The first global ontological standard for ethically driven robotics and automation systems

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

Published in:
IEEE Robotics and Automation Magazine

DOI:
10.1109/MRA.2021.3117414

Published: 10/12/2021

Document Version
Peer reviewed version

Prestes, E., Houghtaling, M. A., Goncalves, P. J. S., Fabiano, N., Ulgen, O., Fiorini, S. R., Murahwi, Z., Olszewska, J. I., & Haidegger, T. (2021). The first global ontological standard for ethically driven robotics and automation systems. *IEEE Robotics and Automation Magazine*, *28*(4), 120–124. https://doi.org/10.1109/MRA.2021.3117414

IEEE 7007 the First Global Ontological Standard for Ethical Robotics and Automation Systems

Edson Prestes, Michael Houghtaling, Paulo Gonçalves, Nicola Fabiano, Ozlem Ulgen, Sandro R. Fiorini, Zvikomborero Murahwi, Joanna Isabelle Olszewska, and Tamás Haidegger

In the complex and rapidly evolving field of Artificial Intelligence (AI) and robotics, 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 [3]. Development of AI and robotics within an ethical framework enables anticipating future contexts of application, and articulation of uses that do not yet exist. Ethical considerations help to create the much desired relationship between technology and human values and address the impacts the technology can have thereby addressing issues of trust, safety, security, data privacy, algorithmic bias. The need for an ethical framework is urgent because of the increasing adoption and use of autonomous and intelligent systems in many domains such as healthcare, education, financial and insurance services. Ethically aligned technology has a clear role in supporting the achievement of the UN Sustainable Development Goals (SDGs) [4][5].

In 2016, the IEEE established its IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems with the aim to ensure that every stakeholder involved in the design, development and management of Autonomous and Intelligent Systems (A/IS) is educated, trained and empowered to prioritise Ethical considerations so that these technologies are advanced for the benefit of humanity. One of the efforts conducted by initiative focuses on the development of soft laws (e.g., standard, guidelines) to help shaping the responsible development and use of A/IS. With this aim, the IEEE RAS/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 scrutinised by the global community in 2021 and it is currently in the final steps for its official approval by the IEEE Standards Association.

Regulatory Frameworks

There are various international regulatory initiatives in the area of emerging technologies with impact on AI and robotics [6]. Current international regulatory requirements are contained in a combination of non-legally binding ethical standards, frameworks, and guidelines, as well as legally binding instruments [7]. Examples of ethical standards include the 2019 OECD Recommendation on AI, 2019 G20 Human-Centred AI Principles, 2019 EU Ethics Guidelines for Trustworthy AI, 2019 Recommendations of UN SG's HLP on Digital Cooperation, 2019 IEEE EAD, 2015 UN Sustainable Development Goals, and the BS 8611:2016, ethical design and application of robots. More recently, in 2021, there has been elaboration of the draft of the very first international normative instrument by UNESCO on Ethics of Artificial Intelligence. Examples of legal requirements from international, regional,

and/or national bodies include the 2016 EU GDPR, 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 ECPAIS [8] (Ethical Certification Programme for Autonomous and Intelligent Systems) is a world first in setting ethical standards for 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 non-human 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 use.

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, the IEEE 7007 Standard is the first global ontological standard that contains concepts, definitions, and axioms which are necessary to establish ethical methodologies for the design, development, and deployment of AI and robotics.

IEEE 7007 Ontologies for Ethically Driven Robotics and Automation Systems Working Group

The IEEE 7007 WG is under the umbrella of the IEEE SA P7000's series devoted to ethics in A/IS. In this scope several working groups were formed, fifteen 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; to align the ethics and engineering communities to understand how to pragmatically design and implement these systems in unison; and to develop a precise communication framework among global experts of different domains that includes Robotics, Automation and Ethics.

To attain the above goals, the IEEE 7007 WG developed a set of ontologies for knowledge representation to represent the domain in a more precise way. As a result, the IEEE 7007 Standard contains a set of ontologies that represents the following domains: norms and ethical principles; data privacy and protection; transparency and accountability; and ethical violation management. The IEEE 7007 WG has followed a development life-cycle in order to ensure the development of a quality standard. The development of its standard was a complex process, and thus required a dedicated life cycle. For this purpose, IEEE 7007 developed an agile, collaborative, and iterative methodology called Robotic Standard Development Life Cycle (ROSADev) [1].

The usefulness of ontologies in standardization is two-fold. On one hand, standardization processes are set to produce a body of knowledge that reflects consensus 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 a product of the standardization process that can be used directly in data processing and automatic reasoning. As an example, one can cite IEEE 1872-2015 [2], which set forth to establish clear definitions to common terms in Robotics and Automation.

IEEE 7007 Standard

Top Level Ontology

As a core ontology, the Ethically driven Robotics and Autonomous Systems (ERAS) ontology represents a mid-level 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 in the "ERAS top-level concepts" model using 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 top-level ontology in other contexts.

Norms and Ethical Principles Ontology

The Norms and Ethical Principles (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 the concepts such as Norms, Ethical Theory, Situation Plan Repertoire, Agent Plans, Plan Actions, Agent Actions as well as corresponding relationships such as "selects plans from", "subscribes to", "satisfies", and "constrains plans for". **Figure X3 depicts a brief and partial graphical view of a subset of the NEP terms with a UML class diagram. See Sidebar / Table T1 for example CLIF axioms defining a few of the concepts and relationships shown in Figure X3.**

Data Privacy and Protection Ontology

The Data Privacy and Protection (DPP) Ontology represents concepts and relationships between diverse agents, entities, and organisations which may be involved at different stages in data gathering, processing, transfer, retention, and storage, and in which autonomous systems may be deployed. Thus, DPP Ontology represents concepts like natural person, caregiver, Data Protection Authority, the Controller, and Authorised Accessor; so as the different types and processing of personal data (e.g., health data, economic data, social data) and the corresponding data process access. The Data Protection and Privacy principles were also included in the standard, like privacy by design, data protection by design, data protection by default, human rights by design.

It is crucial to represent this domain formally, because of the relevance of the existing regulations worldwide about data protection and privacy. 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 action. 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 behaviour, intentions, perceptions, goals, and constraints in a manner that permits authorized users and collaborating agents to understand its past and expected future behaviour. 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”.

Ethical Violation Management

The Ethical Violation Management ontology subdomain presents axioms to formalize 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.

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 failure to conform to the norm’s rules of behavior relevant to 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 and 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, liability, civil and criminal liability [10], have influenced the content of the standard. The parameters between accountability and responsibility is 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 only applicable when the autonomous system is a member of a human directed team and when an action by the system caused a norm violation.

Conclusions

The IEEE 7007 Standard 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 deal with norms and ethical principles, transparency and accountability, ethical violation management and data privacy and protection. It is expected that this work have significant impact worldwide being used to teach ethical design; for both human and institutional capacity building in the domain of Ethics of AI; to create computational ethically aligned systems; to create a taxonomy to support the elaboration of public policies; to strengthen digital cooperation across Nations applied together with the other members of the IEEE P7000 family.

References

[1] Olszewska, J.I., Houghtaling, M., Goncalves, P.J.S., Fabiano, N., Haidegger, T., Carbonera, J.L., Patterson, W. R., Ragavan, S.V., Fiorini, S.R., Prestes, E. Robotic Standard Development Life Cycle in Action. Journal of Intelligent & Robotic Systems volume, 98, 119–131 (2020).

[2] IEEE Std 1872-2015, IEEE Standard Ontologies for Robotics and Automation

[3] Khamis, A., Meng, J., Wang, J., Azar, A. T., Prestes, E., Li., H., Hameed, I.A., Takács, A., Rudas,I.J., Haidegger, T., Robotics and Intelligent Systems Against a Pandemic. Acta Polytechnica Hungarica 18, no. 5 (2021).

[4] Khamis, A., Li, H, Prestes, E., Haidegger, T., AI: a key enabler of sustainable development goals, Part 1 [industry activities]." IEEE Robotics & Automation Magazine 26, no. 3 (2019): 95-102.

[5] Khamis, A., Li, H, Prestes, E., Haidegger, T., AI: A key enabler for sustainable development goals: Part 2 [industry activities]." IEEE Robotics & Automation Magazine 26, no. 4 (2019): 122-127.

[6] Jacobs, T., Veneman, J., Virk, G.V., Haidegger, T. The flourishing landscape of robot standardization [industrial activities]. IEEE Robotics & Automation Magazine 25, no. 1 (2018): 8-15.

[7] Ulgen, O., “User Rights and Adaptive A/IS – From Passive Interaction to Real Empowerment” in HCII Conference Proceedings, in LNCS Series, (Springer 2020), R. A. Sottilare and J. Schwarz (Eds.): HCII 2020, LNCS 12214, 205–217, 2020.
[8] IEEE ECPAIS Expert Focus Group Process Report: Accountability Criteria Development Process" (16 December 2019); Ulgen O., "IEEE ECPAIS Certification Criteria for Autonomous & Intelligent Systems A/IS: Accountability Certification Requirements" (16 December 2019)
[9] ISO/IEC 24707:2018, Information technology Common Logic (CL) A framework for a family of logic-based languages
[10] Ulgen O., “A ‘Human-Centric and Lifecycle Approach’ to Legal Responsibility for AI” (2021) 26(2) Communications Law Journal 96-107.