HSBOOSTER WORKSHOP 2 - AI

Ethical Assurance of Technology through Standardisation & Certification

Prof A G Hessami, IEEE CertifAIEd Vice Chair & Process Architect

04 July 2024

IEEE AI ETHICS

How can we address social responsibility in embedded autonomous intelligent systems?

A G Hessami

Overview of the Talk

- IEEE Introduction and Global Initiative

- P70xx series of AI Technology Standards

- The Ethics Certification Programme (CertifAIEd)

- Value Proposition & Way Forward

STANDARDS

ABOUT IEEE SA

Developing market relevant open standards and solutions:

Advancing global technologies and technology platforms

Promoting innovation

Protecting public safety, health & wellbeing

Contributing to a sustainable future

IEEE SA BY THE NUMBERS

1500+
STANDARDS
& PROJECTS

380+
CORPORATE
MEMBERS

7500+
INDIVIDUAL
MEMBERS

34,000+
GLOBAL PARTICIPANTS

180+
GLOBAL AGREEMENTS

Definitions

Ethics

- A branch of knowledge that deals with moral principles (that govern a person's behaviour)

Morals

- Considerations of right & wrong behaviour (acceptable in a particular society/culture)

Autonomous System

- one that can make decisions without human intervention.

Definitions

Autonomous Operations & Intrinsic Capabilities

▪ make decisions without any requirement for prior human validation of the consequences

of these decisions,

▪ act upon these decisions through interaction with the real (physical) world,

▪ adapt and modify both the decision making processes itself and the consequential

interactions with the real world through a process of learning.

Context – Strategic Technology of Future

Released
In Development
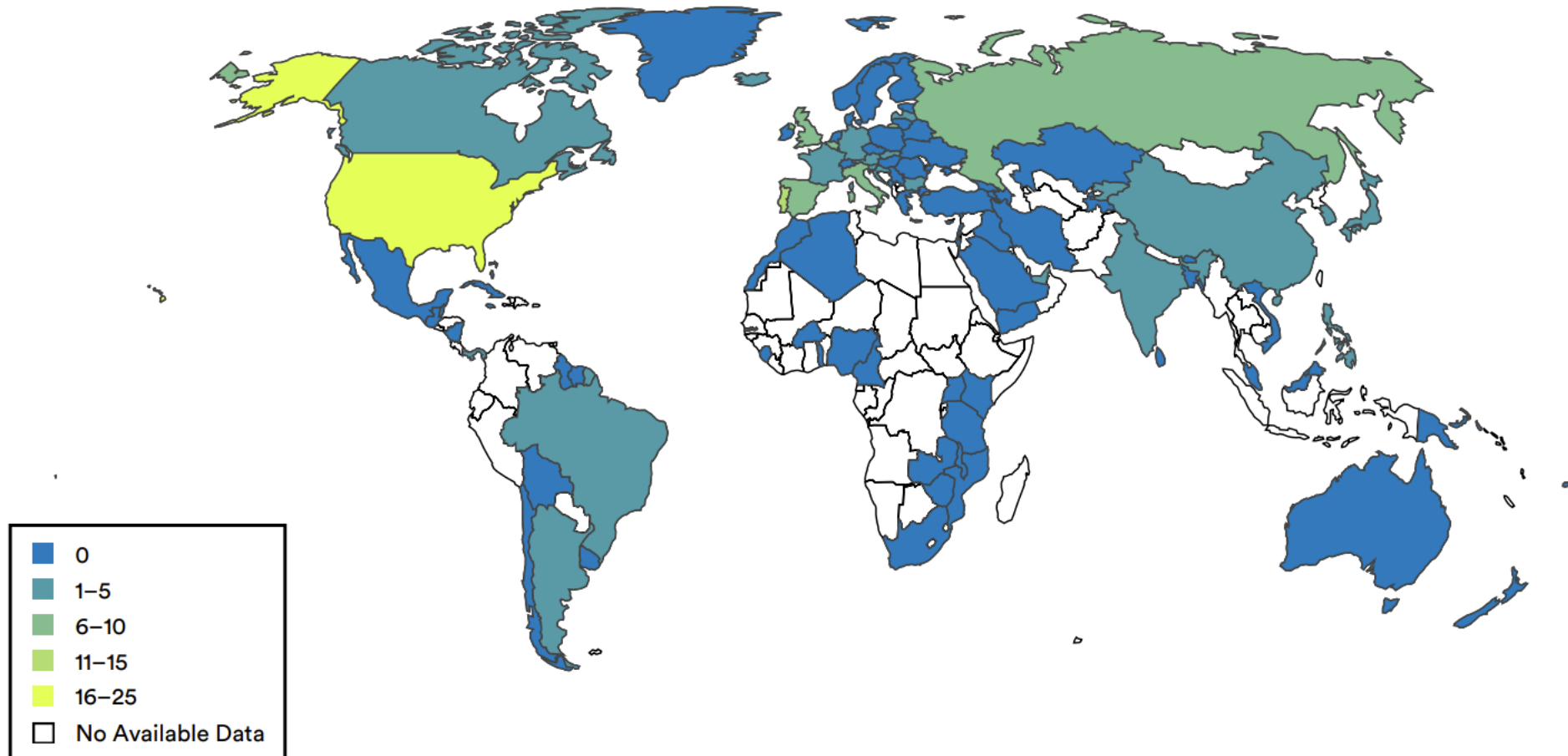Not Released
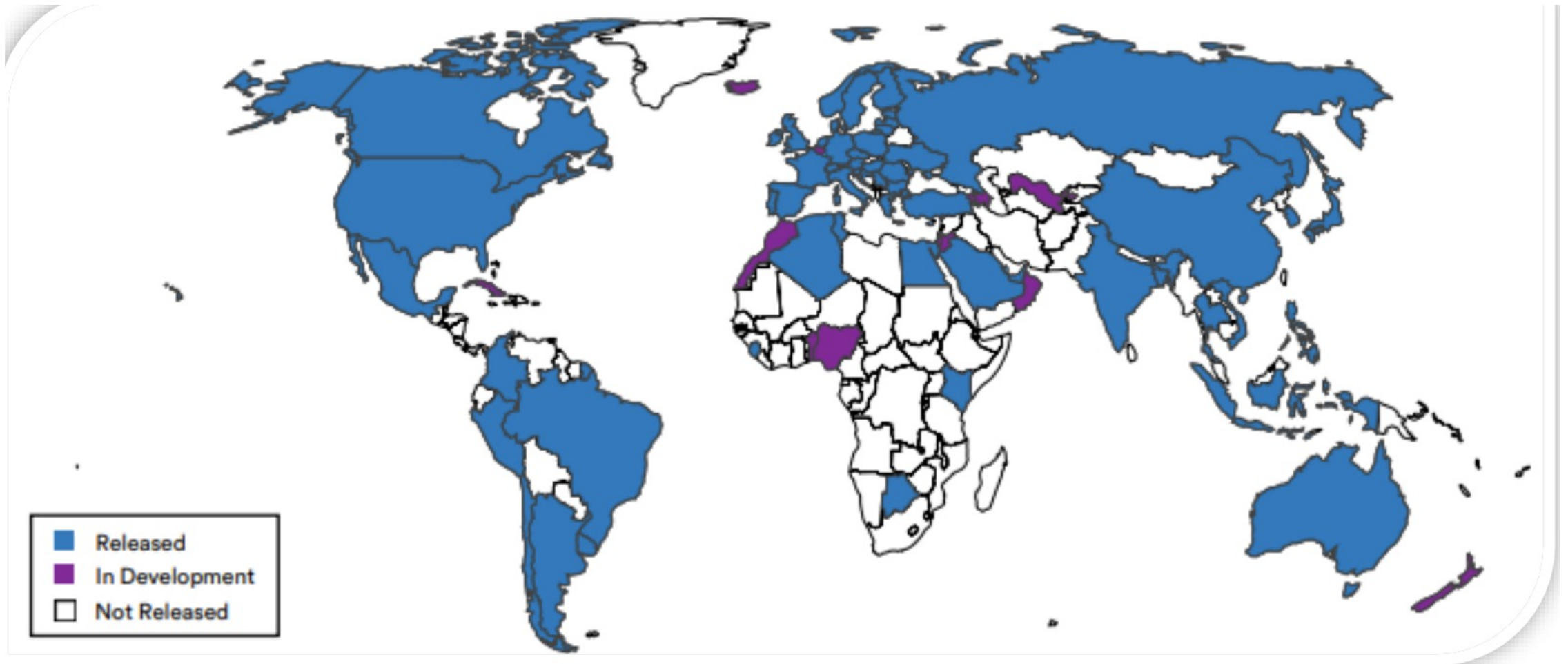
62 countries published an AI strategy

Context – AI Related Laws 2016-2022

Context

Sam Altman, the CEO of OpenAI

➢ regulators and society need to be involved with the technology to guard against negative consequences.

But despite the dangers, it could also be "the greatest technology humanity has yet developed"

Technology Ethics

I. Algorithmic Bias

II. Transparency & Explainability

III. Accountability

IV. Privacy

V. Responsible Governance

VI. Fairness

VII. Safety, Security & Dependability
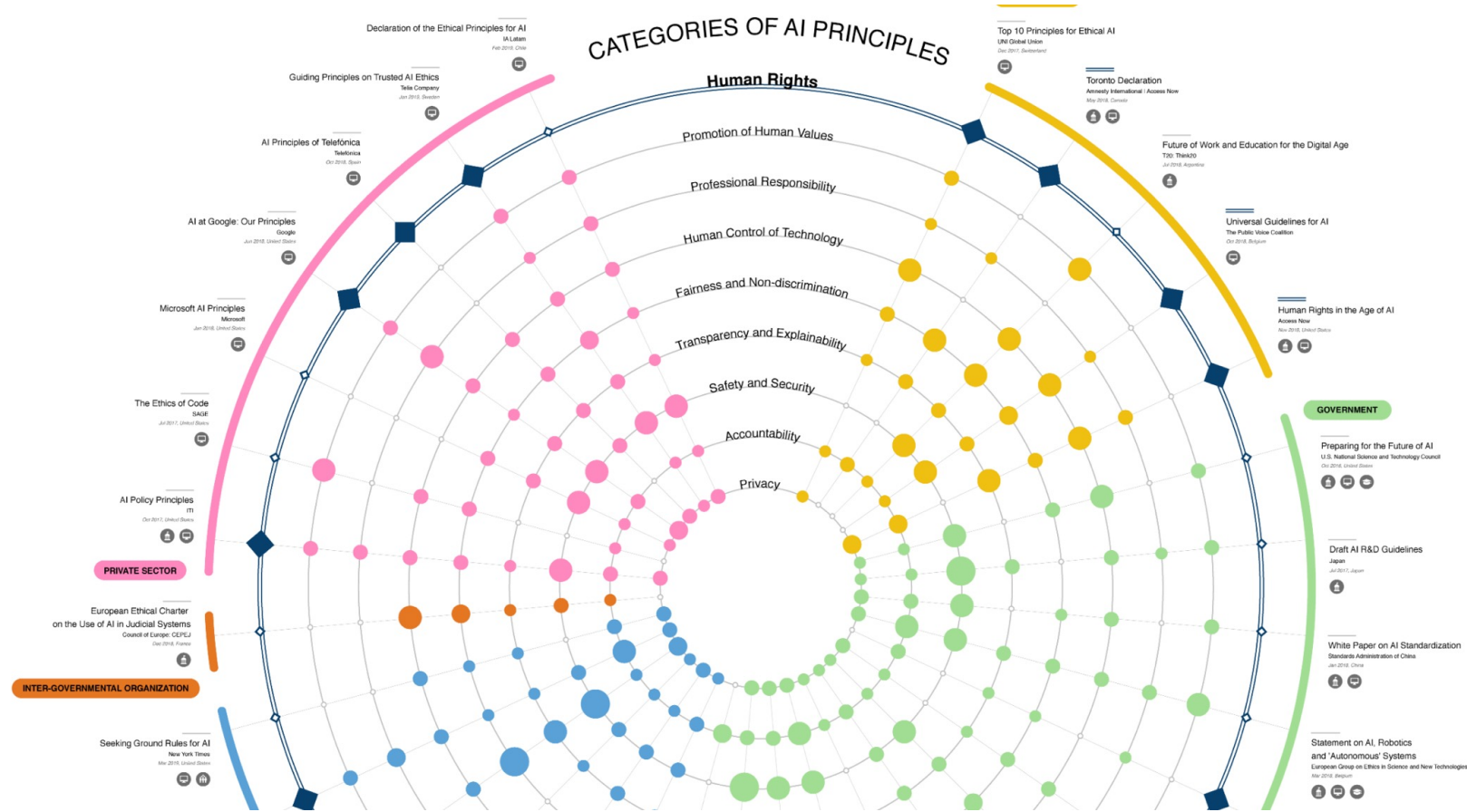
Principal Ethics Theories

➢ **Consequentialism/Utilitarianism (JS Mill)**

       Deals with Happiness & Well being

       Everyone ought to act to bring greatest happiness for greatest No. of People

➢ **Deontological/Duty (E Kant-Categorical Imperative)**

       put yourself as a universal law maker and whether the object will influence people, destroy, threaten or create values?

       Also WD Ross on prima facie duties covering Fidelity, Reparation, Gratitude, Promoting Max Good & non-Maleficence

       Top Management's Personal Ethics impacting on decisions to align with stakeholders' expectations

➢ **Virtue Ethics (Aristotle)**

    ➢ What stakeholders are affected? Virtues are character qualities borne by persons

    ➢ How is virtuous behavior impacted?

    ➢ Vice is opposite to virtue but core principle is to "be good" hence Virtue

IEEE SA IMPACT STANDARDS

AIS ETHICS STANDARDS

- **IEEE 7000-2021™ -** Standard for Model Process for Addressing Ethical Concerns During System Design

- **IEEE 7001-2021™ -** Standards for Transparency of Autonomous Systems

- **IEEE 7002-2022™ -** Standard for Data Privacy Process

- **IEEE P7003™ -** Standard for Algorithmic Bias Considerations

- **IEEE P7004™ -** Standard for Child and Student Data Governance

- **IEEE 7005-2021™ -** Standard for Transparent Employer Data Governance

- ~~**IEEE P7006™ -** Standard for Personal Data Artificial Intelligence (AI) Agent~~

- **IEEE 7007-2021™ -** Ontological Standard for Ethically Driven Robotics and Automation Systems

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

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

- **IEEE 7010-2020™ -** IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being

- **IEEE 7010.1™ -** IEEE Recommended Practice for Environmental Social Governance (ESG) and Social Development Goal (SDG) Action Implementation and Advancing Corporate Social Responsibility

- **IEEE P7011™ -** Standard for the Process of Identifying and Rating the Trustworthiness of News Sources

- **IEEE P7012™ -** Standard for Machine Readable Personal Privacy Terms

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

Red font indicates approved standard.

The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)

Developing metrics and processes towards the implementation of a certification methodology addressing transparency, accountability and algorithmic bias

The Ethics Certification Program for Autonomous and Intelligent Systems (CertifAIEd) 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. CertifAIEd intends to offer a process and define a series of marks by which organizations can seek certifications for their processes around the AIS products, systems, and services they provide.

For More Information:
https://engagestandards.ieee.org/ieeecertifaied.html

The Ethics Certification Program for Autonomous Intelligent Systems

Key Objectives

1) Develop SMART Criteria for Assessment of Ethical Properties

2) Focus on Ethical:

   o Accountability;

   o Algorithmic Bias;

   o Transparency;

   o Privacy;

   o Fairness;

   o Governance.

3) Develop Eco-system for Conformity Assessment & Certification

The Ethics Certification Program for Autonomous Intelligent Systems

Current Status

➢ The three original ECPAIS Certification Criteria suites finalized

➢ Pilot and Tailoring Projects in Community Services, Financial Services, ...

➢ Criteria suites for **Ethical Privacy & Responsible Governance** added

➢ Conformity Assessment Tools and Methodologies

  o Product/Service Ethics Risk/Reward Profiling
  o Ethical Issues Register
  o Case for Ethics
  o Assessment and Ethics Progression Report
  o Eco-system Competence Criteria, Assessment & Management System

CASE FOR ETHICS
ECPAIS TRANSPARENCY

DATE

PRODUCT/SYSTEM NAME

CLIENT ORGANISATION NAME

DIGITAL
ETHICS

AG Hessami

I1D1-Sept. 2020

The Ethics Certification Program for Autonomous Intelligent Systems

The Training Programme

I.   Tech Ethics & ECPAIS Certification for Senior Enterprise Executives
  ❖   *½ Day Tech Ethics scene setting & Overview of IEEE Services*

II.   Tech Ethics & ECPAIS Certification for Technology & Design Experts
  ❖   *3 Days in depth ECPAIS Ethics Certification & Satisfaction Criteria*
  ❖   *Certificate as ECPAIS Accredited Assessor*

III.   Tech Ethics & ECPAIS Certification for Advisors and Consultants
  ❖   *1 Day EAD Overview and ECPAIS applications*
  ❖   *Certificate as ECPAIS Accredited Advisor*

IEEE CertifAIEd™ Ontological Specifications - Ethical Privacy

A contextual set of values pertaining to privacy and satisfaction of a framework of expectations
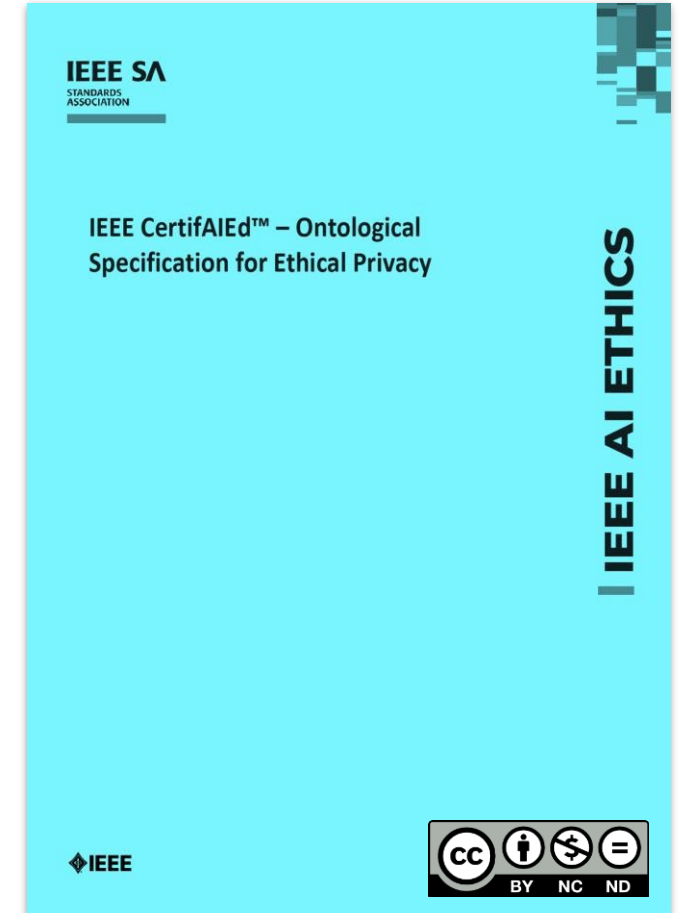
Principal Drivers & Inhibitors

Driver Influences

- Organizational Governance, Capabilities & Maturity
- Clarity and Consistency of AIS Operations
- Ethical Architecture, Design and Dev. Of AIS
- Human Oversight and Enforcement in AIS
- End user Awareness of AIS & Empowerment
- Maintaining Ethical Privacy Integrity
- (Ethical) Decommissioning

Inhibitory Influences

- Overreaching and Overfitting
- Authoritarian and Compulsory Pressures
- Accidental/Incidental Exposure
- Malicious Exposure
- Systemic Vulnerability

The IEEE Ontological Specifications for Ethical Privacy is being released using a Creative Commons license
https://engagestandards.ieee.org/ieeecertifaied.html

IEEE SA
STANDARDS
ASSOCIATION

IEEE CertifAIEd™ – Ontological
Specification for Ethical Privacy

IEEE AI ETHICS

IEEE CertifAIEd™ Ontological Specifications - Algorithmic Bias

Automated recommendations and predictions that disproportionately favour a stakeholder entity over another.

Principal Drivers & Inhibitors

Driver Influences

- Organizational Governance, Capabilities & Maturity
- Clarity of AIS Operations
- Context Alignment
- Justified Use of Protected Characteristics
- System Behavior Monitoring
- Maintaining (Acceptable) Bias Profile

Inhibitory Influence

- Lack of Process Transparency

IEEE SA
STANDARDS
ASSOCIATION

IEEE CertifAIEd™ – Ontological
Specification for Ethical Algorithmic Bias

IEEE AI ETHICS

◆IEEE

IEEE CertifAIEd Ontological Specifications - Ethical Transparency

A contextual set of values pertaining to transparency and the satisfaction of a framework of expectations

Principal Drivers & Inhibitors

Driver Influences

- ◼ Organizational Governance, Capabilities & Maturity
- ◼ Clarity and Consistency of AIS Operations
- ◼ Awareness of AIS Interaction
- ◼ Confidence in System Behaviour
- ◼ Accessible and Fair Control & Feedback
- ◼ Upholding Ethical Transparency Integrity

Inhibitory Influences

- ◼ Behavioural Obfuscation
- ◼ Concern with Liability
- ◼ Protection of Trade Secrets

The IEEE Ontological Specifications for Ethical Transparency is being released using a Creative Commons license
https://engagestandards.ieee.org/ieeecertifaied.html

IEEE SA
STANDARDS ASSOCIATION

IEEE CertifAIEd™ – Ontological
Specification for Ethical Transparency

IEEE AI ETHICS

IEEE CertifAIEd Ontological Specifications-Ethical Accountability

A contextual set of values pertaining to accountability and the satisfaction of a framework of expectations concerning responsibility taking for actions, omissions and ethical outcomes
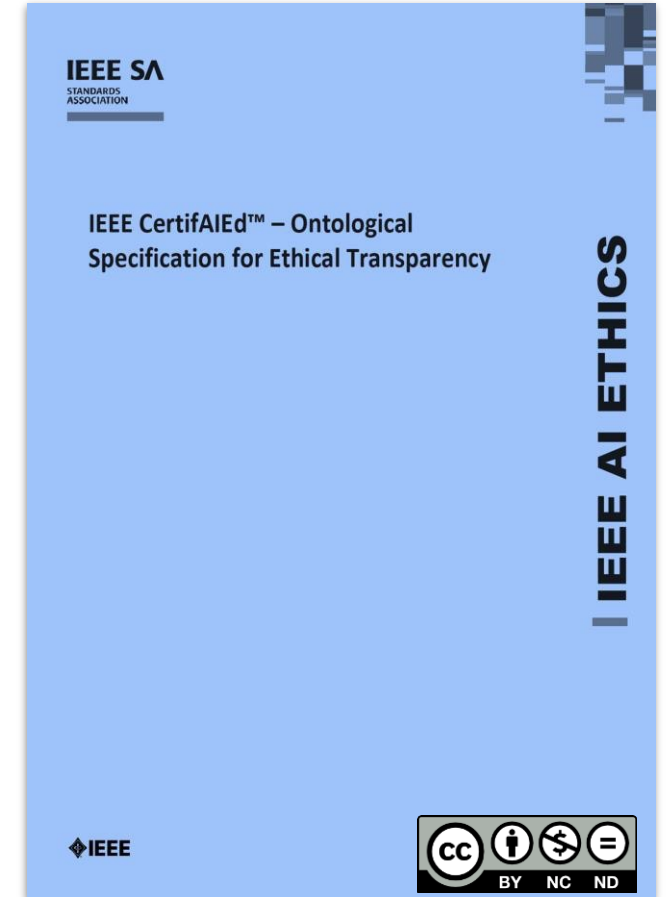
Principal Drivers & Inhibitors

Driver Influences

- Organizational Governance, Capabilities & Maturity
- Clarity and Consistency of AIS Operations
- Human Oversight
- User Interactions
- Upholding Ethical (Accountability) Profile

Inhibitory Influences

- Random and Systematic Errors
- Rubber Stamping
- Inadequate or Non-Existent Records

IEEE SA
STANDARDS ASSOCIATION

IEEE CertifAIEd™ – Ontological Specification for Ethical Accountability

IEEE AI ETHICS
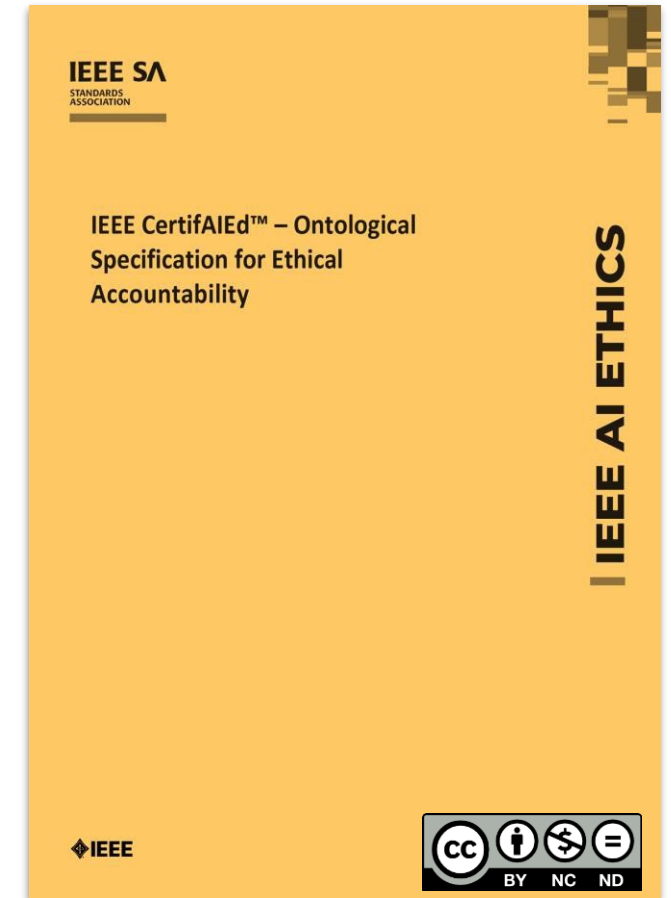
◆IEEE

IEEE CertifAIEd Ontological Specifications - Responsible Governance

Human centric, considerate, and enduring collaborative governance

Principal Drivers & Inhibitors

Driver Influences

- Long Term Social and Natural Capital Driven Strategy
- Sustainable Reinvestment
- Aligned Incentives
- Legal, Ethical and Societal Impact of AIS
- Nurturance
- Empowered Accountability
- Ongoing Stakeholder Participation & Collaboration
- Ongoing Oversight and Redress

Inhibitory Influences

- Fiduciary Norms
- Poor Public Mandate
- Partisanship
- Competent Resource Constraints
- Perception or Actual Apathy and Powerlessness

IEEE SA STANDARDS ASSOCIATION

IEEE CertifAIEd™ – Ontological Specification for Responsible Governance of AIS

IEEE AI ETHICS

◆IEEE

The Agile Ethics Criteria Profiling for Sector Applications

IEEE CertifAIEd: The Process

STEP 1
ENQUIRY

STEP 2
ETHICAL PROFILING

STEP 3
ASSESSMENT

STEP 4
CERTIFICATION AND MARK

IEEE's Ethical Assurance Global Ecosystem

Regulation through Complete Eco-system

- Suite of IEEE70xx Standards
- Suite of Principal Ethics Characterisation Criteria
- An Agile Ethics Profiling Process
- An Assessment & Certification Process
- A Comprehensive AIS Ethics Assessor Training
- Strategic Partnership with Professional Trainers
- A Comprehensive Practitioner Competence Regime
- A pool of lobal AIS Ethics Assessors
- Strategic Partnership with Certification Bodies

- **IEEE 7000-2021™** - Standard for Model Process for Addressing Ethical Concerns During System Design
- **IEEE 7001-2021™** - Standards for Transparency of Autonomous Systems
- **IEEE 7002-2022™** - Standard for Data Privacy Process
- **IEEE P7003™** - Standard for Algorithmic Bias Considerations
- **IEEE P7004™** - Standard for Child and Student Data Governance
- **IEEE 7005-2021™** - Standard for Transparent Employer Data Governance
- ~~IEEE P7006™ - Standard for Personal Data Artificial Intelligence (AI) Agent~~
- **IEEE 7007-2021™** - Ontological Standard for Ethically Driven Robotics and Automation Systems

- **IEEE P7008™** - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
- **IEEE P7009™** - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
- **IEEE 7010-2020™** - IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being
- **IEEE 7010.1™** - IEEE Recommended Practice for Environmental Social Governance (ESG) and Social Development Goal (SDG) Action Implementation and Advancing Corporate Social Responsibility
- **IEEE P7011™** - Standard for the Process of Identifying and Rating the Trustworthiness of News Sources
- **IEEE P7012™** - Standard for Machine Readable Personal Privacy Terms
- **IEEE P7014** - Standard for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems

Regulation

Daily Comment

Congress Really Wants to Regulate A.I., But No One Seems to Know How

Yet another hearing—this one with OpenAI's Sam Altman—has come after a new technology with the possibility to fundamentally alter our lives is already in circulation.

The New Yorker, May 2023

UK AI Strategy

➢ Core Concepts in AI Regulation, Adaptability & Autonomy

➢ UK regulation Overarching Principle,

  ➢ Ensuring AI is used safely.

  ➢ Ensuring AI is technically secure, and functions as designed.

  ➢ Ensuring AI is appropriately transparent and explainable.

  ➢ Embedding fairness into AI.

  ➢ Responsible governance.

  ➢ Ensuring clarity of redress or contestability.

The Ethics Certification Program for Autonomous Intelligent Systems
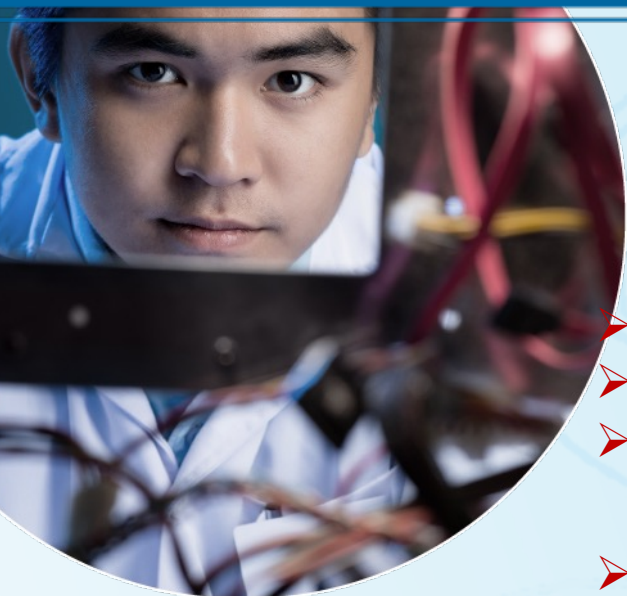
Ethics Certification Value Proposition

I.   Uncertainties with AIS implementations require a Precautionary Approach

II.   Fundamental Human Rights as well as Ethical Values can be undermined by AIS

III.   Responsible Innovation & Governance demand Ethical and CSR alignment

IV.   Independent Competent scrutiny demonstrates verified Duty of Care

V.   Independent Scrutiny identifies Strengths and Shortfalls

VI.   Structured & Comprehensive suites of Criteria enhance integrity & scope of scrutiny

VII.   CertifAIEd provides the mechanisms for enhanced confidence in the societal outcomes

VIII.   Certification is a basis for the Public & Civil Society trust in products and services
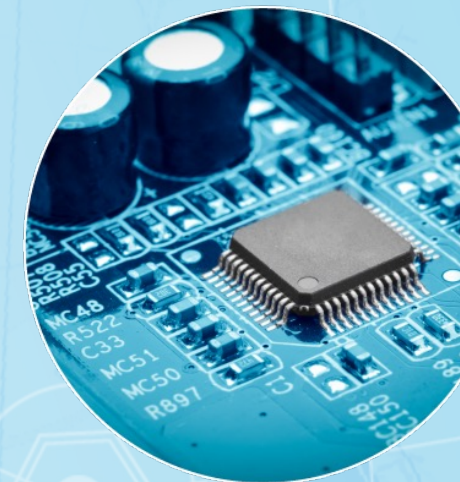
Participating to IEEE CertifAIEd

- If you are interested in joining the IEEE CertifAIEd ecosystem in your capacity as an assessor, a certification organization, a training institute, or otherwise, please reach out to us by filling the form at https://engagestandards.ieee.org/ieeecertifaied.html

- If you are interested in learning about CertifAIEd, please contact us at certifaied@ieee.org

- To demonstrate your commitment to AI Ethics, you can go to https://engagestandards.ieee.org/IEEE-AI-Ethics-Support-Badge.html

Thank you!

- https://ethicsinaction.ieee.org/
- https://engagestandards.ieee.org/ieeecertifaied.html
- https://beyondstandards.ieee.org/new-creative-commons-paper-addresses-ethical-hurdles-to-contact-tracing-adoption/

- https://engagestandards.ieee.org/ieee-7000-2021-for-systems-design-ethical-concerns.html

- https://standards.ieee.org/ieee/7000/6781/
- https://beyondstandards.ieee.org/the-ieee-certifaied-framework-for-ai-ethics-applied-to-the-city-of-vienna/
- https://standards.ieee.org/initiatives/autonomous-intelligence-systems
- https://standards.ieee.org/initiatives/artificial-intelligence-systems/

Ali Hessami
VC & Process Architect, IEEE AI Ethics Certification Programme
A.G.Hessami@IEEE.Org

thank you