GC REAIM
GLOBAL COMMISSION ON RESPONSIBLE AI IN THE MILITARY DOMAIN

Foreign, Commonwealth & Development Office

GC REAIM Expert Policy Note Series
Effective Governance Through Precise Common Understanding

Edson Prestes, Nayat Sánchez-Pi and Maria Vanina Martinez

April 2025

POWERED BY The Hague Centre for Strategic Studies

GC REAIM Expert Policy Note Series

Effective Governance Through Precise Common Understanding

Authors: Edson Prestes, Nayat Sánchez-Pi, and Maria Vanina Martinez

April 2025

Cover photo: Unsplash

The Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM) is an initiative of the Government of the Netherlands that was launched during the 2023 REAIM Summit on Responsible Artificial Intelligence in the Military Domain in The Hague. Upon request of the Dutch Ministry of Foreign Affairs, the Hague Centre for Strategic Studies acts as the Secretariat of the Commission.

The GC REAIM Expert Policy Note Series was funded by the Foreign, Commonwealth and Development Office (FCDO) of the United Kingdom. GC REAIM Experts maintained full discretion over the topics covered by the Policy Notes. The contents of the GC REAIM Expert Policy Note series do not represent the views of the Global Commission as a whole. The Policy Notes are intended to highlight key issues related to the governance of AI in the military domain and provide policy recommendations.

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies
Email: info@hcss.nl
Website: www.hcss.nl

Introduction

One of the basic requirements to create effective hard and soft instruments is to have a good and detailed taxonomy that defines unambiguously all the main entities of the real-world domain of interest. This precise taxonomy will allow us to communicate, explicitly and without multiple interpretations, about concepts that are at the core of these hard and soft instruments. This implies that if one is referring to a particular instance of a concept, all stakeholders involved in this dialogue will understand exactly what is being stated. In addition, they will facilitate the enforcement of the laws and certification process in case of Artificial Intelligent (AI) systems. Although the benefits are clear, precise taxonomies are far from being obtained through a simple and direct process. Achieving a taxonomy involves multiple rounds of negotiations among different stakeholders to reach a consensus, mainly when these stakeholders come from different parts of the world and have different expertise, possibly highlighting different aspects they judge relevant in regulatory debates.

Despite the fact that this process is complex, we, as a global society, need to move the needle gradually, engaging empathetically and doing our best to reach a consensus. Without a precise taxonomy we take the risk of finding ourselves in situations where incorrect and unfair applications of the aforementioned instruments will occur, causing problems of a different nature, be it the incorrect application of the law to systems that were wrongly classified or by not properly covering all systems of interest. For instance, consider an instrument that prohibits all artificial intelligence systems that have a particular behaviour or have crossed a “red line”. To check the compliance with this prohibition, first, one needs to have in hand tools to verify if the system that is being assessed is based on AI. Depending on the way we define what AI means, only a subset of AI systems will be considered. For example, if a taxonomy defines AI systems as needing to have the capacity to learn from the data provided by the user, only systems with this capacity will be taken into consideration, leaving aside classes of the systems that do not learn. Similarly, if this prohibition focuses on behaviours explicitly stated in the system, it will leave out those that emerge from the execution of simpler pre-programmed behaviours. This is the case of algorithms that simulate ant colonies. In this case, each ant in the colony has a simple and individual behaviour, however, as a group, the ant colony can exhibit complex and emergent behaviours not considered before, like building nests, finding food, and so on. Within the military domain, a clear analogy of this could be in trying to decide the compliance of a swarm of drones that perform different surveillance tasks in a determined region, the under- or over specification and classification of such systems could lead to missed opportunities or, otherwise, preventable risks.

This short paper advocates for the need for a more scientific approach to provide a solid foundation to the discussion on regulatory frameworks for AI that is based on methodologies from Computer Science. One of the main and direct benefits is to have a better and clearer picture of the domain to create realistic and effective governance instruments.

Loose Definitions and Problems for Regulations

Using taxonomies with vague or imprecise definitions is common in several domains. This often happens because they are expressed in natural language, e.g. English, that is inherently ambiguous and consequently passive to multiple interpretations. Sometimes this vagueness is intentionally created with the aim not to impose strict boundaries and give flexibility to accommodate instances of different natures that may appear in the middle and long-term future. In addition, both vagueness and imprecision are influenced by non-technical aspects which depend on the point of view of the person or group that is proposing them. As a simple example, let us make a quick reflection on the term artificial intelligence. Although AI is an overly mentioned term, there is no consensus on what AI means. Some proposals for defining AI can be found on Norvig and Russell.[1]

Currently, AI is commonly used as a synonym for systems powered by the implementation of a (set of) method (s) from a very specific branch of Computer Science called Artificial Intelligence that aims to computationally simulate human intelligence. In this case, AI is used for both denoting systems and domains. In addition, one can refer to a system as exhibiting artificial intelligence based on its observed behaviour, without looking into how the system was developed. As one can quickly observe, AI is an overloaded term that may express different intentions, as a system, emergent property, or a research domain. Considering AI as a domain allows us to understand that AI is a rich ecosystem of investigation that includes a variety of approaches and techniques. A good example of potential categorization of AI can be found at.[2]

For benefits of clarity and exemplification, Figure 1 shows Artificial Intelligence as a subdomain of Computer Science that encompasses Machine Learning which in turn has Deep Learning as a subdomain. Thus, among all possible methods and classes of algorithms, there are specific methods used for creating AI systems. Similarly, among all possible Machine Learning classes of algorithms, there is a specific one called Deep Learning. It is worth noticing, this representation is simplistic and does not contain all existing subdomains.

[1] Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Fourth edition., Pearson Series in Artificial Intelligence (Hoboken, NJ: Pearson, 2021).
[2] EIT Community Artificial Intelligence, 'Creation of a Taxonomy for the European AI Ecosystem | EIT', 15 September 2021, https://www.eit.europa.eu/library/creation-taxonomy-european-ai-ecosystem.

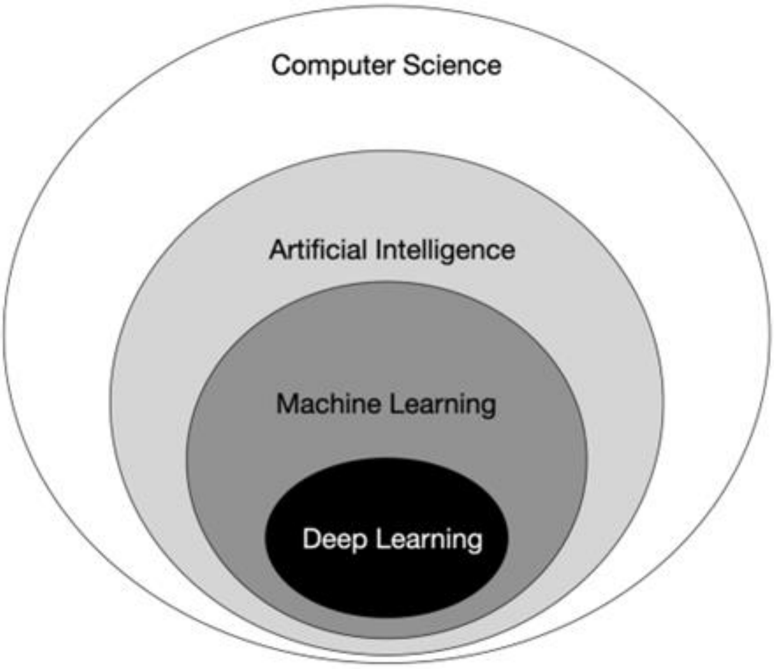

Figure 1: Simplified view of the AI domain.

In regulatory settings, vagueness, imprecision, malleability and incorrect uses of the term AI must be avoided to transmit the intended message. If two people have different understandings of the domain, they will always find counterexamples that will delay any consensus. Their individual view can be correct, but the lack of a common and precise taxonomy will make any negotiation difficult because of simply the misuse of the correct terminology.

Consider for instance the situation where AI is used as synonymous to Machine Learning (ML), or as Deep Learning (DL), in a debate. If the intention is to create a specific law that affects all systems that have the capacity to "learn", the interchangeable use of AI instead of ML or DL will lead this discussion in the wrong direction since some experts will argue this law will not be effective, because some systems do not have the capacity to learn and will be subject to the constraints or regulations imposed by this law (see grey region in Figure 1). A simple example is an Expert System that incorporates a set of rules designed by a human expert on a particular domain in the design time of the system. It has "intelligence" and is an example of an AI based system. A similar situation will happen if ML is used as synonymous to DL regarding transparency. DL is a very specific subdomain in the ML ecosystem focused on Neural Networks (NNs), which are well known for their opaque nature. In NNs, the stored knowledge is represented by multidimensional vectors the meaning of which is not easy to understand. However, in the ML domain, there is a class of algorithms called decision trees that are used for, e.g., classification, and have a higher degree of transparency.

Different views of the domain may also imply difficulties to trade globally or have global coordination around the domain. Consider the situation where country A defines AI systems, for regulatory purposes, as any computational system that can produce an output based on a particular input; whereas country B defines AI systems as those that can mimic human intelligence. From Figure 1, country A's definition for AI systems encompasses all systems defined in the Computer Science domain, while country B considers AI systems as all systems that belong to the Artificial Intelligence region that include systems in the grey, dark grey and black areas. A basic requirement for international trade is that systems developed in the export country are in accordance with the regulations of the import country. Using the previous example, AI systems developed by country A can face some challenges in entering the market of country B, because only a subset of the systems has the properties required by country B to be considered as AI systems. The incompatibility of classification may lead to diminishing the market values of goods in international trade.

Semantically Rich Models

As AI-based systems are rapidly growing in size and complexity, it becomes paramount to have clear definitions that can represent current systems but also can accommodate this diversity to better represent what will come in the future. Creating good taxonomies demands specific methodologies from Ontology Engineering to investigate the domain in a holistic way and produce models that describe the main entities and their relationships in a formal and unambiguous way.

IEEE Robotics and Automation Society (IEEE RAS) has conducted a very intense and distinctive process in this direction. Since 2010, IEEE RAS has embarked on creating standards focused on ontologies to enable unambiguous communication among humans, artificial systems and among artificial systems and humans. In Computer Science and related areas, an ontology is defined as a formal and explicit specification of a shared conceptualization, which can be processed by a computer. Ontologies represent the conceptual structure of a domain of interest through a set of classes, relationships between these classes, properties (or attributes) that characterize them, and axioms that impose constraints on the possible interpretations of classes, relationships, and properties providing additional semantics and information to clarify the intended meaning of the concepts and relationships.

The very first IEEE RAS standard based on ontologies was the IEEE 1872-2015 Ontologies for Robotics and Automation[3] as a result of the process that started in 2010. Another example is the IEEE 7007-2021 Ontological Standard for Ethically Driven Robotics and Automation Systems[4] which proposes a set of well-founded ontologies with vocabulary and definitions for describing the components and dependencies of ethically driven systems. This standard has also been advocated by the Brazilian Government to be used as a foundation to create confidence building measures in the GGE on LAWS meetings.[5] Due to the relevance of both standards, their respective working groups were the recipients of the international IEEE Emerging Technology Award in 2015 and 2021. Though these proposals represent an advancement towards creating precise definitions in the area, they encode a relatively small subset of terms that are generic and common to a wide variety of application domains. Further efforts are needed for specific domains such as the one that concerns us here: the use of AI in the military domain.

---

[3] IEEE, 'IEEE Standard Ontologies for Robotics and Automation', *IEEE Std 1872-2015*, April 2015, 1–60, https://doi.org/10.1109/IEEESTD.2015.7084073.
[4] IEEE, 'IEEE Ontological Standard for Ethically Driven Robotics and Automation Systems', *IEEE Std 7007-2021*, November 2021, 1–119, https://doi.org/10.1109/IEEESTD.2021.9611206.
[5] Brazil. *The Possible Uses of IEEE's First Global Ontological Standard for Ethically Driven Robotics and Automation Systems as a Building Block for Confidence Building Measures Regarding the Responsible Development and Use of Autonomous Weapons Systems.* Working paper CCW/GGE.1/2024/WP.1, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Convention on Certain Conventional Weapons, March 1, 2024. https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2024)/CCW_GGE1_2024_WP1.pdf.

Due to their richness in terms of clarity and semantics, ontologies, and also Ontological Standards, can be used in different ways, such as:

- a guide for teaching how to design an AI system;
- a reference for policy makers and governments to draft AI related policies;
- a common vocabulary to enable communication among government agencies and other professional bodies around the world;
- a part of decision-making during investment in companies and technologies;
- a framework to create systems that can act according to the concepts defined by the ontologies;
- a basis for information transfer among different artificial and non-artificial agents;
- a core component to facilitate understanding of system behaviours.

Ontological Representation Examples

Figure 2 shows the unified modelling language (UML) diagram of the main concepts and relationships found in the Norms and Ethical Principles (NEP) ontology that appears on the IEEE 7007 Standard. Although this notation can be a bit technical, it is not difficult to understand.

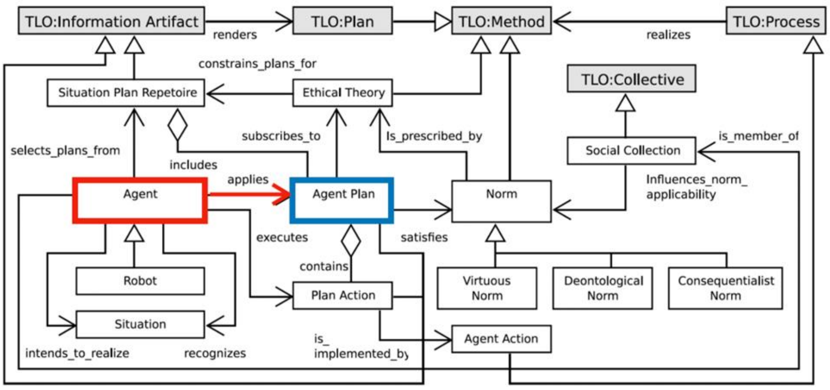

Figure 2: UML Diagram of the main concepts and relationships in NEP.

In this diagram, rectangles represent concepts that are relevant for the domain. For instance, a system powered by artificial intelligence can be considered as an Agent, if this system "can act autonomously and produce changes in their situated environment" (informal definition for Agent available in the IEEE 7007 Std.). Thus, the agent concept is relevant for our discussion and appears highlighted in red in the diagram. To act in the environment, an Agent needs a Plan which can be, e.g., simple and reactive or complex and deliberative. For instance, a robot can deviate to the left every time it detects an obstacle on its right (simple plan), or can plan how to reach a position in the environment avoiding obstacles or specific regions of interest (complex plan). Thus, the concept of Plan is also relevant and is highlighted in blue. The connection between these two concepts is done via a relationship called "applies", since an agent "applies" a plan during the execution of this task. This relationship is depicted in this diagram as a red arrow in the diagram. The tip of the arrow defines the way that one should read the relationship. For the "applies" relationship, the arrow goes from Agent to Agent Plan, then the relationship is read as "Agent applies Agent Plan." Thus, one can now observe that:

- Agent Plan must satisfies Norm
- Social Collection influences applicability of Norm

Diamond symbols indicate aggregation, i.e., that a concept is composed of others. In the diagram, any Agent Plan consists of one or more Actions which are implemented by the robot using specific commands defined by its programming library. Triangle symbols indicate inheritance, i.e., when a broad concept is specialized into a specific one. In the diagram, the concept of Robot is a specialized version of the Agent concept. This implies all Robots are also Agents having the same Agents characteristics, but some Agents may not be Robots. For instance, in the Data Privacy and Protection (DPP) ontology that is part of IEEE 7007, the concept of Person is a specialization of Agent that is not a robot. Using this kind of representation, it is possible to understand the connections and dependencies between concepts. For instance, a Robot, which is a specialization of Agent, can recognize situations present in the Environment and apply plans which can be composed of a set of Actions. All plans satisfy norms which can be of the following types: virtuous, deontological or consequentialist. Independent of the type of the norm, it can be influenced by a social collection, which the agent is part of. Social collection can be a government, a community, a company, and so on. More details can be found at IEEE 7007 Standard. All of these concepts are general, since they indicate how the knowledge is structured to accommodate different perspectives.

These types of conceptualizations also allow for information sharing and reuse. Figure 3 shows a high level description of tasks in the military domain, establishing the goals, legal requirements, functional area they belong to and in which level of command they are executed. Note that tasks are performed by agents (annotation 1..n in the relationship indicates at least one agent) and can have several agents that are effectively responsible for the outcome of the task. An agent is not specified in depth, however, it would perfectly fit with the ontology in Figure 2, that describes agents and other components related to agents independently of the application domain.

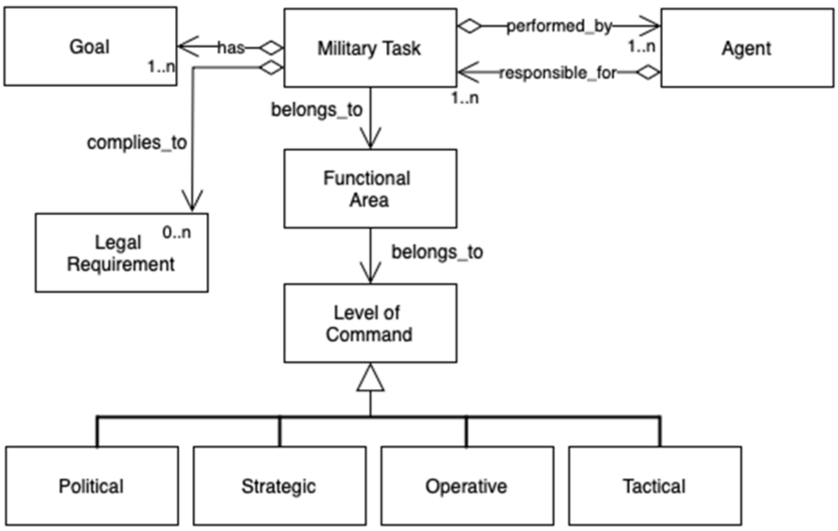

Figure 3: UML Diagram depicting a high level conceptualization of the components and concepts associated with military task.[6]

[6] This partial ontology was elaborated based on the interpretation of the authors of existing literature; Sarah Grand-Clément, 'Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain', 10 November 2023, https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain/; Merel Ekelhof and Giacomo Persi Paoli, 'The Human Element In Decisions About The Use Of Force', 31 March 2020, https://unidir.org/publication/the-human-element-in-decisions-about-the-use-of-force/.

For instance, assume a particular robotic system used for military purposes such as one that performs tasks of logistical support or surveillance. In order to design, understand or validate such a system, it is necessary to identify which concepts better represent the property of the system being taken into consideration. In this case, the system will be a specialization of a robot, situations, actions, plans, and the different types of norms that such plans satisfy need to be specialized accordingly. Using the ontologies one could understand the high-level aspects of the systems behaviour. Finally, the connection with Figure 1 is also very clear. An ontological representation of the domain would allow to identify and make explicit all the properties that are needed to consider an algorithm as belonging to the class of ML or DL. In addition, it would allow us to more precisely identify the dependencies that exist among classes of AI approaches through explicit relationships. These properties and relationships would not hinder advances in the domain or the development of new methods. They would provide a solid foundation to understand not only the current picture of the domain, but also how it is being expanded.

Conclusions

An ontology is a machine-processable artifact that captures a common understanding of the conceptual structure of a given domain. As a consequence, it can be incorporated into a computational system and be used in multiple ways besides being used as a taxonomy. Recent progress shows ontologies being used for enabling transparent mechanisms in wheeled robots to understand the situations faced by them as well as its decisions; to make heterogenous machines exchange information about a particular task; for tracking the source of a particular failure in an industrial robot; and so on. Using ontologies in the Responsible AI in the Military Domain process can enable a smooth communication among different stakeholders because of the constraints imposed by the model to precisely define and describe concepts. It can also allow these stakeholders to pose questions and automatically make inferences from a set of properties of interest to identify the best concepts that fit them when ontologies are used as the core of a computational system. In a governance framework, an ontology-based system would allow sharing information among different players with different levels of granularity about situations, norms, laws, and so on.

We conclude this policy brief with a list of practical recommendations:

- Ontological standards should be used as the main source of information to ground the debates on AI systems in the military domain.
- Domain ontologies must be developed to cover concepts not represented by existing ontological standards.
- Standards Developing Organizations must be invited and included in the discussion on regulations on the domain.
- Technical Standards must be developed in parallel with the discussion on regulations on AI in the military domain. Outputs produced must be shared.
- Centralized and trustful source of information must be created.

About the Authors

Prof. Dr. Edson Prestes

Edson Prestes is Full Professor at Institute of Informatics of the Federal University of Rio Grande do Sul, Brazil. He is leader of the Phi Robotics Research Group and CNPq Research Fellow. He received his BSc in Computer Science from the Federal University of Pará (1996), Amazon, Brazil, and MSc (1999) and PhD (2003) in Computer Science from Federal University of Rio Grande do Sul, Brazil. Edson is Senior Member of the IEEE Robotics and Automation Society (IEEE RAS) and IEEE Standards Association (IEEE SA).

Dr. Nayat Sánchez-Pi

Dr. Nayat Sánchez-Pi is a leading expert in Artificial Intelligence with a distinguished international career spanning Europe and Latin America. She serves as Director and CEO of Inria Chile, the Chilean centre of the French National Institute for Digital Science and Technology. Dr. Sánchez-Pi holds a Ph.D. in Artificial Intelligence from the University Carlos III of Madrid, for which she received the Extraordinary PhD Award.

Dr. Maria Vanina Martinez

Dr. Maria Vanina Martinez obtained her PhD at University of Maryland College Park and her postdoctoral studies at Oxford University. Since 2023, she is tenured scientist at the Artificial Intelligence Research Institute (IIIA-CSIC) in Barcelona. In 2018, she was selected by IEEE Intelligent Systems as one of the ten prominent researchers in AI to watch. She received the 2021 Argentine National Academy of Exact, Physical and Natural Sciences Stimulus Award in the area of Engineering Sciences and the 2023 Konex Award in Science and Technology.

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies