represents the collective input of over one hundred global thought leaders from academia, science, government and corporate sectors in the fields of Artificial Intelligence, ethics, philosophy, and policy.

◈IEEE   v.1
ETHICALLY ALIGNED DESIGN

◈IEEE   OVERVIEW
ETHICALLY ALIGNED DESIGN

Ethically Aligned Design
A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems

Version 1

- Released December, 2016 as a Creative Commons doc / RFI for public input

- Created by over 100 Global AI/Ethics experts, in a bottom up, globally open and transparent process

- Eight Committees / Sections

- Contains over eighty key Issues and Candidate Recommendations

- Designed as the "go-to" resource to help technologists and policy makers prioritize ethical considerations in AI/AS

Version 1 - For Public Discussion

◆IEEE
Advancing Technology for Humanity

ETHICALLY ALIGNED DESIGN

A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems

◆IEEE
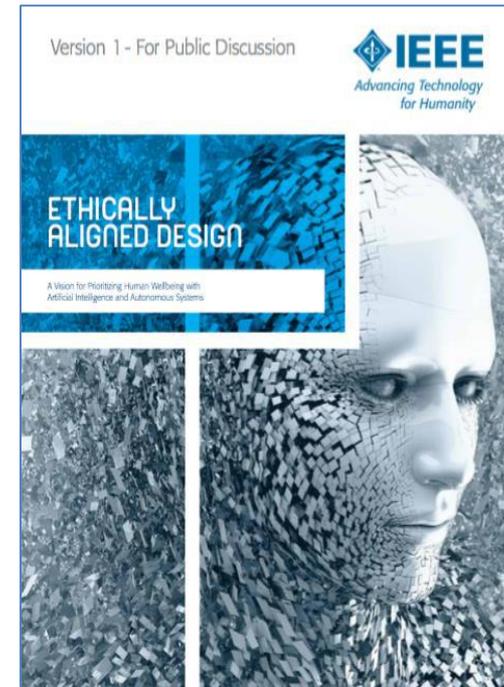
Ethically Aligned Design, v1
Economics/Humanitarian Issues Committee

Technologies, methodologies, and systems that aim to reduce human intervention in our day- to-day lives are evolving at a rapid pace and are poised to transform the lives of individuals in multiple ways. The aim of the Economics/ Humanitarian Issues Committee is to identify the key drivers shaping the human-technology global ecosystem and address economic and humanitarian ramifications, and to suggest key opportunities for solutions that could be implemented by unlocking critical choke points of tension. The goal of the Committee's recommendations is to suggest a pragmatic direction related to these central concerns in the relationship of humans, their institutions and emerging information-driven technologies, to facilitate interdisciplinary, cross-sector dialog that can be more fully informed by expert, directional, and peer-guided thinking regarding these issues.

Sample Issues:

- Misinterpretation of AI/AS in media is confusing to the public.
- Automation is not typically viewed only within market contexts.
- The complexities of employment are being neglected regarding robotics/AI.
- Technological change is happening too fast for existing methods of (re)training the workforce.

◆IEEE

Ethically Aligned Design, v1
Law Committee

The early development of AI/AS has given rise to many complex ethical problems. These ethical issues almost always directly translate into concrete legal challenges—or they give rise to difficult collateral legal problems. The Law Committee feels there is much work for lawyers in this field that, thus far, has attracted very few practitioners and academics despite being an area of pressing need. Lawyers need to be part of discussions on regulation, governance, domestic and international legislation in these areas so the huge benefits available to humanity and our planet from AI/AS are thoughtfully stewarded for the future.

Sample Issues:

- How can we improve the accountability and verifiability in autonomous and intelligent systems?
- How can we ensure that AI is transparent and respects individual rights?
- For example governments may be using AI in a way that impinges on the rights of their citizens instead of protecting those rights.
- How can AI systems be designed to guarantee legal accountability for harms caused by these systems?

IEEE

Ethically Aligned Design
A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems

Version 1 - For Public Discussion

IEEE
Advancing Technology
for Humanity

ETHICALLY
ALIGNED DESIGN

A Vision for Prioritizing Human Wellbeing with
Artificial Intelligence and Autonomous Systems
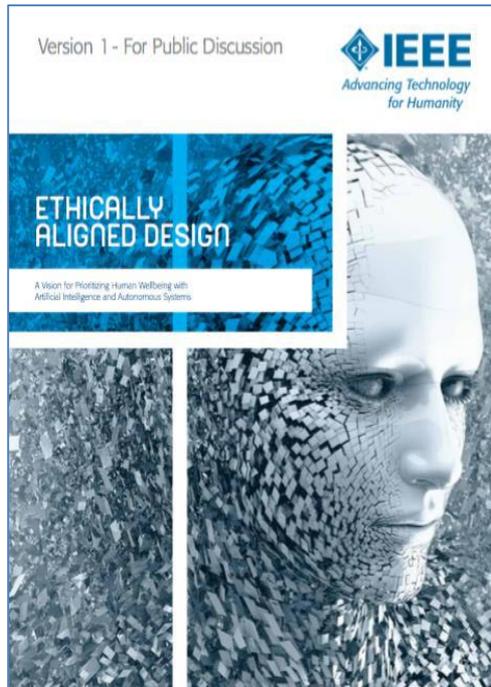
Cover Design TBD

Version 2

- Launching toward end of 2017 as a Creative Commons doc / RFI for second round of public input

- Created by over 250 Global AI/Ethics experts, in a bottom up, transparent, open and increasingly globally inclusive process

- Will incorporate over 200 pages of feedback from public RFI and new Working Groups from China, Japan, Korea and Brazil

- Thirteen Committees / Sections

- Will contains over one hundred twenty key Issues and Candidate Recommendations

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

◆IEEE

IEEE-SA Standards Projects

- **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 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 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-SA Standards Projects

- **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 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-SA Standards Projects

- **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 which are necessary to establish ethically driven methodologies for the design of Robots and Automation Systems.

- *Several new Standards projects pending approval by end of June, 2017.*

Thank you!

The Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

Contact: John C. Havens, Executive Director
john.havens.us@ieee.org / 917-597-3323