Engineering Technology Open Access Journal

Juniper Publishers key to the Researchers

Research Article
Volume 5 Issue 4 - October 2023
DOI: 10.19080/ETOAJ.2023.05.555668

Eng Technol Open Acc

The IEEE-SA CertiFAIEd Ethics Certification Framework for Autonomous Intelligent Systems(AIS) Enabling Fully Integrated and Coordinated Ethical Implementations of AIS – An Illustration Through a Proposed Implementation of the IEEE7007-2021 Ontological Standard for Ethically Driven Robotics and Automation Systems

Zvikomborero Murahwi*

University of Johannesburg, Auckland Park, Johannesburg, South Africa

Submission: June 30,2023; **Published:** October 30, 2023

***Corresponding author:** Zvikomborero Murahwi, University of Johannesburg, UJ Campus, Bunting Rd, Auckland Park, Johannesburg, 2092, South Africa, Email id: zviko.murahwi@ictprojectsadvisory.com

Abstract

We are currently experiencing a rapid increase in adoption and implementation of Artificial Intelligence enabled systems, products and services which traditionally was driven by a desire to achieve higher financial returns and increased operational efficiency for competitive advantage. Because of some serious harms and risks associated with some AI implementations, we are also now seeing a shift towards serious considerations of human safety and values in AI implementations. Standardization, regulation and policy development are increasingly becoming important to support and facilitate a balanced incorporation of values in AI Systems and Services implementations so as to control and even eliminate the any harms and risks (in AI implementations). Whilst a lot has been achieved in developing standards, regulation and policy documents, there are still challenges in putting these into action because of inadequacies in existing frameworks and supporting technologies. This article introduces IEEE-SA's Certified Ethics Framework for Autonomous Intelligent Systems and illustrates how the framework provides a platform for seamless fully integrated implementations of Ethically aligned AI Systems, Services and Products. IEEE-SA's award winning IEEE7007-2021 [1] Ontological Standard for Ethically Driven Robotics and Automation Systems is used as an example to illustrate how the Certified framework would adequately facilitate and support implementations of AI systems based on this standard. The writer will also show how the IEEE Certified framework derives its strengths in its ability to be adapted to support most standards, policy and even regulation implementations in any settings where ethical considerations are key. It can also be proved that the Certified framework can adequately support even those settings where Ethical considerations are absent.

Keywords: IEEE-SA; IEEE; Artificial Intelligence (AI); Autonomous Systems; Autonomous Intelligent Systems (AIS); Measurement; Design

Introduction

This paper has been produced independent of the IEEE and is based on the writer's own knowledge and experience gained, following some intense training in the Certified assessment and application, and has been followed by practice leading to certification as Certified AI Ethics Assessor in the Programme. The IEEE-SA Certified Programme has developed sets of criteria for determining conformance of Autonomous Intelligent Systems (AIS) in the domains of Accountability, Transparency, Privacy, Algorithmic Bias (elimination), and most recently Governance of AI. Whilst the Criteria have been developed with the primary goal of addressing concerns with Ethical issues around these domains and beyond, the model based and data centric approach to the

development of these criteria facilitates expeditious adaptation and verification of a product providing explanations and detail which may be required to explain activity, actions and results / outputs from AIS and related services in many application settings including Financial Services, Healthcare and Security Services. The Certified criteria can also be adapted to work with other frameworks in the design and development of AIS. This paper gives an overview of the IEEE-SA Certified framework criteria for Transparency, Accountability, Privacy and Algorithmic Bias and in doing this further demonstrates how these can be adapted for use in a specific system setting in a specified context of use or operation, including situations where there are very specific requirements.

Introducing IEEE-Sa Certified Framework Criteria

An overview of the IEEE-SA Framework Criteria Suites

For purposes of Ethical verification and certification of AIS, IEEE-SA have developed sets of criteria for Transparency, Accountability, Algorithmic Bias, and Privacy assessment. Criteria for Governance of AI have just been completed but are yet to be published. The criteria are specially designed and structured in a way that enhances integrity and widens the scope of scrutinizing any AIS or AI product. The requirements for satisfying criteria are grouped into Normative and Instructive (Informative) classes. Satisfying Normative requirements is mandatory. A criterion has one or two levels of decomposition, where a level is defined through a Goal. A Goal can be a Driver or Inhibitor. A Driver positively impacts a criterion, and an Inhibitor negatively impacts the criterion. Goals – influence decision making and are defined in consultation with subject matter experts. For a Goal

i. Ethical Foundational Requirements to achieve the goal/ factor are defined.

ii. Ethical Foundational Requirements are classed as Normative or Informative.

iii. The level at which a requirement is assessed is specified (The level can be Low Impact, Medium Impact or High Impact)

iv. The evidence that should be provided to prove that the Ethical Foundational Requirement has been met is defined.

v. Method of measuring how well evidence provided meets the standard requirement(s) is defined (Can be binary or scalar (e.g., 1-5))

vi. Duty Holders to drive implementation of the requirement(s) are identified and specified. (Criteria currently identify with the following duty holders: Regulators, Developers, System Integrators, System Operators and System Maintainers)

(Figure 1) is an illustration of the above concepts, in (Figure 2) is example criteria for Transparency, and (Figure 3) criteria for Algorithmic Bias (Similar criteria exist for Accountability, Privacy and Governance of AI)

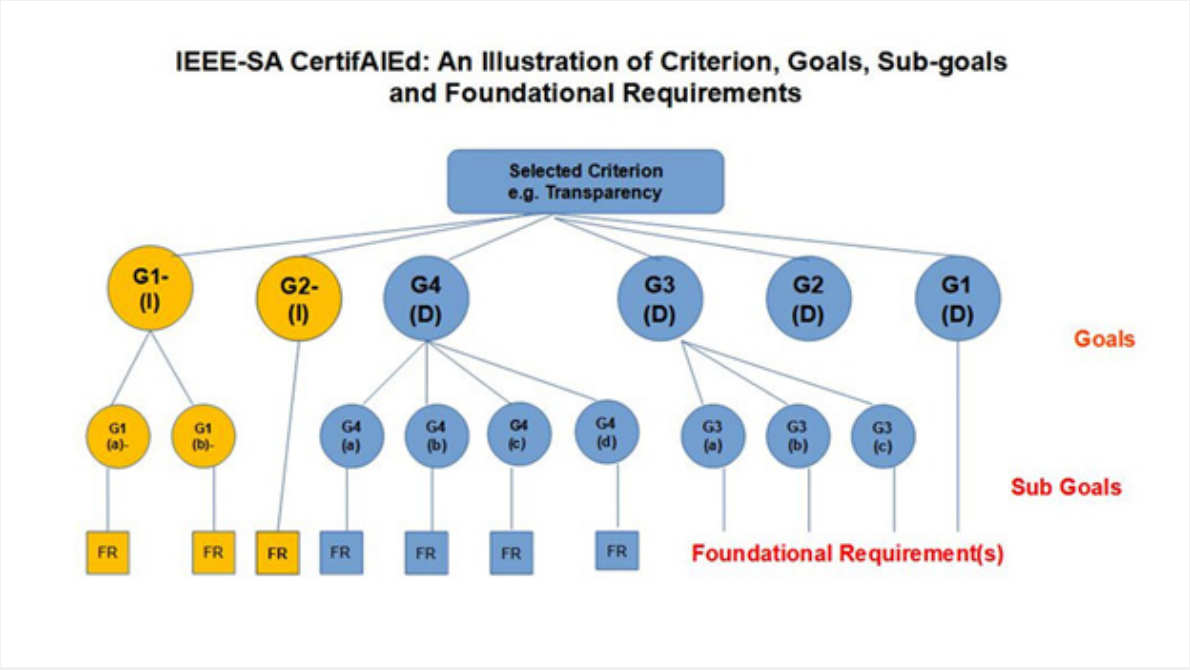

Figure 1: Illustrating CertifAEId Framework Criteria.

How to cite this article: Zvikomborero M. The IEEE-SA CertiFAIEd Ethics Certification Framework for Autonomous Intelligent Systems(AIS) Enabling Fully Integrated and Coordinated Ethical Implementations of AIS – An Illustration Through a Proposed Implementation of the IEEE7007-2021 Ontological Standard for Ethically Driven Robotics and Automation Systems. Eng Technol Open Acc 2023; 5(4): 555668. DOI: 10.19080/ETOAJ.2023.05.555668

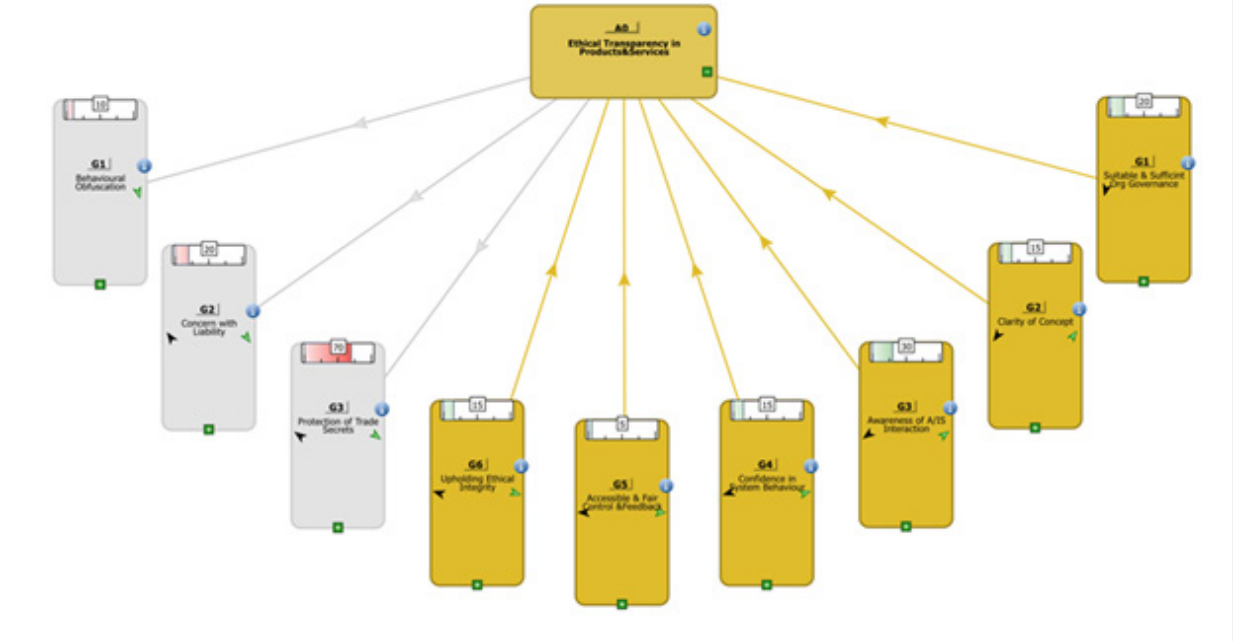

Figure 2: Transparency Criteria Schema, Goals: Drivers (gold), and Inhibitors(gray)).

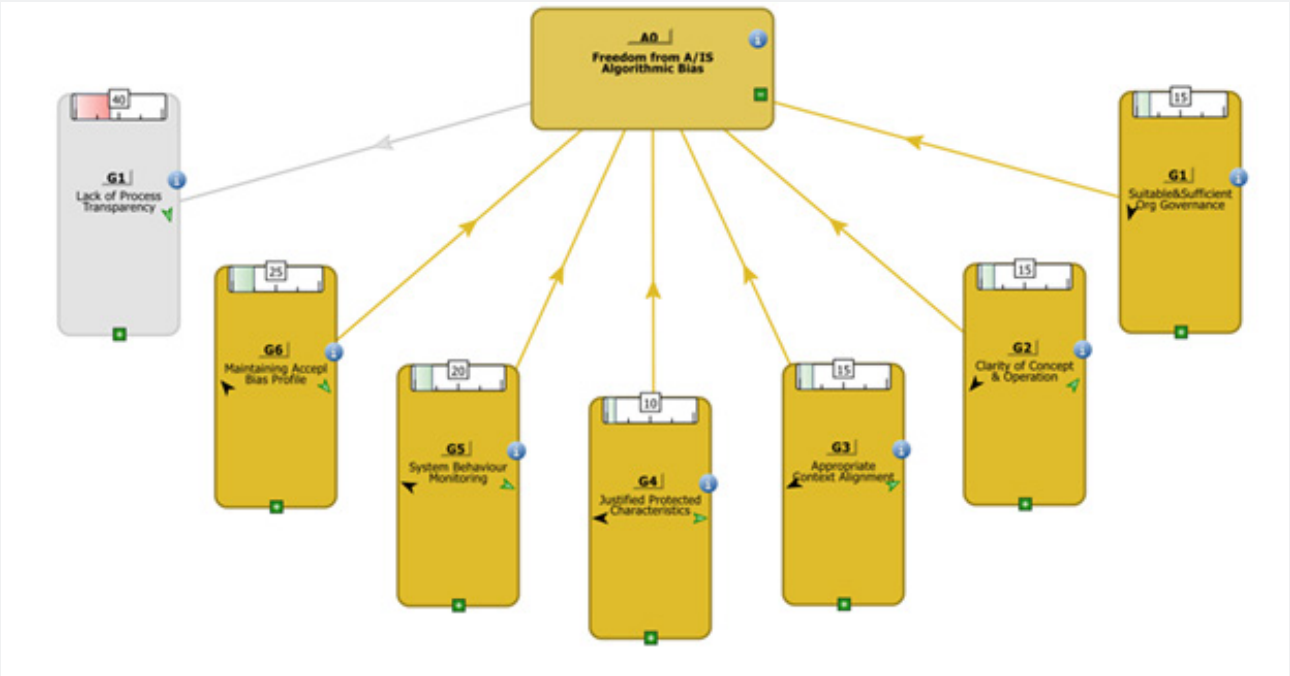

Figure 3: Algorithmic Bias Schema Goals (Drivers(gold) and Inhibitors(gray)).

Determination of Criteria Suite(S) to Use

In a given situation, the criteria to use is determined from a profiling exercise during which relevant stakeholders through existing documentation or through a new analysis exercise determine the Concept of Operations (ConOps), and Context of Operations of the AIS or AI product to identify key issues and risks which need to be addressed during development and operation or use of the System of Interest (SOI).

How to cite this article: Zvikomborero M. The IEEE-SA CertiFAIEd Ethics Certification Framework for Autonomous Intelligent Systems(AIS) Enabling Fully Integrated and Coordinated Ethical Implementations of AIS – An Illustration Through a Proposed Implementation of the IEEE7007-2021 Ontological Standard for Ethically Driven Robotics and Automation Systems. Eng Technol Open Acc 2023; 5(4): 555668. DOI: 10.19080/ETOAJ.2023.05.555668

Applying the Selected Suite(S) of Criteria

The graphical criteria suite is developed through creative workshops with subject matter experts, and subsequently converted to a matrix of Foundational Requirements. The Matrix allows for detailed specification of a goal, the goal's Foundational Requirement(s), specification of the Foundational requirements as Normative or Informative, Foundational Requirement application level as low, medium or high impact, Duty Holder or Stakeholder responsible for implementing Foundational Requirement, evidence that shall be produced to show that a Foundational Requirement has been fulfilled, Measurement or Rating on a given scale (say 1 to 5) of the level of satisfaction with the evidence provided. These specifications are summarized in (Table 1).

In (Table 2) are example specifications.

Table 1: Selected Criteria Suite Matrix Specification.

Criteria Suite Item	Specific Use Case (e.g. Transparency)
Goal	Use Case Goal
Foundational Requirement (FR)	Use Case Foundational Requirement
FR Normative or Instructive	Normative or Instructive
Certification Level	Low Impact/Medium Impact/ High Impact
Duty Holder	Regulator/Developer/
Acceptable Evidence For FR implementation	UC FR Evidence
Evidence Measurement	Scaled / Binary

Table 2: Example Matrix Extract.

Use Case Goal	Use Case FR	Use case Normative or Informative	Use Case Certification Level	Duty Holder	Use Case FR Evidence	Evidence Measurement or Rating
Transparent System Op	The organization to ensure system actions are recorded	Normative	HI	Developer	Audit trailed financial transactions	5-Fully Documented Audit Trails 4–Detailed but not well documented Audit Trail 3-Audit trail hard to followed 2-Audit trail has no detailed 1-No Audit trail

Implementation of requirements, is currently through 5 main Duty Holders as follows:

i. **Regulators**: For Governance related Governance Goals and Requirements.

ii. **Developers:** For Systems, Data and Software Engineering related Goals and Requirements.

iii. **Integrators:** For trans-functional and systems and data Goals and Requirements.

iv. **Operators**: For System Operations Goals and Requirements.

v. **Maintainers:** For all Systems, Data and Software maintenance related Goals and Requirements.

t`12

Criteria Suite Implementation for verification and certification purposes is currently through 3 main Stakeholders as follows (They would normally check work done by the above Duty Holders):

i. **Developers**: an organization developing a new system using internal or external resources, or a software developer developing software for a client.

ii. **Assessors:** Trained to do or lead assessments for verification and certification purposes.

iii. **Certifiers:** Provide independent verification and certification from the work done by an assessor.

Proposed Application of the IEEE-Sa Certified Framework and Criteria Suites to IEEE7007 -2021 Based Systems, Products and Services

Adapting the Framework Criteria Suites

IEEE 7007-2021 [1] defines four (4) Ontologies for Ethically Driven Robotics and Automation Systems [2] in 4 Domains as follows:

i. Norms and Ethical Principles which formalizes aspects of ethical theories and principles that characterize the norms of expected behaviours for norm aware agents and autonomous systems.

How to cite this article: Zvikomborero M. The IEEE-SA CertiFAIEd Ethics Certification Framework for Autonomous Intelligent Systems(AIS) Enabling Fully Integrated and Coordinated Ethical Implementations of AIS – An Illustration Through a Proposed Implementation of the IEEE7007-2021 Ontological Standard for Ethically Driven Robotics and Automation Systems. Eng Technol Open Acc 2023; 5(4): 555668. DOI: 10.19080/ETOAJ.2023.05.555668

ii. Data Protection and Privacy which formalizes relevant concepts and relationships characterizing the data protection and privacy rules and regulations that shall be observed and upheld by ethical agents and autonomous systems.

iii. Transparency and Accountability which formalizes the concepts and relationships necessary to enable ethical autonomous systems with capabilities to provide informative explanations for plans and associated action.

iv. Ethical Violation Management which formalizes concepts and relationships associated with capabilities to detect, assess, and manage ethical violations in autonomous system behaviour. In addition to ethical violation conceptualizations, this subdomain also addresses concepts and relationships governing accountability, responsibility, and legal notions of personhood for agents.

For more detail about and/or to obtain a copy of the IEEE7007-2021 [1] Standard, please visit the IEEE Standards Association website. It is proposed that each domain is handled separately. The Certified and its criteria suites would be customized to align with terminology and specific requirements, including derivations for Duty Holders, of each one of the 4 Ontological Domains.

Adapting Foundational Requirements Specification of the Selected Criteria Suite

The current approach and method can be adopted as is, but some items such as Measurement or Rating of evidence can be adapted e.g., in binary situation where it's a "YES" or "NO".

Adapting Implementation Structures, Roles and Responsibilities: Stakeholders and Duty Holders

This is where there would be most of the variations as the currently defined roles are for delivering verification and certification. The following would be potential adaptations/ variations:

i. Variations can be made for specific inclusion of Testers to the team of Developer, Assessor and Certifier.

ii. The Duty Holder classes could be changed slightly to have Regulator, Developer, Implementer, Operator and Maintainer with the Implementer taking up the role of Integrator as well.

What Would be the Benefits of Adopting the Framework and its Criteria Suites?

i. The IEEE-SA Certified framework potentially provides a fully integrated platform for addressing issues which may arise at any level: from governance through development right down to operations and maintenance. It creates some common and shared understanding of the System(s) of Interest.

ii. The framework puts traceability to between origin of, occurrence of, and solutions to issues.

iii. The framework enables coordination in resolving issues. Stakeholder relations (internal and external) can be adequately addressed [3].

iv. By simultaneously considering Drivers and Inhibitors, the framework and its criteria suites helps to put more substance into explaining situations, and decisions and actions taken in any situation, bearing in mind that one would normally want to strengthen a driver, and minimize an Inhibitor.

Conclusion

i. Autonomous Intelligent Systems (AI) are here and their use in many aspects of our lives is expected to grow exponentially.

ii. Regulation, Policy and Standards like IEEE7007-2021 [1] for critical life enabling, life serving and lifesaving systems will increasingly become important to regulate the use of AI and control or even eliminate potential risks and harms.

iii. Frameworks like IEEE-SA' Certified for Ethics Certification of AIS which facilitate and support standardization and regulation in AIS must be developed, matured, maintained and made available so that the benefits of AIS are realized and appreciated by All.

By providing a framework and platform for thorough scrutiny of AIS or any product, system or service, the IEEE-SA Certified framework and its suite of criteria provide capability and support for fully integrated implementations of IEEE7007-2021 [1] and similar standards and regulation in engineering, implementing and operating AIS. The framework and the approach to its implementation could soon become an important component in engineering systems, products and services in the age of AI.

References

1. IEEE (2021) 7007-2021 IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems.
2. Michael Houghtaling A, Sandro Fiorini R, Nicola Fabiano, Paulo Goncalves JS, Zvikomborero Murahwi, et al. (2001) Standardizing an Ontology for Ethically Aligned Robotic and Autonomous Systems, pp. 1-5.
3. Meeri Haataja (2020) The ethics certification program for autonomous and intelligent systems (ECPAIS).

How to cite this article: Zvikomborero M. The IEEE-SA CertiFAIEd Ethics Certification Framework for Autonomous Intelligent Systems(AIS) Enabling Fully Integrated and Coordinated Ethical Implementations of AIS – An Illustration Through a Proposed Implementation of the IEEE7007-2021 Ontological Standard for Ethically Driven Robotics and Automation Systems. Eng Technol Open Acc 2023; 5(4): 555668. DOI: 10.19080/ETOAJ.2023.05.555668

DOI: 10.19080/ETOAJ.2023.05.555668

How to cite this article: Zvikomborero M. The IEEE-SA CertiFAIEd Ethics Certification Framework for Autonomous Intelligent Systems(AIS) Enabling Fully Integrated and Coordinated Ethical Implementations of AIS – An Illustration Through a Proposed Implementation of the IEEE7007-2021 Ontological Standard for Ethically Driven Robotics and Automation Systems. Eng Technol Open Acc 2023; 5(4): 555668. DOI: 10.19080/ETOAJ.2023.05.555668