IEEE SA
STANDARDS ASSOCIATION

IEEE CertifAIEd™ – Ontological Specification for Ethical Accountability

IEEE AI ETHICS

IEEE

Abstract: The IEEE CertifAIEd™ criteria for certification in ethical accountability are discussed in this ontological specification. Providing actionable methods to granularly assess and benchmark systems and organizations in their ethical performance is the goal of this work. Original methods of analyzing the respective drivers and inhibitors that influence the emergence of a quality of ethics, in this case accountability, are utilized by the certification methodology. The creation of the certification process is discussed, along with its intended implementation. An overview of the criteria schema and example criteria are also provided. This certification process has been designed to generate a tailorable and scalable system for the development of conformity assessment and certification for emergent ethical features of autonomous intelligent systems (AIS). The contents of this ontological specification are designed to be broadly applicable to a wide variety of domains and use-cases as well as providing flexibility through up to three levels of criteria, enabling a deeper and more sophisticated certification process where necessary.

Keywords: accountability, autonomous intelligent systems, ethics

The Institute of Electrical and Electronics Engineers, Inc.
3 Park Avenue, New York, NY 10016-5997, USA

TRADEMARKS AND DISCLAIMERS

IEEE believes the information in this publication is accurate as of its publication date; such information is subject to change without notice. IEEE is not responsible for any inadvertent errors.

The ideas and proposals in this specification are the respective author's views and do not represent the views of the affiliated organization.

Notice and Disclaimer of Liability Concerning the Use of IEEE SA Documents

This IEEE Standards Association ("IEEE SA") publication ("Work") is not a consensus standard document. Specifically, this document is NOT AN IEEE STANDARD. Information contained in this Work has been created by, or obtained from, sources believed to be reliable, and reviewed by members of the activity that produced this Work. IEEE and the IEEE Conformity Assessment Program (ICAP) members expressly disclaim all warranties (express, implied, and statutory) related to this Work, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; quality, accuracy, effectiveness, currency, or completeness of the Work or content within the Work. In addition, IEEE and the ICAP members disclaim any and all conditions relating to: results; and workmanlike effort. This document is supplied "AS IS" and "WITH ALL FAULTS."

Although the ICAP members who have created this Work believe that the information and guidance given in this Work serve as an enhancement to users, all persons must rely upon their own skill and judgment when making use of it. IN NO EVENT SHALL IEEE SA OR ICAP MEMBERS BE LIABLE FOR ANY ERRORS OR OMISSIONS OR DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS WORK, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Further, information contained in this Work may be protected by intellectual property rights held by third parties or organizations, and the use of this information may require the user to negotiate with any such rights holders in order to legally acquire the rights to do so, and such rights holders may refuse to grant such rights. Attention is also called to the possibility that implementation of any or all of this Work may require use of subject matter covered by patent rights. By publication of this Work, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. The IEEE is not responsible for identifying patent rights for which a license may be required, or for conducting inquiries into the legal validity or scope of patents claims. Users are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. No commitment to grant licenses under patent rights on a reasonable or non-discriminatory basis has been sought or received from any rights holder.

This Work is published with the understanding that IEEE and the ICAP members are supplying information through this Work, not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought. IEEE is not responsible for the statements and opinions advanced in this Work.

Participants

At the time this specification was completed, the IEEE CertifAIEd™,[1] Accountability Expert Working Group had the following membership:

Dr. Ozlem Ulgen, *Chair*
Ali Hessami, *Technical Editor*

Alka Kumar
Amandeep Gill
Aurelie Jacquet
Gisele Waters
Matt Newman
Ali Hessami
Sara Spinelli
Yohko Hatada

The Accountability Expert Focus Group

The work of CertifAIEd™ was largely driven by the efforts of three expert focus groups, their appointed leads, and support from the Chair. The Accountability Expert Focus Group (AEFG) was formed of volunteers from diverse backgrounds and experience including legal, computer science, technological, organizational, safety, human factors, auditing, and fiscal. However, other experts were invited to complement gaps identified in the profile of the AEFG. The AEFG held 16 ideas capture workshops in developing the ethical accountability schema, a graphical representation of factors that positively or negatively influence ethical accountability, which is set out in Annex A.

[1] IEEE CertifAIEd™ is a trademark owned by The Institute of Electrical and Electronics Engineers, Incorporated.

Introduction

The advent of automation during the industrial revolution brought about societal and business benefits in large-scale production, consistency, quality, and efficiencies that made commodities affordable. One key feature of most automation systems is the existence of human in the loop (HITL) at some stage providing oversight and control on critical aspects of the process or production. The development of *learning* machines that can perform specific tasks without using explicit instructions is now the foundation of autonomous intelligent systems (AIS) proliferating pervasively in all facets of industry, service provision, and governance. These machines rely on patterns and inductive or deductive inference, thereby raising the prospect of autonomous decision-making (ADM) by algorithmic learning systems (ALS), or ADM/ALS.

ADM/ALS offers the possibility of reducing and ultimately removing the human agent from operation, control, and supervisory roles, thereby reducing costs and potential errors while processing a much larger number of transactions offering higher service levels. While this brings savings, efficiencies, and business benefits, the removal of the human agent from the control and oversight loop brings about uncertainties and concerns regarding trustworthiness, fairness, explicability, and rationality of the automated decisions.

The uncertainties and societal concerns over ethicality and trustworthiness of ADM/ALS in all walks of life, especially in high-risk environments such as transportation, healthcare, financial, and public services, pose a formidable challenge to the uptake and innovation in deployment of the AIS-based solutions. There is thus a desire to regulate the implementation of ADM/ALS in order to provide a safety net and assurance about potential risks and societal harms that may ensue in the course of pursuing the perceived benefits.

From a broader ethical perspective, three key areas of concern in development and deployment of ADM/ALS relate to accountability, transparency, freedom from unacceptable algorithmic bias/fairness, privacy, and responsible governance. To this end, the IEEE Standards Association (SA) has developed a suite of detailed criteria for evaluation, conformity assessment, and certification of these properties of ADM/ALS products and services through CertifAIEd™. This program is a key facet of the IEEE SA's Global Initiative and Ethically Aligned Design portfolio.

Contents

1. Overview

1.1 Scope

The IEEE ethics certification criteria for assurance of many ethical facets of the development and deployment of autonomous intelligent systems (AIS) constitute an extensive hierarchical suite, developed by a panel of competent experts through a model-based creative process. The criteria suite for accountability comprises articulation of pertinent critical factors at two levels of hierarchy: Level 1 and Level 2. The Level 1 and Level 2 criteria collectively constitute the entire ethical accountability suite for the purposes of conformity assessment and certification. This ontological specification provides insight into and specification of Level 1 accountability factors to disseminate and enhance the understanding of IEEE's ethics certification criteria.

The ethics criteria suites are also developed from a general ethics perspective. The development strategy and deployment approach for these criteria provide an efficient and pragmatic approach for customization of a given suite for application-specific context and requirements. This is referred to as *profiling* and, in practice, the generic ethical accountability suite can be customized into many profiles appropriate to the requirements, terminology, context, and priorities of a given sector, culture, or application vertical. This specification examines the generic ethics for ethical accountability.

1.2 Purpose

This ontological specification discusses the development and specification of accountability conformity assessment and certification criteria of the IEEE CertifAIEd™.[2]. The criteria are applicable to all ethical accountability concerns within the context of AIS.

2. Definitions, acronyms, and abbreviations

2.1 Definitions

For the purposes of this document, the following terms and definitions apply.

ethical accountability: A contextual set of values pertaining to accountability and the satisfaction of a framework of expectations concerned with taking responsibility for actions, omissions, and outcomes and their ethical consequences (such as justice, redress, preservation of autonomy, self-determination, self-selected communities/locum and intimacies, and where issues of dignity and wellbeing in the use of technology are pertinent).

NOTE 1—Ethics is human focused, so accountability is human centric and intended to keep the human in the loop. Various dimensions—such as geographic, cultural, and ethnic—are included.

[2] IEEE CertifAIEd™ is a trademark owned by The Institute of Electrical and Electronics Engineers, Incorporated.

NOTE 2—Accountability is multidimensional and varies with roles and responsibilities based on organizational structure, position, delegation, degree of management, organizational culture and geographic location, and societal culture.

NOTE 3—Norms describe right and wrong actions that lead to judgments of good or evil persons or actions made by or on behalf of persons.

NOTE 4—Accountability overlaps with, and is largely complementary to, the aspects enforced and protected by localized regulation (i.e., sector-specific regulation) and law.

2.2 Acronyms and abbreviations

ADM	autonomous decision-making
AIS	autonomous intelligent system(s)
ALS	algorithmic learning system
EFR	ethical foundational requirement

3. Stakeholders

The key stakeholders of the ethical accountability of AIS are the following entities: developers, system/service integrators, system/service operators, maintainers, regulators, and the end users (see 6.3 on duty holders).

NOTE 1—An entity can be an individual, a single organization, or a group of collaborating individuals and organizations. The above labels for the five groups of stakeholders are generic and can be mapped in terms of activities and influence against the AIS life cycle but with overlapping activities. A single entity may assume multiple roles, that is, a developer may also fulfill and complete system design, integration, and maintenance.

NOTE 2—End users are a legitimate class of stakeholders, but there are no requirements placed on this group in these criteria.

4. Context

The IEEE CertifAIEd™ has been designed to generate a tailorable and scalable system for the development of conformity assessment and certification for emergent ethical features of AIS. This program developed ethical criteria for transparency, accountability, and algorithmic bias during an earlier phase and then ethical privacy in a subsequent phase. The current focus is on ethical accountability criteria that go beyond legal stated requirements of accountability and complement the legally enforceable protection measures. During explorations, it became clear how multifaceted and complex the issue of accountability is and how it extends beyond the notion of being held liable for some act or omission due to compliance failings as currently denoted in the law. Also noteworthy is that not all jurisdictions approach accountability in their respective legal systems in the same way; therefore, there was more of a need to identify this suite of criteria to help organizations assess and conform to accountability.

At the commencement of the exploratory and creative approach to the development of the principal concepts and formulation of the criteria, accountability was broadly defined as in 2.1.

As AIS are increasingly interwoven in human daily existence, they are increasing the risk of: under- and over- reliance on such systems; fewer critical challenges to or interrogations of such systems; and further avoidance of responsibility-taking and deference to such systems. This leads to compromising impacts—in often unknown and insidious ways—on human ethical privacy, ethical transparency, autonomy, dignity, and the propensity for growth of human flourishing. With respect to AIS, special attention is warranted because AIS have—superior to that of any human or human organization—an ability to glean insight from vast amounts of data. As a result, AIS have the potential to warp human input and output channels in ways that humans (individually and in groups) will not be able to defend but for which they cannot avoid responsibility and potential liability.

As such, the IEEE CertifAIEd™ accountability criteria suite comprises a holistic and systemic set of factors required in decision-making, rulemaking, enforcement, redress, operational governance (including organizational compliance culture), and, most importantly, human capacity and behavior across not only the AIS life cycle but with assumptions and dependencies from the wider AIS ecosystem as well.

The criteria have also sought to emphasize the importance of contextual understanding, culture, and continuous monitoring to ensure appropriateness and timeliness of interventions. Furthermore, for the purposes of accountability, this suite of ethical criteria reflects an effort to have responsibility remain with the humans and human organizations involved in the actions bringing AIS into being as it is still considered premature to preassign any such responsibilities to the AIS themselves.

The criteria take into account real-world scenario applications as well as capturing new and emerging concerns about accountability for AIS. Innovative aspects of the criteria include the following:

a) application to supply chain operations, including business operations, dealings, and third parties;
b) minimum assessment requirements comprising 1) sector risks, including web-based global operation risks, 2) potential harms/adverse impacts from AIS, 3) end-user needs, and 4) supply chain awareness;
c) overall legal compliance taking account of cross-jurisdictional reach and sector-specific AIS operations;
d) early warning system or messages for a dynamic or learning system;
e) “black box” scenario protocol;
f) user pre-use information and opt-out mechanism; and
g) emergency response mechanism for random and systematic errors.

5. Accountability factors

In considering what goals/factors contribute to the quality of ethicality—in addition to the classical identification of contributory factors—we recognized a need, supported by the adopted methodology, to map those goals/factors that would detract from it also. These are referenced as *drivers* and *inhibitors*, respectively, in the accountability schema (see Annex A). The rationale being many real-world constraints can frustrate well-meaning objectives due to issues of human resourcing, management, technological limitations, and cultural change.

5.1 Drivers of accountability

The five supportive influencing factors (drivers) impacting accountability are the following:

a) *Organizational governance, capability, and maturity:* This driver goal deals with the organization's capability, maturity, governance processes, and political will/good faith for accountability assurance.

b) *Clarity and consistency of AIS operations:* This driver goal seeks to ascertain a clear definition and the articulation and communication of the concepts and results of operation in various intended environments for AIS products, services, or systems to the relevant stakeholders.

c) *Human oversight:* This driver goal identifies whether or not an organization is upholding a holistic approach to human agents being able to understand an AIS product, service, or system behavior in order to be able to intervene as appropriate. This includes being able to set up a process to deny continuation of activity and assess the context to ensure timely corrective action. Furthermore, in implementing human oversight, the organization should be mindful of and mitigate against harmful or detrimental human intervention.

d) *User interactions:* This driver goal seeks to ascertain how potential users are being made aware of the existence and functions of an AIS element within products, services, or systems in the context of use and how they are being empowered to sufficiently understand and make decisions regarding the use of such systems. This may also identify opportunities to vet or veto the system prior to use, during use, or to opt-out after use, also opening up opportunities for the user to challenge AIS decision-making.

e) *Upholding ethical profile:* This driver goal seeks to ascertain whether efforts to maintain an ethical profile of AIS products, services, or systems are in place with respect to accountability requirements and criteria/behaviors, mapped to the organizational ethical policies and values.

5.2 Inhibitors of accountability

The three constraining influencing factors (inhibitors) impacting accountability are:

a) *Random and systematic errors:* This inhibitory goal considers that AIS will always have random errors in their workings that could create risks/negative impacts. Therefore, an organization needs a structure and process in place to manage random errors and mitigate first- and second-order impacts on stakeholders and users. Human-related systematic errors due to both omission and commission during the life cycle are included.

b) *Rubber stamping:* This inhibitory goal considers the impact of a cursory approach to accountability, an approach that follows the letter rather than the spirit of an ethical requirement, lending itself to a situation where minimal effort is exercised to attain conformity or the appearance of conformity in the sense of ethics washing.

c) *Inadequate or nonexistent records:* This inhibitory goal considers the lack of adequate records or the nonexistence of records—concerning the AIS life-cycle process, product, service, or system development, deployment, and operation—which, if existent, would serve to demonstrate accountability claims and/or identify who is responsible for the proper operation of the AIS.

Explanation of the goals and associated requirements, requisite evidence, and scale of measurement are depicted in Annex A (schema) and Annex B (criteria).

6. Accountability certification criteria

6.1 Accountability ethical foundational requirements (EFRs)

The ethical accountability schema, in conjunction with the accountability ethical foundational requirements (EFRs), enables the assessment and auditing of organizations and their autonomous intelligent technologies for accountability with clear criteria that can be turned into a scoring mechanism. As a model-based approach, the schema captures both negative and positive aspects (inhibitors and drivers, respectively) of accountability for AIS with ease of reference. It represents an efficient means of real-time creative knowledge capture as well as operating as the foundation for development of ethical accountability requirements.

The detailed accountability EFRs are depicted in Annex B.

6.2 Normative and instructive accountability EFRs

The accountability EFRs contain a series of expected behavioral norms and instructions on how to enact aspects of the certification, without going into specifics where not strictly necessary, to preserve flexibility of implementation within a bounded set of principles. In this spirit, the accountability EFRs depicted in Annex A (schema) and Annex B (criteria) are classed into *normative* (mandatory) and *instructive* (recommended) for the purposes of conformity assessment against the suite of accountability certification criteria.

6.3 Duty holders of the accountability EFRs

The accountability EFRs depicted in Annex A (schema) and Annex B (criteria) are additionally noted against the specific group of duty holders for the purposes of conformity assessment. The principal groups are as follows:

— *Developer (D):* The entity (see NOTE 1—Clause 3) that designs and develops a component (product) or system for general or specific purpose/application. This could be as a result of a developer's own instigation or response to the market or a client requirement. The developer is responsible for the ethical assurance of the generic or application-specific product or system and associated supply chain.

— *(System/service) Integrator (I):* The entity that designs and assures a solution through integrating multiple components, potentially from different developers, and tests, installs, and commissions the whole system in readiness for delivery to an operator. The system delivery may take place over several stages. The integrator is usually the duty holder for total system assurance and certification, safety, security, reliability, availability, sustainability, and so forth. For this, it may rely on the certification or proof of ethics from various developers or the supply chain.

— *(System/service) Operator (O):* The entity that has a duty, competences, and capabilities to deliver a service through operating a system delivered by an integrator.

— *Maintainer (M):* The entity tasked with conducting required monitoring, preventive or reactive servicing and maintenance, and required upgrades to keep the system operational at an agreed service level. Maintainer could also be charged with abortion of maintenance and disposal of the system.

— *Regulator (R):* The entity that enforces standards and laws for the protection of life, property, or the natural habitat through imposing duties and accreditation/certification.

6.4 The levels of ethical accountability certification

Three main levels of assessment of conformity are established, depending on the scale of risks posed and the impact of the AIS on health, welfare, safety, and ethical values of stakeholders. The levels are:

1) *Baseline, low impact (LI):* The smallest subset of accountability EFRs is applicable for conformity assessment.
2) *Compliant, medium impact (MI):* A larger set of accountability EFRs than baseline is applicable for conformity assessment.
3) *Critical, high impact (HI):* Any AIS product, service, or system that presents a likelihood of injury or harm to well-being, health, safety, security, privacy, and welfare must satisfy all accountability EFRs.

The level of certification is determined through a risk-profiling exercise on the product, service, or system that takes place as the first phase of the conformity assessment activities.

6.5 Required evidence

These are the types and quantity of evidence items required to satisfy the stated requirements. A single requirement may relate to one or many items of objective evidence for evaluation of the degree to which the requirement is met (satisfaction).

6.6 Evaluation of evidence

This evaluation of evidence comprises a suitable scale of measurement and scoring of the evidence. A two-tier approach to the measurement of the evidence items is adopted as follows:

a) Top-level finding: No critical findings in the detailed normative requirements/areas requiring attention for improvement.

b) Overall score: On a 1 to 5 scale (based on aggregate of satisfying sublevel goals):

5- Excels baseline requirements

4- Sustains baseline requirements

3- Meets baseline requirements (pass mark)

2- Needs improvement

1- Does not meet requirements

A score of 3 is generally considered to be a sufficient pass mark for most cases. However, certain elements that represent a particularly strong risk or that operate in a mission-critical capacity may require a higher score to be considered sufficient.

NOTE 1—The scale of evaluation and the typical pass mark shall be appropriate to the criticality of the requirement and the nature of the evidence and may vary for each accountability EFR.

NOTE 2—Each accountability EFR can have its own bespoke units, measurement scale, and benchmark for evaluation appropriate to its nature. The 1 to 5 scoring adopted is the default for all accountability EFRs in Annex B and can be modified as appropriate to the nature of the evidence.

6.7 The constraints of accountability certification

The certification cannot cover every potential eventuality, and it has a period of validity during which it will become progressively less effective in upholding the quality of accountability if not updated periodically to match changes in technology, culture, law, consumer standards or expectations, and practices. Eventually, without update, the certification may drift from contemporary realities and established best practices.

It will be important to make regular updates and amendments to the underlying concept schema where appropriate. The team has attempted to think ahead in terms of a foreseeable time horizon, future proofing (to some degree) the criteria and certification through discussion of technologies or practices that may be prototyped presently but are not yet in common deployment or in line with established norms and best practices.

Annex A

AIS ethical accountability schema

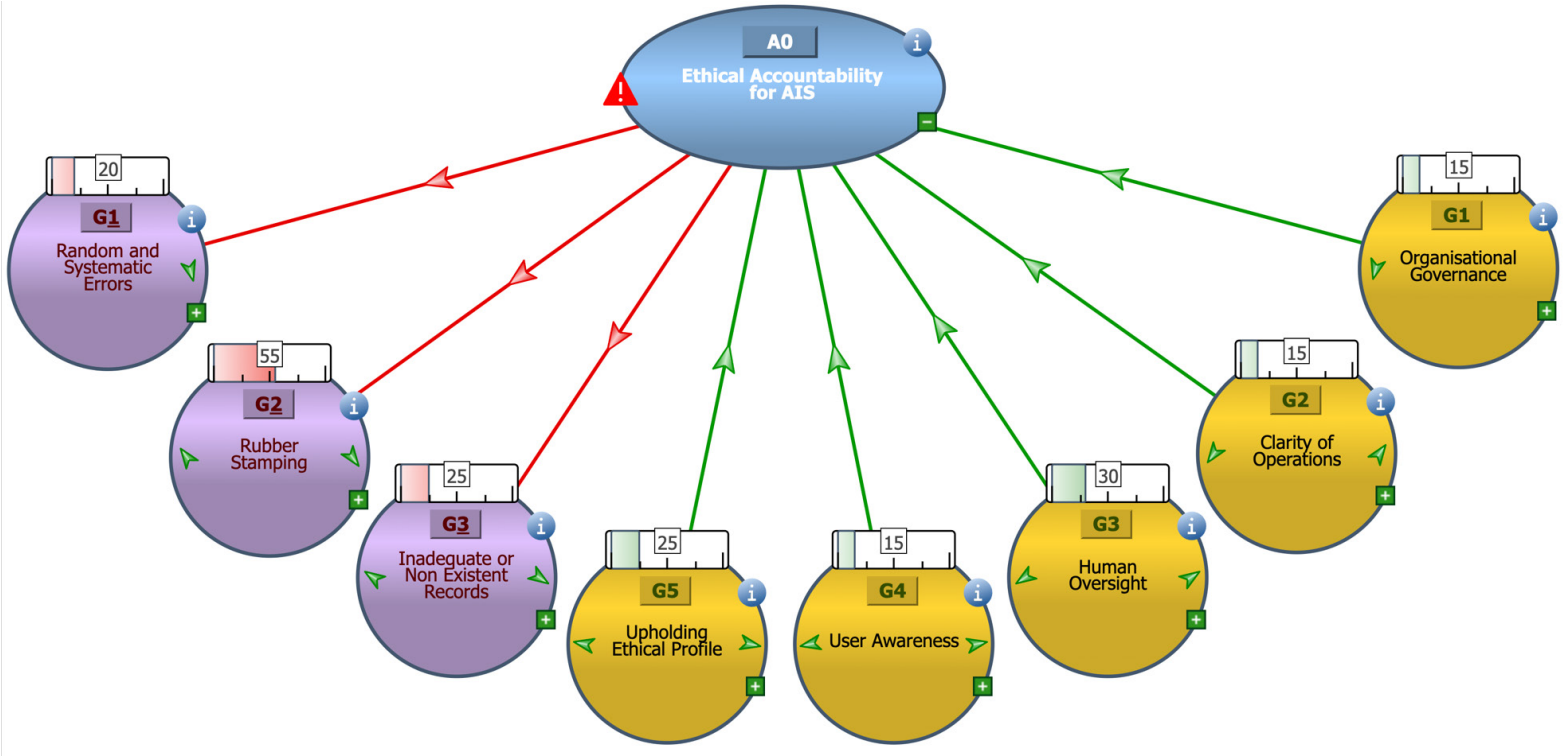

Figure A.1—Drivers and inhibitors of AIS ethical accountability.

Annex B

Ethical accountability certification criteria

Accountability schema goal description	Accountability ethical foundational requirements (EFRs)	Normative/ instructive	Cert level LI, MI, HI	Duty holder D, I, O, M, R	Required evidence	Evidence measurement and pass mark
G1 - Organizational governance, capability, and maturity The organization's capability, maturity, governance processes, and political will/good faith for ethical assurance.	The following accountability ethical foundational requirements shall be fulfilled for the product, system, or service by the relevant duty holders: a) Demonstrate that a suitable and sufficient organizational governance framework is in place, reflecting capability, maturity, and processes to ensure legal responsibility and ethical accountability	N	HI	D, I, O, M, R	The following items of evidence fulfill the ethical foundational requirements. a) A copy of the organizational chart showing lines of responsibility and accountability including the supply chain b) Designated positions for risk management, data protection compliance, legal compliance, stakeholder management, and ethical profile management and coordination across all roles c) Minimum assessment requirements comprising: 1. sector risks, including web-based global operation risks; 2. potential harms/adverse impacts from AIS; 3. end-user needs (e.g., privacy); and 4. supply chain awareness and compliance with minimum assessment requirements. d) Implementation of local laws and requirements relevant to above minimum assessment requirements e) Overall legal compliance	Two-tier approach measurement of the evidence items: a) Top-level finding: "No critical findings in the detailed normative requirements"/"areas requiring attention for improvement." b) Overall score: On 1-5 scale (based on aggregate of satisfying sublevel goals) such as: 5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet requirements

Accountability schema goal description	Accountability ethical foundational requirements (EFRs)	Normative/ instructive	Cert level LI, MI, HI	Duty holder D, I, O, M, R	Required evidence	Evidence measurement and pass mark
					(dependent on cross-jurisdictional reach and sector-specific operations of AIS) f) Engagement and participation in industry initiatives	
G2 - Clarity of operations Transparency as existing for the AIS product, service, or system in terms of clear definition and articulation of the concepts of operation in various environments.	The following accountability ethical foundational requirements shall be fulfilled for the product, system, or service by the relevant duty holders: a) Demonstrate that all facets and modes of intended operation, as well as the operational environments, stakeholders, and contexts are analyzed under various scenarios and clearly specified for design, deployment, and maintenance modes for the product, service, or system	N	HI	D, I, O, M,	The following items of evidence fulfill the ethical foundational requirements: a) Design specifications b) Operational scenarios specification c) Functional design specification d) Operational manuals and guidelines	Two-tier approach to encourage adoption: a) Top-level finding: "no critical findings in the detailed requirements" / "areas requiring attention for improvement" b) Organizational readiness finding: on 1-5 scale (based on aggregate of satisfying sublevel goals) such as: 5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet requirements
G3 - Human oversight Human agents should be able to understand the AIS product, service, or system behavior in order to be able to intervene and	The following accountability ethical foundational requirements shall be fulfilled for the product, system, or service by the relevant duty holders: a) Demonstrate human	N	HI	D, I, O, M, R	The following items of evidence fulfill the ethical foundational requirements: a) Implementation of explanation tools from the design phase to enable human understanding and	Two-tier approach measurement of the evidence items: a) Top-level finding: "No critical findings in the detailed normative

Accountability schema goal description	Accountability ethical foundational requirements (EFRs)	Normative/ instructive	Cert level LI, MI, HI	Duty holder D, I, O, M, R	Required evidence	Evidence measurement and pass mark
set up a process to deny continuation of activity and assess the context; in implementing human oversight, the organization should be mindful of and mitigate against harmful or detrimental human intervention.	oversight capability throughout the product, service, or system life cycle, including any human intervention mechanism, and mitigation of potential harmful effects from human intervention				review of AIS b) Reasonable and proportionate human review (e.g., AI chips for fire alarms or thermostats require only focused review at product design and manufacture, then post-deployment product liability for faults, whereas more sophisticated AIS such as healthcare robots, algorithmic banking, insurance, and human resources (HR) decision-making require full life-cycle human review) c) Human intervention protocol d) "Black box" scenario protocol e) Management sign-off of intended AIS use and risk implications f) Organization chart of lines of responsibility and accountability throughout AIS life cycle	requirements"/"areas requiring attention for improvement." b) Overall score: On 1-5 scale (based on aggregate of satisfying sublevel goals) such as: 5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet requirements
G4 - User interactions Potential users being aware of the existence of an AIS element within the product, service, or system in the context of use, and being empowered to make decisions of the use of such systems.	The following accountability ethical foundational requirements shall be fulfilled for the product, system, or service by the relevant duty holders: Demonstrate the existence of mechanisms/procedures to enable: a) Pre-use awareness of the type of product, service, or system they are interacting with, including whether there is an AIS element b) The user to opt out of	N	HI	D, I, O, M, R	The following items of evidence fulfill the ethical foundational requirements: a) Reasonable and proportionate information to enable user awareness b) Specific mechanism for user pre-use information (e.g., product specification; T&C; web pop-up box) c) Mechanism for user acknowledgment/consent of pre-use information d) Opt-out provision (e.g., speak to	Two-tier approach measurement of the evidence items: a) Top-level finding: "No critical findings in the detailed normative requirements"/"areas requiring attention for improvement." b) Overall score: On 1-5 scale (based on aggregate of satisfying sublevel goals) such as:

Accountability schema goal description	Accountability ethical foundational requirements (EFRs)	Normative/ instructive	Cert level LI, MI, HI	Duty holder D, I, O, M, R	Required evidence	Evidence measurement and pass mark
	using the product/ service/system c) The user to challenge an AIS decision effectively and efficiently d) The user to understand the full terms and conditions (T&C) that apply to any interactions e) The user to revisit the site at a later date to understand previous interactions				human operator) e) Mechanism for user to challenge AIS decision	5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet requirements
G5 - Upholding ethical profile Efforts to maintain an ethical profile of the AIS product, service, or system with respect to accountability requirements and criteria/behaviors.	The following accountability ethical foundational requirements shall be fulfilled for the product, system, or service by the relevant duty holders: a) Demonstrate that efforts are put in place to include accountability criteria/behaviors as part of the AIS ethical profile b) Mapping algorithmic AIS ethical profile to the organizational ethical policies and values	N	HI	D, I, O, M, R	The following items of evidence fulfill the ethical foundational requirements: a) Ethical issues register b) Tailored organizational ethical policy statement c) Documents explaining the risk management and strategic response actions in case of malfunctions d) Section on website explaining AIS ethical profile that demonstrates the human operator's capability to challenge algorithmic decision-making e) Audit reports f) External studies/reports (if any) g) Interviews with employees, agents, business partners, supply chain operators, and (where relevant) clients	Two-tier approach measurement of the evidence items: a) Top-level finding: "No critical findings in the detailed normative requirements"/"areas requiring attention for improvement." a) Overall score: On 1-5 scale (based on aggregate of satisfying sublevel goals) such as: 5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet

Accountability schema goal description	Accountability ethical foundational requirements (EFRs)	Normative/ instructive	Cert level LI, MI, HI	Duty holder D, I, O, M, R	Required evidence	Evidence measurement and pass mark
						requirements
G1b - Random and systematic errors AIS will always have random errors in its workings that could create risks/negative impacts, so the organization needs a structure and process in place to manage random errors and mitigate first- and second-order impacts on stakeholders and users; includes human-related systematic errors due to omissions and commission during life cycle.	The following accountability ethical foundational requirements shall be fulfilled for the product, system, or service by the relevant duty holders: Demonstrate a robust structure and process in place for managing random errors, including human-related systematic errors of omission and commission during life cycle, and mitigating first- and second-order impacts on stakeholders and users. In particular: a) Measure efficiency and relevance of tests in place throughout the AIS life cycle to ensure robustness of AIS b) Measure the number and level of incidents that have occurred and how well they have been addressed c) Measure the efficiency and speed at which the organization can implement its emergency response (e.g., "red button" procedure") d) Measure the efficiency and speed of mitigation procedures	N	HI	D, I, O, M, R	The following items of evidence fulfill the ethical foundational requirements: a) Tests and test records b) Records of incidents measured c) "Red button procedure" or some other emergency response mechanism d) Risk matrix and emergency procedures that are updated on a regular basis e) Mitigation procedures and strategy	Two-tier approach measurement of the evidence items: a) Top-level finding: "No critical findings in the detailed normative requirements"/"areas requiring attention for improvement." b) Overall score: On 1-5 scale (based on aggregate of satisfying sublevel goals) such as: 5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet requirements
G2b - Rubber stamping	The following accountability ethical foundational requirements shall be	N	HI	D, I, O, M, R	The following items of evidence fulfill the ethical foundational requirements:	Two-tier approach measurement of the evidence items:

Accountability schema goal description	Accountability ethical foundational requirements (EFRs)	Normative/ instructive	Cert level LI, MI, HI	Duty holder D, I, O, M, R	Required evidence	Evidence measurement and pass mark
Cursory approach to accountability in letter rather than spirit aimed at satisfying a minimal effort for conformity.	fulfilled for the product, system, or service by the relevant duty holders: a) Avoid rubber stamping of ethics ("ethics washing") by having an ethics policy not only in name but also capable of being implementable, measurable, and publicly communicable and by having regular board-level consideration of ethics				a) Organizational charts b) Annual reports c) Questionnaires from business partners and interested third parties where the responses become evidence for not rubber stamping d) Minutes of board meetings on ethics issues e) Designated person responsible for implementing ethics policy f) Maintenance of an ethical issues register g) Consideration of values and ethics during recruitment of new employees, agents, business partners, supply chain operators, and subsequent training and embarking on new projects	a) Top-level finding: "No critical findings in the detailed normative requirements"/"areas requiring attention for improvement." b) Overall score: On 1-5 scale (based on aggregate of satisfying sublevel goals) such as: 5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet requirements
G3b - Inadequate or non-existent records Lack of adequate records or nonexistent records for the process, product, service, or system to demonstrate accountability claims.	The following accountability ethical foundational requirements shall be fulfilled for the product, system, or service by the relevant duty holders: a) Demonstrate maintenance of adequate records for the process, product, service, or system development, deployment and operation in order to identify who is responsible for the proper operation of the AIS	N	HI	D, I, O, M, R	The following items of evidence fulfill the ethical foundational requirements: a) Documented accountability best practice that explains the necessity and rationale for choices and compromises made which are in line with risk assessment b) Ongoing review and update of system performance c) Recording of system errors and malfunctions, with	Two-tier approach measurement of the evidence items: a) Top-level finding: "No critical findings in the detailed normative requirements"/"areas requiring attention for improvement." b) Overall score: On 1-5 scale (based on aggregate of satisfying sublevel goals)

Accountability schema goal description	Accountability ethical foundational requirements (EFRs)	Normative/ instructive	Cert level LI, MI, HI	Duty holder D, I, O, M, R	Required evidence	Evidence measurement and pass mark
					accompanying remedial action recorded d) Ongoing review and update of legal and ethical standards requirements e) Remedial process for failures in responsibility and accountability f) Clear documentation of who has responsibility for proper delivery of various stages of AIS creation, maintenance, and operation g) Logs of what was done with the AIS, by whom, how, and when	such as: 5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (typical pass mark) 2- Needs improvement 1- Does not meet requirements

Annex C

Bibliography

The following sources and public domain frameworks have been consulted for the verification, coverage, integrity, quality, and currency of the certification criteria independently developed in CertifAIEd™.

[B1] *The Age of Digital Interdependence*, Report of the UN Secretary-General's High-level Panel on Digital Cooperation, United Nations, Jun. 2019.[3]

[B2] "Ethics Guidelines for Trustworthy AI," High-Level Expert Group on Artificial Intelligence (AI HLEG), European Commission, Apr. 2019.[4]

[B3] "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Apr. 4, 2019.[5]

[B4] Floridi, L., J. Cowls, T. C. King, "How to design AI for social good: Seven essential factors," *Science and Engineering Ethics*, vol. 26, pp.1771–1796, 2020.

[B5] "G20 AI Principles," *G20 Ministerial Statement on Trade and Digital Economy*, Annex, Jun. 2019.[6]

[B6] Madary, M., and Thomas K. Metzinger, "Real virtuality: A code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology," *Frontiers in Robotics and AI*, vol. 3, no. 3, Feb. 19, 2016.

[B7] OECD/LEGAL/0449, *Recommendation of the Council on Artificial Intelligence*, May 21, 2019.[7]

[B8] "The State of AI Ethics," Montreal AI Ethics Institute, Jan. 2021.[8]

[3] United Nations publications are available from the United Nations website (https://www.un.org).
[4] European Commission publications are available from the Futurium website (https://futurium.ec.europa.eu/en).
[5] IEEE publications are available from the Institute of Electrical and Electronics Engineers, 445 Hoes Lane, Piscataway, NJ 08854-4141, USA (http://standards.ieee.org).
[6] Available from https://www.mofa.go.jp/files/000486596.pdf.
[7] Organisation for Economic Co-operation and Development publications are available from the OECD website (https://www.oecd.org/).
[8] Available from https://montrealethics.ai/wp-content/uploads/2021/01/The-State-of-AI-Ethics-Report-January-2021.pdf.

IEEE SA
STANDARDS ASSOCIATION

IEEE CertifAIEd™

http://engagestandards.ieee.org/ieeecertifaied.html

Connect with us on:

Twitter: twitter.com/ieeesa
Facebook: facebook.com/ieeesa
LinkedIn: linkedin.com/groups/1791118
Beyond Standards blog: beyondstandards.ieee.org
YouTube: youtube.com/ieeesa

standards.ieee.org
Phone: +1 732 981 0060

IEEE

A CHAMPION FOR
IEEE AI Ethics