# Standardizing an Ontology for Ethically Aligned Robotic and Autonomous Systems

Michael A. Houghtaling, *Member, IEEE*, Sandro Rama Fiorini, Nicola Fabiano, *Member, IEEE*, Paulo J. S. Gonçalves, *Senior Member, IEEE*, Ozlem Ulgen, Tamás Haidegger, *Senior Member, IEEE*, Joel Luís Carbonera, Joanna Isabelle Olszewska, *Senior Member, IEEE*, Brian Page, *Life Senior Member, IEEE*, Zvikomborero Murahwi, *Member, IEEE*, and Edson Prestes, *Senior Member, IEEE*

***Abstract*—Domain-specific ontologies support system design and can establish a framework for fulfilling user-level, safety, or ethical requirements. The IEEE 7007-2021 Ontological Standard for ethically driven robotics and automation systems is the first industry standard to introduce a structure of ontologies concerning robot ethics and related fields, such as data privacy, transparency, responsibility, and accountability, offering a systems science approach to support the ethically aligned design of complex cyber–physical systems (CPSs) and robots particularly. This article provides a comprehensive overview of the main ontological commitments composing the foundation of the standard, the rationale behind their development, together with use cases of applications. Future directions for ethically aligned robotics and artificial intelligence (AI)-based systems along IEEE 7007-2021 are outlined, taking into account the exponentially growing fields of service and medical robotics.**

***Index Terms*—Accountability, automation, ethics, ontology, privacy, responsibility, robotics, standards, transparency.**

## I. Introduction

THE PAST years have seen an increasing global push for considering applied ethics in the design (ethics-by-design), in relation to the development, deployment, and usage of artificial intelligence (AI)-based applications in general. Intelligent systems, systems of systems, and robotics and automation (R&A) domains are affected predominantly [1]. In particular, in 2016, one could witness the major effort conducted by IEEE entitled *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems* [2]. This initiative brought together experts from various societies, application domains, and geographies, to reflect on the key issues and propose high-level ethical principles and recommendations for creating ethically aligned intelligent cyber–physical system (CPS) [3].

One of the activities originating from the Global Initiative was the IEEE 7000 Series,[1] which consists of a set of standardization projects aimed at developing industry standards dealing with aspects of ethics in design and operation of intelligent systems. This article relates to the IEEE 7007 Project and extends previous work [4], providing a detailed description of the IEEE 7007-2021 Ontological Standard for ethically driven R&A systems (ERASs), which was published officially by IEEE in November 2021.[2]

As AI-based systems are rapidly growing in size and complexity, it becomes paramount to have clear definitions of the system components with ethical implications, thus communication about ethical notions becomes less error-prone, either by artificial agents or by humans and institutions. The IEEE 7007-2021 proposes a set of well-founded ontologies, called Ontologies for ERAS, with vocabulary and definitions for describing the components and dependencies of ethically driven systems.

In computer science and related areas, an ontology is a formal and explicit specification of a shared conceptualization [5]. In line with this, an ontology is a machine-processable artifact that captures a common understanding of the conceptual structure of a given domain, by specifying the types of entities that are supposed to exist according to a community and that makes explicit the assumptions made by the practitioners of that community regarding those domain entities. Ontologies represent the conceptual structure of a domain through a set of classes of domain entities, relationships between these classes, properties (or attributes) that characterize them, and axioms that impose constraints on the possible interpretations of classes, relationships, and properties.

Manuscript received 27 March 2023; revised 29 August 2023; accepted 25 October 2023. This article was recommended by Associate Editor F. Sahin. *(Corresponding author: Tamás Haidegger.)*

Michael A. Houghtaling, retired, was with IBM Systems and Technology Group, Tucson, AZ 85750 USA. He resides in Tucson, AZ, USA.

Sandro Rama Fiorini is with IBM Research, Rio de Janeiro 20031-170, Brazil.

Nicola Fabiano is with the Studio Legale Fabiano, University of Ostrava, 00183 Rome, Italy.

Paulo J. S. Gonçalves is with IDMEC, Instituto Politécnico de Castelo Branco, 6000-084 Castelo Branco, Portugal.

Ozlem Ulgen is with the School of Law, University of Nottingham, NG7 2RD Nottingham, U.K.

Tamás Haidegger is with the University Research and Innovation Center, Óbuda University, 1034 Budapest, Hungary (e-mail: haidegger@ieee.org).

Joel Luís Carbonera and Edson Prestes are with the Informatics Institute, Federal University of Rio Grande do Sul, Porto Alegre 90010-150, Brazil.

Joanna Isabelle Olszewska is with the School of Computing and Engineering, University of the West of Scotland, G72 0LH Glasgow, U.K.

Brian Page is with Visiontech Communications, Spokane Valley, WA 99216 USA.

Zvikomborero Murahwi is with the Department of Research and Development, Gratia ICT Projects Advisory, Johannesburg 2196, South Africa.

Digital Object Identifier 10.1109/TSMC.2023.3330981

[1] https://ethicsinaction.ieee.org/p7000/

[2] The IEEE Ethical standards are available fee under the IEEE GET program (https://ieeexplore.ieee.org/browse/standards/get-program/page/series?id=93).

ERAS includes five ontologies that deal with aspects ranging from norms and agent actions, to privacy, transparency, norm violation, and responsibility. It extends from IEEE 1872-2015 standard [6] that introduced a collection of ontologies about central concepts in R&A, chief among which is the core ontology for R&A (CORA) [7].

Ontologies provide a formal artifact on which to base tasks, such as agent communication, data exchange, and high-level reasoning. Such capabilities have already been leveraged in various systems science and R&A applications [8], [9], [10], [11]. Given their systematic approach to vocabulary definition and representation, ontologies inherently constitute efficient bases for industry standards. This attribute motivates the development of formal standards using ontologies, both in the R&A and other fields [12], [13].

The formal vocabulary defined by the ERAS standard enables unambiguous dialog between researchers and stakeholders across many diverse disciplines and communities involved with or affected by ethical R&A systems. The facilitation of such dialog will be extremely important as R&A and other information and communication technologies (ICTs) are applied in support of resolving the complex social and technological challenges confronting the world.

In this regard, the unique contributions of the ERAS standard are twofold: first, it may influence regulations in the field of ethical R&A systems, and second, the deployment of R&A systems conform with the standard may contribute as ICT artifacts supporting important social-technological goals, such as the 17 UN Sustainable Development Goals described in [14] and [15]. Other novel aspects are the contributions toward the development of trustworthy autonomous systems and the provision of guidelines for verification, certification, and assurance of autonomous AI systems.

This article aims to provide a comprehensive overview of the IEEE 7007-2021, to guide the reader through the logic and structure of the ERAS, with examples highlighting possible applications. It is expected that beyond R&A research groups, development teams in the industry will also find it applicable, and the methodology provided will become a highly applied standard across the CPS domains. More specifically, the ERAS is composed of the following ontologies.

1) *Top-Level Ontology (TLO):* An ontology with fundamental commitments used by other ontologies in the standard.
2) *Norms and Ethical Principles (NEPs):* A core ontology defining the main principles involved in ethics and ethical behavior, such as norms, plans, and actions.
3) *Data Privacy and Protection (DPP) Ontology:* An ontology detailing notions related to privacy and protection, inspired by current regulations on the topic.
4) *Transparency and Accountability (TA) Ontology:* An ontology for the characterization of autonomous system behaviors involved with composing and providing informative explanations for system plans and agent actions.
5) *Ethical Violation Management (EVM) Ontology:* An ontology for characterizing situations in which agents fail to conform with prescribed norms.

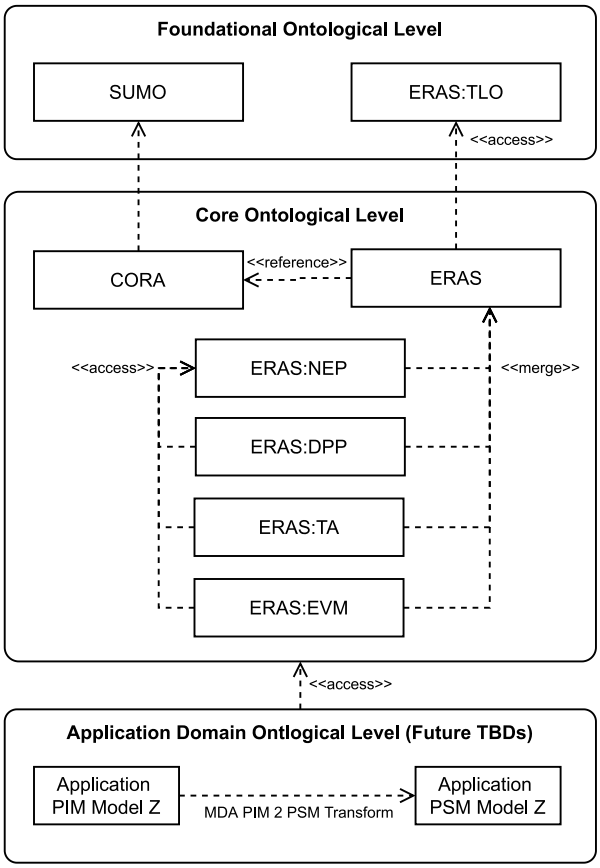



Fig. 1. Overview of ontologies and their dependencies in the IEEE 7007.

This article is organized as follows. Section II describes the main ontological commitments of ERAS and its components. Section III presents an example use case and Section IV finishes with final remarks.

## II. Ontologies for Ethically Driven R&A Systems

The domain of ethically driven R&A is quite complex, encompassing fields with distinct methodologies and viewpoints, going from Engineering to Law. To cope with the formalization of a relevant vocabulary and ontological commitments into a coherent structure, the Working Group applied three strategies. The first was to design the ERAS ontology as a middle-level core ontology positioned between a top foundational level and a bottom application domain level. The second was to partition the domain into four separate subdomains. The third was the application of the model-driven architecture (MDA) methodology as applied to ontological engineering [16], [17]. The MDA architecture entails four modeling levels: 1) an M3 meta-metamodel level; 2) an M2 metamodel level; 3) an M1 model level; and 4) an M0 instance model level. M1 level models are framed from a "separation of concerns" perspective regarding commitments that distinguish between platform-independent models (PIMs) and platform-specific models (PSMs).

The four ERAS subdomain ontologies depicted in Fig. 1 are composed as M1 level PIMs with no commitments to specific implementation technologies. That is, each ERAS subdomain

TABLE I
EXAMPLES OF INFORMAL DEFINITIONS AND AXIOMS

| Term | Informal Definition | Axioms |
|---|---|---|
| Method | A Description subcategory that classifies entities which are abstract descriptions of Occurrent Process actions to produce some result. | (forall (x)<br>(if (Method x)<br>(Description x)))<br><br>(forall (x)<br>(if (Method x)<br>(exists (y)<br>(and<br>(Process y)<br>(describes x y)<br>(realizes y x))))) |
| Plan | A Method subcategory that classifies entities which specify a sequence of processes intended to satisfy a specified purpose or goal for an Agent by affecting changes in the Agent's physical situation. | (forall (x)<br>(if (Plan x)<br>(Method x)))<br><br>(forall (a p m e)<br>(if (and<br>(Agent a)<br>(Process p)<br>(participates-in a p)<br>(Plan m)<br>(Physical e)<br>(not (= e p ))<br>(realizes p m)<br>(affects p e))<br>(selected-by m a))) |

is defined as ontologies where no specific M0 model level instances are specified. The terminology for each ERAS subdomain is defined as a collection of hierarchical classes for concepts with binary and ternary relationships between the classes. The core formally specified ERAS commitments provide a framework for application-specific models to elaborate and extend the ERAS M1 PIM ontology models as shown in the lower part of Fig. 1. Application-specific PIM ontology models extend the core models with relevant M1 concepts and relationship commitments for the application, which are then transformed into PSM commitments for selected technology and implementation platforms. This separation of concerns facilitates the focus on "what" before making commitments to "how" as well as enabling the shared reuse of the ERAS core commitments across multiple "to be determined" application domains concerned with ethically driven autonomous systems.

As described previously, the principal aspects of an ontology are the formal axiomatic definitions for the shared conceptualizations and semantic commitments specified in the ontology. To facilitate a complete formalization of the ERAS core vocabulary, the ERAS standard also includes formal axioms for a TLO which consists of a foundational set of concepts and relationships. The TLO terminology is discussed further in a subsequent section of this article.

The IEEE 7007-2021 is organized in the usual format for IEEE standards. It is composed mainly of normative definitions expressed in the common logic interchange format (CLIF, [18]), augmented with informal textual descriptions along with informative UML diagrams (Table I presents two examples of informal definitions and axioms). Table II lists

TABLE II
STATISTICS ON ONTOLOGY CONSTRUCTS IN IEEE 7007

| Subdomains | # of Concepts | # of Relations | # of Axioms |
|---|---|---|---|
| TLO | 27 | 19 | 25 |
| NEP | 40 | 42 | 45 |
| DPP | 36 | 40 | 41 |
| TA | 21 | 23 | 30 |
| EVM | 13 | 24 | 33 |

the number of concepts, relationships, and semantic axioms for each subdomain and the TLO ontology. The standard also includes informative definitions for each term defined in the ontologies, as well as four example use cases, employed in the domain analysis discussed previously.

### A. ERAS Top-Level Ontology

A frequently employed framework for ontology composition partitions relevant descriptions and formal semantic definitions into three levels, a top foundational level, a central core level, and a lower domain application level for one or more application-specific set of conceptual commitments. The ERAS ontology is positioned as a middle core ontology. As shown in Fig. 1, the ERAS ontology is composed of four subdomain ontologies, each of which is defined with normative CLIF axioms and informative UML class diagrams. Further, the core-level semantic axioms defined in the ERAS subdomain ontologies refer to or "access" a relevant set of top-level foundational axioms. This insures complete and consistent semantic definitions specifying the ontological commitments for each ERAS subdomain. The TLO provides a minimal set of such foundational commitments (Fig. 2). Whereas other foundational ontologies attempt to prescribe broad, general-purpose conceptualizations and terminology applicable to many lower-level ontologies and divergent applications, the TLO is intended only for providing complete CLIF formalizations of ERAS subdomain axioms. However, to facilitate possible alignments with other foundational ontologies, the selected TLO categories and commitments are very similar to those found in SUMO [19], GFO [20], UFO [21], and KR [22].

At top level, the TLO ontology separates entities on their existence in space–time. Physical entities exist in space–time, including every-day objects and events (Fig. 2). Abstract entities include propositional entities, such as descriptions, in addition to properties and collectives. As with other top-level ontologies, continuants denote physical objects, including agents and information artifacts (e.g., books), as well as situations. Situations are aggregates of other entities and relationships representing a particular part of reality that can be perceived by some agent. TLO concepts also distinguish between processes (e.g., music concerts, cooking, and kicking), from events, which are parts of processes. Additional TLO concepts to provide context and semantic commitments for other ERAS ontology subdomains include the abstract concepts of Methods and Plans. Examples of informal definitions and CLIF axioms for the terms Method and Plan are provided in Table I.

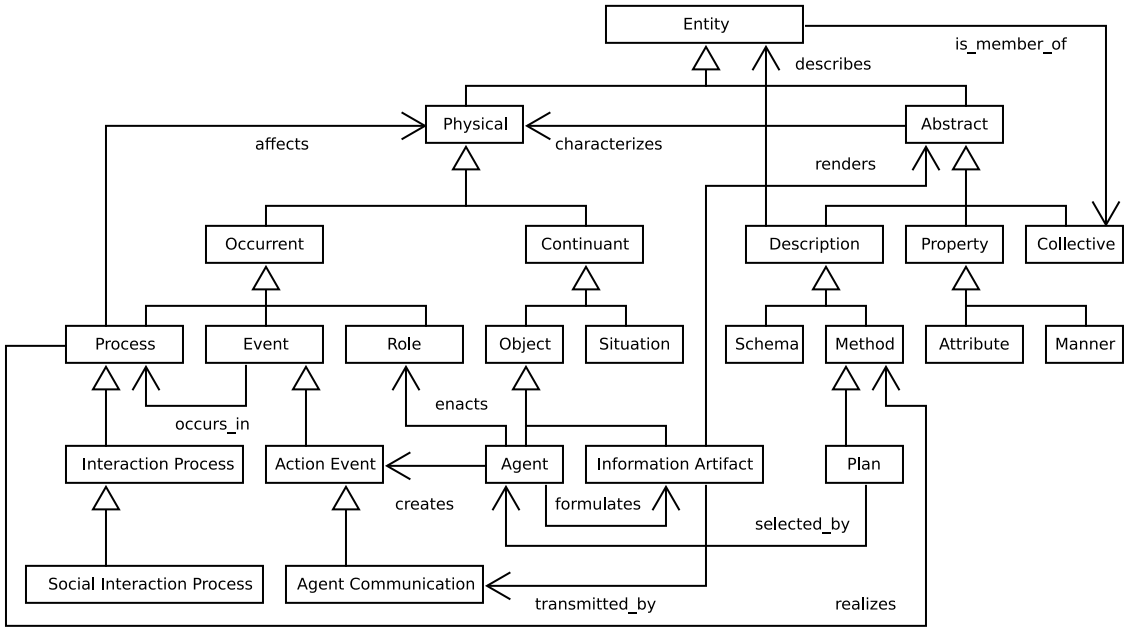



Fig. 2. UML diagram of the main concepts and relations in ERAS TLO.

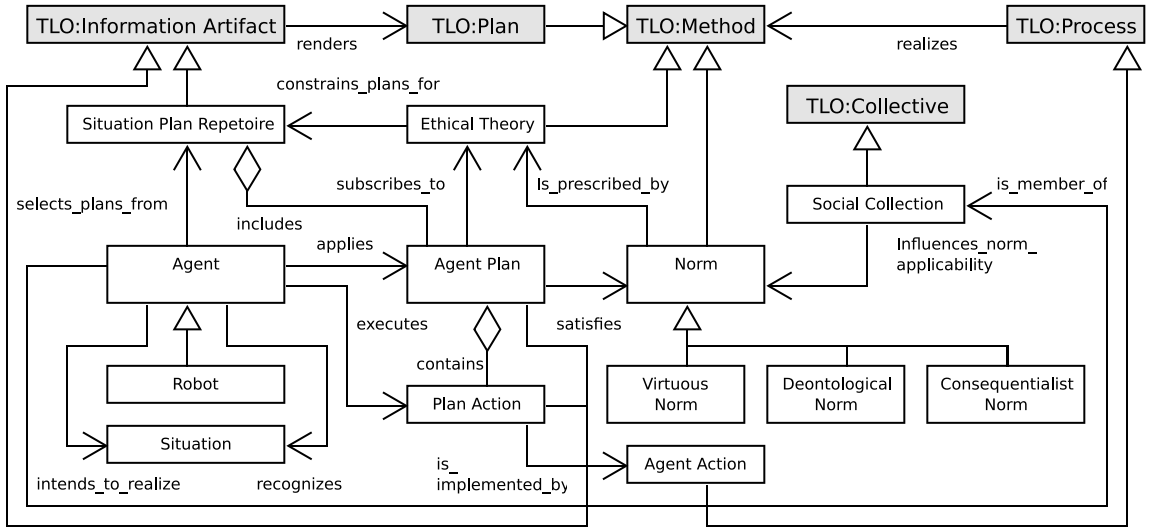



Fig. 3. UML diagram of main concepts and relations in NEPs subdomain ontology.

### *B. Norms and Ethical Principles Ontology*

The NEPs Ontology formalizes a vocabulary for the unambiguous description and communication of ethical behaviors expected from agents that claim or attempt to adhere to ethical theories and the norms associated with them. As shown in Fig. 3, both norms and ethical theories are methods, which are descriptions of processes that reflect fundamental ontological commitments in NEP: norms, such as rules and laws, to denote how agents should or should not act.

Within that context, the NEP terminology describes a general process, where agents act by executing actions that implement plans, satisfying norms prescribed by ethical theories. These plans may be a part of provisioned plan repertoires constrained by ethical theories, from which agents may select relevant plans. For example, a care robot could be programmed for selecting plans from a repertoire that is constrained by medical duty theory of ethics, such that actions implemented by the robot could be restricted to plans prescribed by this theory. Note that NEP does not describe *what* constitutes ethical behavior in a given context. It just tries to state the minimal set of ontological commitments necessary to *represent* a situation where an agent follows or not some ethical theory (see Section II-E for representation of ethical violation).

The NEP Situation category extends the TLO Situation category to represent an agent's intentions and context. An agent's perception of its environment is represented as a situation and is used as a basis to select and apply plans deemed relevant for the current recognized situation.

The NEP subdomain's model and vocabulary enable the description of scenarios where agents select plans which entail norms that should be satisfied by the plan actions contained in

the selected plans. The type of norms involved depends upon the ethical theory or theories subscribed to by agent plans that are selected by agents based on perceived situations.

Note that the plan selection can also be constrained by ethical theories and that such ethical constraints might exist even in simple reactive robots. For example, consider a simple industrial robot that is programmed to halt operation as soon as it detects movement in its operating area. It can be said to be selecting plans constrained by some sort of deontological, normative ethics implemented in its code-base at design stage.

NEP also assumes that social collections, like communities or companies, might influence norm applicability and consequently agent actuation in the environment, similar to what happens with humans when they are interacting with others in different environments. For instance, social rules employed in a familiar context are different from those applied in a business environment.

Several NEP concepts and relations defined in the standard are not shown in Fig. 3. These include categories, such as Agent Role, Action Rationale, Ethical Dilemma, and subclasses for the TLO Agent Communication concept. The Agent Role category is used to associate capabilities, rights, and obligations to an agent enacting the role. For example, an Agent assigned and enacting a Caregiver role would have the responsibility of representing and protecting the person dependent upon the caregiver. The Action Rationale category attributes logical justifications for a plan action. The principle of autonomy could be the logical rationale for an elderly home care agent to acquiesce to the elder's refusal to take medicine at a prescribed time. Agent Communication subcategories consist of Explanation, Query, Answer, and Task Assignment concepts to distinguish between the different illocutionary forms used by agents, such as in an interaction to enable system transparency. The Ethical Dilemma category accounts for cases where two or more norms associated with a plan action conflict or where none of the choices for normative behavior is deemed unambiguously acceptable. Refer to the IEEE 7007-2021 for more details.

### C. Data Privacy and Protection Ontology

The DPP Ontology specifies a vocabulary for describing ethical behavior related to proper treatment and use of personal data by robotic and autonomous systems (Fig. 4). The DPP subdomain focuses primarily on ethical aspects since many legal principles are currently defined in regulations like GDPR (EU Regulation n. 2016/679.[3])

In addition, the concepts expressed in existing regulations about data protection and privacy throughout the world were used as background and context for the conceptualizations and formal axioms defined for the DPP subdomain. Since this area of law is becoming increasingly complex and regulated, it is expected that it will continue to change rapidly with advances in technology and cultural expectations. As a consequence, the DPP terminology was centered around core concepts and relationships, expected to be general enough to accommodate current and future developments in the DPP domain. Still, specific domain applications need to take into account discrepancies between the vocabulary incorporated in this ontology and their possibly distinct interpretations across local, regional, and national jurisdictions. Expected ethical behavior for robotic and autonomous systems, in this domain requires that agent plans for the systems abide by the DPP constraints in effect for their area of operation and as they pertain to their interactions with people.

One of the conceptual categories in DPP is Personal Data. An instance of personal data is an information artifact which is about some person. The IEEE 7007-2021 defines further subcategories of personal data, such as health and economic data. In this view, any digital file that encodes information about someone is considered personal data.

Personal data might be collected in "personal data transaction," as part of an "data access process" which is administered by a data "controller" according to some provisioned data "access policy." Note that data access consent is captured by the ternary relation "consents to." For example, John (Person) might consent to share a pdf of his medical record (Personal Data) with Hospital *X* (Controller) as part of Hospital *X* admittance process (Data Access Process), in which John sends the pdf file by email (Personal Data Transaction).

Also, note that the data access policy implements some data access principle. A data access principle defines general guidelines about implementation of data access in a given context. Those include principles, such as privacy by design, data protection by design, data protection by default, and human rights by design.

Other parts of the DPP ontology relate to common data access roles in privacy regulations around the world, such as the distinction between "data controller" and "data processor."

### D. Transparency and Accountability Ontology

The transparency of robotic and automated systems is defined as the extent to which such a system discloses the processes or parameters that relate to its functioning [23]. Such transparency makes possible the discovery of why and how a system decides to behave appropriately, or not, for its situated environment [2], [24].

The definition of transparency has been elaborated in several works [25], [26], in different contexts, such as intelligent systems [27] and human–robot interactions [28], as well as in standards such as IEEE 7001-2021 [29]. However, the IEEE 7007-2021 standard proposes for the first time an ontological definition of the transparency concept [4].

The TA Ontology formalizes a vocabulary for the characterization of autonomous system behaviors involved with composing and providing informative explanations for system plans and agent actions. Some of the associated concepts and relationships involved are depicted in Fig. 5.

Ethically aware agents would normally be expected to have the ability to be transparent in their interactions with other agents. This means they should be capable of communicating intentions, perceptions, and goals in a manner that permits authorized users and collaborating agents to understand past, present, and future behaviors. The Explanation category, as

[3]http://data.europa.eu/eli/reg/2016/679

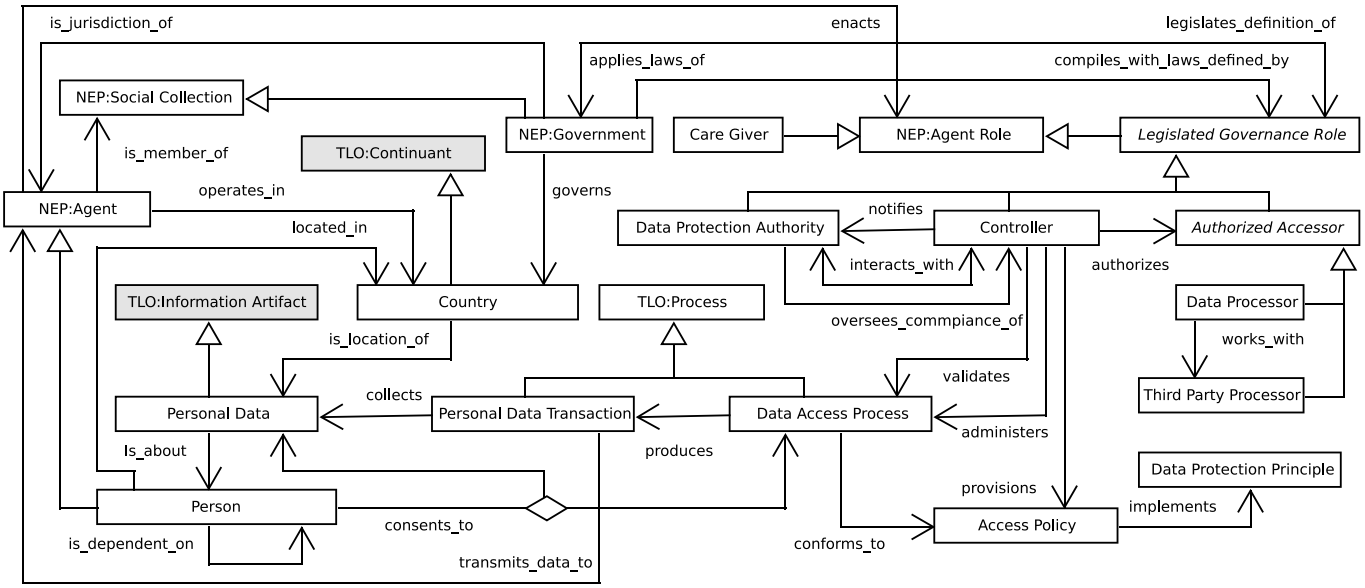



Fig. 4. UML diagram of main concepts in DPP ontology.

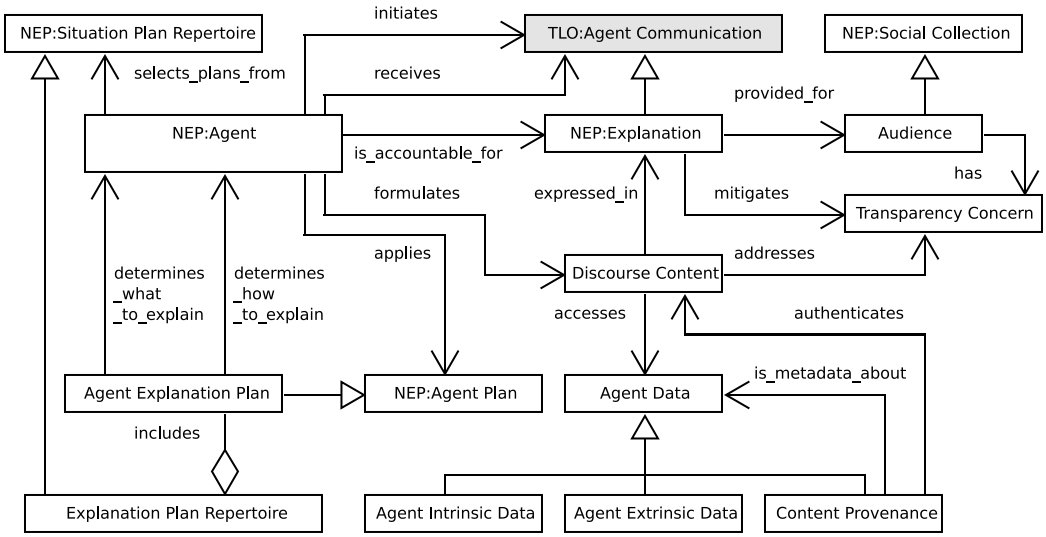



Fig. 5. UML diagram of main concepts in TA ontology.

a subcategory of TLO Agent Communication, classifies this exchange of information as a response to requests received from an external agent that initiates the agent communication action event. The explaining agent formulates the discourse content that is "expressed in" the explanation and "is accountable for" the information contained within the Explanation.

To formulate the requested explanation, the agent has access to an explanation plan repertoire which includes relevant explanation plans. The agent selects an appropriate explanation plan which will "determine what to explain" and "determine how to explain." The selected explanation plan takes into account the transparency concerns of the audience to which the explanation is provided. The semantics of discourse content and explanation is given by three relationships: "provided for," "mitigates," and "addresses." As depicted in Fig. 5, an explanation is "provided for" an audience, and it "mitigates" the Transparency Concerns that an Audience may have. The discourse content "expressed in" the explanation "addresses" those transparency concerns. For instance, an ethically aware medicine delivery robot (Agent) might be accountable for responding to a request from a head nurse (Audience) to explain (Explanation) why a certain medicine was not delivered to a patient on time (Transparency Concern).

In order to compose the discourse content to be "expressed in," the robot may access a wide variety of data relevant to the circumstances of the request, such as *intrinsic data*, *extrinsic data*, and *content provenance*. Agent intrinsic data refers to data that is generated by or composed about the Agent. Examples include plan data, action execution traces, interaction traces, and agent static data (e.g., user manuals, design specifications, principles of operations, or verification metrics). Each of these subcategories is formally defined in the IEEE 7007-2021 Std. Agent extrinsic data classifies data that is not directly about or affiliated with the agent but

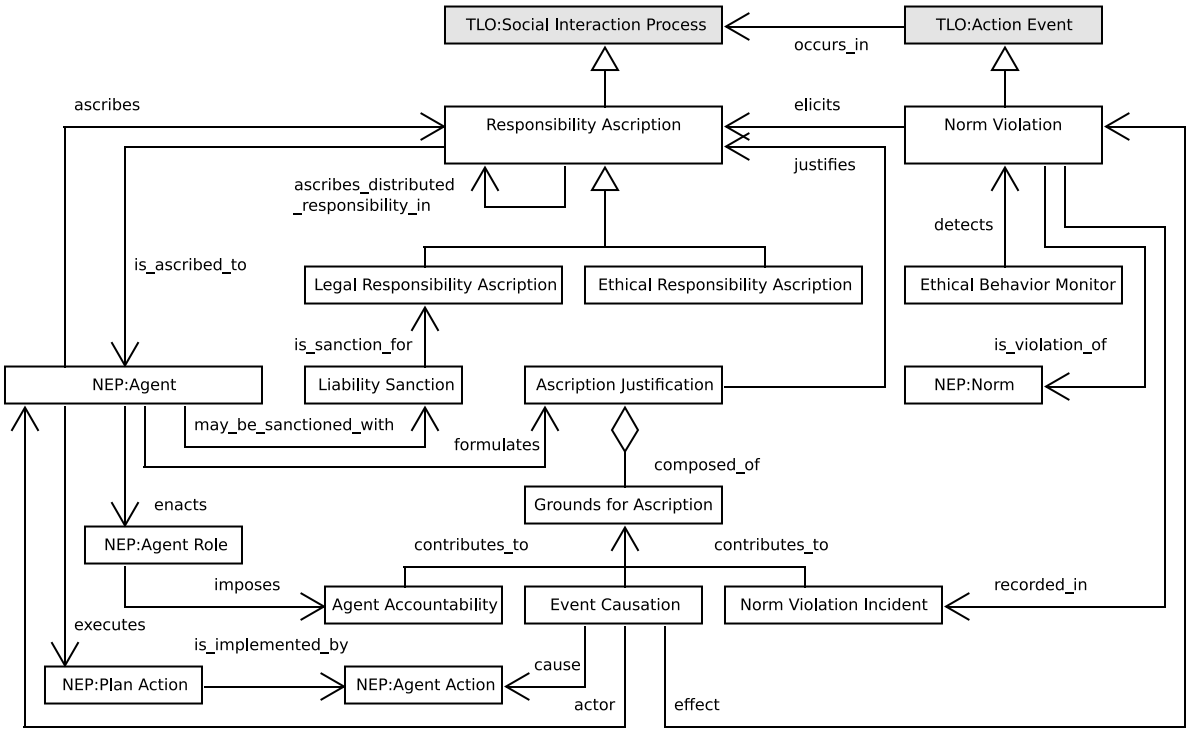



Fig. 6. Partial UML diagram of EVM ontology.

which is about external world circumstances in the agent's situated environment. These include, for example, data about current weather and news articles. Content provenance data refers to metadata information about the other pieces of agent data regarding the sources, agents, and processes involved in the generation and composition of the information formulated in the discourse content. This information authenticates the formulated discourse content and is used to assess the quality, reliability, and trustworthiness of the subject data provided in an agent's explanation.

Returning to the medicine delivery robot case, the explanation could communicate the facts that another nurse canceled the order (Intrinsic Data), or that the medicine was not available (Extrinsic Data), as well as providing metadata such as medicine inventory files or nurse logs (Content Providence).

Note that the agent creating and providing explanations can also be a human–robot designer. In that case, the explanation might be an interaction between a designer and a regulator on why a given autonomous system behaved in a particular way. In that case, agent data is about the autonomous system itself.

### *E. Ethical Violation Management Ontology*

The EVM Ontology (Fig. 6) formalizes a vocabulary for the characterization of circumstances where ethically committed agents fail to successfully conform to the norms prescribed by the ethical theory governing the plans applied by the agent and which consequently, results in unethical behavior. When such a situation occurs, one or more norm violation action events can be detected by an ethical behavior monitor. An ethical monitor is an object that can be realized by an agent system component or by another agent (e.g., a robot or human).

Once detected, a norm violation elicits a responsibility ascription that is a social interaction process that assigns responsibility for a norm violation to an agent by another agent or agency acting in an authoritative role. Since a norm violation can be either a violation of an ethical norm or a legal norm, the responsibility ascription process has specializations for each type. A legal responsibility ascription may have a liability sanction associated with it.

Let us assume, a given care robot that fails to promptly dispense medication for its user at the prescribed time might be found to have violated a bio-ethical norm of nonmaleficence by a human operator (Ethical Behavior Monitor). The responsibility for this particular violation event might then be attributed to the robot designer company, which might be justified in a number of ways.

Ascription of responsibility for a norm violation is justified by an ascription justification. This is an information artifact that the agent or agency participating in the responsibility ascription formulates and that is composed of one or more grounds for ascription. Grounds for ascription are given by a collection of factual circumstances, causal events, agent accountability, and legal or ethical obligations that contribute to the norm violation. More specifically, agent roles assigned to or enacted by agents impose accountability descriptions. These are descriptions, such as age, physical state, mental state, capabilities, intentions, knowledge, role responsibilities, and authority.

Another major contributor to grounds for ascription is event causation. The EVM ontology defines Event Causation as a subcategory of the TLO Interaction Process which classifies the constituent process entities that identify the actor, the

action that caused the violation, and the resulting norm violation effect.

Coming back to the previous example on the care robot, the justification for ascribing responsibility to the robot designer might be justified on the grounds that an incorrect assembly of a robot component caused a malfunction, which generated the violation incident (Event Causation).

The features described in the EVM subdomain strengthen requirements for transparency, accountability, and responsibility including associated Legal notions in Robotic and Autonomous Systems. It also addresses explainability in the behavior of Robotic and Autonomous Systems. Consequently, concerns often raised around the management of potential harms and unethical behavior in Robotic and Autonomous Systems are adequately addressed with appropriate semantic commitments. Formalizing ontological categories such as Agent Plan, Agent Action and Ethical Behavior Monitor provides a consensus vocabulary to describe aspects of system failures due to both commission and omission on the part of the autonomous system. The EVM further addresses multiple Agencies with distributed responsibility to accommodate situations involving System of Systems.

The EVM ontology includes the characterization of other commitments not shown in Fig. 6. For instance, it allows different responsibility levels for autonomous systems based on the socio-technology governance maturity level achieved by a government [30], [31]. Note, however, that the EVM ontological commitments prevent the characterization of an autonomous system acting as a single agent from being ascribed responsibility for any type of norm violation. The reason why an autonomous system acting as a single agent cannot be ascribed responsibility for any norm violation relates to ethical and legal considerations. From a legal perspective, autonomous systems do not have the legal personality to be able to make claims, have claims made against them, or suffer any punitive measures for norm violations. Rules contained in law are predicated on human conduct and action and attributable to natural persons. Where the law provides for legal personality of nonhumans, such as corporate entities, this is because the latter are able to act in the legal sphere due to humans working within and representing them. This would not be the case for autonomous systems operating on their own [32]. From an ethical perspective, an autonomous system does not possess moral agency to be able to distinguish right from wrong, or what is harmful and what is not. No moral wrong or harm could be inflicted on the system, and there would be no effective moral or legal sanctions against it [32], [33]. Since humans are involved in the design, development, and deployment of autonomous systems, any wrongs or harms caused by such systems can be attributed to existing natural and legal people. 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.

Consider again the example care robot that fails to deliver medication at a prescribed time. A distributed responsibility ascription could ascribe both the designers of the robot and the agents involved with the delivery and configuration of the system. This is an important aspect given that last-mile-delivery robotic systems in the medical domain have recently been developed and deployed without thorough testing, while the medical clearance requirements are mandatory [34], [35].

## III. Example Scenario and Use Case Mapping

The following emergency response (EMR) scenario adapted from [36] illustrates a conceptual mapping example for an application that adopts the ontological commitments of the IEEE 7007-2021 standard [37].

The scenario involves an autonomous agent deployed with human teams responding to flooding event emergencies. The robot assistant facilitates human team members by prompting or recommending plans and actions appropriate for perceived situations. The plans and actions recommended by the robot are constrained by the following two ethical norms.

1) *Obligation Norm:* It is obligatory to deploy emergency services to an area experiencing a high risk of flooding.
2) *Prohibition Norm:* It is forbidden to evacuate people into an unsafe area.

The commitments of the ERAS ontology are applied in this example by composing an M1 model that extends the ERAS M1 model with two NEP Agent subclasses, three associated NEP Agent Plan subclasses, and two TLO concept subclasses. Fig. 7 illustrates the model subclass extensions. Depicted are the subclass relationships connecting the nine application domain concepts with relevant NEP and TLO subdomain concepts.

This application domain extension of the core NEP ontology enables the design of an EMR robotic assistant to be provisioned with suitable agent plans capable of advising human team members during flooding events. Such plans would suggest when and where to send support resources and when and where to evacuate residents in high-risk areas. Selection and execution of these agent plans differ from conventional BDI autonomous designs in that the robotic agent in this case is constrained by the two example ethical norms.

To complete the example and to illustrate an associated mapping of ERAS CLIF axioms to OWL DL, a partial M0 model expressed in OWL2 Manchester syntax [38] accompanies the M1 extended model. The Appendix presents mapped axioms for the extended class definitions with a few selected instances. The two ethical norms for the scenario are mapped as SWRL[4] rules [39] and are also shown in the Appendix. To facilitate the formalization of the EMR M0 assertions and SWRL rules, four additional relationships are also defined.

## IV. Discussion

The publication of the IEEE 7007-2021 Ontological Standard for ERASs provides the first ontology standard documenting a consensus agreement among a diverse community of stakeholders on the concepts and terminology for the target domain of ethically aligned autonomous systems. It documents each term in the vocabulary with normative

[4]SWRL is an acronym for Semantic Web Rule Language.

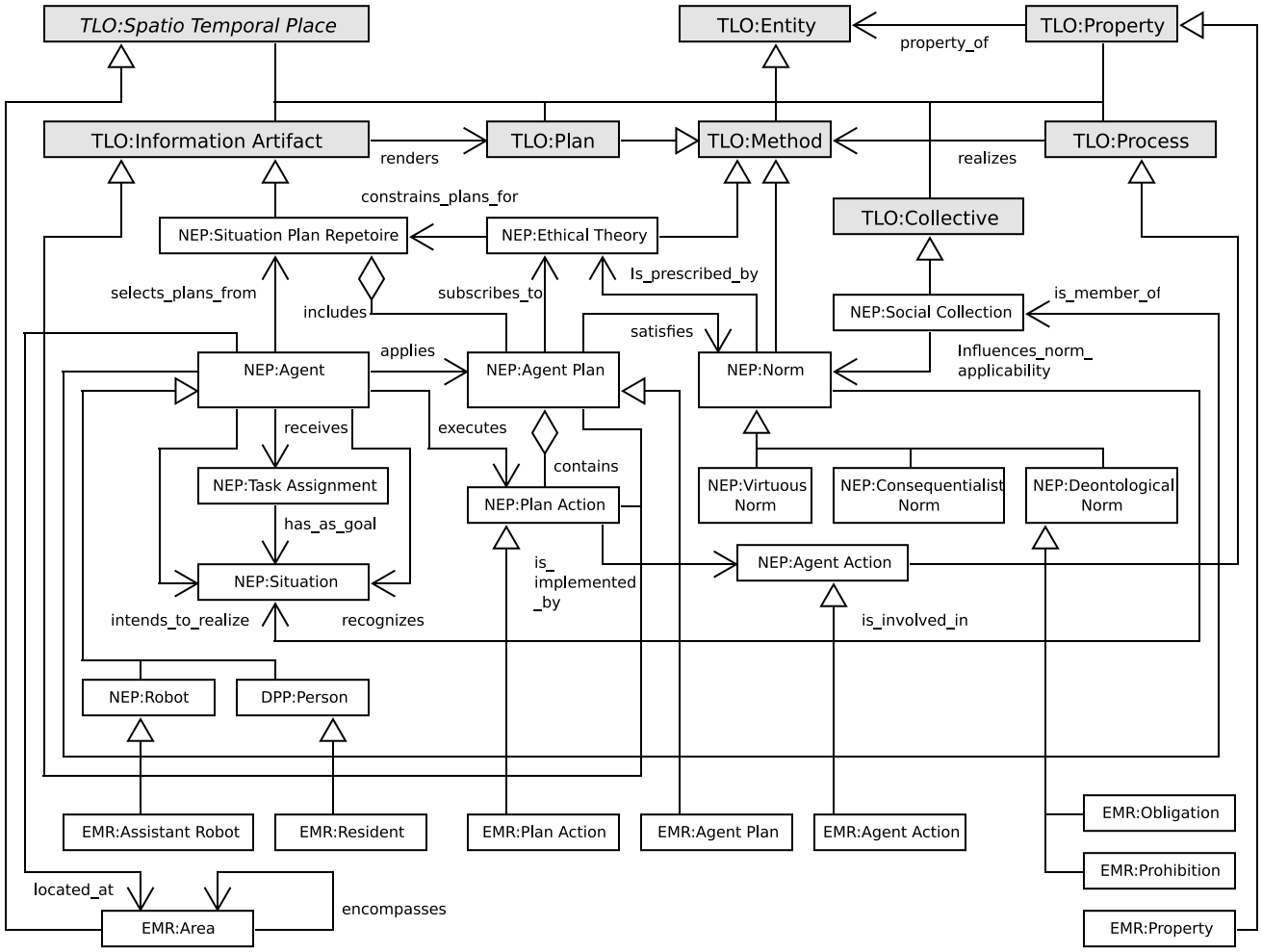



Fig. 7. Partial UML diagram for EMR robot application.

axioms expressed in a first-order logic notation (CLIF) and it augments its formalization of concepts and relationships with informative natural language definitions and accompanying graphical models using UML class diagrams.

The resulting vocabulary and semantic commitments enable abstract descriptions characterizing what it means for autonomous systems to exhibit ethical behavior. However, ERAS does not prescribe specific instances of ethical norms or affiliated ethical theories, nor does it define how to implement such systems. This is important since the target domain is emergent with extensive work in progress or in plan. As a consequence, ERAS is framed with the expectation that the conceptualizations will need to evolve and adapt to future development and regulatory constraints as the technologies, norms, and social principles evolve and change.

In addition to the many academic papers cited in the bibliography of the standard, several other sources of background information contributing to the formulation of the ERAS ontological commitments included legally binding instruments, nonlegally binding ethical standards, governmental, and NGO frameworks and guidelines.[5]

[5]Such as legally binding instruments [40], 2016 EU GDPR, 2019 OECD Recommendation on AI, 2019 G20 Human-Centered AI Principles, 2019 EU Ethics Guidelines for Trustworthy AI, 2019 IEEE EAD, 2015 UN Sustainable Development Goals, BS 8611:2016, ethical design and application of robots, UNESCO's Recommendation on Ethics of AI, and possibly others.

A computational system built on the top of the ERAS ontologies will be able to reason and make inferences about the instances of the concepts and relationships represented by the ontologies and provide answers to questions.

1) What are the ethical dilemmas faced by the system and how to address them?
2) What are the norms and principles that need to be satisfied by the agent plan?
3) What norms and principles are of particular interest of a specific social collection?
4) What access policy was used to process a particular data?
5) What are the data used in a particular agent explanation?
6) What sanction is ascribed to an agent from a norm violation that led to negative event causation?
7) Who is responsible in case of norm violations?

Answers to questions like the aforementioned provide a better understanding of a system's technology and, consequently, improves users trust in it. Such answers also enable analysis and scrutiny of the system thus supporting audits, as well as the communication and transmission of information among different stakeholders.

In relation to the latter, an ontological framework like ERAS can help to foster cooperation with different countries, particularly important where there are cross-jurisdictional applications and implications of using such systems. The basic

component of any governing framework is data. Data refers to a collection of values that the global community judges relevant about subjects, situations, scenarios, laws, stakeholders, environment, etc. Data can be divided into raw data and information. Raw data refers to any data that is directly observed and/or collected, while information refers to any data processed by a transaction that aims to make inference, elicit patterns, make estimations, and so on. All these data can be collected and represented by ERAS ontologies.

With data from different scenarios, one can better understand issues and solutions and use them to elaborate recommendations, public policies, and soft and hard laws at national, regional, and global levels.

In this sense, a governing framework, grounded on ERAS ontologies, would have several benefits.

1) It can store and represent data in a standard format to be used in different scenarios and places. A country or region can access the data collected and analyzed by other countries and/or regions.
2) It enables democracy fostering citizen participation via the communication channels to get data locally and shared globally.
3) It can be used to share and exchange information for technical and nontechnical purposes, e.g., sharing best practices or information about a particular subject.
4) It can be used to provide information tailored for different audiences.
5) It can enable precise and unambiguous communication among different stakeholders.
6) It can be used for human and institutional capacity building. Issues and recommendations can be used for police-makers' capacity building and also for training of new professionals in the digital domain.

It is expected that the ERAS ontologies can serve as reference guide for future normative and legal policies developed by stakeholder communities in the domain. For stakeholders involved with robotic and autonomous system life cycles it will be important to develop certification and verification processes to strengthen the public's confidence that such systems will cause no harm, that their limitations are known, and that there will be human accountability regarding their use. Combining the ontology commitments of the IEEE 7007-2021 Std. with certification methods identified by the IEEE ECPAIS (Ethical Certification Program for Autonomous and Intelligent Systems) Standard will help establish policies and processes for the ethical certification of products, services, and systems employing AI and robotic systems.

One of the objectives of the ERAS ontology standard is to facilitate its adoption by domain-specific applications that seek to adhere to its ethical commitments as formalized in CLIF axioms. Since many such systems will likely be developed by designers familiar with Ontology Web Language (OWL) a mapping from CLIF axioms to OWL DL axioms facilitates this objective Because OWL DL is less expressive than the FOL semantics of CLIF the mapping will only be approximate. Equivalent OWL DL expressions can be defined for ERAS concepts and relationships but not for all of the CLIF semantic axioms listed in Table II.

The certification of autonomous system implementations based on the ethical principles entailed in the ERAS standard will also require verification of capabilities not specifically referenced in the vocabulary of the standard. These include security, privacy, and safety features so that such systems can be protected from unauthorized access, from hacking and installation of malicious components, and from the theft of information. In addition, system and user information should be protected with the application of current encryption standards throughout the life cycle of the system. Therefore, while a demonstration of expected ethical behaviors is required, it is also important that the ethically aligned autonomous systems demonstrate safe, reliable, secure, and trustworthy operations in order to secure public trust prior to system deployments.

## V. Conclusion

The IEEE 7007-2021 Standard is arguably a bold, initial step toward documenting a common view and understanding of what it means for Robotic and Autonomous Systems to be imbued with ethical behaviors, nevertheless, there is still substantial work remaining. Even though this standardization is expected to help homogenize the view on the field, the public view and normative frameworks on the topic are likely to evolve and change, which will probably impact the IEEE 7007-2021 in the long run. Second, it is expected that other research projects may extend and specialize the ontologies in the IEEE 7007-2021 toward specific subdomains and applications. Finally, better validation of the ontologies in the current standard will be achieved, as applications using it are developed and deployed. A mapping of the ERAS CLIF axioms into approximate OWL DL axioms for each ERAS ontology has been completed and will be made available at a future date.

## Appendix

### A. Example OWL Axioms for NEP EMR Extended Model

The EMR use case introduced in Section III is augmented here with selected axiom examples and comments explaining the defining OWL expressions. Fig. 8 compares corresponding axioms for the ERAS concepts *Entity*, *Norm*, and *Agent Plan* as they are defined in CLIF and then mapped to OWL. Fig. 9 presents example OWL axioms specifying relationships, and concepts for the M1 application level entities mapped from the relevant ERAS subdomain ontologies.

The EMR concepts are specified using OWL class—subclass constructs. The EMR subclass associations define respective TLO, and NEPs core ERAS concepts as superclasses of the EMR classes. For example, the EMR Agent Plan is a subclass of the NEP Agent Plan.

Relationships are defined by specifying the respective domain and range classes for the relationship. Thus, the *constraints* relationship associates an instance of the NEP Deontological Norm with an instance of EMR Agent Action.

Fig. 9 also shows the OWL instance M0 definitions for some of the OWL classes as well as asserting some factual

**Example CLIF and OWL Concept Axioms**

| Concept | CLIF | OWL |
|---|---|---|
| Entity | (forall (x) (Entity x)) | CLASS: tlo:Entity |
| Norm | (forall (x) (if (Norm x)<br>(tlo:Method x))) | CLASS: nep:Norm SubClassOf: tlo:Method |
| AgentPlan | (Forall (x) (if (AgentPlan x)<br>(tlo:InformationArtifact x))) | Class: nep:AgentPlan<br>SubClassOf: tlo:InformationArtifact |

Fig. 8. CLIF and OWL axioms for ERAS concepts.

**Example EMR OWL Relationship Definitions**

```
ObjectProperty: emr:has_property
                Domain: tlo:Entity
                Range:  emr:Property

ObjectProperty: emr:encompasses
                Domain: emr:Area
                Range:  emr:Area

ObjectProperty: emr:constrains
                Domain: nep:DeontologicalNorm
                Range:  emr:AgentAction

ObjectProperty: emr:destination
                Domain: emr:AgentAction
                Range:  emr:Area
```

**Example EMR OWL Class Definitions**

```
Class: emr:Obligation  SubClassOf: nep:DeontologicalNorm
Class: emr:Prohibition SubClassOf: nep:DeontologicalNorm
Class: emr:Area        SubClassOf: tlo:SpatioTemporalPlace

Class: emr:Property       SubClassOf: tlo:Property
Class: emr:AgentPlan      SubClassOf: nep:AgentPlan
Class: emr:PlanAction     SubClassOf: nep:PlanAction
Class: emr:AgentAction    SubClassOf: nep:AgentAction
Class: emr:AssistantRobot SubClassOf: nep:Robot
```

**Example SWRL Obligation Rule**

```
Rule: emr:AssistantRobot( robbie),
      emr:Resident( ?b),
      emr:Area( ?w),
      tlo:located_at( ?b, ?w).
      emr:has_property(?w, high_risk),
      emr:AgentAction( send_support),
      emr:PlanAction( prepare_evac),
      nep:is_implemented_by( prepare_evac, send_support),
      emr:constrains( assistance_required, send_support)
      -> nep:state( assistance_required, nep:activated).
         emr:has_property( send_support, emr:enabled),
         nep:executes( robbie, prepare_evac)
```

**Example EMR OWL Instance Definitions**

```
Individual: emr:high_risk  Types: emr:Property
Individual: emr:unsafe     Types: emr:Property
Individual: emr:enabled    Types: emr:Property
Individual: emr:disabled   Types: emr:Property

Individual: area1   Types: emr:Area
            Facts:  emr:encompasses emr:wobegon
Individual: wobegon Types: emr:Area
            Facts:  emr:has_property emr:high_risk
Individual: area2   Types: emr:Area
            Facts:  has_property emr:unsafe

Individual: robbie  Types: emr:AssistantRobot
Individual: bob     Types: dpp:Person
            Facts:  tlo:located_at emr:wobegon

Individual: send_support  Types: emr:AgentAction
Individual: prepare_evac  Types: emr:PlanAction
            Facts:  nep:is_implemented_by emr:send_support

Individual: evac_the_area Types: emr:AgentAction
            Facts:  emr:destination emr:area2

Individual: assistance_required Types: emr:Obligation
            Facts:  emr:constrains emr:send_support

Individual: safe_evacuation     Types: emr:Prohibition
            Facts:  emr:constrains emr:evac_the_area
```

**Example SWRL Prohibition Rule**

```
Rule: emr:Area( ?us),
      emr:has_property(?us, unsafe),
      emr:AgentAction( evac_the_area),
      emr:destination( evac_the_area, ?us),
      emr:constrains( safe_activation, evac_the_area)
      -> nep:state( safe_evacuation, nep:activated).
         emr:has_property( evac_the_area, emr:disabled)
```

Fig. 9. Example OWL axioms for EMR model.

relationships for the created instances. As an example, the "area1" and "wobegon" individuals are defined as instances of the EMR Area class where "area1" encompasses the "wobegon" area. and where "wobegon" has the property of "high risk."

The last two portions of Fig. 9 depict the SWRL rules that specify the commitments for the two deontological norms in the example scenario. Each rule is composed with a conjunction of facts as preconditions that when true establish the post-condition predicates as facts in the OWL application ontology. The predicates with "*?*" prefixes represent free variables that may be bound or grounded to OWL instances during a reasoning process.

In the example Prohibition rule, the rule specifies that when there is an Area instance that has the property of "unsafe" and when that area instance is the destination of an Agent Action instance labeled "*evac the area*" and when there is a Prohibition Norm instance labeled "safe evacuation" that constrains the Agent Action, the consequent facts stipulating that the state of the Prohibition Norm is "activated" and that

the "*evac the area*" Agent Action is "*disabled*" are to be asserted.

## ACKNOWLEDGMENT

This article is not linked any particular funding, the authors were able to dedicate time to this project thanks to the support of their respective institutions. The authors recognize the strong support of the whole IEEE P7007 Working Group.

## REFERENCES

[1] E. Tunstel et al., "Systems science and engineering research in the context of systems, man, and cybernetics: Recollection, trends, and future directions," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 5–21, Jan. 2021.

[2] R. Chatila, K. Firth-Butterflied, J. C. Havens, and K. Karachalios, "The IEEE global initiative for ethical considerations in artificial intelligence and autonomous systems [standards]," *IEEE Robot. Autom. Mag.*, vol. 24, no. 1, p. 110, Mar. 2017.

[3] "Ethically aligned design," IEEE Standards Association. 2019. [Online]. Available: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v1.pdf

[4] E. Prestes et al., "The first global ontological standard for ethically driven robotics and automation systems [standards]," *IEEE Robot. Autom. Mag.*, vol. 28, no. 4, pp. 120–124, Dec. 2021.

[5] R. Studer, V. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *Data Knowl. Eng.*, vol. 25, nos. 1–2, pp. 161–197, 1998.

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

[7] E. Prestes et al., "Towards a core ontology for robotics and automation," *Robot. Auton. Syst.*, vol. 61, no. 11, pp. 1193–1204, 2013.

[8] S. Brunner, M. Kucera, and T. Waas, "Ontologies used in robotics: A survey with an outlook for automated driving," in *Proc. IEEE Int. Conf. Veh. Electron. Saf. (ICVES)*, 2017, pp. 81–84.

[9] M. A. Cornejo-Lupa, R. P. Ticona-Herrera, Y. Cardinale, and D. Barrios-Aranibar, "A survey of ontologies for simultaneous localization and mapping in mobile robots," *ACM Comput. Surveys*, vol. 53, no. 5, pp. 1–26, 2020.

[10] F. Muhlenbach, "A methodology for ethics-by-design AI systems: Dealing with human value conflicts," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, 2020, pp. 1310–1315.

[11] D. Cavaliere, V. Loia, A. Saggese, S. Senatore, and M. Vento, "Semantically enhanced UAVs to increase the aerial scene understanding," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 3, pp. 555–567, Mar. 2019.

[12] *Industrial Automation Systems and Integration—Integration of Life-Cycle Data for Process Plants Including Oil and Gas Production Facilities—Part 1: Overview and Fundamental Principles*, ISO Standard 15926-1:2004, 2004.

[13] *Geographic Information—Ontology—Part 4: Service Ontology*, ISO Standard 19150-4:2019, 2019.

[14] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: State-of-the-art, needs and perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2389–2406, 3rd Quart., 2018.

[15] T. Haidegger et al., "Robotics: Enabler and inhibitor of the sustainable development goals," *Sustain. Prod. Consumption*, to be published. [Online]. Available: https://doi.org/10.1016/j.spc.2023.11.011

[16] D. Gašević, D. Djurić, and V. Devedžić, *Model Driven Architecture and Ontology Development*. Heidelberg, Germany: Springer, 2006. [Online]. Available: https://link.springer.com/book/10.1007/3-540-32182-9

[17] D. Djurić, D. Gašević, and V. Devedžić, "MDA standards for ontology development," in *Ontologies*, vol. 15. Boston, MA, USA: Springer, 2007, pp. 215–264. [Online]. Available: https://link.springer.com/chapter/10.1007/978-0-387-37022-4_8

[18] *Information Technology—-Common Logic (CL)—A Framework for a Family of Logic-Based Languages*, ISO/IEC Standard 24707:2018, 2018.

[19] I. Niles and A. Pease, "Towards a standard upper ontology," in *Proc. Int. Conf. Formal Ontol. Inf. Syst.*, 2001, pp. 2–9.

[20] H. Herre, "General formal ontology (GFO): A foundational ontology for conceptual modelling," in *Theory and Applications of Ontology: Computer Applications*. Dordrecht, The Netherlands: Springer, 2010, pp. 297–345. [Online]. Available: https://link.springer.com/chapter/10.1007/978-90-481-8847-5_14

[21] G. Guizzardi, A. B. Benevides, C. M. Fonseca, D. Porello, J. P. A. Almeida, and T. P. Sales, "UFO: Unified foundational ontology," *Appl. Ontol.*, vol. 17, no. 1, pp. 167–210, 2022.

[22] J. F. Sowa, *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Pacific Grove, CA, USA: Brooks/Cole Publ. Co., 1999.

[23] A. Spagnolli, L. E. Frank, P. Haselager, and D. Kirsh, "Transparency as an ethical safeguard," in *Proc. Int. Workshop Symbiotic Interact.*, 2018, pp. 1–6.

[24] S. Lakhmani, J. Abich, D. Barber, and J. Chen, "A proposed approach for determining the influence of multimodal robot-of-human transparency information on human-agent teams," in *Proc. Int. Conf. Augmented Cogn.*, 2016, pp. 296–307.

[25] A. Weller, "Transparency: Motivations and challenges," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham, Switzerland: Springer Int. Publ., 2019, pp. 23–40. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-28954-6_2

[26] S. Larsson and F. Heintz, "Transparency in artificial intelligence," *Internet Policy Rev.*, vol. 9, no. 2, pp. 1–16, 2020.

[27] J. Olszewska, "Designing transparent and autonomous intelligent vision systems," in *Proc. 11th Int. Conf. Agents Artif. Intell.*, 2019, pp. 850–856.

[28] F. Cantucci and R. Falcone, "Towards trustworthiness and transparency in social human–robot interaction," in *Proc. IEEE Int. Conf. Human-Mach. Syst. (ICHMS)*, 2020, pp. 1–6.

[29] *IEEE Standard for Transparency of Autonomous Systems*, IEEE Standard 7001-2021, 2022.

[30] R. van den Hoven van Genderen, "Legal personhood in the age of artificially intelligent robots," in *Research Handbook on the Law of Artificial Intelligence*, W. Barfield and U. Pagallo, Eds. Cheltenham, U.K.: Edward Elgar Publ., 2018, pp. 213–250.

[31] 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*. Singapore: Springer 2018, pp. 15–55. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-13-2874-9_2

[32] O. Ulgen, "A human-centric and lifecycle approach to legal responsibility for AI," *Commun. Law J.*, vol. 26, no. 2, pp. 97–108, 2021.

[33] L. Floridi and J. Sanders, "On the morality of artificial agents," *Minds Mach.*, vol. 14, no. 3, pp. 349–379, Aug. 2004.

[34] A. Khamis et al., "Robotics and intelligent systems against a pandemic," *Acta Polytechnica Hun.*, vol. 18, no. 5, pp. 13–35, 2021.

[35] T. Haidegger, S. Speidel, D. Stoyanov, and R. Satava, "Robot-assisted minimally invasive surgery—Surgical robotics in the data age," *Proc. IEEE*, vol. 110, no. 7, pp. 835–846, Jul. 2022.

[36] F. Meneguzzi, O. Rodrigues, N. Oren, W. W. Vasconcelos, and M. Luck, "BDI reasoning with normative considerations," *Eng. Appl. Artif. Intell.*, vol. 43, pp. 127–146, Aug. 2015.

[37] *IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems*, IEEE Standard 7007-2021, 2021.

[38] "OWL 2 Web ontology language Manchester syntax (second edition)," W3C Working Group Note 11, W3C, Cambridge, MA, USA, 2012. [Online]. Available: https://www.w3.org/TR/owl2-manchester-syntax/

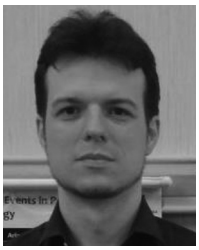
[39] A. Lawan and A. Rakib, "The semantic Web rule language expressiveness extensions—A survey," 2019, arXiv:1903.11723.

[40] O. Ulgen, "User rights and adaptive a/IS –from passive interaction to real empowerment," in *Adaptive Instructional Syst.*, R. A. Sottilare and J. Schwarz, Eds. Cham, Switzerland: Springer Int. Publ., 2020, pp. 205–217. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-50788-6_15

**Michael A. Houghtaling** (Member, IEEE) received the Bachelor of Science degree from the U.S. Air Force Academy, Air Force Academy, CO, USA, in 1966, and the M.S. degree in computer science from the University of Arizona, Tucson, AZ, USA, 1975.

His graduate studies at the University of Arizona included a minor in computational linguistics with a focus on formal logic, semantics, and transformational grammars. His professional career spanned 35 years as a Software Engineer with IBM (retiring), Tucson, in 2014 as a Senior Software Engineer. His projects and interests include knowledge-based software engineering, model-based reasoning, system usage models supporting statistical testing protocols, and complex event processing agents for distributed system diagnostics and maintenance. Application of industry standards in these domains revealed the importance of formal ontological models supporting stakeholder collaboration which motivates the focus of his continuing activities.

Mr. Houghtaling is a Senior Member of ACM, and an AAAI Member.



**Sandro Rama Fiorini** received the Ph.D. degree in computer science from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

He is a Research Scientist with IBM Research, Rio de Janeiro, Brazil. His work spans theoretical and applied topics in knowledge representation and knowledge graphs, particularly in geology, robotics and, more recently, material design.

Dr. Fiorini was the Vice-Chair of the EDARR Working Group in charge of writing IEEE 7007–2021 standard, also being heavily involved in the writing of IEEE 1872–2015 standard.

**Nicola Fabiano** (Member, IEEE) graduated in Law from the University of Bari, Bari, Italy (top grade and magna cum laude). He received the postgraduate degree in civil law (civil law specialist) from the University of Camerino, Camerino, Italy.

He is currently an Italian Lawyer, entitled to represent clients before the Italian High Courts, and an Adjunct Professor with Ostrava University, Rome, Italy. He is the former President of the San Marino Data Protection Authority from 2019 to 2021 and also the former National Expert for the Republic of San Marino on the "Consultative Committee of the Council of Europe of the Convention for the Protection of Individuals about Automatic Processing of Personal Data (Convention No. 108)," and the "Ad hoc Committee on Artificial Intelligence (CAHAI)" of the Council of Europe. He has been the Government of the Republic of San Marino's Advisor for drafting legislation on personal data protection from November 2017 to April 2018. He is an author of scientific articles, essays, books, and papers.

Mr. Fabiano is a winner of several awards. He is a member of the Italian National Bar (Consiglio Nazionale Forense—CNF) Privacy Commission and a Technical Expert at Accredia (The Italian Accreditation Body) Certification and Inspection Department. He is also a frequent speaker at international conferences to present the findings of some of his research related to the most innovative and technological solutions and their impact on natural persons and, hence, on data protection and privacy. He created the Data Protection and Privacy Relationships Model based on high-mathematics and set theory. He is certified as a Data Protection Officer and Privacy Assessor (UNI 11697:2017), a Security Manager (ICT)—UNI 11506:2017, and a Information Security Management Systems Professionals—ISO 27021:2017.

**Paulo J. Sequeira Gonçalves** (Senior Member, IEEE) was born in Covilhã, Portugal, in 1972. He received the M.Sc. and Ph.D. degrees in mechanical engineering (control, automation, robotics, and industrial informatics scientific area) from the University of Lisbon, Lisbon, Portugal, in 1998 and 2005, respectively.

He is currently an Associate Professor with the Department of Electrotechnical and Industrial Engineering, Polytechnical University of Castelo Branco (IPCB), Castelo Branco, Portugal. He is a Senior Researcher with the Institute of Mechanical Engineering, Instituto Superior Técnico, University of Lisbon. He is responsible for the Robotics and Intelligent Equipment Laboratory of IPCB. His main research interests include ontologies, computational intelligence, robotics, industrial automation, computer vision, and in particular visual servo control of robots. He is the author or coauthor of over 150 journal articles, books, book chapters, and conference articles across the various domains described above.

Prof. Sequeira Gonçalves received the "Scientific Merit Award" from IPCB in 2017 and 2022, for his merits. He is currently serving as the Vice-Chair of the Portuguese Chapter, IEEE Standards Association, IEEE Industrial Electronics, and IEEE Computational Intelligence, where he served as the Chair of the Portuguese Chapter. He is currently the Chair of the IEEE P1872.3 Standard for Ontology Reasoning on Multiple Robots, and was an Officer of the working groups who worked toward IEEE RAS related standards, P1871, and P7007. Those working groups have received the IEEE SA "Emerging Technology Award" in 2015 and 2021. He is an Active Member of the IEEE Robotics and Automation Society.

**Ozlem Ulgen** (Member, IEEE) is an Associate Professor of Law with the School of Law, University of Nottingham, Nottingham, U.K. She specializes in the law, ethics, and regulation of AI and robotics in the civilian and military sectors. Her research focuses on protecting human agency, rights, and attribution of legal responsibility. She has numerous publications on Kantian ethics and human dignity in AI and robotics, cosmopolitan ethics in warfare, and the law and ethics of autonomous weapons. She is the author of "A 'Human-Centric and Lifecycle Approach' to Legal Responsibility for AI" *Communications Law Journal* (2021).

Dr. Ulgen is involved in international law-making and standard-setting work for the UN, UNESCO, IEEE, and acts as an Academic Legal Expert to the UN Group of Governmental Experts on Legal Autonomous Weapons Systems. She is an EPSRC SPRITE+ Expert Fellow and served as an Expert Member of IEEE P7007 and P7000. She served on the IEEE Committee for Classical Ethics in Autonomous and Intelligent Systems, and was a contributor to the chapter on "Classical Ethics in A/IS" in The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems (IEEE Report, April 2019).

**Tamás Haidegger** (Senior Member, IEEE) received the M.S.E.E. and M.S.B.M.E degrees from the Budapest University of Technology and Economics, Budapest, Hungary, and the Ph.D. degree in medical robotics.

He is currently a Professor with Óbuda University, Budapest, where he is also the Director of the University Research and Innovation Center and the Technical Lead of Medical Robotics Research. He is a Bolyai Fellow of the Hungarian Academy of Sciences, Budapest, also holding the title of Consolidator Research of Óbuda University. Besides, he is a Research Area Manager with the Austrian Center of Medical Innovation and Technology (ACMIT), Wiener Neustadt, Austria, working on surgical simulation and training. He is the author and coauthor of over 250 scientific papers, books, articles across the various domains of biomedical engineering, with over 3000 independent citations to his work, and he holds ten patents. His main fields of research are medical technologies, control/teleoperation of surgical robots, image-guided therapy, and digital health technologies.

Dr. Haidegger received the Gabor Dennis Award, the MTA Bolyai Plaque, and the Kalmar Award among other recognitions, for his merits. He is the Co-Editor-in-Chief of *Acta Polytechnica Hungarica* and an Associate Editor of the IEEE TRANSACTIONS ON MEDICAL ROBOTICS AND BIONICS and the *IEEE Robotics and Automation Magazine*. He is an Active Member of the IEEE Robotics an Automation Society (serving as an Associate VP), IEEE SMC, IEEE EMBC, IEEE SA, and euRobotics aisbl, holding leadership positions in the IEEE Hungary Section as well. He has been running a professional blog on medical robotic technologies for over 15 years: http://surgrob.blogspot.com.

**Joel Luís Carbonera** received the bachelor's degree in computer science from the University of Caxias do Sul, Caxias do Sul, Brazil, in 2007, and the master's and Doctoral degrees in artificial intelligence from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil, in 2012 and in 2016, respectively.

He is currently an Adjunct Professor with the Federal University of Rio Grande do Sul. He has worked as a Scientist with IBM Industrial Laboratory, Brazil. He has experience in the area of artificial intelligence with an emphasis on knowledge engineering, ontology engineering, and machine learning. He also has experience in developing ontologies and knowledge-based systems for the Petroleum Geology domain and developing ontology-based standards for the Robotics and Automation area. His main current research interest is investigating cognitively inspired approaches to knowledge representation and reasoning.

Dr. Carbonera has served as a Deputy Coordinator of the IEEE 1872.2-Standard for Autonomous Robotics Ontology working group and as a Contributor to the IEEE 1872.1-Robot Task Representation working group. He was also a Coordinator of the standardization area of the Southern Brazil chapter of the IEEE Robotics and Automation Society. He was a member of IEEE's Robotics and Automation Working Group, which developed the first ontology-based standard for knowledge representation in the robotics and automation domain. He was also a member of the IEEE 7007-Ontological Standard for Ethically Driven Robotics and Automation Systems working group, which developed a standard for representing knowledge about ethical notions for robotics and automation. He is currently also a member of the research group "Computing Systems for Petroleum E&P," of the BDI Group (UFRGS intelligent database group).

**Joanna Isabelle Olszewska** (Senior Member, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees from École polytechnique fédérale de Lausanne, Lausanne, Switzerland, and the Ph.D. degree from University College London, London, U.K.

She is a British Computer Scientist. She is a Lecturer (Ass. Prof.) with the University of the West of Scottland, Glasgow, U.K., and leads research in Algorithms and Softwares for Trustworthy Intelligent Vision Systems. She is an author of 100+ peer-reviewed publications and one book.

Dr. Olszewska holds several awards, e.g., ESWA Outstanding Reviewer Award, and distinctions, e.g., ACM Distinguished Speaker, and she has been a supervisor for 60 B.Sc. (Hons.)/M.Sc./Ph.D. students, including the supervision of the Best B.Sc. Thesis awarded by the GCHQ Prize and the supervision of the Best M.Sc. Dissertation awarded by the AGD Prize. She stands on the IEEE Artificial Intelligence Standard Committee, she is part of the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, and she is the Co-Chair of the IEEE Technical Committee on Verification of Autonomous Systems. She has given talks, e.g., at conferences, such as ICRA, at events, such as the Canadian Mathematical Society (CMS) Anniversary Meeting, at ENS Paris. She has been a TPC member of over 100 international conferences, such as IJCAI and has chaired over 70 conference/workshop sessions, e.g., at IROS. She has been appointed as the Guest Editor of the *Knowledge Engineering Review Journal* (Cambridge University Press), and an Associate Editor for the *Frontiers in Artificial Intelligence* and the *Machine Learning with Applications* (Elsevier). She has contributed to 11 ISO/IEC/IEEE standards in various roles, e.g., the Vice-Chair of ISO/IEC/IEEE P41062. She is a C.Eng. and C.Sci. and a Fellow of BCS and HEA.

**Brian Page** (Life Senior Member, IEEE) received the Bachelor of Science degree (with Distinction) in electrical engineering from the University of Calgary, Calgary, AB, Canada.

He has a long and distinguished career in the field of electrical engineering, spanning almost half a century. Throughout his extensive professional journey, he has consistently showcased his expertise in the complex domain of system integration, particularly in the context of large High Availability Telecom and IT networks. As a key member of the Calgary 1988 and Atlanta 1996 Olympic organizing committee, responsible for significant portion of the design, implementation and operation of the games technology systems. He worked on leading edge complex projects operating from DC to light plus a dabble in Radio Astronomy.

Mr. Page is a Life Senior Member of IEEE, who has contributed to numerous IEEE published standards P7002, P7007, P23026, mhealth1752 as well as additional standards under development.

**Zvikomborero Murahwi** (Member, IEEE) received the B.S. degree in computer science and mathematics, and the master's degree in project management (information systems management).

He is currently consulting independently specializing in AI Ethics and Explainable AI. He actively participates in AI Ethics standardization with both IEEE and ISO and was the Technical Editor for IEEE7010-2020 and a WG Secretary for the development of IEEE7000-2021.

Mr. Murahwi is an IEEE-SA Member and an IEEE Certified Lead Assessor and a Trainer.

**Edson Prestes** (Senior Member, IEEE) received the B.Sc. degree in computer science from the Federal University of Pará Amazon, Belém, Brazil, in 1996, and the M.Sc. and Ph.D. degrees in computer science from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil, in 1999 and 2003, respectively.

He is a Full Professor with the Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil. He is a Leader of the Phi Robotics Research Group and a CNPq Research Fellow. Throughout his career, he has worked on several initiatives related to standardization, robotics, artificial intelligence, and ethics of artificial intelligence in academia, industry, international, and multilateral organizations.

Prof. Prestes is a member of the Global Future Council on the Future of Artificial Intelligence and of the Cross-Global Future Council Working Group on the G20 Digital Agenda at World Economic Forum; a South America Ambassador at IEEE TechEthics; the Chair of the IEEE RAS/SA 7007—Ontologies for Ethically Driven Robotics and Automation Systems Working Group; the Vice-Chair of the IEEE RAS/SA Ontologies for Robotics and Automation Working Group; a Former Member of the United Nations Secretary-General's High-Level Panel on Digital Cooperation; and a Former Member of the UNESCO Ad Hoc Expert Group for the Recommendation on the Ethics of Artificial Intelligence. He is a Senior Member of the IEEE Robotics and Automation Society and IEEE Standards Association. Additional information can be found at www.inf.ufrgs.br/ prestes/ or https://www.linkedin.com/in/edson-prestes/.