

# TRUSTWORTHY AI FROM PRINCIPLES TO PRACTICE

---

FEBRUARY 2022

**Dr Clara Neppel**  
Senior Director European Operations

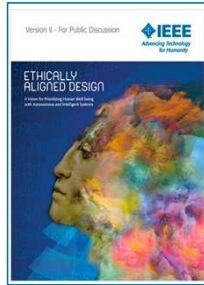
# ADVANCING TECHNOLOGY FOR HUMANITY

## ABOUT IEEE

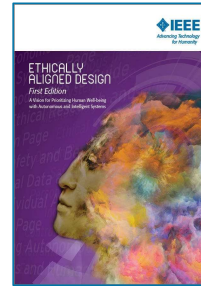
- 400,000+ Members
- 160+ Countries
- 46 Technical Societies and Councils
- 1900+ Annual Conferences
- Global Humanitarian Efforts
- Developing market relevant open standards and solutions



EAD, Version 1  
December, 2016



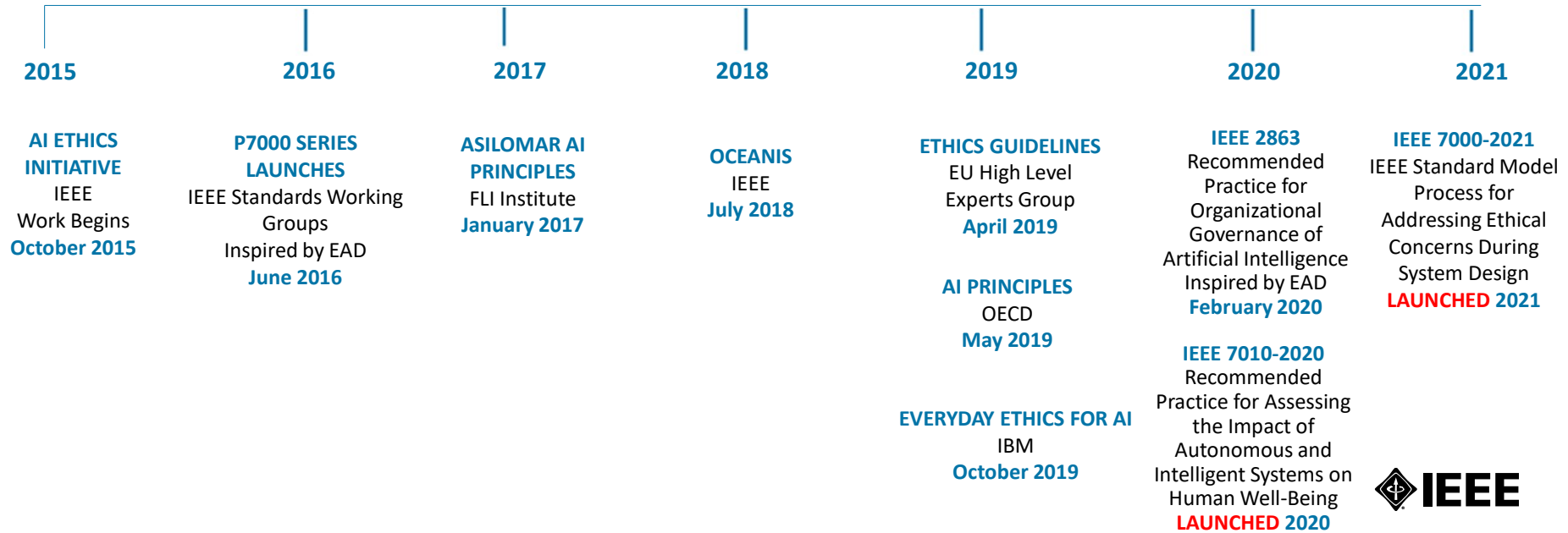
EAD, Version 2  
December 2017



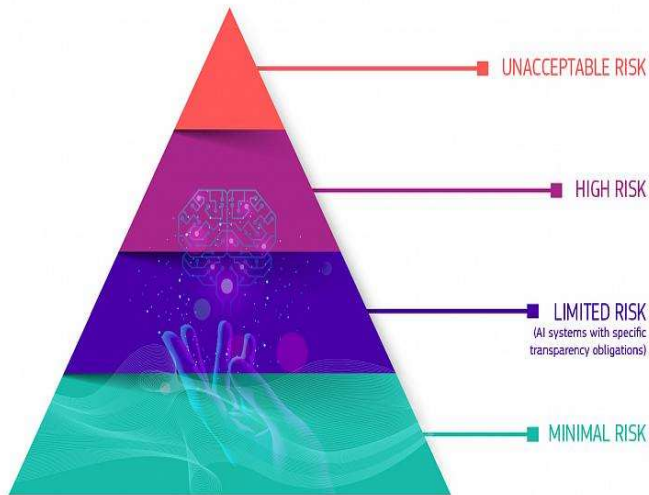
EAD, Edition 1  
March 2019



BKCenter Chart  
January, 2020



# EU AI Act



“Mandatory requirements applicable to the design and development of certain AI systems before they are placed on the market that will be further operationalised through harmonised **technical standards**”

**High Risk** - horizontal mandatory requirements and **conformity assessment**

**Limited Risk** - Transparency obligations

**Minimal Risk** - Voluntary codes of conduct



# Make the implicit explicit and verifiable for the entire AI ecosystem



## STANDARDS

Developing a growing series of standards that promote innovation, foster interoperability and honor human values.

- **Technical**
- **Socio-technical**



## CERTIFICATION

Developing metrics and processes towards the implementation of a certification methodology.

- **Transparency**
- **Accountability**
- **Algorithmic bias**



## GOVERNANCE

Support responsible Artificial Intelligence Systems innovation through proper governance mechanisms.

- **Business**
- **Cities**
- **Public bodies**

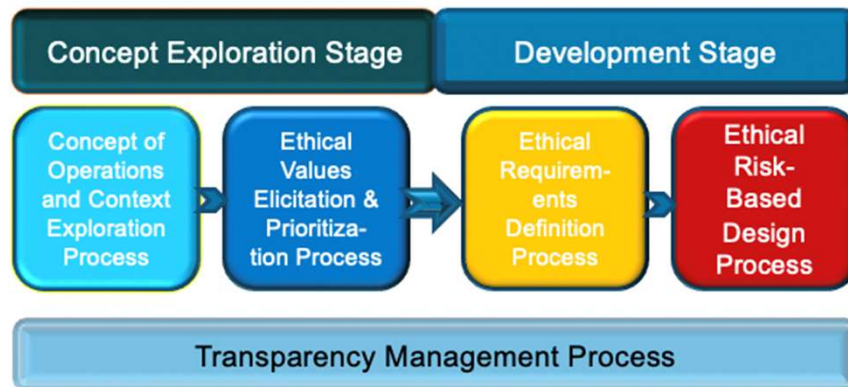
<https://standards.ieee.org/initiatives/artificial-intelligence-systems>

# Selected Projects of Interest

Completed	Under Development	New Projects	Standards+ Projects
<ul style="list-style-type: none"> <li>• IEEE 7000 - Standard for Model Process for Addressing Ethical Concerns During System Design</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P2863 - Recommended Practice for Organizational Governance of Artificial Intelligence</li> <li>• IEEE P7002-Standard for Data Privacy Process</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7015 - Data and AI Literacy, Skills, and Readiness Working Group</li> </ul>	<ul style="list-style-type: none"> <li>• <b>AI Impact Use Cases</b></li> <li>• <b>Synthetic Data</b></li> <li>• <b>Trustworthy Autonomous Driving</b></li> </ul>
<ul style="list-style-type: none"> <li>• IEEE 7001 - Standards for Transparency of Autonomous Systems</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7003 - Standard for Algorithmic Bias Considerations</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7030 - Recommended Practice for Ethical Assessment of Extended Reality (XR) Technologies</li> </ul>	
<ul style="list-style-type: none"> <li>• IEEE 7005-Standard for Transparent Employer Data Governance</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7004 - Standard for Child and Student Data Governance</li> </ul>		
<ul style="list-style-type: none"> <li>• IEEE 7007 - Ontological Standard for Ethically Driven Robotics and Automation Systems</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7008 - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems</li> </ul>		
<ul style="list-style-type: none"> <li>• IEEE 7010 - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7009 - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems</li> </ul>		
<ul style="list-style-type: none"> <li>• IEEE 2089 - IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7011 -Standard for the Process of Identifying and Rating the Trustworthiness of News Sources</li> </ul>		
<ul style="list-style-type: none"> <li>• IEEE 3652.1 - Guide for Architectural Framework and Application of Federated Machine Learning</li> </ul>	<ul style="list-style-type: none"> <li>• IEEE P7012 - Standard for Machine Readable Personal Privacy Terms</li> <li>• IEEE P7014 - Standard for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems</li> </ul>		

# IEEE 7000™- 2021

## Addressing Ethical Concerns During System Design



- Requires the active engagement of direct and indirect stakeholders.
- Requires engaging a ‘Value Lead’ who accompanies the ethical alignment of systems end-to-end and over time.
- Stresses cultural and ethical context sensitivity.
  - It does not bias value exploration through value principles lists.
  - It embraces value principles lists (such as human rights) to complete the value mission.
- Better value proposition



# The Yoma Story

## Yoma went from an AI-driven talent calculation machine to a community platform for mutual and local support of African youth

### Before (very first idea November 2019): **AI focus**

- Young peoples' data is combined and aggregated to calculate individuals' 'talent scores' with an AI engine hosted in Germany
- Young people are represented through its AI-based talent score and homogenized, comparable profiles are created
- All data providers can pull talent scores from young people.
- "Diamonds in the rough" are contacted to be channeled into innovation hubs to support African business

### After (summer 2021): **Bottom-up Youth support**

- Online marketplace with opportunities to develop skills, engage with a community and access employment
- Achievements and personal growth are verified using blockchain and added to a digital CV, which they can share with peers and employers.
- This allows youth to build an alternative trust profile, which enhances employability and allows for more informed matching with the labour market.
- Young people have maximum privacy and control and agency
- "Diamonds in the rough" become mentors for other young people to provide local community support AND can respond to African businesses if they want





IEEE offers the 7000™ standard free of charge online to read.

The screenshot shows the IEEE Standards Association website. At the top left is the IEEE SA logo. To the right is a search bar labeled "Search this website". Below the logo is a navigation menu with "Standards", "Programs & Services", "Practice Areas & Focuses", and "Get Involved". On the right side of the navigation bar is a "MAC ADDRESS" button. The main content area features two tabs: "Standard" and "Active". The title of the page is "IEEE 7000-2021 - IEEE Approved Draft Model Process for Addressing Ethical Concerns During System Design". Below the title are two buttons: "BUY THIS STANDARD" and "ACCESS VIA SUBSCRIPTION".

<https://standards.ieee.org/standard/7000-2021.html>

#### Explore This Standard

- **Standard Details**
- Additional Resources
- More

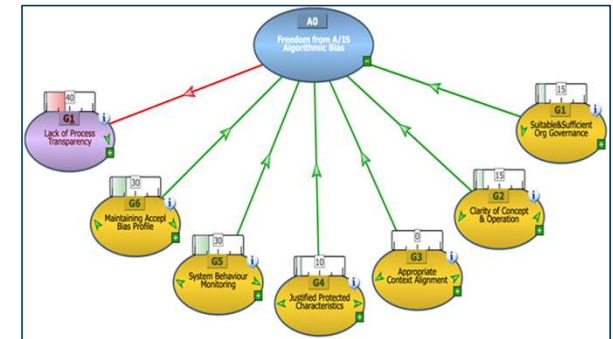
#### Standard Details

The standard establishes a set of processes by which organizations can include consideration of human ethical values throughout the stages of concept exploration and development. This standard supports management and engineering in transparent communication with selected stakeholders for values elicitation and prioritization. It involves traceability of ethical values through an operational concept, value propositions, and value dispositions in the system design. The standard describes processes that provide for traceability of ethical values in the concept of operations, ethical requirements, and ethical risk-based

# IEEE CertifAIEd

The Mark of AI Ethics

- **Certification is a basis for trust in products and services**
- **Independent Scrutiny**
  - ✓ Identifies Strengths and Shortfalls
  - ✓ Demonstrates Duty of Care
- **Adaptive and agile**
  - ✓ Assessment effort proportionate to the risk profile
  - ✓ Rapid tailoring to the needs of a sector
- **Demonstrate implementation towards the proposed EU AI regulation**





City of Vienna Earns IEEE AI Ethics Certification Mark

Towards a  
Positive Vision for  
Humanity

TRUST

