The First Global Ontological Standard for Ethically Driven Robotics and Automation Systems

By Edson Prestes, Michael A. Houghtaling, Paulo J.S. Gonçalves, Nicola Fabiano, Ozlem Ulgen, Sandro Rama Fiorini, Zvikomborero Murahwi, Joanna Isabelle Olszewska, and Tamás Haidegger

In the complex and rapidly evolving fields of artificial intelligence (AI) and robotics, the elaboration of ethical concerns, considerations, and requirements helps elucidate the nature of technology's reach and impact on society where there is a legal void. Thus, establishing ethics in AI and robotics is fundamental to identifying their potential risks and benefits, especially in our pandemic-wrecked world [1].

The development of AI and robotics within an ethical framework enables the anticipation of future application contexts and articulation of uses that do not yet exist. Ethical considerations help to create a much-desired relationship between technology and human values and address the impacts a technology can have, thereby addressing issues of trust, safety, security, data privacy, and algorithmic bias. The need for an ethical framework is urgent because of the increasing adoption and use of autonomous and intelligent systems (A/ISs) in many domains, such as health care, education, finance, and insurance services. Ethically aligned technology has a clear role in supporting the achievement of the United Nations (UN) Sustainable Development Goals (SDGs) [2], [3].

In 2016, IEEE established its Global Initiative on Ethics of Autonomous and Intelligent Systems with the aim of ensuring that every stakeholder involved in the design, development, and management of A/ISs is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity. One of the efforts conducted by this initiative focuses on the development of soft laws (e.g., standards and guidelines) to help shape the responsible development and use of A/ISs.

With this aim, the IEEE Robotics and Automation Society (RAS)/Standards Association (SA) 7007 Ontologies for Ethically Driven Robotics and Automation Systems Working Group (IEEE 7007 WG) was established in 2017. During the past four years, this group has been working to create an ontological standard to enable the development of ethically driven robotics and automation systems. This standard was scrutinized by the global community in 2021, and it was officially approved by the IEEE SA on 24 September 2021. Due to the relevance of this standard, the IEEE 7007 WG has been selected as a recipient of the IEEE SA Emerging Technology Award "for developing an innovative ontological standard on the ethics of artificial intelligence" (see "RAS Standard Receives IEEE SA Emerging Technology Award!").

Regulatory Frameworks

There are various international regulatory initiatives in the area of emerging technologies with an impact on AI and robotics [4]. Current international regulatory requirements are contained in a combination of nonlegally binding ethical standards, frameworks, and guidelines as well as legally binding instruments [5].

Examples include the 2019 OECD *Recommendation on AI*; 2019 G20 *Human-Centered AI Principles*; 2019 European Union (EU) *Ethics Guidelines for Trustworthy AI*; 2019 recommendations of the UN Secretary-General's High-Level Panel on Digital Cooperation; 2019 IEEE *Ethically Aligned Design*; 2015 UN SDGs; and BS 8611:2016, *Ethical Design and Application of Robots*. More recently, in 2021, there has been an elaboration of the draft of the very first international normative instrument by the UN Educational, Scientific, and Cultural Organization on the ethics of AI. Examples of legal requirements from international,

RAS Standard Receives IEEE SA Emerging Technology Award!

The IEEE 7007 WG has been selected as a recipient of the IEEE SA Emerging Technology Award "for developing an innovative ontological standard on the ethics of artificial intelligence." The chair of the group is Edson Prestes, the vice chair is Sandro Fiorini, the technical editors are Mike Houghtaling and Babita Ramlal, and the secretary is Paulo J.S. Gonçalves.

The IEEE SA Emerging Technology Award is awarded for the initiation, advancement, or progression of a new technology through the IEEE SA open consensus process. Further information about the award, including a list of past recipients, may be found at https://standards.ieee.org/about/awards/etech/index.html.

Digital Object Identifier 10.1109/MRA.2021.3117414
Date of current version: 9 December 2021

regional, and/or national bodies include the 2016 EU General Data Protection Regulation, bilateral and multilateral treaties, and the 2018 Council of Europe "Modernised Convention for the Protection of Individuals With Regard to Automatic Processing of Personal Data" (Convention 108+).

The IEEE Ethics Certification Program for Autonomous and Intelligent Systems [6] is a world first in setting standards for the ethical certification of products, services, and systems deploying AI and robotics in the public and private sectors. Certification is essential to guarantee that these technologies operate as expected when they are interacting with human and nonhuman agents. For stakeholders involved directly and indirectly in the lifecycle of AI and robotics systems, certification guarantees that these systems will cause no harm, their limitations are known, and there will be human accountability and responsibility for their use. This, in turn, fosters greater societal confidence in the technology's utilization.

Different from these frameworks, the standard developed by the IEEE 7007 WG has a formal and ontological representation that can be used not only as a foundation to elaborate public policies but also to create computational systems. In fact, IEEE Standard 7007 is the first global ontological standard that contains the concepts, definitions, and axioms that are necessary to establish ethical methodologies for the design, development, and deployment of AI and robotics.

IEEE 7007 WG

The IEEE 7007 WG is under the umbrella of the IEEE SA P7000 series devoted to ethics in A/IS. In this scope, several WGs were formed—15 to date—to deliver a broad range of standards and/or recommended practices. Among the goals of the IEEE 7007 WG are to

- establish a set of definitions and their relationships that will enable the development of robotics and automation systems in accordance with worldwide ethics and moral theories
- align the ethics and engineering communities to understand how to pragmatically design and implement these systems in unison
- develop a precise communication framework among global experts of different domains, including robotics, automation, and ethics.

To attain these goals, the IEEE 7007 WG developed a set of ontologies for representing the domain in a more precise way. As a result, IEEE Standard 7007 contains a set of ontologies that represents norms and ethical principles (NEP), data privacy and protection (DPP), transparency and accountability, and ethical violation management (EVM). The 2development of this standard was a complex process requiring a dedicated lifecycle. For this purpose, the IEEE 7007 WG developed an agile, collaborative, and iterative methodology called the *robotic standard development lifecycle* [7].

The usefulness of ontologies in standardization is twofold. On the one hand, standardization processes are set to produce a body of knowledge that reflects a consensual view of practitioners around a topic, defining, among other aspects, a standard knowledge structure in a domain, including common concepts, relationships, and attributes. Ontologies and their methods provide a formal approach to that aspect of the standardization process, which is expected to produce a sounder standard. On the other hand, the ontologies themselves, as formal artifacts, can be seen as products of the standardization process that can be used directly in data processing and automatic reasoning. As an example, one can cite IEEE 1872-2015 [8], which set forth to establish clear definitions for common terms in robotics and automation.

IEEE 7007 Ontological Standard for Ethically Driven Robotics and Automation Systems

Top-Level Ontology

As a core ontology, the ethically driven robotics and autonomous systems (ERAS) ontology represents a midlevel set of formalizations and commitments that are platform independent and intended to fit between an upper top-level or foundational ontology and lower-domain and application-specific ontologies. While some potential users of the standard may intend to align the ERAS core formalizations with existing top-level ontologies specific to their application domain, other user communities will only require a minimal top-level set of conceptualizations to complete the formalization of the concepts, terms, and commitments axiomatized in the ERAS ontology.

For that purpose, the four ERAS subdomain ontologies are augmented with axioms sufficient to complete the definitions and commitments expressed in the core ERAS models. These axioms are expressed formally using the Common Logic Interchange Format (CLIF) [9]. The ERAS top-level ontology (ERAS-TLO) formalizations define a minimal set of terms deemed relevant to the characterization of ethically oriented agents and autonomous systems. It is not intended to be applicable as a TLO in other contexts.

NEP Ontology

The NEP ontology subdomain formalizes the terminology and ontological commitments associated with ethical theories and principles that characterize the norms of expected behaviors for norm-oriented agents and autonomous systems. This includes axioms for concepts, such as norms, ethical theory, situation plan repertoire, agent plans, plan actions, and agent actions as well as the corresponding relationships, such as "selects plans from," "subscribes to," "satisfies," and "constrains plans for." Figure 1 depicts a brief and partial view of a subset of the NEP terms with a Unified Modeling Language (UML) class diagram.

DPP Ontology

The DPP ontology represents concepts and relationships among the diverse agents, entities, and organizations that may be involved at different stages in data gathering, processing, transfer, retention, and storage and in which autonomous systems may be deployed. Thus, the DPP ontology represents concepts like the natural person, caregiver, data protection authority, controller, and authorized accessor as well as the different types and processing of personal data (e.g., health data, economic data, and social data) and corresponding data process access. DPP principles, like privacy by design, data protection by design, data protection by default, and human rights by design, were also included in the standard.

It is crucial to represent this domain formally because of the relevance of the existing regulations worldwide about DPP. In addition, evaluating the impact of driven robotics and automation systems on personal data and, hence, on the processing of personal information is essential to the regulation of A/IS. As stated in the standard, "Data privacy is a highly complex and increasingly regulated area of law, in which the regulatory regime is rapidly evolving. No standard can provide unconditional consistency with all applicable laws and regulations, which continue to change rapidly in this area, and may also vary at the local, state and regional level. Users of this Standard are responsible for keeping apprised of such laws and regulations."

Transparency and Accountability Ontology

The transparency and accountability ontology subdomain formalizes the vocabulary and ontological commitments relevant for terms capable of expressing the concepts and relationships necessary to enable ethical autonomous systems with capabilities that provide informative explanations for plans and associated actions. Ethically aware agents require the ability to be transparent in their interactions with other agents. An agent qualifies as an autonomous transparent agent if it is enabled with an always-available mechanism capable of reporting its behavior, intentions, perceptions, goals, and constraints in a manner that permits authorized users and collaborating agents to understand its past and expected future behaviors. To express these capabilities, this ontology includes axioms for concepts, such as explanation, agent explanation plan, explanation plan repertoire, discourse content, agent data, transparency concern, audience, and content provenance, along with corresponding relationships, such as "determines what to explain," "determines how to explain," "formulates," "expressed_in," "authenticates," and "is accountable for."

EVM

The EVM ontology subdomain presents axioms to formalize the terminology associated with capabilities to detect, assess, and manage ethical and legal norm violations occurring within or generated by autonomous system behavior. This includes concepts such as norm violation, norm violation incident, responsibility ascription, ascription justification, grounds for ascription, agent accountability, event causation, liability sanction, and ethical behavior monitor. Figure 2 presents a partial view of the EVM concepts and relationships in a UML class diagram.

During an ethically aware agent's interaction with the environment and other agents, some norms can be violated. A norm violation is an action event reflecting a failure to conform to the norm's rules of behavior relevant to

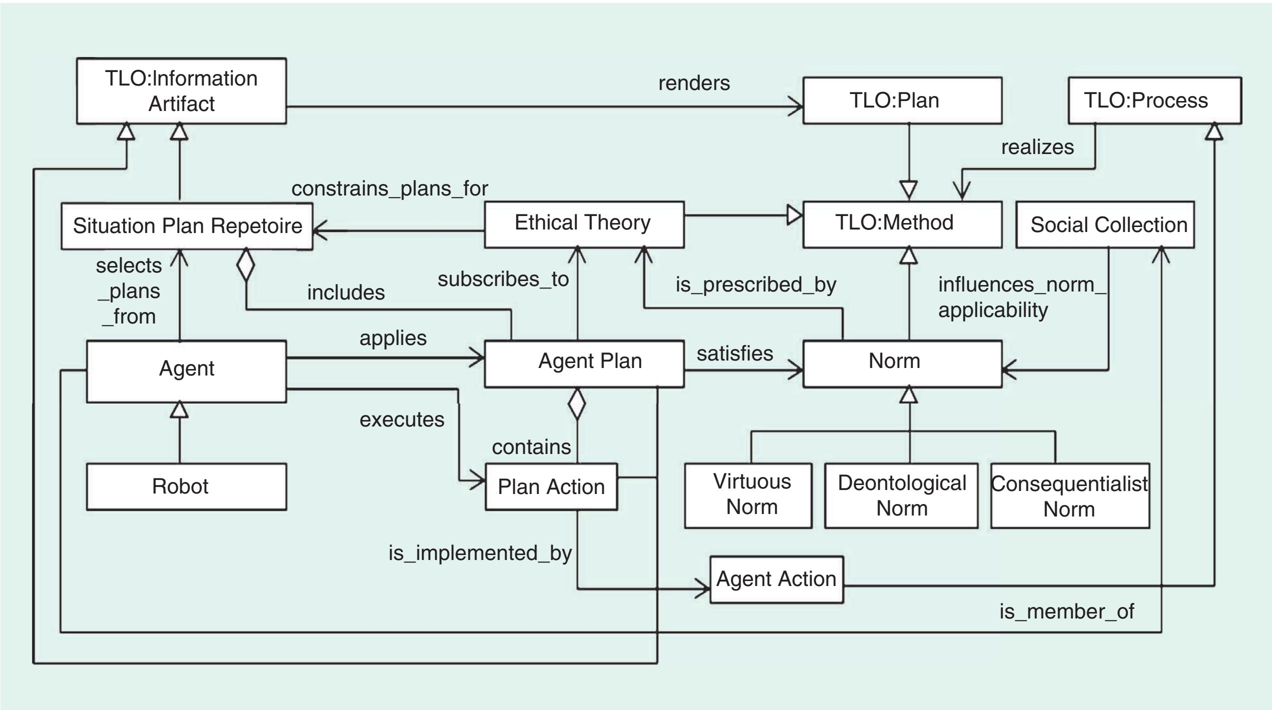

Figure 1. A partial UML model of the ERAS NEP ontology. UML: Unified Modeling Language.

the agent's situation. Agent system components or other agents providing an ethical behavior monitoring service may detect and record norm violations using norm violation incident information artifacts. A norm violation elicits a responsibility ascription process as a social interaction process to identify those responsible for the violation. A responsibility ascription process that results in the ascription of responsibility to one or more agents is justified by an ascription justification information artifact. This category represents the collection of facts formulated and asserted by an authoritative agent or agency to ascribe responsibilities for ethical or legal norm violations. It is composed of constituent grounds for ascription information artifacts.

Ethical violation as well as transparency and accountability ontologies identify accountability and legal responsibility as important real-world concepts impacting AI and robotics. Legal responsibility and its manifestations in terms of culpability as well as civil and criminal liability [10], [11] have influenced the content of the standard. The parameters between accountability and responsibility are also reflected with use of terminology that conveys a spectrum of potential agents who may be held responsible (e.g., partial or distributed responsibility).

An important observation here is that the EVM core axioms restrict autonomous system agent responsibility ascription to a set of specific system ethical norm violations and when human agents are involved in the collective distributed responsibility chain. Autonomous systems cannot be ascribed any responsibility for legal norm violations. An autonomous system acting as a single agent cannot be ascribed responsibility for any type of norm violation. Distributed responsibility is applicable only when the autonomous system is a member of a human-directed team and when an action by the system caused a norm violation.

Conclusions

IEEE Standard 7007 is the first global ontological standard elaborated to establish ethical methodologies for the design, development, and deployment of A/IS. It contains a set of ontologies that represents, explicitly and formally, core concepts that are relevant to dealing with NEP, transparency and

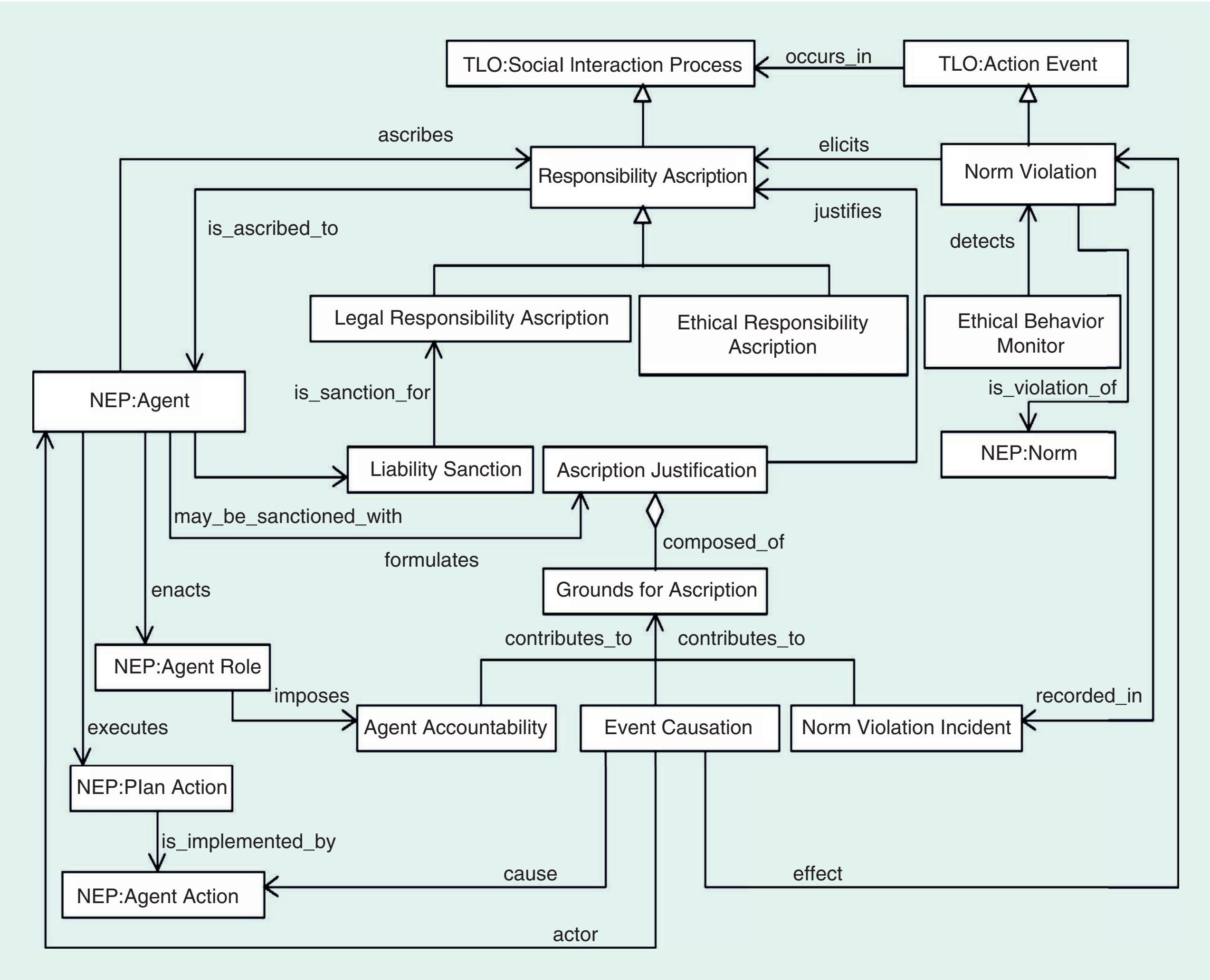

Figure 2. A partial UML model of the ERAS EVM ontology.

accountability, EVM, and DPP. It is expected that this work has a significant impact worldwide in being used to teach ethical design; for both human and institutional capacity building in the domain of the ethics of AI; to create computational ethically aligned systems; to create a taxonomy to support the elaboration of public policies; and to strengthen digital cooperation across nations applied together with the other members of the IEEE P7000 family.

References

[1] A. Khamis et al., "Robotics and intelligent systems against a pandemic," *Acta Polytechnica Hungarica*, vol. 18, no. 5, 2021. doi: 10.12700/APH.18.5.2021.5.3.

[2] A. Khamis, H. Li, E. Prestes, and T. Haidegger, "AI: A key enabler of sustainable development goals, Part 1 [Industry Activities]," *IEEE Robot. Autom. Mag.*, vol. 26, no. 3, pp. 95–102, 2019. doi: 10.1109/MRA.2019.2928738.

[3] A. Khamis, H. Li, E. Prestes, and T. Haidegger, "AI: A key enabler for sustainable development goals: Part 2 [Industry Activities]," *IEEE Robot. Autom. Mag.*, vol. 26, no. 4, pp. 122–127, 2019. doi: 10.1109/MRA.2019.2945739.

[4] T. Jacobs, J. Veneman, G. V. Virk, and T. Haidegger, "The flourishing landscape of robot standardization [Industrial Activities]," *IEEE Robot. Autom. Mag.*, vol. 25, no. 1, pp. 8–15, 2018. doi: 10.1109/MRA.2017.2787220.

[5] O. Ulgen, "User rights and adaptive A/IS – From passive interaction to real empowerment," in *Proc. Int. Conf. Human-Comput. Interaction*, R. A. Sottilare and J. Schwarz, Eds. Cham, 2020, vol. 12214, pp. 205–217. doi: 10.1007/978-3-030-50788-6_15.

[6] "The ethics certification program for autonomous and intelligent systems (ECPAIS)," IEEE SA, Piscataway, NJ. https://standards.ieee.org/industry-connections/ecpais.html

[7] J. I. Olszewska et al., "Robotic standard development life cycle in action," *J. Intell. Robot. Syst.*, vol. 98, no. 1, pp. 119–131, 2020. doi: 10.1007/s10846-019-01107-w.

[8] *IEEE Standard Ontologies for Robotics and Automation*, IEEE Standard 1872-2015.

[9] "*Information Technology-Common Logic (CL)-A Framework for a Family of Logic-Based Languages*," ISO/IEC 24707:2018, International Organization for Standardization, Geneva, Switzerland, July 2018.

[10] O. Ulgen, "A human-centric and lifecycle approach to legal responsibility for AI," *Commun. Law J.*, vol. 26, no. 2, pp. 96–107, 2021.

[11] R. van den Hoven van Genderen, "Do we need new legal personhood in the age of robots and AI?" in *Robotics, AI and the Future of Law*, M. Corrales, M. Fenwick, and N. Forgó, Eds. Singapore: Springer-Verlag, 2018, pp. 15–55.

Edson Prestes, Informatics Institute, Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90040-060, Brazil. Email: edson.prestes@ieee.org.

Michael A. Houghtaling, IBM–retired, Arizona, USA. Email: michael-h@acm.org.

Paulo J.S. Gonçalves, IDMEC, Instituto Politécnico de Castelo Branco, Castelo Branco, 6000-084, Portugal. Email: paulo.goncalves@ipcb.pt.

Nicola Fabiano, Studio Legale Fabiano, International Institute of Informatics and Systemics, Florida, USA. Email: nicola.fabiano@ieee.org.

Ozlem Ulgen, School of Law, University of Nottingham, Nottingham, NG7 2RD, U.K. Email: ulgeno@hotmail.co.uk.

Sandro Rama Fiorini, IBM Research, Paraíso, São Paulo, 04007-005, Brazil. Email: srfiorini@ibm.com.

Zvikomborero Murahwi, Independent IT Consultant, South Africa. Email: zviko.murahwi@ictprojectsadvisory.com.

Joanna Isabelle Olszewska, University of the West of Scotland, Glasgow, G72 0LH, U.K. Email: joanna.olszewska@ieee.org.

Tamás Haidegger, University Research and Innovation Center, Óbuda University, Budapest, 1034, Hungary. Email: haidegger@ieee.org.

Digital Object Identifier 10.1109/MRA.2021.3117465