IEEE SA STANDARDS ASSOCIATION

Verifying Ethics in AI-based solutions

AI without Ethics is not sustainable. How do you implement and verify AI Ethics? How do you demonstrate it, to enable adoption of your solutions?

Abstract

AI is now an intrinsic part of solution development. Like any new technology, it brings risks and exposures that need to be addressed to generate the trust needed for market adoption. Yet most organizations have a nascent understanding of AI-based ethical considerations, and there is a need for regulations and established legal guidance in the development and assessment of AI-based solutions. IEEE CertifAIEd™ criteria address many ethical facets of the development and deployment of autonomous intelligent systems (AIS). The IEEE CertifAIEd mark provides organizations with the visibility that demonstrates their commitment at safeguarding transparency, accountability, algorithmic bias, and privacy to build trust in their AIS.

The need for Ethics in AI

AI has taken center stage in our lives. It is now pervasively used in technologies that surround us and that influence every aspect of our daily routines. Today's news echoes how consumers are becoming more aware of the impact of algorithms. Much of the impact is beyond technical concerns but around ethical dimensions of their AI experience. More is needed to secure against the risks created by the unintended consequences of AI optimizations, which are often amplified by the automation and efficiency gain brought in by the AI. Unethical unintended consequences of those solutions are not always readily apparent, yet they can be prevalent and impactful on society. Factoring ethics in AIS will help protect, differentiate, and grow product adoption.

The challenges of factoring in Ethics in AIS

While capabilities for technical development and deployments of AIS are widely available, the same level of capabilities are not readily formalized for AI Ethics. The possible lack of a clear AI Ethics focused value proposition, and the lack of a contextual model and guidance on what constitutes an ethical solution impede an organization's efforts to integrate AI Ethics into its development. Even when ethical criteria are obvious or well-defined, the translation of such high-level values into actionable evaluation criteria is not straightforward. The thoroughness of the exercise is dependent on internal practices and personal competencies of the individuals involved. Further, such AI Ethics practices vary, limiting their re-usability outside of the organization and possibly the option for independent verification.

AI Ethics is a nascent discipline. Most organizations and product development teams are either unaware of or unclear on how to reach out to AI Ethics professionals for help in their solution development. There exist today international institutions' policy recommendations such as those of the OECD or UNESCO[1], or the Global Partnership on AI Framework. For example, the EU AI Act is planned for launch in 2023 with enforcement to be enacted shortly thereafter. Given that product development lifecycles are roughly between a year or two, organizations would likely want to address their AI Ethics development needs early in their endeavors. IEEE CertifAIEd™ offers such criteria and methodology for the assessment and certification of an AIS against ethical risks.

[1] https://oecd.ai/en/ai-principles, https://en.unesco.org/artificial-intelligence/ethics#drafttext,

CertifAIEd: IEEE's AI Ethics Certification Program

IEEE's CertifAIEd program objective is to enable, enhance and reinforce *trust* through AI Ethics specifications, training, criteria, and certification. It stems from the rationale that an entity benefits from an independent ethical evaluation and certification of its AIS. The IEEE CertifAIEd Mark communicates additional confidence for entities that have their AIS ethically aligned with expected and consistent behaviors and conveys trust to customers and consumers.

IEEE CertifAIEd outlines how ethical concepts are influenced by driving and detrimental factors. These factors are derived from a schema of possibly several hundred criteria and sub-criteria that serve for the AIS assessment. Ethical foundational requirements for this criterion are specified, their impact is identified, and the required evidence is outlined. The provision of evidence against the enunciated requirements allows an entity to testify about the adherence of its AIS to specific ethical criteria.

The IEEE CertifAIEd™ Criteria

From the ethical perspective, four key areas of concern in development and deployment of solutions relate to *Privacy, Accountability*, *Transparency,* and freedom from unacceptable *Algorithmic Bias*.

Transparency relates to the criteria and values embedded in a system design and the openness and disclosure of choices made for development and operation. This applies to the entire ethical AI context of application for the product or service under consideration such as data sets and not restricted to technical and algorithmic aspects alone.

By way of example, the high-level principles for the IEEE CertifAIEd Transparency criteria decompose into driving and detrimental factors as depicted in Figure 1.

IEEE CertifAIEd Transparency Criteria

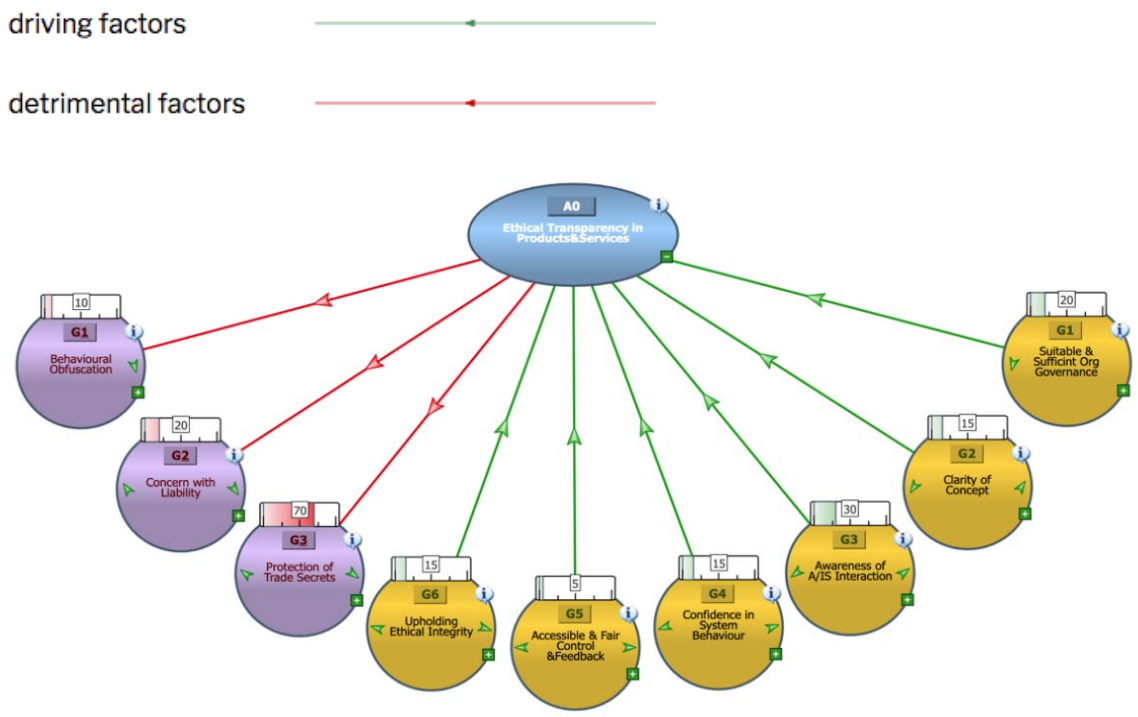

Figure 1

Accountability considerations concern the duty for individuals and institutions involved in the design, development, or deployment of ethical AI solutions to remain responsible for the behavior of the system as long as its integrity is respected. This assumes that the solution's AI autonomy and learning capacities are the result of algorithms and computational processes designed by humans who should remain responsible for their outcomes. A key driver in accountability is explicit, sufficient, and proper documentation and traceability for system design, development, and deployment.

Algorithmic bias relates to systematic errors and repeatable undesirable behaviors in an AI solution that create unfair outcomes, such as granting privileges to one group of users over others where they're expected to be neutral and unbiased. This can emerge due to many factors, from the design of the algorithm influenced by pre-existing cultural or institutional practices, the decisions relating to the way data is classified, collected, selected, or used to train the algorithm, the unanticipated context of application and even presentational aspects emerging from search engines and social media.

Privacy relates to the private sphere of life and public identity of an entity (individual, group, community) upholding dignity. It extends beyond the notion of compliance with privacy as currently denoted in the law, to ethical privacy, defined as a contextual set of values pertaining to privacy and the satisfaction of a framework of expectations (preservation of autonomy, self-determination, and self-selected communities/locum and intimacies).

The IEEE CertifAIEd™ – Ontological Specifications for Ethical Transparency, Accountability, Privacy and Algorithmic Bias offer a first level of insight into the criteria and are each published under a Creative Commons BY-NC-ND 4.0 license. These are extracted from the comprehensive IEEE licensed material that includes the details on the several hundred criteria.

Organizations that download the Ontological Specifications can indicate their commitment to AI Ethics via a sticker that is provided together with the documents. The Ontological Specifications can be accessed at https://engagestandards.ieee.org/ieeecertifaied.html.

The IEEE CertifAIEd™ Process

The IEEE CertifAIEd process is a collaboration between an organization's development team and IEEE authorized experts in which the organization's AIS is assessed, recommendations are made for needed adjustments, and verification and certification ensue, leading to a mark attesting to the AIS's capability to fulfill applicable requirements.

The IEEE CertifAIEd process consists of three core phases: (1) an exploratory profiling phase where the AIS ethical scope, complexity, and impact are clearly defined and agreed upon between the parties; (2) an assessment phase during which the organization's AIS is evaluated against the identified criteria and evidence requirements provided; and (3) an independent certification phase, resulting in the issuance of a CertifAIEd mark, attesting to the AIS's conformance to ethical criteria.

What to Expect

Experts are available to help guide you through the steps leading to an IEEE CertifAIEd mark:

1 ENQUIRY

An IEEE Authorized Assessor(s) meets with you to discuss your goals, expected outcomes, and any additional considerations for the assessment, such as timeframe, budget, and participating personnel. The product, service, or system to be assessed is identified and its concept of operations is described. During this stage, the project scope for the assessment and certification are agreed upon.

2 ETHICAL PROFILING

The IEEE Authorized Assessor(s) works with you to develop an Ethical Risk and Reward Profile, where both parties agree on the human / socio-technical values affected by the product, and their relative impacts to human, societal, and environmental well-being. Based on this activity, the product's ethical risk profile and the applicable criteria suite(s) are identified.

For more information, contact us at certifaied@ieee.org.

3 ASSESSMENT

The Assessor harvests and provides you with the appropriate criteria set based on your product's ethical risk profile and criteria suite(s). You then collect and submit evidence that demonstrates that your product meets each of the criteria provided. The Assessor helps you to clarify the ethical foundational requirements associated with each of the criteria, and provides feedback on the evidence provided. Based on the evidence and feedback, you submit a Case for Ethics document that details how your product demonstrates conformance with the applicable criteria.

4 CERTIFICATION AND MARK

Upon completion of an assessment, an IEEE Authorized Certifier conducts an independent and comprehensive review of the Case for Ethics document and related material and provides a detailed Assessment Report which may include suggestions for improvement. The Certifier validates the results then issues a certificate indicating that your product, service, or system has met the relevant ethical criteria. The Certifier grants you the IEEE CertifAIEd mark and adds you to the CertifAIEd registry.

The IEEE CertifAIEd™ Ecosystem

IEEE supports entities by providing a trained ecosystem of independent professionals and collaborators to assist in the assessment and certification of the ethical dimensions of AIS. The list of these assessment and certification collaborators will be reflected in the IEEE CertifAIEd ecosystem registry.

How to engage with IEEE CertifAIEd

If you are interested in leveraging IEEE CertifAIEd for your solutions or in your organization, please contact certifAIEd@ieee.org or go to the IEEE CertifAIEd website[2]

"IEEE has laid the groundwork for AI Ethics based on principles and standards created by hundreds of our volunteers over the past five years, which are already having a global impact. IEEE CertifAIEd represents our continued evolution of the AI Ethics ecosystem by establishing a program to inspire trust and a means towards responsible implementation of AI systems that demonstrates an organization's commitment to upholding human values, dignity, and well-being, and to respecting, protecting, and preserving fundamental human rights.[3]"

-Konstantinos Karachalios, Managing Director of IEEE SA.

"Data security and data protection must be at the forefront when using AI from the very beginning. That's why we relied on international expertise (from IEEE) during the development of the software and had our program ethically certified.[4]"

-Deputy Director General for the City of Vienna, Peter Weinelt.

Technical References

- IEEE CertifAIEd™ – Ontological Specification for Ethical Privacy
- IEEE CertifAIEd™ – Ontological Specification for Ethical Algorithmic Bias
- IEEE CertifAIEd™ – Ontological Specification for Ethical Transparency
- IEEE CertifAIEd™ – Ontological Specification for Ethical Accountability

[2] https://engagestandards.ieee.org/ieeecertifaied.html

[3] City of Vienna Earns IEEE AI Ethics Certification Mark; Reinforcing Commitment to Digital Humanism Strategy

[4] ibid