IEEE SA
STANDARDS ASSOCIATION

IEEE CertifAIEd™ – Ontological Specification for Ethical Transparency

IEEE AI ETHICS

IEEE

Abstract: The IEEE CertifAIEd™ criteria for certification in ethical transparency 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 transparency, 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 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: autonomous intelligent systems, ethics, transparency

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 CertifAIEd™ Transparency Expert Working Group had the following membership:

Eleanor (Nell) Watson, *Chair*
Jonathan Barr, *Vice Chair*
Ali Hessami, *Technical Editor*

Lillie Beiting	Minna Mustakallio	Jon Stokes
Martin Clancy	Tista Saha	Kenneth Thorson
Steven Etteninger	Sarah Spinelli	Frith Tweedie
Val Goddard	Shikha Srivastava	Gerlinde Weger
Markus Kalliola	Alejandro Saucedo	Ali Hessami

The Transparency Expert Focus Group

The work of CertifAIEd™,[1] was largely driven by the efforts of expert focus groups, their appointed leads, and support from the Chair and the Program Manager. The Transparency Expert Focus Group (TEFG) was formed of volunteers from many different backgrounds and experiences, including legal, computer science, technological, organizational, safety, human factors, auditing, and fiscal.

[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 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, 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 developed 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 ethical transparency comprises articulation of pertinent critical factors at three levels of hierarchy: Level 1, Level 2, and Level 3. The three levels of criteria collectively constitute the entire ethical transparency suite for the purposes of conformity assessment and certification. This ontological specification provides insight into and specification of Level 1 ethical transparency 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 transparency 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 transparency.

1.2 Purpose

This ontological specification discusses the development and specification of ethical transparency conformity assessment and certification criteria of IEEE CertifAIEd™.[2] The criteria are applicable to all ethical transparency 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 transparency: A contextual set of values pertaining to transparency and the satisfaction of a framework of expectations (preservation of autonomy, self-determination, and self-selected communities/locum and intimacies).

NOTE 1—Ethics is human focused, so ethical transparency is human centric/anthropomorphic.

NOTE 2—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 3—Ethical transparency overlaps with, and is largely complementary to, the aspects enforced and protected by law.

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

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 transparency 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 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 early phase and then ethical privacy in a subsequent phase. The current focus is on ethical transparency criteria that go beyond legal stated requirements of transparency and complement the legally enforceable protection measures. During explorations, it became clear how multifaceted and complex the issue of transparency is and how it extends beyond the notion of compliance with transparency as currently denoted in the law. Also noteworthy is that not all jurisdictions approach transparency 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 ethical transparency.

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

As such, the CertifAIEd™ ethical transparency criteria suite comprises a holistic and systemic set of factors required in decision-making, rulemaking, enforcement, redress, operational governance, 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.

5. Ethical transparency factors

In considering what goals/factors contribute to the quality of transparency—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 transparency 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 ethical transparency

The six supportive influencing factors (drivers) impacting ethical transparency 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 ethical transparency 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 the intended environments for AIS products, services, and systems to the relevant stakeholders.

c) *Awareness of AIS interaction:* This driver goal identifies whether an end user will be immediately made aware if they are interacting with an AIS agent that functions in a manner that a reasonable person might confuse for a human being.

d) *Confidence in system behavior:* This driver goal emphasizes the quality of having complete confidence in total system behavior. This may be achieved through simulation, prediction, examination, and so forth of hypothetical scenarios in advance of the fact.

e) *Accessible and fair control and feedback:* 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 on the use of such systems. This may also identify where there is a disadvantage to the end user due to a lack of suitable alternative options.

f) *Upholding ethical transparency integrity:* This driver goal looks at efforts to maintain an ethical profile of AIS products, services, or systems with respect to transparency requirements and criteria/behaviors across the AIS life cycle and beyond.

5.2 Inhibitors of ethical transparency

The three constraining influencing factors (inhibitors) impacting ethical transparency are as follows:

a) *Behavioral obfuscation*: This inhibitory goal relates to the use of technologies that minimize their apparent spillover effects (externalities in economics terms), such as pollution, whether by intentional design or incidental omission due to the challenges of adequately detecting, accounting for, and managing externalities. It is also concerned with attempts to deceive or manipulate humans in any way.

b) *Concern with liability*: This inhibitory goal considers the service provider's awareness of potential risk exposure and delivery of the bare minimum of information (or an inadequate amount) to manage the risk. This could include legal, commercial, financial, and human intervention dimensions.

c) *Protection of trade secrets*: This inhibitory goal considers the potential for organizations to seek to protect their intellectual property (IP) through insufficient transparency or obfuscation of processes, functions, and capabilities.

Explanation of the goals and associated requirements, requisite evidence, and scale of measurement are depicted in Annex B.

6. Ethical transparency certification criteria

6.1 Transparency ethical foundational requirements (EFRs)

The ethical transparency schema, in conjunction with the transparency ethical foundational requirements (EFRs), enables the auditing of organizations and their autonomous intelligent technologies for ethical transparency 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 ethical transparency 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 transparency requirements.

The detailed transparency EFRs are depicted in Annex B.

6.2 Normative and instructive transparency EFRs

The transparency 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, in order to preserve flexibility of implementation within a bounded set of principles. In this spirit, the transparency EFRs depicted in Annex B are classed into *normative* (mandatory) and *instructive* (recommended) for the purposes of conformity assessment against the suite of ethical transparency certification criteria.

6.3 Duty holders of the transparency EFRs

The transparency EFRs depicted in Annex B 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 a 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 transparency 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:

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

6.7 The constraints of ethical transparency certification

The certification process cannot cover every potential eventuality. Changes in technology, culture, law, consumer standards, and practices may diminish its effectiveness or applicability to support the quality of ethical transparency. Eventually, without update, the certification may drift from contemporary realities and established best practices.

Therefore, it will be important to make regular updates and amendments to the underlying concept schema where appropriate. The IEEE CertifAIEd™ team has forecast potential technological and cultural developments for a foreseeable time horizon, thereby future proofing the criteria and certification as far as possible. This has been accomplished 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 transparency schema

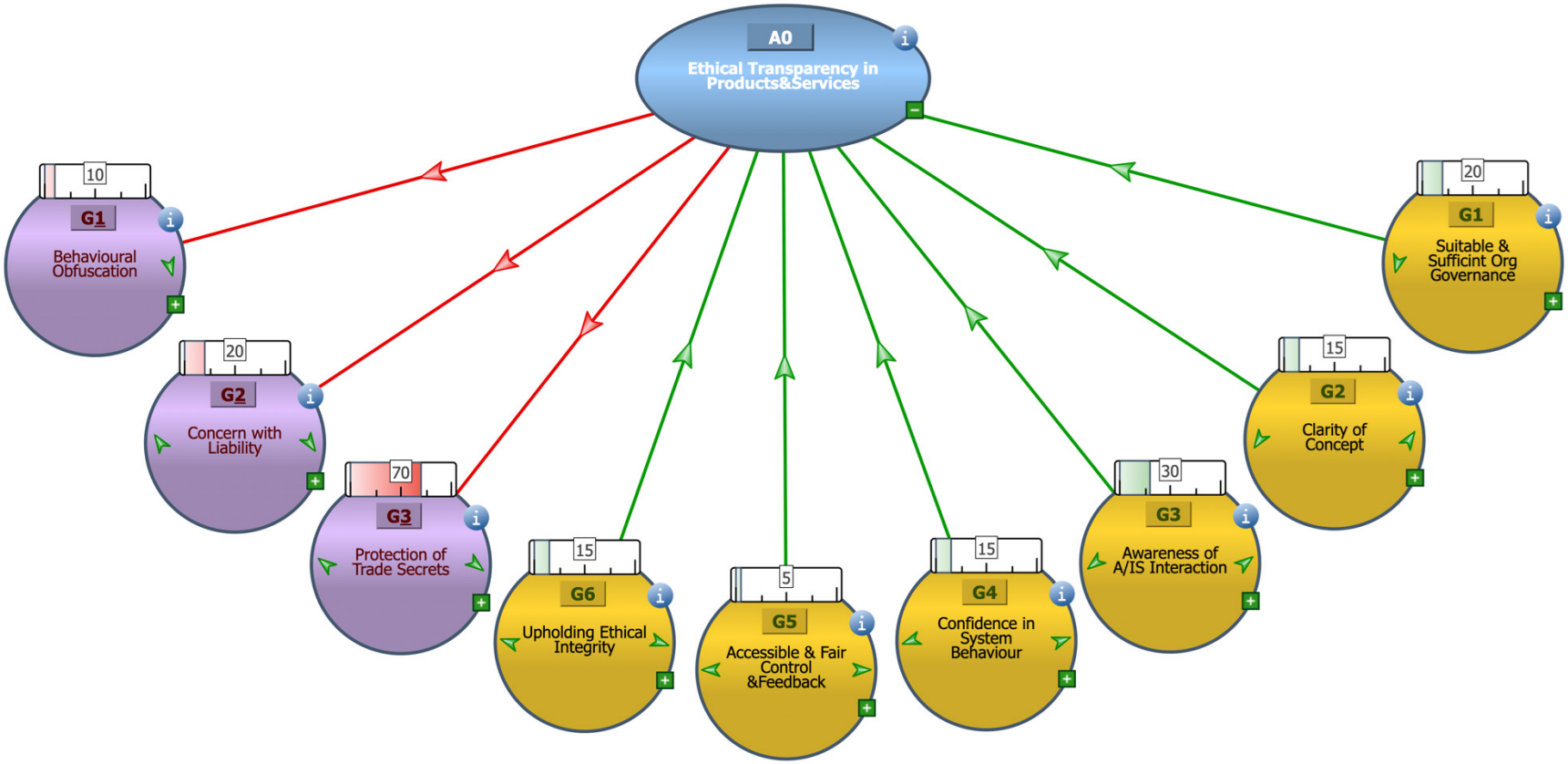

Figure A.1— Drivers and inhibitors of AIS ethical transparency.

Annex B

Ethical transparency certification criteria

Transparency schema goal description	Transparency 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 This comprises the capability, maturity, and intent of the development organization in having the right motivation and resources, processes, and so forth to achieve transparency.	The following privacy ethical foundational requirements shall be fulfilled for the product, system, or service by the 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	LI	D, I, O, M, R	The following item(s) shall be presented as evidence for conformity against the transparency requirement(s): a) Organization 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	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 (pass mark) 2- Needs improvement 1- Does not meet requirements

Transparency schema goal description	Transparency 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
					above minimum assessment requirements. e) Overall legal compliance (dependent on cross-jurisdictional reach and sector-specific operations of AIS). f) Engagement and participation in industry initiatives.	
G2- Clarity of operations Detailed description of the total level of performance, capabilities, and behavioral features of a product, service, or system.	The duty holder shall fulfil the following transparency requirement(s): a) Demonstrate a system design overview that is open, accessible, and takes user needs into account and is well documented. A precis of the design shall be made accessible to the public. b) Specify the concepts of operation for development, trials, and global contexts of use that would assume and include the operational environment. c) Where possible, simulate the concepts and contexts of operations as modeled and validate these in advance of the design efforts. d) Modeling of interactions	N N I	LI LI LI	D, I, O, M, R D, I, O, M, R D, I, O, M, R	The following item(s) shall be presented as evidence for conformity against the transparency requirement(s): a) Abstract overview of the system, context of operation, and the original concepts of product/system deployment in the operating environment including: 1. Design specifications 2. Operational scenarios specification 3. Functional design specification 4. Operational manuals and guidelines	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 (pass mark) 2- Needs improvement 1- Does not meet requirements

Transparency schema goal description	Transparency 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
	(e.g., UML), and examples of various parameters and environments shall be carried out to further clarify the concept of operation.	I	LI	D, I, O, M, R		
G3- Awareness of AIS interaction An end-user must be aware if they are interacting with an AIS agent that functions in a manner that a reasonable person might confuse for a human being.	The duty holder shall fulfill the following transparency requirement(s): a) Ensure user awareness of the type of product, service, or system they are interacting with, including whether there is an AIS element b) The user is able to opt out of using the product, service, or system c) The user able to challenge an AIS decision effectively and efficiently	N N N	LI LI LI	D, I, O, M, R D, I, O, M, R D, I, O, M, R	The following item(s) shall be presented as evidence for conformity against the transparency requirement(s): a) Reasonable and proportionate information to enable user awareness b) Specific mechanism for user pre-use information (e.g., product specification; terms and conditions (T&C); web pop-up box) c) Mechanism for user acknowledgment/consent of pre-use information. d) Opt-out provision (e.g., speak-to-human operator) e) Mechanism for user to challenge AIS decision	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 (pass mark) 2- Needs improvement 1- Does not meet requirements
G4- Confidence in system behavior The quality of having	The duty holder shall fulfill the following transparency requirement(s):				The following item(s) shall be presented as evidence for conformity against the transparency requirement(s):	Two-tier approach measurement of the evidence items:

Transparency schema goal description	Transparency 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
complete confidence in total system behavior. This may be achieved through, simulation, prediction, and so forth.	a) Design a system that has a consistent and predictable operation behavior in various environments	N	LI	D, I, O, M, R	a) The user manual of the product capturing the system installation requirements and system behavior under various conditions including deviations and corrective actions	a) Top-level finding: "No critical findings in the detailed normative requirements"/"areas requiring attention for improvement"
	b) Ensure conformance to the system requirements during product installation	N	LI	D, I, O, M, R	b) The user manual also covering actions required from the end user in case of deviations	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- <u>Meets baseline requirements (pass mark)</u> 2- Needs improvement 1- Does not meet requirements
	c) Clearly communicate product's transparency confidence upholding design and features to the users	N	LI	D, I, O, M, R	c) The accuracy of the various AI subsystems of the product and the overall accuracy of the system	
	d) Devise mechanisms to check and log aberrations/deviations in the system behavior	N	LI	D, I, O, M, R	d) Documentation or recording of consensus algorithm execution	
	e) Design the system to take corrective actions in scenarios of behavior deviations	N	L	D, I, O, M, R	e) Records with immutable/indelible forms of information supporting consistent system behavior	
	f) Update users about the actions required in scenarios of behavior deviations	N	LI	D, I, O, M, R	f) Logs with input to the AIS along with the corresponding outcomes especially to record deviations	
	g) Log system behavior/outcome for every input and send logs periodically to a central server for audit	N	LI	D, I, O, M, R	g) Audit reports of system behavior with regard to time	
	h) Regularly communicate any changes in product behavior to end users	N	LI	D, I, O, M, R		

Transparency schema goal description	Transparency 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
G5 - Accessible control and feedback This is comprised of stakeholders understanding how to take back meaningful control from an AIS or influencing factors upon it, enabling end users to reliably and meaningfully opt in or out.	The duty holder shall fulfill the following transparency requirement(s): a) Design a system that allows its end users visibility and discretion over the usage of their data in this system and its network(s) b) Clearly communicate internal and external usage of end user data, including data sharing with third parties c) Avoid bias and discrimination in AIS architecture, and ensure accessibility for persons with disabilities d) Regularly communicate to end users changes and updates to the AIS model that impact data exchange, storage, usage, or security	N N N N	LI LI LI LI	D, I, O, M, R D, I, O, M, R D, I, O, M, R D, I, O, M, R	The following item(s) shall be presented as evidence for conformity against the transparency requirement(s): a) Overview of the system model and mapping indicating where an end user can consent or opt out b) Consent documentation before, during, and after usage of the AIS system c) Communication policies for system changes, access changes, and storage security protocols d) Reasonable and proportionate information to enable user awareness. e) Specific mechanism for user pre-use information (e.g., product specification; T&C; web pop-up box) f) Mechanism for user acknowledgment/consent of pre-use information g) Opt-out provision (e.g., speak to human operator) h) Mechanism for user to challenge AIS decision i) Communication of major shareholders of the organization deploying the AIS	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 (pass mark) 2- Needs improvement 1- Does not meet requirements
G6 - Upholding ethical integrity	The duty holder shall fulfill the following transparency requirement(s):				The following item(s) shall be presented as evidence for conformity against the transparency requirement(s):	Two-tier approach measurement of the evidence items:

Transparency schema goal description	Transparency 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
This goal is concerned with upholding the primacy of transparency as a concern throughout the life of the AIS. This also caters to changes in ethical norms or technology that may invalidate prior assumptions.	a) Demonstrate that efforts are put in place to include accountability criteria/behaviors as part of the AIS ethical profile b) Mapping an algorithmic AIS ethical profile to the organizational ethical policies and values	N N	LI LI	D, I, O, M, R D, I, O, M, R	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	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 (pass mark) 2- Needs improvement 1- Does not meet requirements
G1b - Behavioral obfuscation This relates to AIS and autonomous systems minimizing their apparent spillover effects (externalities in economics terms), such as pollution, whether by intentional design or incidental omission due to the challenges of adequately detecting, accounting for,	The duty holder shall fulfill the following transparency requirement(s): a) All system behaviors that may affect third parties are taken note of, correctly logged, and no attempt to cover them up is made; any such obfuscation should be disclosed, and a plan of action taken to minimize the effects.	N	LI	D, I, O, M, R	The following item(s) shall be presented as evidence for conformity against the transparency requirement(s): a) Notes and logs of all system behaviors that may affect third parties; no evidence of attempts to cover them up; disclosures of any such obfuscation; and a documented plan of action to minimize the effects.	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:

Transparency schema goal description	Transparency 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
and managing externalities.						5- Excels baseline requirements 4- Sustains baseline requirements 3- Meets baseline requirements (pass mark) 2- Needs improvement 1- Does not meet requirements
G2b - Concern with liability The service provider's awareness of potential risk exposure, and delivery of bare minimum (or inadequate) information to manage the risk. This could include legal, commercial, financial, and human intervention dimensions.	The duty holder shall fulfill the following transparency requirement(s): a) Transparency should be given priority over concern for legal exposure at all levels of the organization b) Adequate transparency in user documents c) User manual stating organization and stakeholder responsibilities clearly. d) Presence of transparency-related legal cases of product.	N N N N	LI LI LI LI	D, I, O, M, R D, I, O, M, R D, I, O, M, R D, I, O, M, R	The following item(s) shall be presented as evidence for conformity against the transparency requirement(s): a) Possessing adequate insurance where applicable b) Details of cases where insurance was claimed, especially with regard to ethical issues c) Legal counsel shall make a precommitment to give primacy to transparency d) An absence of any court cases related to transparency of product e) User manual with stakeholders' responsibilities/liabilities in various scenarios	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 (pass mark) 2- Needs improvement 1- Does not meet requirements

Transparency schema goal description	Transparency 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
G3b - Protection of trade secrets Enterprises' desire to protect their intellectual property (IP) through insufficient transparency or obfuscation.	The duty holder shall fulfill the following transparency requirement(s): a) Organizations shall not use protection of trade secrets/IP as a basis to minimize/avoid transparency.	N	LI	D, I, O, M, R	The following item(s) shall be presented as evidence for conformity against the transparency requirement(s): a) Documented transparency best practice that explains the necessity and rationale for choices and compromises made, which should be in line with prioritizing transparency over IP protection	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 (pass mark) 2- Needs improvement 1- Does not meet requirements
END						

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] Benschop, Thijs, Cathrine Machingauta, and Matthew Welch, "Statistical Disclosure Control for Microdata: A Practice Guide," Jul. 26, 2021.[4]

[B3] Brownlee, Jason, "A gentle introduction to k-fold cross-validation," *Machine Learning Mastery Blog*, last modified Aug. 3, 2020.[5]

[B4] Brownlee, Jason, "Why do you get different results on different runs of an algorithm with the same data?" *Machine Learning Mastery Blog*, last modified Aug. 5, 2016.[6]

[B5] Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy, and Dario Amodei, "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," Future of Humanity Institute, Report, Feb. 2018.[7]

[B6] "Ethics Guidelines for Trustworthy AI," High-Level Expert Group on Artificial Intelligence (AI HLEG), European Commission, Apr. 2019.[8]

[B7] "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.[9]

[B8] Floridi, Luciano, "Soft ethics and the governance of the digital" *Philosophy & Technology*, vol. 31, no. 1, pp. 1–8, 2018.

[B9] Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena, "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and Machines,* vol. 28, no. 4, pp. 689–707, 2018.

[B10] 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.

[B11] "G20 AI Principles," *G20 Ministerial Statement on Trade and Digital Economy*, Annex, Jun. 2019.[10]

[3] United Nations publications are available from the United Nations website (https://www.un.org).
[4] Available from https://buildmedia.readthedocs.org/media/pdf/sdcpractice/latest/sdcpractice.pdf.
[5] Available from https://machinelearningmastery.com/k-fold-cross-validation/.
[6] Available from https://machinelearningmastery.com/randomness-in-machine-learning/.
[7] Available from https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.
[8] European Commission publications are available from the Futurium website (https://futurium.ec.europa.eu/en).
[9] IEEE publications are available from the Institute of Electrical and Electronics Engineers, 445 Hoes Lane, Piscataway, NJ 08854-4141, USA (http://standards.ieee.org).
[10] Available from https://www.mofa.go.jp/files/000486596.pdf.

[B12] Hajkowicz, Stefan, Sarvnaz Karimi, Tim Wark, Caron Chen, M. Evans, Natalie Rens, Dave Dawson, Andrew Charlton, Toby Brennan, Corin Moffatt, Sriram Srikumar, and K.J. Tong, "Artificial intelligence: Solving problems, growing the economy and improving our quality of life" Commonwealth Scientific and Industrial Research Organisation (CSIRO), 2019.[11]

[B13] Howell, David, "Resampling Statistics: Randomization and the Bootstrap." University of Vermont, Oct 14, 2015.[12]

[B14] 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.

[B15] Malgieri, Gianclaudio, and Giovanni Comandé, "Why a right to legibility of automated decision-making exists in the general data protection regulation" *International Data Privacy Law*, vol. 7, no. 4, pp. 243–265, Nov. 13, 2017.

[B16] Muralidhar, Krishnamurty, and Rathindra Sarathy, "Data shuffling—A new masking approach for numerical data," *Management Science*, vol. 52, no. 5, pp. 658–670, 2006.

[B17] OECD/LEGAL/0449, *Recommendation of the Council on Artificial Intelligence*, May 21, 2019.[13]

[B18] "OECD Due Diligence Guidance for Responsible Business Conduct," OECD, 2018.

[B19] "OECD Guidelines for Multinational Enterprises," Update 2011, OECD, 2011.

[B20] Prabhu , "Understanding hyperparameters and its optimization techniques" Towards Data Science (blog), Jul. 3, 2018. [14]

[B21] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (OJ L 119 04.05.2016, p. 1), Apr 27, 2016.[15]

[B22] "Resampling Method" *ScienceDirec*t.[16]

[B23] "The State of AI Ethics," Montreal AI Ethics Institute, Jan. 2021.[17]

[B24] "Sustainable Development Goals," Transforming our World: the 2030 Agenda for Sustainable Development, United Nations, 2015.

[B25] Watson, Eleanor, ECPAIS TEFG Presentation, IEEE Berlin Meeting, Sept. 30, 2019.

[B26] Wong, Jenna, Travis Manderson, Michal Abrahamowicz, David L Buckeridge, and Robyn Tamblyn, "Can hyperparameter tuning improve the performance of a super learner? A case study," *Epidemiology*, vol. 30, no. 4, pp. 521–531, Jul. 2019.

[11] Available from https://apo.org.au/node/268341.
[12] Available from https://www.uvm.edu/~dhowell/StatPages/ResamplingWithR/ResamplingR.html.
[13] Organisation for Economic Co-operation and Development publications available from the OECD website (https://www.oecd.org/).
[14] Available from https://towardsdatascience.com/understanding-hyperparameters-and-its-optimisation-techniques-f0debba07568.
[15] Access to European Union legal documents available from EUR-Lex (https://eur-lex.europa.eu/homepage.html).
[16] Available from https://www.sciencedirect.com/topics/mathematics/resampling-method.
[17] 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