The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (The IEEE Global Initiative) is a program of the IEEE Standards Association with the status of an Operating Unit of The Institute of Electrical and Electronics Engineers, Incorporated (IEEE), the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity with over 420,000 members in more than 160 countries.

©2019 IEEE
19-0002

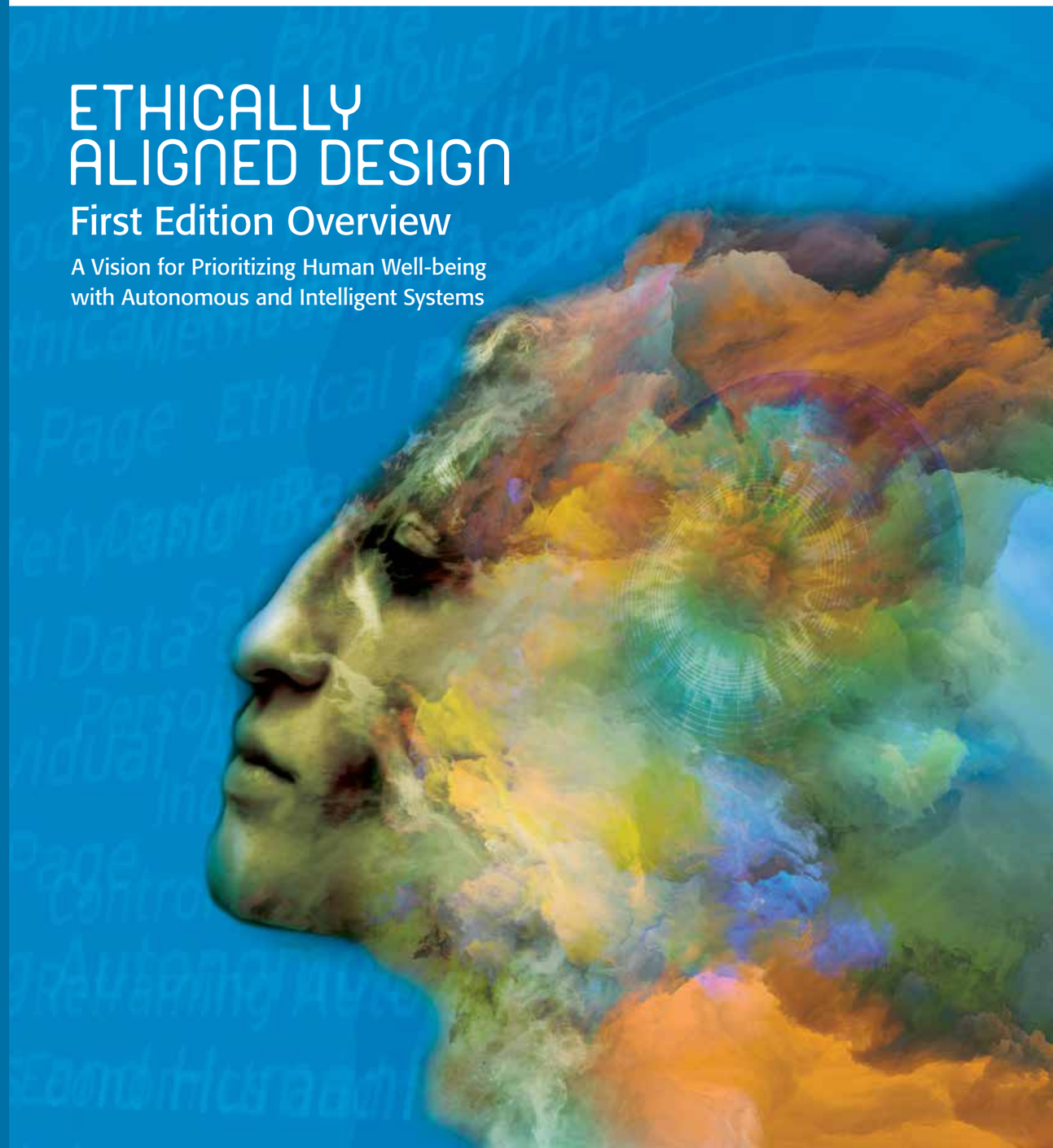
IEEE
Advancing Technology for Humanity

IEEE
Advancing Technology for Humanity

ETHICALLY ALIGNED DESIGN

First Edition Overview

A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems



Praise for *Ethically Aligned Design*

"To create and foster trust between humans and machines, you must understand the ethical resources and standards available for reference during the designing, building, and maintenance of AI. The large-scale focus on AI ethics by groups like The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems... should be mirrored in businesses and working groups of all sizes."

- IBM: Everyday Ethics for Artificial Intelligence; A practical guide for designers & developers, September 2018

"With...Ethically Aligned Design...there could be every chance companies will move beyond good intentions and into standardized practices that factor human well-being into design."

- Mashable

"This document, which will continue to change and grow with the times, encourages those pushing technology forward to consider how such progress might interfere with ethical concerns. The nature of the publication, that it is 'living', shows how well its creators at IEEE understand the ever-changing nature of the AI field."

- Futurism

"How often do you come across a paper from the tech community that says, 'Human well-being is the highest virtue for a society, and human flourishing begins with conscious contemplation'? That is the tech community that I recognize, a community of intelligent and articulate people with a genuine desire to make the world a better place. A community for which I am proud to belong."

- Australia's Chief Scientist, Dr. Alan Finkel, IEEE Sections Congress, August 2017

"As the tech world pushes forward with the development of artificial intelligence, The Institute of Electrical and Electronics Engineers (IEEE) is asking everyone to pause and consider the ethical ramifications."

- ZDNet

"The organization has published a framework document it's hoping will guide the industry toward the light—and help technologists build benevolent and beneficial autonomous systems, rather than thinking that ethics is not something they need to be worrying about."

- TechCrunch

Introduction

As the use and impact of autonomous and intelligent systems (A/IS) become pervasive, we need to establish societal and policy guidelines in order for such systems to remain human-centric, serving humanity's values and ethical principles. These systems must be developed and should operate in a way that is beneficial to people and the environment, beyond simply reaching functional goals and addressing technical problems. This approach will foster the heightened level of trust between people and technology that is needed for its fruitful use in our daily lives.

To be able to contribute in a positive, non-dogmatic way, we, the techno-scientific communities, need to enhance our self-reflection. We need to have an open and honest debate around our explicit or implicit values, including our imaginary[1] around so-called "Artificial Intelligence" and the institutions, symbols, and representations it generates.

Ultimately, our goal should be "eudaimonia", a practice elucidated by Aristotle that defines human well-being, both at the individual and collective level, as the highest virtue for a society. Translated roughly as "flourishing", the benefits of eudaimonia begin with conscious contemplation, where ethical considerations help us define how we wish to live.

Whether our ethical practices are Western (e.g., Aristotelian, Kantian), Eastern (e.g., Shinto, 墨家/School of Mo, Confucian), African (e.g., Ubuntu), or from another tradition, honoring holistic definitions of societal prosperity is essential versus pursuing one-dimensional goals of increased productivity or gross domestic product (GDP). Autonomous and intelligent systems should prioritize and have as their goal the explicit honoring of our inalienable fundamental rights and dignity as well as the increase of human flourishing and environmental sustainability.

The goal of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems ("The IEEE Global Initiative") is that *Ethically Aligned Design* will provide pragmatic and directional insights and recommendations, serving as a key reference for the work of technologists, educators, and policy makers in the coming years.

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

[1] The symbols, values, institutions, and norms of a societal group through which people imagine their lives and constitute their societies.

creative commons
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 United States License.

Executive Summary

I. Purpose of *Ethically Aligned Design*

Autonomous and intelligent technical systems are specifically designed to reduce the necessity for human intervention in our day-to-day lives. In so doing, these new systems are also raising concerns about their impact on individuals and societies. Current discussions include advocacy for a positive impact, such as optimization of processes and resource usage, more informed planning and decisions, and recognition of useful patterns in big data. Discussions also include warnings about potential harm to privacy, discrimination, loss of skills, adverse economic impacts, risks to security of critical infrastructure, and possible negative long-term effects on societal well-being.

Because of their nature, the full benefit of these technologies will be attained only if they are aligned with society's defined values and ethical principles. Through this work we intend, therefore, to establish frameworks to guide and inform dialogue and debate around the non-technical implications of these technologies, in particular related to ethical aspects. We understand "ethical" to go beyond moral constructs and include social fairness, environmental sustainability, and our desire for self-determination.

Our analyses and recommendations in *Ethically Aligned Design* address values and intentions as well as implementations, both legal and technical. They are both aspirational, what we hope or wish should happen, and practical, what we—the techno-scientific community and every group involved with and/or affected by these technologies—could do for society to advance in positive directions.

Chapters in Ethically Aligned Design, First Edition

1. From Principles to Practice
2. General Principles
3. Classical Ethics in A/IS
4. Well-being
5. Affective Computing
6. Personal Data and Individual Agency
7. Methods to Guide Ethical Research and Design
8. A/IS for Sustainable Development
9. Embedding Values into Autonomous and Intelligent Systems
10. Policy
11. Law

II. General Principles

The ethical and values-based design, development, and implementation of autonomous and intelligent systems should be guided by the following General Principles:

1. **Human Rights**
   A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.

2. **Well-being**
   A/IS creators shall adopt increased human well-being as a primary success criterion for development.

3. **Data Agency**
   A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.

4. **Effectiveness**
   A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.

5. **Transparency**
   The basis of a particular A/IS decision should always be discoverable.

6. **Accountability**
   A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.

7. **Awareness of Misuse**
   A/IS creators shall guard against all potential misuses and risks of A/IS in operation.

8. **Competence**
   A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

creative commons

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 United States License.

Ethically Aligned Design, First Edition Overview

III. Ethical Foundations

Classical Ethics

By drawing from over two thousand five hundred years of classical ethics traditions, the authors of *Ethically Aligned Design* explored established ethics systems, addressing both scientific and religious approaches, including secular philosophical traditions, to address human morality in the digital age. Through reviewing the philosophical foundations that define autonomy and ontology, this work addresses the alleged potential for autonomous capacity of intelligent technical systems, morality in amoral systems, and asks whether decisions made by amoral systems can have moral consequences.

IV. Areas of Impact

A/IS for Sustainable Development

Through affordable and universal access to communications networks and the Internet, autonomous and intelligent systems can be made available to and benefit populations anywhere. They can significantly alter institutions and institutional relationships toward more human-centric structures, and they can address humanitarian and sustainable development issues resulting in increased individual societal and environmental well-being. Such efforts could be facilitated through the recognition of and adherence to established indicators of societal flourishing such as the United Nations Sustainable Development Goals so that human well-being is utilized as a primary success criteria for A/IS development.

Personal Data Rights and Agency Over Digital Identity

People have the right to access, share, and benefit from their data and the insights it provides. Individuals require mechanisms to help create and curate the terms and conditions regarding access to their identity and personal data, and to control its safe, specific, and finite exchange. Individuals also require policies and practices that make them explicitly aware of consequences resulting from the aggregation or resale of their personal information.

Legal Frameworks for Accountability

The convergence of autonomous and intelligent systems and robotics technologies has led to the development of systems with attributes that simulate those of human beings in terms of partial autonomy, ability to perform specific intellectual tasks, and even a human physical appearance. The issue of the legal status of complex autonomous and intelligent systems thus intertwines with broader legal questions regarding how to ensure accountability and allocate liability when such systems cause harm. It is clear that:

- Autonomous and intelligent technical systems should be subject to the applicable regimes of property law.
- Government and industry stakeholders should identify the types of decisions and operations that should never be delegated to such systems. These stakeholders should adopt rules and standards that ensure effective human control over those decisions and how to allocate legal responsibility for harm caused by them.
- The manifestations generated by autonomous and intelligent technical systems should, in general, be protected under national and international laws.
- Standards of transparency, competence, accountability, and evidence of effectiveness should govern the development of autonomous and intelligent systems.

Policies for Education and Awareness

Effective policy addresses the protection and promotion of human rights, safety, privacy, and cybersecurity, as well as the public understanding of the potential impact of autonomous and intelligent technical systems on society. To ensure that they best serve the public interest, policies should:

- Support, promote, and enable internationally recognized legal norms.
- Develop government expertise in related technologies.
- Ensure governance and ethics are core components in research, development, acquisition, and use.
- Regulate to ensure public safety and responsible system design.
- Educate the public on societal impacts of related technologies.

creative commons
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 United States License.

V. Implementation

Well-being Metrics

For autonomous and intelligent systems to provably advance a specific benefit for humanity, there need to be clear indicators of that benefit. Common metrics of success include profit, gross domestic product, consumption levels, and occupational safety. While important, these metrics fail to encompass the full spectrum of well-being for individuals, the environment, and society. Psychological, social, economic fairness, and environmental factors matter. Well-being metrics include such factors, allowing the benefits arising from technological progress to be more comprehensively evaluated, providing opportunities to test for unintended negative consequences that could diminish human well-being. A/IS can improve capturing of and analyzing the pertinent data, which in turn could help identify where these systems would increase human well-being, providing new routes to societal and technological innovation.

Embedding Values into Autonomous and Intelligent Systems

If machines engage in human communities as quasi-autonomous agents, then those agents must be expected to follow the community's social and moral norms. Embedding norms in such quasi-autonomous systems requires a clear delineation of the community in which they are to be deployed. Further, even within a particular community, different types of technical embodiments will demand different sets of norms. The first step is to identify the norms of the specific community in which the systems are to be deployed and, in particular, norms relevant to the kinds of tasks that they are designed to perform.

Methods to Guide Ethical Research and Design

To create autonomous and intelligent technical systems that enhance and extend human well-being and freedom, values-based design methods must put human advancement at the core of development of technical systems. This must be done in concert with the recognition that machines should serve humans and not the other way around. System developers should employ values-based design methods in order to create sustainable systems that can be evaluated in terms of not only providing increased economic value for organizations but also of broader social costs and benefits.

Affective Computing

Affect is a core aspect of intelligence. Drives and emotions such as anger, fear, and joy are often the foundations of actions throughout our lives. To ensure that intelligent technical systems will be used to help humanity to the greatest extent possible in all contexts, autonomous and intelligent systems that participate in or facilitate human society should not cause harm by either amplifying or dampening human emotional experience.

The Mission and Results of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

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.

To advance toward this goal, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems brought together more than a thousand participants from six continents who are thought leaders from academia, industry, civil society, policy, and government in the related technical and humanistic disciplines to identify and find consensus on timely issues surrounding autonomous and intelligent systems.

By "stakeholder" we mean anyone involved in the research, design, manufacture, or messaging around intelligent and autonomous systems—including universities, organizations, governments, and corporations—all of which are making these technologies a reality for society.

creative commons
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 United States License.

Ethically Aligned Design, First Edition Overview

From Principles to Practice—Results from our Work to Date

In addition to the creation of *Ethically Aligned Design*, The IEEE Global Initiative, independently or through the IEEE Standards Association, has directly inspired the following works:

- **The launch of the IEEE P7000™ series of approved standardization projects**

  (see page 12). This is the first series of standards in the history of the IEEE Standards Association that explicitly focuses on societal and ethical issues associated with a certain field of technology.

- **Artificial Intelligence and Ethics in Design**

  These ten courses are designed for global professionals, as well as their managers, working in engineering, IT, computer science, big data, artificial intelligence, and related fields across all industries who require up-to-date information on the latest technologies. The courses explicitly mirror content from *Ethically Aligned Design*, and feature numerous experts as instructors who helped create *Ethically Aligned Design*.

  More information can be found at:
  innovationatwork.ieee.org/courses/artificial-intelligence-and-ethics-in-design/

- **The creation of an A/IS Ethics Glossary**

  We received a great deal of feedback on the need for definitions and recommendations for key terms in *Ethically Aligned Design*. To that end, we created a companion Glossary. The Glossary features more than two hundred pages of terms that help to define the context of A/IS ethics for multiple stakeholder groups, specifically: engineers, policy makers, philosophers, standards developers, and computational disciplines experts. It is currently in its second iteration and has also been informed by the IEEE P7000™ standards working groups.

  Download the Glossary at:
  standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_glossary.pdf

- **The launch of OCEANIS**

  The IEEE Standards Association, inspired by the work of The IEEE Global Initiative, has contributed significantly to the establishment of The Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS). It is a global forum for discussion, debate, and collaboration for organizations interested in the development and use of standards to further the creation of autonomous and intelligent systems. OCEANIS members are working together to enhance the understanding of the role of standards in facilitating innovation, while addressing problems that expand beyond technical solutions to addressing ethics and values.

  More information can be found at:
  ethicsstandards.org

- **The launch of ECPAIS**

  The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) has the goal to create specifications for certification and marking processes that advance transparency, accountability, and reduction in algorithmic bias in autonomous and intelligent systems. ECPAIS intends to offer a process and define a series of marks by which organizations can seek certifications for their processes around the A/IS products, systems, and services they provide.

  More information can be found at:
  standards.ieee.org/industry-connections/ecpais.html

- **The launch of CXI**

  The Council on Extended Intelligence (CXI) was directly inspired by the work of The IEEE Global Initiative and the work of MIT Media Lab around "Extended Intelligence". CXI was launched jointly by the IEEE Standards Association and MIT Media Lab. Its mission is to proliferate the ideals of responsible participant design, data agency, and metrics of economic prosperity, prioritizing people and the planet over profit and productivity. Membership includes thought leaders from the EU Parliament and Commission, the UK House of Lords, the OECD, the United Nations, local and national administrations, and renowned experts in economics, data science, and multiple other disciplines from around the world.

  More information can be found at:
  globalcxi.org

- **The launch of EADUC**

  The Ethically Aligned Design University Consortium (EADUC) is being established with the aim to reach every engineer at the beginning of their studies. It should help them prioritize values-driven, applied ethical principles at the core of their work. Working in conjunction with philosophers, designers, social scientists, academics, data scientists, and the corporate and policy communities, EADUC also has the goal that *Ethically Aligned Design* will be used in teaching at all levels of education globally as the new vision for design in the algorithmic age.

- **The launch of "AI Commons"**

  The work of The IEEE Global Initiative has delivered key ideas and inspiration that are rapidly evolving toward establishing a global collaborative platform around A/IS. The mission of AI Commons is "to gather a true ecosystem to democratize access to AI capabilities and thus to allow anyone, anywhere to benefit from the possibilities that AI can provide". In addition, the group will be working "to connect problem owners with the community of solvers, to collectively create solutions with AI". The ultimate goal is "to implement a framework for participation and cooperation to make using and benefiting from AI available to all".

  More information can be found at:
  www.aicommons.com

creative commons
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 United States License.

Ethically Aligned Design, First Edition Overview

IEEE P7000™ Approved Standardization Projects

The IEEE P7000™ series of standards projects under development represents a unique addition to the collection of over 1,900 global IEEE standards and projects. Whereas more traditional standards have a focus on technology interoperability, functionality, safety, and trade facilitation, the IEEE P7000 series addresses specific issues at the intersection of technological and ethical considerations. Like its technical standards counterparts, the IEEE P7000 series empowers innovation across borders and enables societal benefit.

For more information or to join any working group, please see the links below. Committees that authored *Ethically Aligned Design*, as well as other committees within IEEE, that created specific working groups are listed below each project.

- **IEEE P7000™,** IEEE Standards Project Model Process for Addressing Ethical Concerns During System Design
  Inspired by Methodologies to Guide Ethical Research and Design Committee, and supported by IEEE Computer Society
  www.computer.org

- **IEEE P7001™,** IEEE Standards Project for Transparency of Autonomous Systems
  Inspired by the General Principles Committee, and supported by IEEE Vehicular Technology Society
  vtsociety.org

- **IEEE P7002™,** IEEE Standards Project for Data Privacy Process
  Inspired by The Personal Data and Individual Agency Control Committee, and supported by IEEE Computer Society
  www.computer.org

- **IEEE P7003™,** IEEE Standards Project for Algorithmic Bias Considerations
  Supported by IEEE Computer Society
  www.computer.org

- **IEEE P7004™,** IEEE Standards Project for Child and Student Data Governance
  Inspired by The Personal Data and Individual Agency Control Committee, and supported by IEEE Computer Society
  www.computer.org

- **IEEE P7005™,** IEEE Standards Project for Employer Data Governance
  Inspired by The Personal Data and Individual Agency Control Committee, and supported by IEEE Computer Society
  www.computer.org

- **IEEE P7006™,** IEEE Standards Project for Personal Data AI Agent Working Group
  Inspired by The Personal Data and Individual Agency Control Committee, and supported by IEEE Computer Society
  www.computer.org

- **IEEE P7007™,** IEEE Standards Project for Ontological Standard for Ethically Driven Robotics and Automation Systems
  Supported by IEEE Robotics and Automation Society
  www.ieee-ras.org

- **IEEE P7008™, I**EEE Standards Project for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
  Inspired by the Affective Computing Committee, and supported by IEEE Robotics and Automation Society
  www.ieee-ras.org

- **IEEE P7009™,** IEEE Standards Project for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
  Supported by IEEE Reliability Society
  rs.ieee.org

- **IEEE P7010™,** IEEE Standards Project for Well-being Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems
  Inspired by the Well-being Committee, and supported by IEEE Systems, Man and Cybernetics Society
  ieeesmc.org

- **IEEE P7011™,** IEEE Standards Project for the Process of Identifying and Rating the Trustworthiness of News Sources
  Supported by IEEE Society on Social Implications of Technology
  technologyandsociety.org

- **IEEE P7012™,** IEEE Standards Project for Machine Readable Personal Privacy Terms
  Supported by IEEE Society on Social Implications of Technology
  technologyandsociety.org

- **IEEE P7013™,** IEEE Standards Project for Inclusion and Application Standards for Automated Facial Analysis Technology
  Supported by IEEE Society on Social Implications of Technology
  technologyandsociety.org

creative commons
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 United States License.

Who We Are

About IEEE

IEEE is the largest technical professional organization dedicated to advancing technology for the benefit of humanity, with over 420,000 members in more than 160 countries. Through its highly cited publications, conferences, technology standards, and professional and educational activities, IEEE is the trusted voice in a wide variety of areas ranging from aerospace systems, computers, and telecommunications to biomedical engineering, electric power, and consumer electronics.

To learn more, visit the IEEE website: www.ieee.org.

About the IEEE Standards Association

The IEEE Standards Association (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 1,900 active standards and over 650 standards under development.

For more information, visit the IEEE-SA website: standards.ieee.org.

About The IEEE Global Initiative

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (The IEEE Global Initiative) is a program of The Institute of Electrical and Electronics Engineers (IEEE), the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity with over 420,000 members in more than 160 countries. **More information can be found at: standards.ieee.org/industry-connections/ec/autonomous-systems.html.**

The IEEE Global Initiative provides the opportunity to bring together multiple voices in the related technological and scientific communities to identify and find consensus on timely issues. **Names of experts involved in the various committees of The IEEE Global Initiative can be found at: standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ec_bios.pdf**

IEEE makes all versions of *Ethically Aligned Design* available under the Creative Commons Attribution-Non-Commercial 4.0 United States License. Subject to the terms of that license, organizations or individuals can adopt aspects of this work at their discretion at any time. It is also expected that *Ethically Aligned Design* content and subject matter will be selected for submission into formal IEEE processes, including standards development and education purposes. **Information about Creative Commons Licensing can be found at: creativecommons.org/licenses/by-nc/4.0/us/**

The IEEE Global Initiative and *Ethically Aligned Design* contribute, together with other efforts within IEEE, such as IEEE TechEthics™, (techethics.ieee.org), to a broader effort at IEEE to foster open, broad, and inclusive conversation about ethics in technology.

creative commons

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 United States License.