



Standardization Activities for Ethics of AI

Edson Prestes
Professor
Head, ϕ -Robotics Research Group
Informatics Institute
Federal University of Rio Grande do Sul
Brazil

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

<http://ethicsinaction.ieee.org>

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

- Videos & Webinars
- News & Events
- Ethically Aligned Design, Version 1, Translations and Reports
- Download Ethically Aligned Design, Version 2

VIEW THE COMPLETE LIST OF IC ACTIVITIES

ABOUT

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.

- [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.](#)
- [Learn how to join the process of developing the final version of Ethically Aligned Design.](#)

ETHICS IN ACTION

We've launched the second version of *Ethically Aligned Design!* [View Launch Details.](#)

Ethically Aligned Design, Version 2

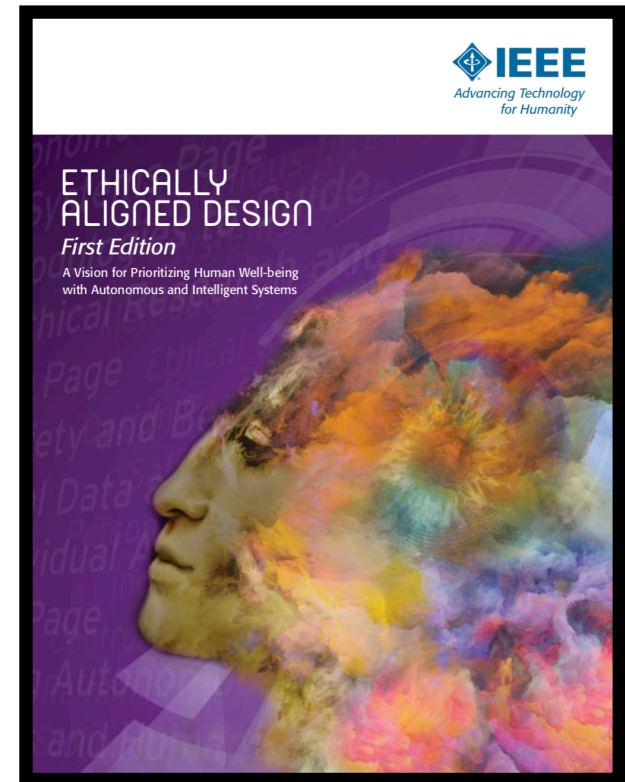
Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (A/IS) represents the collective input of [several hundred participants](#) from six continents who are thought leaders from academia, industry, civil society, policy and government. The goal of *Ethically Aligned Design* is to advance a public discussion about how we can establish ethical and social implementations for intelligent and autonomous systems and technologies, aligning them to defined values and ethical principles that prioritize human well-being in a given cultural context.



[Download Complete Document](#)



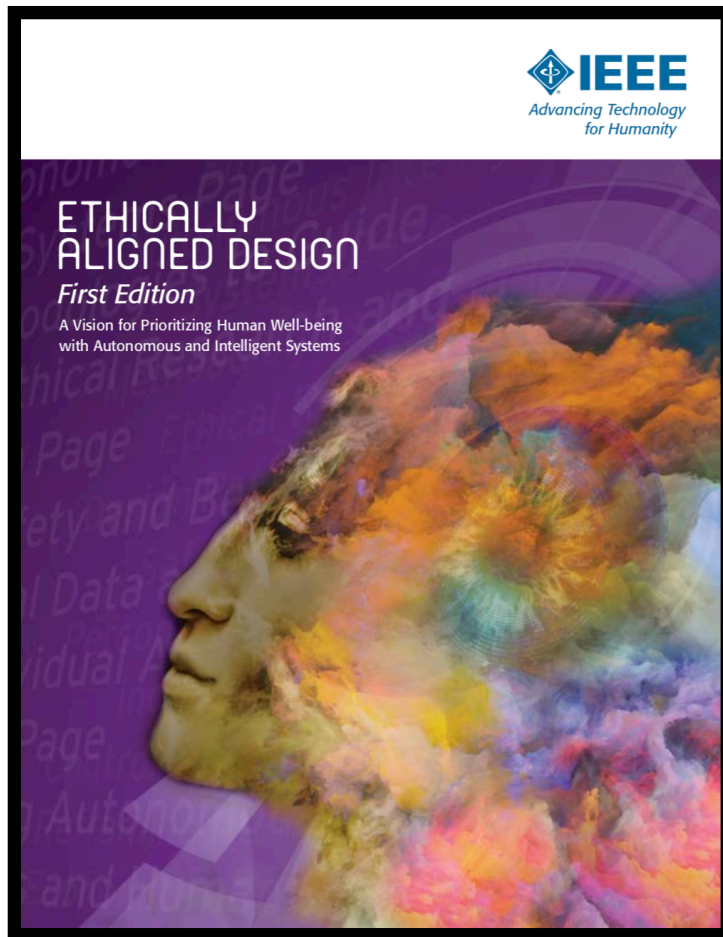
[Download Overview](#)



Ethically Aligned Design 1st Edition was Created by more than 700 global experts focused on the pragmatic instantiation of human-centric, values-driven design.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

<http://ethicsinaction.ieee.org>



Introduction	2
Executive Summary	3-6
Acknowledgements	7-8
<i>Ethically Aligned Design</i>	
From Principles to Practice	9-16
General Principles	17-35
Classical Ethics in A/IS	36-67
Well-being	68-89
Affective Computing	90-109
Personal Data and Individual Agency	110-123
Methods to Guide Ethical Research and Design	124-139
A/IS for Sustainable Development	140-168
Embedding Values into Autonomous and Intelligent Systems	169-197
Policy	198-210
Law	211-281
<i>About Ethically Aligned Design</i>	
The Mission and Results of The IEEE Global Initiative	282
From Principles to Practice—Results of Our Work to Date	283-284
IEEE P7000™ Approved Standardization Projects	285-286
Who We Are	287
Our Process	288-289
How the Document was Prepared	290
How to Cite <i>Ethically Aligned Design</i>	290
Key References	291

IEEE Position Statement

Ethical Aspects of Autonomous and Intelligent Systems

*Approved by the
IEEE Board of Directors (24 June 2019)*

IEEE supports the inclusion of ethical considerations in the design and deployment of autonomous and intelligent systems.

Autonomous and Intelligent systems (A/IS) are systems that are capable of adaption and learning based on feedback and data from their environment. A/IS hold great promise to benefit society in applications domains as diverse as transportation, health and social care, environmental preservation, enterprise productivity, communication network optimization, power grid adaptation and management, agriculture, manufacturing, and entertainment. Recent success in machine learning, signal processing, planning algorithms, digital sensing, embedded systems, cloud computing, as well as voice, image and pattern analysis have greatly accelerated application of A/IS. They hold great promise to benefit society, but they also present potential new social, legal and ethical challenges, with corresponding new requirements to address issues of systemic risk, diminishing trust, privacy challenges and issues of data transparency, ownership and agency.

IEEE Standardisation Initiatives

P7000 - Model Process for Addressing Ethical Concerns During System Design;

P7001 - Transparency of Autonomous Systems;

P7002 - Data Privacy Process;

P7003 - Algorithmic Bias Considerations

P7004 - Standard for Child and Student Data Governance

P7005 - Standard on Employer Data Governance

P7007- Ontological Standard for Ethically Driven Robotics and Automation Systems

P7008 - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems

P7009 - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems

P7010 - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems

P7011 - Standard for the Process of Identifying & Rating the Trust-worthiness of News Sources

P7012 - Standard for Machine Readable Personal Privacy Terms

P7014 - Standard for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems

P7015 - Standard for Data and Artificial Intelligence (AI) Literacy, Skills, and Readiness

IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

<http://standards.ieee.org/develop/project/7007.html>

This standard establishes a set of ontologies with different abstraction levels that contain concepts, definitions, axioms and use cases which are deemed relevant and appropriate to establish ethically driven methodologies for the design of Robots and Automation Systems.

It was developed by **IEEE 7007 - Ontologies for Ethically Driven Robotics and Automation Working Group**.

Chair: Edson Prestes (UFRGS, Brazil)

Vice-chair: Sandro Fiorini (IBM, Brazil)

Secretary: Paulo J.S. Goncalves, Polytechnic Institute of Castelo Branco, Portugal)

Technical Editor: Michael Houghtaling (Retired IBM Senior Software Engineer, USA)

Technical Editor: Babita Ramlal, Canadian Ministry of Government & Consumer Services, Canada)

More than 100 members from different parts of the globe working from 2017-2021.

IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

<http://standards.ieee.org/develop/project/7007.html>

Ontologies allow to capture and represent consensual knowledge in an explicit and formal way, independently of a particular programming language.

Ontologies are an efficient approach to disambiguate knowledge used among groups of humans, robots, and other artificial systems that share the same conceptualisation.

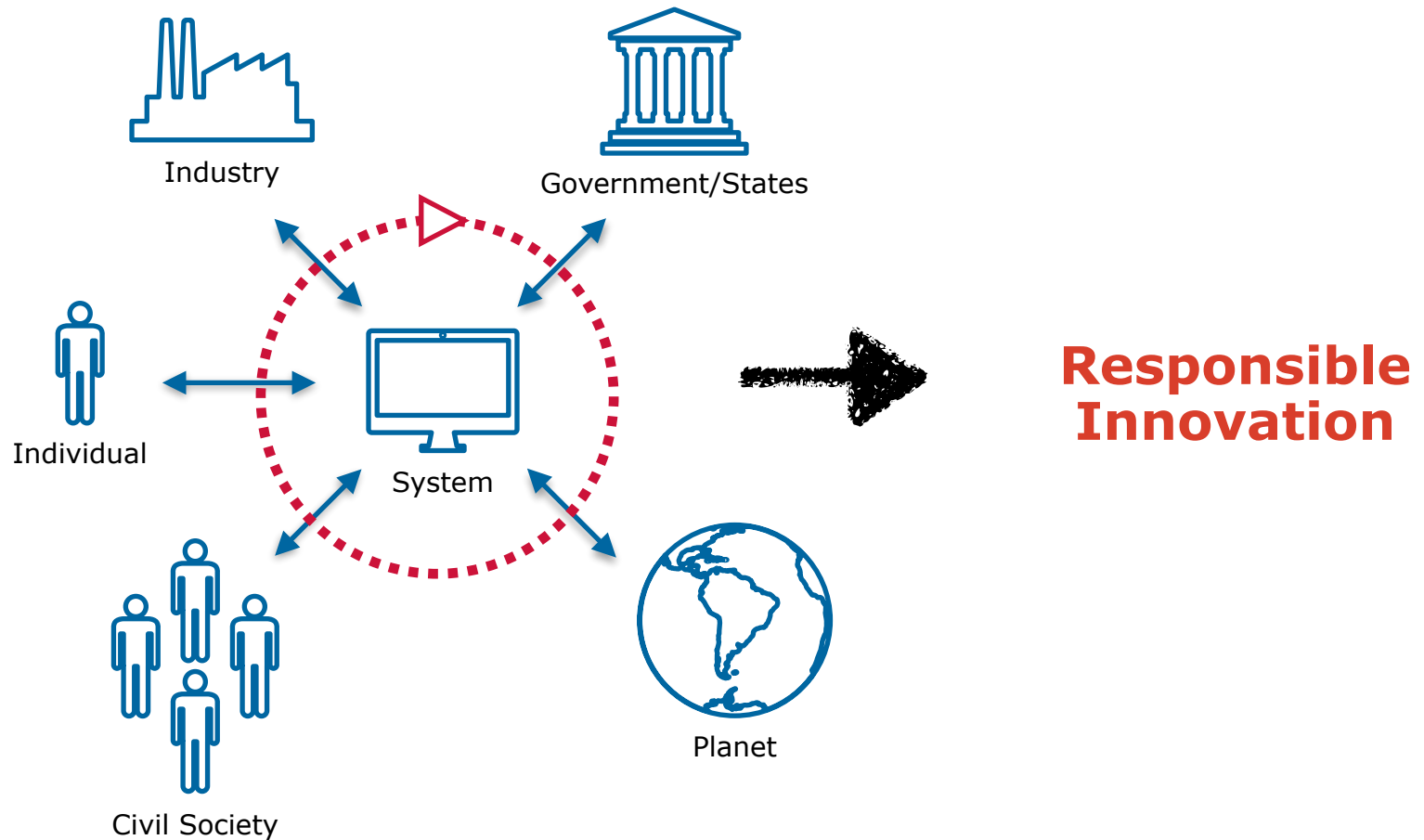
Formal ontologies typically allow us to encode the knowledge in a machine readable format.

The main elements of an ontology are:

- classes which stand for concepts at all granularities;
- relations which stand for associations between concepts;
- formal axioms which constrain and add consistency rules to the concept and relationship structures.

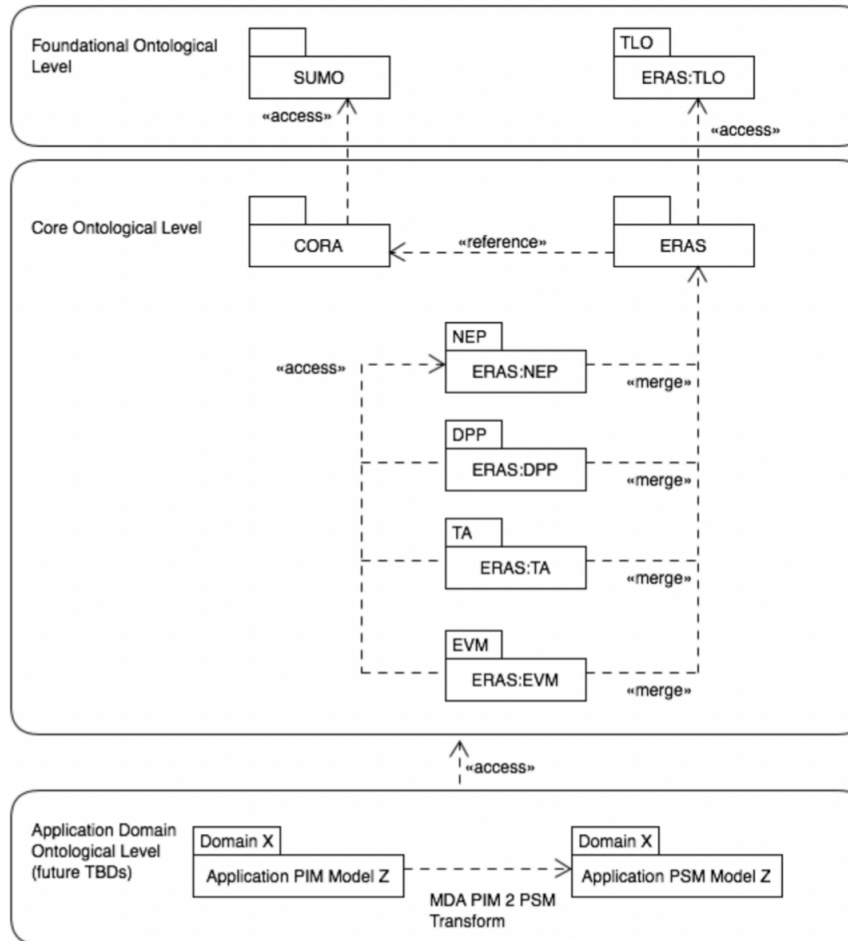
IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

<http://standards.ieee.org/develop/project/7007.html>



IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

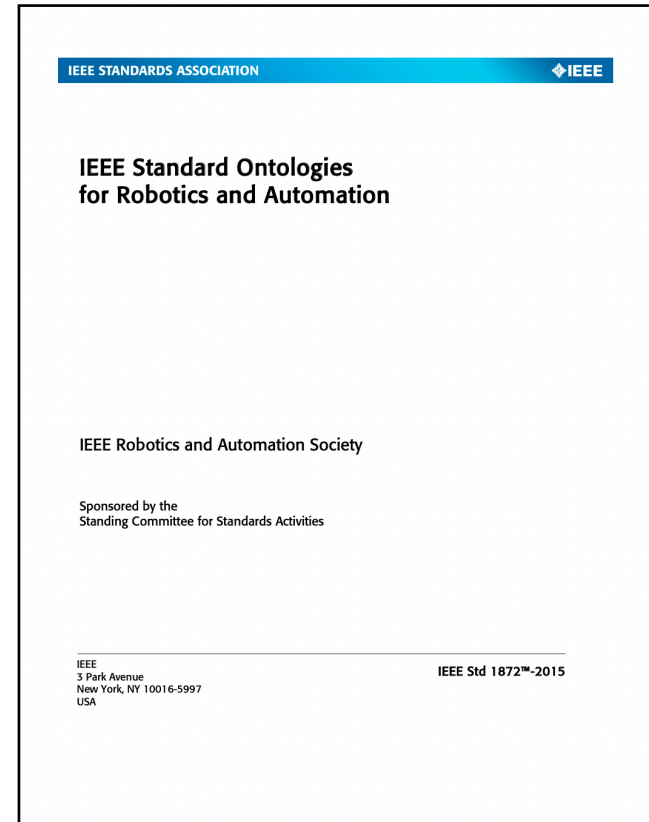
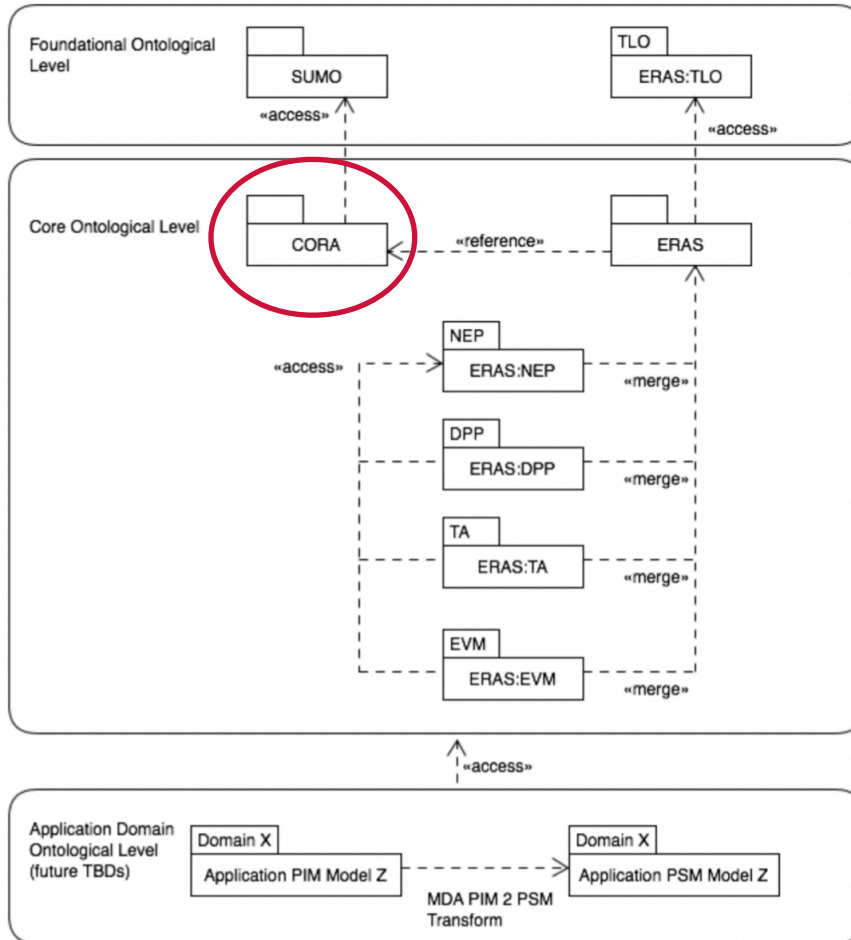
<http://standards.ieee.org/develop/project/7007.html>



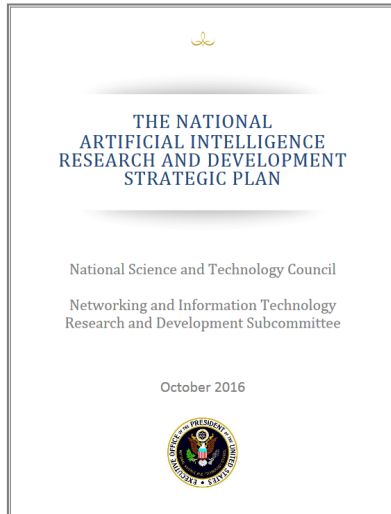
IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

<http://standards.ieee.org/develop/project/7007.html>

First IEEE RAS Standard



IEEE 1872-2015 Standard Ontologies for Robotics and Automation



Mentioned in Former President Obama's **"The National Artificial Intelligence Research and Development Strategic Plan"**

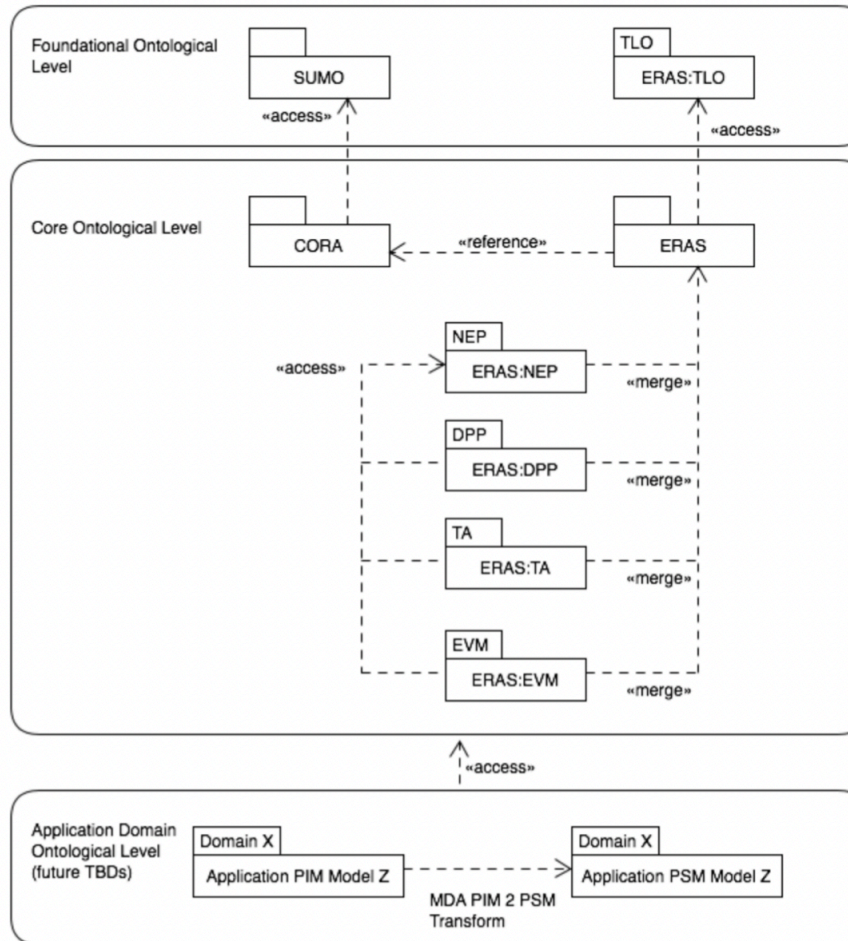
"One example of an AI-relevant standard that has been developed is P1872-2015 (Standard Ontologies for Robotics and Automation), developed by the Institute of Electrical and Electronics Engineers (IEEE). This standard provides a systematic way of representing knowledge and a common set of terms and definitions. These allow for unambiguous knowledge transfer among humans, robots, and other artificial systems, as well as provide a foundational basis for the application of AI technologies to robotics."



In 2015, Working Group was the recipient of IEEE-SA Emerging Technology Award.

IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

<http://standards.ieee.org/develop/project/7007.html>



IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

<http://standards.ieee.org/develop/project/7007.html>

Norms and Ethical Principles (NEP)

This subdomain focuses on concepts and relationships centered around aspects of ethical theories and principles that characterize the norms of expected behaviors for norm aware agents and autonomous systems.

Data Protection and Privacy (DPP)

This subdomain documents the concepts and relationships characterizing the data protection and privacy rules and regulations that should be observed and upheld by ethical agents and autonomous systems.

Transparency and Accountability (TA)

This subdomain captures the concepts and relationships necessary to enable ethical autonomous systems with capabilities that provide informative explanations for past and future contemplated plans and associated action selections.

Ethical Violation Management (EVM)

This subdomain accounts for the set of concepts and relationships associated with capabilities to assess, detect and manage ethical violations in robot behavior. In addition to ethical violation conceptualizations, this subdomain also includes aspects governing accountability, responsibility and legal notions of personhood for agents.

IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

IEEE SA
STANDARDS
ASSOCIATION

IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

IEEE Robotics and Automation Society

Developed by the
Standing Committee for Standards Activities

IEEE Std 7007™-2021

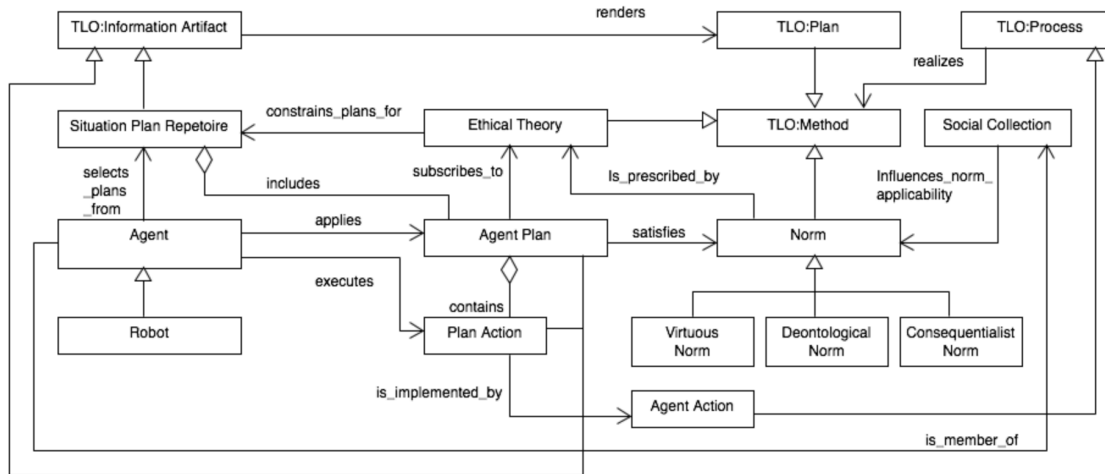


STANDARDS

Contents

1. Overview.....	1
1.1 Scope.....	1
1.2 Purpose.....	1
1.3 Word usage	2
2. Normative references.....	2
3. Definitions, acronyms, and abbreviations	2
3.1 Definitions	2
3.2 Acronyms and abbreviations	3
4. Ontologies for ethically-aligned robotics and automation systems	4
4.1 Conventions	4
4.2 Background	4
4.3 Top-level definitions.....	6
4.4 ERAS top-level concepts	6
4.5 Norms and ethical principles	15
4.6 Data Privacy and Protection	29
4.7 Transparency and Accountability	45
4.8 Ethical Violation Management	56
4.8.1 Axiom Pattern A for Governments with no capacity	67
4.8.2 Axiom Pattern B for Governments achieving an evolving capacity	67
Annex A (informative) Informative definitions	69
A.1 Top-level definitions.....	69
A.2 Norms and Ethical Principles.....	70
A.3 Data Protection and Privacy.....	75
A.4 Transparency and accountability	81
A.5 Ethical Violation Management	86
Annex B (informative) Ontology development.....	89
Annex C (informative) Use cases	91
C.1 Use Case Template.....	91
C.2 Norms and Ethics Use Case: Domestic Personal Assistant Robot	92
C.3 Ethical Violation Management Use Case: Data Privacy and Protection	94
C.4 Transparency use case: autonomous system behavior explanation	96
Annex D (informative) Distributed Responsibility Ascription for Autonomous Systems.....	99
Annex E (informative) Bibliography.....	101

IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems



(forall (x) (if (AgentPlan x)
 (exists (r t a ia n)
 (and
 (SituationPlanRepertoire r)
 (EthicalTheory t)
 (PlanAction a)
 (AgentAction ia)
 (Norm n)
 (includes r x)
 (subscribes_to x t)
 (constrains_plans_for t r)
 (contains x a)
 (is_implemented_by a ia)
 (satisfies x n))))))

Agent plan: A plan subcategory that consists of specifications, partial or complete, for a sequence of agent actions to achieve target goals, objectives, and services to realize agent intentions.

IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

IEEE SA
STANDARDS
ASSOCIATION

IEEE SA EMERGING
TECHNOLOGY AWARD

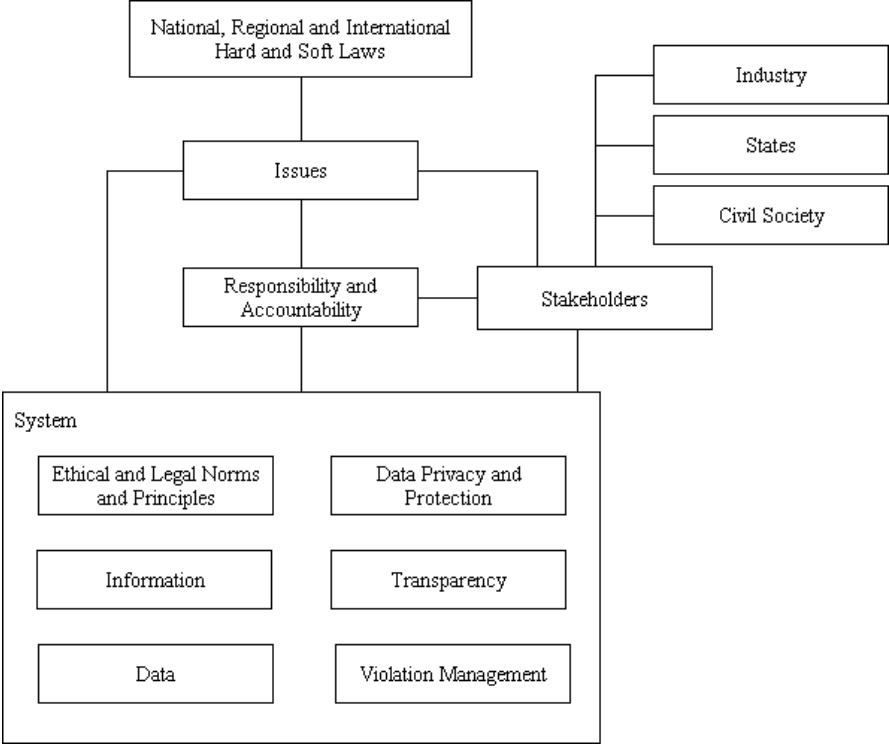
Given to:
IEEE P7007™
Working Group



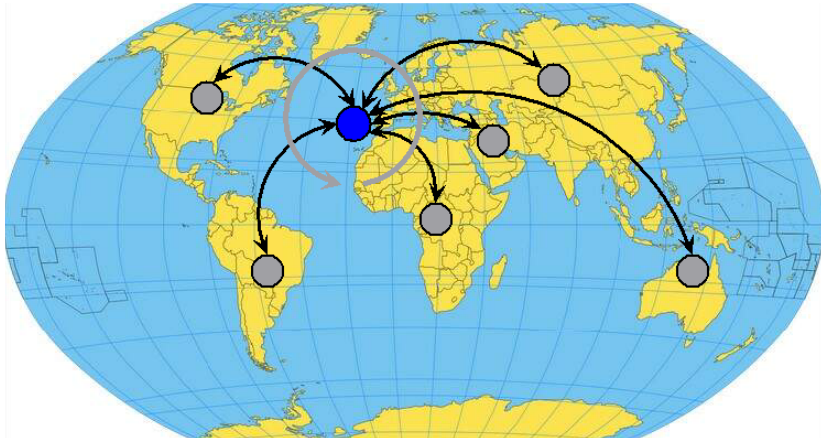
IEEE

For developing an innovative ontological standard on the ethics of artificial intelligence

Multilevel Governing Framework



To be expanded



IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems

Applications

- Guide for teaching ethical design - human and institutional capacity building programs;
- Reference by policy makers and governments to draft AI related policies;
- Common vocabulary to enable the communication among government agencies and other professional bodies around the world;
- Common vocabulary to enable the communication among artificial systems;
- Framework to create ethically-driven systems;
- Framework to strengthen digital cooperation across States - support the development of assessment and decision-making support tools.

PLANET POSITIVE 2030

Imagine the Future We Can Build Together

Join The Campaign

https://bit.ly/3vJRKuy*

Contact : John C. Havens (j.c.havens@ieee.org)

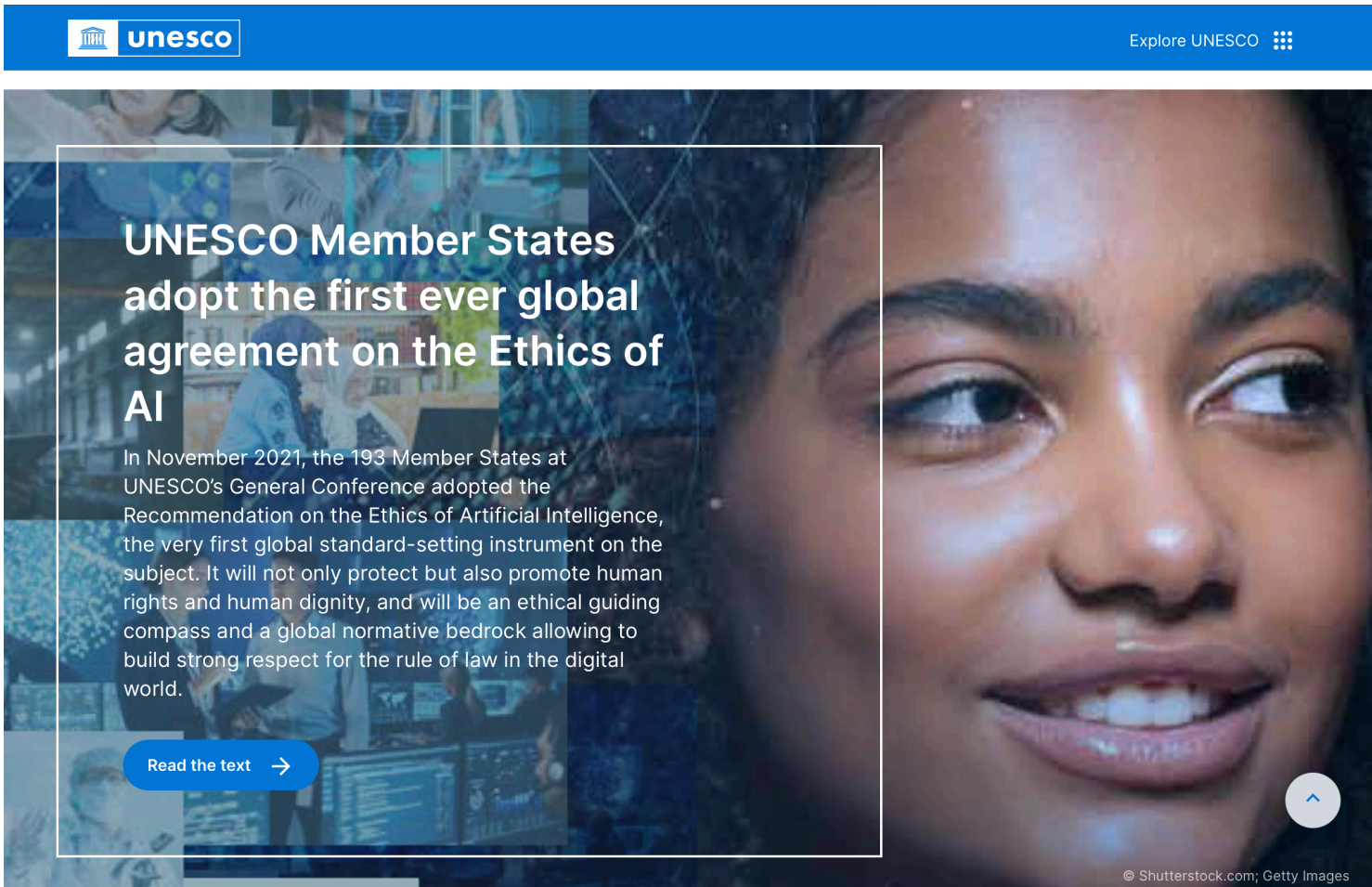



Planet Positive 2030 is an initiative supported by the IEEE Standards Association (IEEE SA) that brings together a global, open community of experts to chart a path for all people to achieve a flourishing future for 2030 and beyond.


Our two “Impossible” Goals:

- Transform society and infrastructure to achieve Planet Positive 2030. (“Planet Positive 2030” – Reduce greenhouse gas emissions to net 50% of 2005 emissions by 2030 and significantly increase regeneration and resilience of earth’s ecosystems.
- Identify the technological solutions we need to design, innovate and deploy to reach Planet Positive 2030.

Recommendation on the Ethics of Artificial Intelligence

A blue banner for UNESCO featuring a collage of images related to technology and AI. On the right side, there is a large, close-up portrait of a young woman with a slight smile, looking off to the side. The background of the banner includes images of people working at computers and abstract digital patterns.

 **unesco**

Explore UNESCO 

UNESCO Member States adopt the first ever global agreement on the Ethics of AI

In November 2021, the 193 Member States at UNESCO's General Conference adopted the Recommendation on the Ethics of Artificial Intelligence, the very first global standard-setting instrument on the subject. It will not only protect but also promote human rights and human dignity, and will be an ethical guiding compass and a global normative bedrock allowing to build strong respect for the rule of law in the digital world.

[Read the text](#) →

↑

© Shutterstock.com; Getty Images

Recommendation on the Ethics of Artificial Intelligence

Objectives of this Recommendation are

- to provide a **universal framework of values, principles and actions to guide States** in the formulation of their legislation, policies or other instruments regarding AI, consistent with international law;
- to guide the **actions of individuals, groups, communities, institutions and private sector companies** to ensure the embedding of ethics in all stages of the AI System life cycle;
- to **protect, promote and respect human rights and fundamental freedoms, human dignity and equality, including gender equality**; to safeguard the interests of present and future generations; to preserve the environment, biodiversity and ecosystems; and to respect cultural diversity in all stages of the AI system life cycle;
- to **foster multi-stakeholder, multidisciplinary and pluralistic dialogue and consensus** building about ethical issues relating to AI systems;
- to promote **equitable access to developments and knowledge in the field of AI and the sharing of benefits**, with particular attention to the needs and contributions of LMICs, including LDCs, LLDCs and SIDS.

Recommendation on the Ethics of Artificial Intelligence

Values

- Respect, protection and promotion of human rights and fundamental freedoms and human dignity;
- Environment and ecosystem flourishing;
- Ensuring diversity and inclusiveness;
- Living in peaceful, just and interconnected societies.

Principles

- Proportionality and do no harm;
- Safety and security;
- Fairness and non-discrimination;
- Sustainability;
- Right to Privacy, and Data Protection;
- Human oversight and determination;
- Transparency and explainability;
- Responsibility and accountability;
- Awareness and literacy;
- Multi-stakeholder and adaptive governance and collaboration;

Recommendation on the Ethics of Artificial Intelligence

Policy Areas

- Ethical Impact Assessment;
- Ethical Governance and Stewardship;
- Data Policy;
- Development and International Cooperation;
- Environment and Ecosystems;
- Gender;
- Culture;
- Education and Research;
- Communication and Information;
- Economy and Labour;
- Health and Social Well-being;

Recommendation on the Ethics of Artificial Intelligence

Some recommendations

- Member States should ensure that AI governance mechanisms are inclusive, transparent, multidisciplinary, multilateral (this includes the possibility of mitigation and redress of harm across borders) and multi-stakeholder. In particular, governance should include aspects of anticipation, and effective protection, monitoring of impact, enforcement and redress.
- Member States should encourage and support researchers to analyse the impact of AI systems on the local labour environment in order to anticipate future trends and challenges. These studies should have an interdisciplinary approach and investigate the impact of AI systems on economic, social and geographic sectors, as well as on human-robot interactions and human-human relationships, in order to advise on reskilling and redeployment best practices.
- Member States should endeavour to employ effective AI systems for improving human health and protecting the right to life, including mitigating disease outbreaks, while building and maintaining international solidarity to tackle global health risks and uncertainties, and ensure that their deployment of AI systems in health care be consistent with international law and their human rights law obligations. Member States should ensure that actors involved in health care AI systems take into consideration the importance of a patient's relationships with their family and with health care staff.

Recommendation on the Ethics of Artificial Intelligence

TIMEFRAME	ROADMAP ACTION
April 2020	First meeting of the 24-member Ad Hoc Expert Group convened to prepare the draft text of the Recommendation.
July-August 2020	Global public online consultation and eleven regional and sub-regional virtual consultations. More than 600 submissions and 50,000 suggestions received on the first draft text.
August-September 2020	Second meeting of the Ad Hoc Expert Group to revise the draft text of the Recommendation based on outcomes of consultations.
September 2020	Draft text of the Recommendation transmitted to Member States for written comments to be received by 31 December 2020.
January-March 2021	UNESCO Secretariat prepares a final report containing the draft text of the Recommendation based on comments and observations from Member States
April 2021	Transmission of the final report containing text of the Recommendation to Member States. First session of the special committee of intergovernmental experts to prepare a final draft of the Recommendation.
June 2021	Second session of the special committee of intergovernmental experts to prepare a final draft of the Recommendation.
Mid-August 2021	Transmission of the final draft of the Recommendation by the special committee of intergovernmental experts to Member States.
Autumn 2021	41st General Conference: Examination and possible adoption of the final draft Recommendation.

Recommendation on the Ethics of Artificial Intelligence



Approved and adopted by acclamation!



Thank you so much for listening me!

Contact Information

Email: edson.prestes@ieee.org.

My homepage: <http://www.inf.ufrgs.br/~prestes>

Linkedin: <https://www.linkedin.com/in/edson-prestes/>