

# Formalization of Ethical Decision Making: Implementation in the Data Privacy of Wearable Robots

Sofia Almpani, National Technical University of Athens, Greece\*

 <https://orcid.org/0000-0003-4823-479X>

Petros Stefanias, National Technical University of Athens, Greece

Panayiotis Frangos, National Technical University of Athens, Greece

## ABSTRACT

As automation in robotics and artificial intelligence is increasing, we will need to automate a growing amount of ethical decision making. However, ethical decision-making raises novel challenges for designers, engineers, ethicists, and policymakers, who will have to explore new ways to realize this task. For example, engineers building wearable robots should take into consideration privacy aspects and their different context-based scenarios when programming the decision-making procedures. This in turn requires ethical input in order to respect norms concerning privacy and informed consent. The presented work focuses on the development and formalization of models that aim at ensuring a correct ethical behavior of artificial intelligent agents, in a provable way, extending and implementing a logic-based proving calculus. This leads to a formal theoretical framework of moral competence that could be implemented in artificial intelligent systems in order to best formalize certain parameters of ethical decision-making to ensure safety and justified trust.

## KEYWORDS

Argumentation, Data Privacy, Decision-making, Ethics, Logic, Logic programming, Wearable Robots

## INTRODUCTION

As autonomous artificial intelligent (AI) systems take up a progressively prominent role in our daily lives, it is undoubtedly that they will sooner or later be called on to make significant, ethically charged decisions and actions (Bringsjord et al., 2006). Over the last years, the issue of ethics in artificial intelligence and robots has gained great attention and many important theoretical and applied results were derived in the perspective of developing ethical systems (Tzafestas, 2018). But how could a robot or any AI agent be considered ethical? Some of the requirements needed are a broad capability to envisage the consequences of its own decisions as well as an ethical policy with rules to test each possible decision/consequence, so as to choose the most ethical scenario (Danaher, 2019; Tzafestas, 2018). The challenge is how we can guarantee that robots will always perform ethically correct behavior as defined by the ethical code declared by their human supervisors.

Academic research and real-life incidents of AI system failures and misuse have indicated the need for employing ethics in software development (Bringsjord et al., 2006). Nevertheless, studies on

DOI: 10.4018/IJEACH.320488

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

methods and tools to address this need in practice are still lacking, resulting in a growing demand for AI ethics as a part of software engineering (Vakkuri et al., 2019). But how can AI ethics be integrated in engineering projects when they are not formally considered? There has been some work on the formalization of ethical principles in AI systems (L. A. Dennis et al., 2015). Previous studies that attempt to integrate norms into AI agents and design formal reasoning systems has focused on: ethical engineering design (Flanagan et al., 2008; Robertson et al., 2019; Winfield et al., 2019; Wynsberghe, 2012), norms of implementation (Hofmann, 2012; Sisk et al., 2020), moral agency (Cunneen et al., 2019; Floridi & Sanders, 2004), mathematical proofs for ethical reasoning (Bringsjord et al., 2006), logical frameworks for rule-based ethical reasoning (Ågotnes & Wooldridge, 2010; Arkin, 2009; Iba & Langley, 2011), reasoning in conflicts resolution (Pereira & Saptawijaya, 2007), and inference to apply ethical judgments to scenarios (Blass & Forbus, 2015).

One of the categories of AI ethics is Ethics by Design, which is the incorporation of ethical reasoning abilities as a part of system behavior, such as in ethical robots (Vakkuri et al., 2019). In this work, if we assume that an AI agent can be capable of ethical agency, the purpose is to enable AI agents to reason ethically (L. A. Dennis et al., 2015). This includes taking into consideration societal and moral norms; to hierarch the respective priorities of norms in various contexts; to explain its reasoning; and to secure transparency and safety (Dignum, 2018). These systems are often established with the purpose to assist ethical decision-making by people, identifying the ethical principles that a system should not violate (L. Dennis et al., 2016).

Moral reasoning is a key issue in AI ethics, and computational formal proofs are perhaps the single most effective tool for determining credible and trustful reasoning (L. Dennis et al., 2016). The Moral extension of the Argumentative Proof Event Calculus (MAPEC) presented in (Almpani et al., 2022) combines the ethical framework from (L. Dennis et al., 2016) and the moral competence from (Malle & Scheutz, 2019) to develop a formal representation of ethical scenarios and integrate moral norms and concepts (See Figure 1). For a detailed description of the initial argumentative proof event calculus (APEC) see (Almpani & Stefaneas, 2017; Almpani et al., 2017).

The presented use case includes ethical considerations relating to the data privacy of wearable robots (WR). The case study in this paper describes the desirable ethical behavior of WRs concerning user's data access. It is discussed how code in robotic architecture affects data and privacy and why such issues should be considered from a formal verification perspective (Lutz & Tamò Larrieux, 2015).

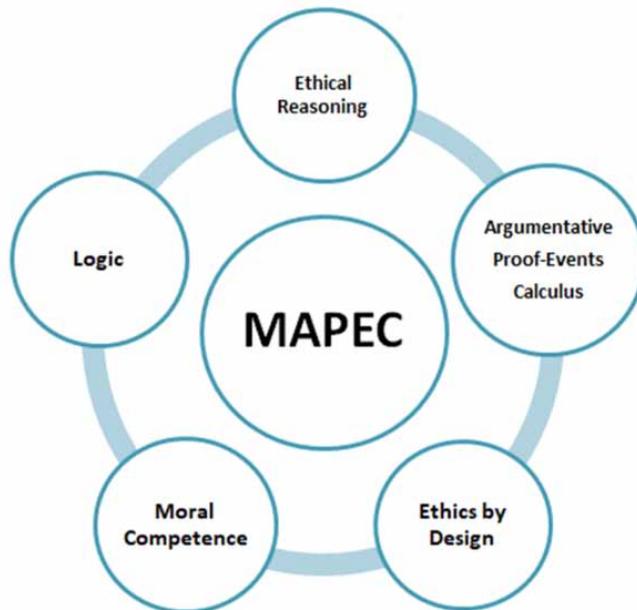
For the realization of this effort, the objectives are:

- to formalize what it means for a system's decision-making to be ethically correct;
- to provide a logical specification with which the system can be built and checked;
- to extend Argumentative Event-Calculus to create an abstract Moral framework (MAPEC) with ethical logic-based argumentation;
- to illustrate a case study concerning the data privacy of WRs to indicate how such an ethical framework can be implemented in computational systems.

## A FORMAL FRAMEWORK FOR ETHICAL CODES

In an autonomous system, it is not aimed to show that an agent always follows the moral thing, but that its actions are taken for the right reasons. In many real life scenarios, it is not easy to provide a complete set of decisions that will cover all situations (L. Dennis et al., 2016). Therefore, the system may have two modes of operation; either it uses its pre-existing set of actions in conditions which are within its anticipated parameters; or when new options appear it acts outside of these parameters based on various available resources that allow to govern its actions using ethical reasoning (L. Dennis et al., 2016).

Figure 1. Research framework of MAPEC



To represent ethical codes and rules it requires an ethical policy, a hierarchy over the rules that are appropriate in different contexts (defining even which rule is more acceptable to violate when no ethical option is available). To demonstrate that a system has the property of making the right decisions (both operationally and ethically), it should be formally specified what the “right decisions” are.

Formal verification (Fisher et al., 2013) includes to prove or disprove that a system is compliant with a requirement determined in a mathematical language, i.e., a “formally specified property” expressed within a linear temporal logic, which in our case allows us to define what decisions should the rational agents made at some specific moment (L. Dennis et al., 2016). Thus, the ethical policy can be formalized in some computational logic  $L$ , whose well-defined formulas and proof theory specify the basic concepts required: the temporal structure, events, actions, sequences, agents, and so on (Bringsjord et al., 2006). The presented methodology proof-theoretically formalizes the ethical policy and implements it, meaning that this methodology encodes not the semantics of the logic  $L$  but its proof calculus (Bringsjord et al., 2006).

Logic-based systems that are capable of dealing with increasing degrees of environmental uncertainty and variability are preferable (Gomila & Müller, 2012) and cognition constitutes a way to deal with an undefined and uncertain world, meaning not necessarily a chaotic one but just a complex one. Argumentation is a tool of cognition that can formalize the science of common sense reasoning on which new types of systems can be engineered (A. Kakas & Michael, 2016).

Therefore, to address the challenge of ensuring ethically correct behavior, a logic-based argumentation approach such as MAPEC (Almpani et al., 2022) is proposed to guarantee that robots only execute events that can be proved ethically acceptable in a human-selected logic, by formalizing an ethical code (Bringsjord et al., 2006).

### **Moral Competence Expressed with an Argumentation-Based Framework**

In an ethical framework, a moral vocabulary allows the agent to represent norms, ethically substantial behaviors, and their judgments (conceptually and linguistically) to fuel the moral communication. It contains: a normative frame referring to the features of norms and to the normatively-supported

qualities of agents; a language of norm violation characterizing attributes of violations and of violators; and a language of responses to violations (Malle & Scheutz, 2019). In our approach, the concept of norms is described with events, extending their context to abstract ethical events. The abstract ethical events present the arguments in a moral debate. The violations are analogous to the counterarguments. The role of ethical agents can be easily depicted as akin to the role of the supporter (or prover) and attacker in our argumentation framework (Almpani & Stefanias, 2017), where the supporter plays the role of the ethical correct agent and the attacker the role of the violator. Their actions are the responses to moral violations with arguments or counterarguments.

Moral communication expresses an agent's efforts to recognize, clarify, or defend norm events, as well as interfere or rectify after a norm violation.

### Definition 1: Abstract Ethical Events

An abstract ethical event is represented with  $e$  and its purpose is to defend an ethical principle  $c$ . The  $c$  can be interpreted also as ‘‘the supporter considers it immoral to permit or cause  $\neg c$  (to happen)’’. The *Abstract Ethical Event* has the same structural components (data  $\Phi$ , warrant  $w$ , ethical claim  $c$ ) as a proof event in APEC (Almpani et al., 2017). Thus, an ethical principle  $c$  is in force when the event concludes to  $c$ , based on the data  $\Phi$  and following the inference rules  $w$ :

$$e \text{ } \zeta \text{ } < \text{communicate } < \Phi, c >, w >$$

where  $e \in E$ ,  $E$  the set of ethical events for the  $c$ . Similarly,  $e^*$  denotes the *violation event*.

**Moral judgment** is the evaluation of the actions relative to norms that leads to the judgment of the temporal state of the moral actions, which includes the predicates *Happens*( $e,t$ ), *Initiates*( $e,f,t$ ), *ActiveAt*( $e,f,t$ ), and *Clipped*( $e,f,t$ ), leading finally either for the ethical principle to be *Valid*( $e,f,t$ ) or to *Terminate*( $e,f,t$ ) (Almpani et al., 2022).

A **system of norms** contains a society's principles for ethical behavior. They guide supporter's arguments and decisions to behave with specific (moral) actions and shape others' (moral) judgments of those behaviors (Malle & Scheutz, 2019). Thus, they establish *an ethical policy with ethical rules*.

### Definition 2: Ethical Policy

An ethical policy  $P$  is a tuple  $P = \langle R, \geq \rangle$  where  $R$  is a finite set of **ethical rules** between the events  $e$ , with  $e \in E$ , and  $\geq$  is a complete (not necessarily strict) priority order on  $R$ . The expression  $e_1 = e_2$  indicates that violating  $e_1$  is equivalently unethical as violating  $e_2$ , while  $e_1 \geq e_2$  denotes that violating  $e_1$  is equally or less unethical to violating  $e_2$ . A special category of ethical event, symbolized as  $e_0$ , is vacuously satisfied, and encompassed in every policy so that  $\forall \theta \in E: e > e_0$ , indicating it is always strictly more unethical to do nothing and permit any of the unethical conditions to happen.

**Moral action** is an event, taking place in compliance with the norms and in specific time, which is accommodated to and harmonized with other social agents (violators or provers) who operate under the same context. The norm violations  $e^*$  of a violator are denoted as **attack**( $e^*,t$ ) events and the ethical proving action of a supporter are denoted as **support**( $e,t$ ), specified both by the time  $t$  to express the temporal sequence of the actions.

### Definition 3: Ethical Actions

Given a certain context  $a$ , an event  $e$ , and an ethical principle  $c$ , an ethical action can be the formulas:

$\text{support}(e,t) (\Rightarrow \perp_a) c$  denoting the actions of a supporter to defend the ethical principle  $c$  with ethical event  $e$  in context  $\alpha$  and at time  $t$ .

$\text{attack}(e^*,t) (\Rightarrow \perp_a \neg c)$ , denoting the actions of a violator to contravene the ethical principle  $c$  with violation  $e^*$  in context  $\alpha$  and at time  $t$ .

## Prioritized Ethical Rules to Define Context-Based Scenarios

Context determines dynamic priorities on the decision policies of the agent (A. C. Kakas & Moraitis, 2003). To be able to reason about scenarios in terms of ethics we need a scenario selection process that uses the ethical policy, which can be represented within argumentation theory. The agent can be in various contexts while deciding which scenario to choose, so the rules from all the contexts need to be considered when implementing a plan. We advocate scenarios that are ethical or at least violate the fewest ethical principles, both in quantity and in severity.

The scenarios are ordered using  $<$  which leads to a complete order over scenarios (L. Dennis et al., 2016). This can describe an agent's ethical policy based on the different contexts with argumentation levels. In the first level we have the rules that refer directly to the domain of the agent, the *object-level decision rules*. In the other *priority levels*, the rules relate to the ethical policy under which the agent generates different possible scenarios that the agent can choose. In the *higher level priority* there are the rules representing the optimal course of action, the more ethical (or less unethical) scenario (A. C. Kakas & Moraitis, 2003).

### Definition 4: Levels of Ethical Rules:

Given a policy  $P = \langle R, \geq \rangle$  and a plan based on the ethical rules  $R$ ,  $V$  is a set of abstract ethical events (including the events  $e$  and the violations  $e^*$  of the ethical principles  $c$ ) defined as:

$$V = \langle e \mid e \in (\Phi, c), e \in E, \text{support}(e, t) \Rightarrow \perp_a \rangle c$$

We define the operation *Higher* for the higher level of ethical scenarios  $L$  based on the set of events  $V$ , as follows:

$$L = \text{Higher}(V) = \{ e \mid e \in V, \text{and } \forall e_i \in V : e \geq e_i \}$$

Consider a set of available, possibly ethical, scenarios  $L_i$  for the different set of  $V_i$ . The scenarios lead to different levels of ethical rules  $L_i$

$\in L$  that satisfies the following properties, to define which available scenario is more ethical (or less unethical). For every  $i, j \in N$ , it holds that  $L_i > L_j$  if at least one of the following holds:

1.  $V_i = \emptyset$  and  $V_j \neq \emptyset$ .
2. for every  $\hat{I} \text{ Higher}(V_j \setminus V_i)$  and every  $\hat{I} \text{ Higher}(V_i \setminus V_j)$
3. for every  $\hat{I} \text{ Higher}(V_j \setminus V_i)$ , and every  $\hat{I} \text{ Higher}(V_i \setminus V_j)$ , while  $|\text{Higher}(V_j \setminus V_i)| < |\text{Higher}(V_i \setminus V_j)|$ .

If none of them holds, then  $L_i$  and  $L_j$  are equally (un)ethical, *i.e.*,  $L_i \sim L_j$ .

The first relation makes sure that the ethical scenarios will always be favored by the unethical ones. The second one guarantees that when the principles that are the same in both scenarios are ignored, then the scenario that defends the most valuable principle is considered "higher" ethical. The third states that when the principles that in each scenario are violated are different, but equally valuable, the plan which violates less in number principles is "higher" ethical.

We can now define a logical property which specifies what it means that the reasoning and the decision-making of an agent are ethical. Informally, we have that whenever an agent selects a scenario,  $L_i$ , then all other applicable scenarios  $L_j$  should be ethically "lower", *i.e.*, that  $L_j < L_i$ .

## IMPLEMENTATION IN DATA PRIVACY OF WEARABLE ROBOTS

*Moral Argumentative Proof-Events Calculus* is a framework to help stakeholders to various AI project build an ethics roadmap in a methodical way. This framework can present ethics foresight early in the deployment procedure, rather than implement it as an auditing or assessment tool. There are *three main stages* in this procedure which includes the interaction of three aspects (agents, ethical principles, and contexts):

1. identify the normative frame and the agents;
2. discover the ethical events and rules; and

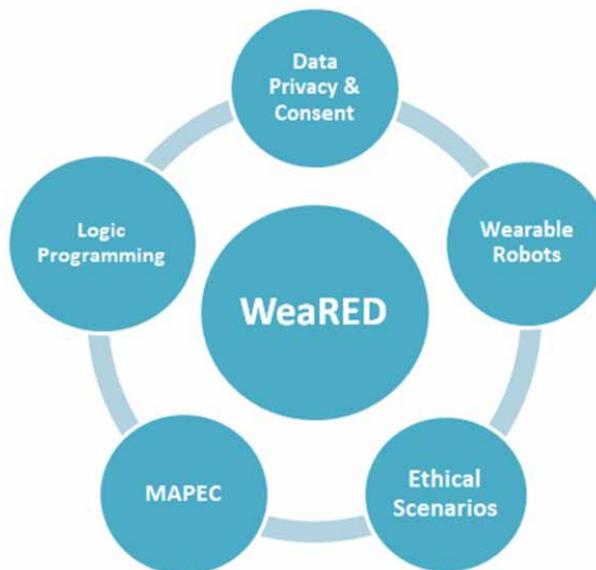
3. prioritize the ethical rules to define the order of scenarios.

To better illustrate the procedure, we walk through a fictional use case of a WR and its privacy dilemmas to demonstrate how it can be applied.

The growth of WRs market (it is expected to record a CAGR of 22.17% over the period 2020 – 2025 (Mordor Intelligence, 2020)) makes it essential to regulate unique privacy challenges that should be addressed, concerning data gathering (Spann, 2016; Wachsmuth, 2018), transfer protocols, standards for consent and exceptions (Consortium, 1996), etc. This implemented use case considers a method for developing verifiable ethical mechanisms for WRs' data privacy (L. A. Dennis et al., 2015).

This system, which will be named **Wearable Robots' Ethics of Data (WeaRED)**, presents a (minor) list of related ethical challenges to outline possible implementations of the above-described formal theoretical framework (See Figure 2). The ethical policy is given by comparing the challenges in terms of how unethical it is to violate them (L. Dennis et al., 2016). The ethical scenarios are context-dependent refinements of the ethical policy.

Figure 2. Elements integrated in implementation of the use case WeaRED



In the *initial stage*, the primary goal is to identify the scope of the ethics analysis and set the scene by **identifying the primary normative frame and the key agents involved**. For example, this use case outlines how an outcome of a data-driven algorithm from a WR is intended to be used, which group of agents may interact with the robot's user and what can be the ethical rules deriving from their potential access in user's data– in our case they can be doctors ( $\mathbf{R}_{\text{doc}}$ ), family ( $\mathbf{R}_{\text{fam}}$ ), coworkers ( $\mathbf{R}_{\text{cow}}$ ), or strangers ( $\mathbf{R}_{\text{str}}$ ).

A list of top ethical principles that are in importance to data access should be included, such as informed consent ( $\mathbf{c}_1$ ), privacy ( $\mathbf{c}_2$ ), and safety ( $\mathbf{c}_3$ ). These ethical values are “communicated” with the **ethical events**:

- $\mathbf{e}_1$  = Share personal data with consent,

- $e_2$  = Don't share personal data,
- $e_3$  = Share personal data without consent,

with  $e_1=e_2$  and  $e_i > e_3$ , for  $i=1,2$ .

*In the second stage*, the framework starts to delve deeper into the analysis by conducting an **exploration of agents' ethical events and the ethical rules in the different contexts**. This step leads to identify what kind of risks and violations are applicable to the primacy stage. WRs are unique in that they are attached to the user, employing many sensors that collect data from brain waves, muscle movement, heart rate, temperature, and so on (Felzmann et al., 2018). This data is collected and processed on board. The possibly beneficial or problematic operations that are related to the data generated and conducted by the WR, might improve care delivery and user's WRs' experience, but might also be confronted with exceptional dangerous situations (Felzmann et al., 2018).

For instance, during regular conditions, such systems are expected to fulfill their decisions within a prearranged ethical framework of rules and protocols. The general principle of data procedures is (General Data Protection Regulations (GDPR), 2016):

*"The explicit and informed, written or recorded consent of the data subject is mandatory for the disclosure, process or transmission of personal data".*

However, in exceptional scenarios, they may choose to disregard their basic goals or break rules to perform with ethical behavior, e.g., to save user's life. Based on technical guidelines for medical data security (Consortium, 1996; Floridi & Sanders, 2004), there is an exception on this general condition stated that:

*"In medical emergencies, where the data subject cannot give consent as in the case of an incapacitated person, on fully regaining his faculties the data subject must be able to withdraw any consent given on his behalf".*

Nevertheless, we need to ensure that this may happen only for justifiably ethical reasons based on the conditions seriousness, which in different time points can be regular ( $t_1$ ), of middle risk ( $t_2$ ), or dangerous ( $t_3$ ). When the WR determines that its user is in danger it requests new *scenarios* from the ethical policy since the current one (i.e., not sharing any data without consent) is no longer *valid*. The ethical guide can produce scenarios based on consisting emergency contingency protocols. In each case we have a different ethical policy. In this case the WR should evaluate the possible feasible scenarios and decide its actions (e.g., whether the WR should withdraw any consent given on his behalf and to whom from the people nearby), leading to the third stage.

In the *final stage*, the **ethical scenarios from the previous stage are prioritized** based on the various contexts. In our case study we have a user that wears a supportive WR in his daily activities, which includes visits in the hospital to check his condition and the condition of the WR ( $\alpha_1$ ), going to work ( $\alpha_3$ ), staying at home ( $\alpha_2$ ) or going outside ( $\alpha_4$ ). In emergency conditions, if a doctor (or in exceptional scenarios anyone nearby) cannot access the data, this can delay important medical decisions and potentially harm the health of the user. We suppose that the system has an emergency system that activates when the personal health data of the user are indicating that the person is in danger (i.e., that the user is unable to provide consent to share the data necessary for others to provide him help), and it evaluates the context-based scenarios created.

For example, in regular conditions the WR should not share personal medical data with a stranger, but if the user is in a situation that his/her life is threatened, it is ethical permissible and preferable to share the necessary information to whoever is near than decide to protect the data instead of user's

life. We propose the general order,  $R_{doc} > R_{fam} > R_{cow} > R_{str}$  with  $R_i > R_j$  meaning that it is less unethical to violate the ethical values referring to  $R_i$  than  $R_j$ , and thus preferable if there is no other ethical choice.

The different scenarios are demonstrated in **Table 1** to show how the different parameters (i.e., health conditions and potential agents to whom data could be shared to) are related to each other in various context-based scenarios.

To create a **computational prototype** of this use case, ethical reasoning was integrated into the logic-based *PSOA RuleML* programming language (Boley, 2015) to illustrate how this ethical thinking can be formalized. *PSOA RuleML* has also been used for the formalization of legal procedures in similar cases, such as Exoskeletons (i.e., a type of WR) (Almpani et al., 2020) and Medical Devices (Almpani et al., 2019, 2018b, 2018a) providing evidence that *PSOA RuleML* is well-suited to express robotic-related decision-making procedures.

**Table 1. Context-based Scenarios in “WeaRED”**

Con Ag	Regular $\in$	Middle risk ( $t_1$ )	Dangerous ( $t_2$ )
$R_{doc}$	$(t_3), i=1,2,3,4$	$support(e_i, t_1) \Rightarrow c_2, i=1...4$	$support(e_i, t_2) \Rightarrow c_2, i=1...4$
$R_{fam}$	$Higher(V_a) > Higher(V_i) > Higher(V_c) > Higher(V_r)$ $support(e_i, t_1) \Rightarrow c_2, i=1,3,4$	$support(e_i, t_1) \Rightarrow c_1, i=1...4$	$attack(e_{1,2}, t_3) \Rightarrow \sim c_2, i=1...4$ $support(e_3, t_3) \Rightarrow c_3$ with $Higher(V_i) = \{e   e \in V : e_3 \geq e_{1,2}\}$
$R_{cow}$	$support(e_i, t_2) \Rightarrow c_2, i=1...4$	$support(e_i, t_1) \Rightarrow c_1,$ $support(e_{1,2}, t_2) \Rightarrow c_2,$ $i=1,2,4$	
$R_{str}$	$support(e_i, t_2) \Rightarrow c_1, i=1...4$	$support(e_i, t_1) \Rightarrow c_1, i=1...4$	

*PSOA* programs may perform deductive reasoning on their atomic beliefs as described in their *PSOA-style reasoning rules* (Boley, 2015) which can indicate that the agent deduces that everything is as in their regular conditions if it is not the case of “emergency” situation (inferred from users’ health data). Otherwise, in “dangerous” conditions an agent needs to identify that deduction should be applied to deduce *supplementary scenarios and decisions* rather than the “regular” one. Based on the above, an agent (i.e., the WR) should: assesses the level of ethical rules to get scenarios annotated with ethical principles; identifies the available scenarios when the “regular” scenario cannot be executed; selects the most ethical scenario from the available set.

The code fragment below encodes the scenarios where the WR might need to share data (or not) with a *coworker*. In this fragment, scenario **VC\_N3** refers to emergency’s cases where the WR doesn’t have the option to share the data in a “higher” scenario (e.g., with a doctor), while scenarios **VC\_N1**, **VC\_N2** is more “regular” scenarios where the WR is either not permitted to share personal data with a coworker or is only permitted to share data with user’s consent. Generally, the scenarios referring to the coworkers are preferred only if the option of a doctor or a relative is not available. In this ethical approach,

the parameter of user's preference can and should be taken into consideration when programming the scenarios' priority. The computational scenarios of the other agents can be similarly formalized.

%WeaRED implementation in PSOA RuleML / Data Privacy of Wearable Robots

```
%Ethical Policy - Rules Coworker (Rcow)
:Support (:DataToCoworker:Consent) :-
    And (:Condition (:MiddleRisk)
        :Context (:Work))
:Support (:DataToCoworker:NoAccess) :-
    Or (And (:Condition (:Regular)
        (:Context (?c)))
        And (:Condition (:MiddleRisk)
            ~:Context (:Work)))
:Attack (:DataToCoworker:WithoutConsent) :-
    And (:Condition (:Dangerous)
        :Context (:Work)
        ~:Share (:DataToDoctor:withoutConsent)
        ~:Share (:DataToFamily:withoutConsent))
```

```
%Scenarios Vc
:Scenario (:VC:N1) :- :Support (:DataToCoworker:Consent)
:Scenario (:VC:N2) :- :Support (:DataToCoworker:NoAccess)
:Scenario (:VC:N3) :- :Attack (:DataToCoworker:WithoutConsent)
% Scenarios priority order
:ScenarioViolate (:VF ?n) :- ~:ScenarioViolate (:VD ?n)
:ScenarioViolate (:VC ?n) :-
Or (~:ScenarioViolate (:VF ?n)
    ~:ScenarioViolate (:VD ?n))
:ScenarioViolate (:VS ?n) :- ~:ScenarioViolate (?v ?n)
This creates an ethical knowledge base as follows:
```

- A database is introduced consisting of a set of ethical rules, creating ethical scenarios in a variety of levels.
- A priority between the scenarios is defined.
- If no (more) ethical scenarios were available for a purpose, the different levels of ethical rules are generated from a context-based ethical policy which annotate the scenarios with ethical rules that can be risked being violated.
- In selecting plans, we prioritized those that are most ethical (according to the order <), leading to the final decision-making.

The work attempts to establish that an ethical policy can be combined within a robot agent in such a way that the dedication to the policy can be *formally* verified and so it can be checked that the agent will always choose the most ethical decisions.

## CONCLUSION

To summarize, this work attempts to develop a proof-theoretically representation of norm scenarios and integrate ethical concepts into a system by developing a logic-based argumentative proving calculus (MAPEC). An example of the application of such a representation is illustrated with the data-privacy-related scenarios of wearable robots. The next step, in future research, is to build algorithms

that can computationally capture ethical cognition and actions with formal decision-making that not only take ethics into consideration when reasoning, but which also can be proved to do so.

### **COMPETING INTERESTS**

All authors of this article declare there are no competing interest.

### **FUNDING**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Funding for this research was covered by the author(s) of the article.

## REFERENCES

- Ågotnes, T., & Wooldridge, M. (2010). Optimal social laws. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, I*, (pp. 667–674). Research Gate. doi: <ALIGNMENT.qj></ALIGNMENT>10.1145/1838206.1838294
- Almpani, S., & Stefaneas, P. (2017). On proving and argumentation. *AIC 2017, 5th International Workshop on Artificial Intelligence and Cognition, Larnaka*. Research Gate.
- Almpani, S., Stefaneas, P., Boley, H., Mitsikas, T., & Frangos, P. (2018a). Computational Regulation of Medical Devices in PSOA RuleML. In C. Benzmüller, F. Ricca, X. Parent, & D. Roman (Eds.), *Rules and Reasoning* (pp. 203–210). Springer International Publishing. doi:10.1007/978-3-319-99906-7\_13
- Almpani, S., Stefaneas, P., Boley, H., Mitsikas, T., & Frangos, P. (2018b). Object-Relational Rules for Medical Devices: Classification and Conformity. In H. Panetto, C. Debruyne, H. A. Proper, C. A. Ardagna, D. Roman, & R. Meersman (Eds.), *On the Move to Meaningful Internet Systems. OTM 2018 Conferences* (pp. 584–591). Springer International Publishing. doi:10.1007/978-3-030-02671-4\_37
- Almpani, S., Stefaneas, P. S., Boley, H., Mitsikas, T., & Frangos, P. (2019). *A Rule-Based Model for Compliance of Medical Devices Applied to the European Market. International Journal of Extreme Automation and Connectivity in Healthcare (IJEACH)*. IGI Global. doi:10.4018/IJEACH.2019070104
- Almpani Sofia, Mitsikas Theodoros, Stefaneas Petros, and Frangos Panayiotis. (2020). ExosCE: A Legal-based Computational System for Compliance with Exoskeletons CE Marking. *Paladyn, Journal of Behavioral Robotics, De Gruyter, 2020*.
- Almpani, S., & Stefaneas, P., & Vandoulakis, I. (2017). *On the Role of argumentation in discovery proof-events. C3GI 2017, 6th International Workshop on Computational Creativity, Concept Invention, and General Intelligence, Madrid, Spain*.
- Almpani, S., Stefaneas, P., & Frangos, P. (2022). Argumentation-Based Logic for Ethical Decision Making. *Studia Humana, 11*(3-4), 46–52. doi:10.2478/sh-2022-0015
- Arkin, R. (2009). Governing Lethal Behavior in Autonomous Robots. In *Governing Lethal Behavior in Autonomous Robots*. doi:10.1201/9781420085952
- Blass, J., & Forbus, K. (2015). Moral Decision-Making by Analogy: Generalizations vs. Exemplars. *AAAI Conference on Artificial Intelligence; Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI. doi:10.1609/aaai.v29i1.9226
- Boley, H. (2015). PSOA RuleML: Integrated Object-Relational Data and Rules. In W. Faber & A. Paschke (Eds.), *Reasoning Web. Web Logic Rules (RuleML 2015)—11th Int'l Summer School 2015, Berlin, Germany, July 31- August 4, 2015, Tutorial Lectures* (Vol. 9203). Springer.
- Bringsjord, S., Arkoudas, K., & Bello, P. (2006). Toward a General Logicist Methodology for Engineering Ethically Correct Robots. *IEEE Intelligent Systems, 21*(4), 38–44. doi:10.1109/MIS.2006.82
- Consortium, S. (Ed.). (1996). *Data Security for Health Care: Technical Guidelines*. Ios Pr Inc.
- Cunneen, M., Mullins, M., Murphy, F., & Gaines, S. (2019). Artificial Driving Intelligence and Moral Agency: Examining the Decision Ontology of Unavoidable Road Traffic Accidents through the Prism of the Trolley Dilemma. *Applied Artificial Intelligence, 33*(3), 267–293. doi:10.1080/08839514.2018.1560124
- Danaher, J. (2019). Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. *Science and Engineering Ethics*. doi:10.1007/s11948-019-00119-x PMID:31222612
- DennisL. A.FisherM.WinfieldA. F. T. (2015). Towards Verifiably Ethical Robot Behaviour. *ArXiv:1504.03592 [Cs]*. <https://arxiv.org/abs/1504.03592>
- Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2016). Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems, 77*, 1–14. doi:10.1016/j.robot.2015.11.012
- Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology, 20*(1), 1–3. Advance online publication. doi:10.1007/s10676-018-9450-z

- Felzmann, H., Kapeller, A., Hughes, A.-M., & Fosch Villaronga, E. (2018). *Ethical, legal and social issues in wearable robotics: Perspectives from the work of the COST Action on Wearable Robots*.
- Fisher, M., Dennis, L., & Webster, M. (2013). Verifying autonomous systems. *Communications of the ACM*, 56(9), 84–93. doi:10.1145/2494558
- Flanagan, M., Howe, D., & Nissenbaum, H. (2008). Embodying values in technology: Theory and practice. *Information Technology and Moral Philosophy*, 322–353. 10.1017/CBO9780511498725.017
- Floridi, L., & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines*, 14(3), 349–379. doi:10.1023/B:MIND.0000035461.63578.9d
- General Data Protection Regulations (GDPR). (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*. GDPR.
- Gomila, A., & Müller, V. (2012). Challenges for Artificial Cognitive Systems. *Journal of Cognitive Science*, 13(4), 453–469. doi:10.17791/jcs.2012.13.4.453
- Hofmann, B. (2012). Ethical Challenges with Welfare Technology: A Review of the Literature. *Science and Engineering Ethics*, 19(2), 389–406. doi:10.1007/s11948-011-9348-1 PMID:22218998
- Iba, W., & Langley, P. (2011). *Exploring Moral Reasoning in a Cognitive Architecture*. Thirty-Third Annual Meeting of the Cognitive Science Society, Expanding the Space of Cognitive Science Proceedings of the 33rd Annual Meeting of the Cognitive Science Society, Boston, USA.
- Kakas, A. C., & Moraitis, P. (2003). Argumentation based decision making for autonomous agents. *Proc. Second International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, Melbourne, Australia*, (pp. 883–890). ACM. doi:10.1145/860575.860717
- Kakas, A., & Michael, L. (2016). Cognitive Systems: Argument and Cognition. *The IEEE Intelligent Informatics Bulletin*, 17, 14–20.
- Lutz, C., & Tamò Larrieux, A. (2015). RoboCode-Ethicists – Privacy-friendly robots, an ethical responsibility of engineers? *ACM Web Science, Oxford*, 7, 2015. doi:10.1145/2786451.2786465
- Malle, B., & Scheutz, M. (2019). Learning How to Behave: Moral Competence for Social Robots Handbuch Maschinenethik, pp. 1–24. doi:10.1007/978-3-658-17484-2\_17-1
- Mordor Intelligence. (2020). *Wearable Robots and Exoskeletons Market—Growth, Trends, and Forecast (2020—2025)*. Mordor Intelligence. <https://www.mordorintelligence.com/industry-reports/wearable-robots-and-exoskeletons-market>
- Pereira, L., & Saptawijaya, A. (2007). Modelling Morality with Prospective Logic. *International Journal of Reasoning-Based Intelligent Systems*, 1(3/4), 99–111. doi:10.1504/IJRIS.2009.028020
- Robertson, L. J., Abbas, R., Alici, G., Munoz, A., & Michael, K. (2019). Engineering-Based Design Methodology for Embedding Ethics in Autonomous Robots. *Proceedings of the IEEE*, 107(3), 582–599. doi:10.1109/JPROC.2018.2889678
- Sisk, B. A., Mozersky, J., Antes, A. L., & DuBois, J. M. (2020). The “Ought-Is” Problem: An Implementation Science Framework for Translating Ethical Norms Into Practice. *The American Journal of Bioethics*, 20(4), 62–70. doi:10.1080/15265161.2020.1730483 PMID:32208070
- Spann, S. (2016). Wearable Fitness Devices: Personal Health Data Privacy in Washington State. *Seattle University Law Review*, 39, 1411.
- Tzafestas, S. (2018). Ethics in Robotics and Automation: A General View. *International Robotics and Automation Journal (IRAT-J)*, 4. 10.15406/iratj.2018.04.00127
- Vakkuri, V., Kemell, K.-K., & Abrahamsson, P. (2019). Implementing Ethics in AI: An industrial multiple case study. *ArXiv, abs/1906.12307*.
- Wachsmuth, I. (2018). Robots Like Me: Challenges and Ethical Issues in Aged Care. *Frontiers in Psychology*, 9, 432. doi:10.3389/fpsyg.2018.00432 PMID:29666601

Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems [Scanning the Issue]. *Proceedings of the IEEE*, 107(3), 509–517. doi:10.1109/JPROC.2019.2900622

Wynsberghe, A. (2012). Designing Robots for Care: Care Centered Value-Sensitive Design. *Science and Engineering Ethics*, 19(2), 407–433. doi:10.1007/s11948-011-9343-6 PMID:22212357

*Sofia Almpani is a postdoctoral researcher at the School of Applied Mathematics and Physical Sciences of National Technical University of Athens (NTUA). Sofia received her PhD titled “Argumentation and rule-based logic in mathematical proving and legal artificial intelligence applications” from the School of Electrical and Computer Engineering (NTUA), Bachelor Degrees from the Faculty of Education School of Primary Education (AUTH) and from the Department of Mathematics (UOA), and the Master Degree “Educational Leadership, Management and Emerging Technologies” (Marconi). Her research interests focus on argumentation, logic programming, AI, Robotics, and STEAM. Orcid: <https://orcid.org/0000-0003-4823-479X>*

*Petros Stefaneas co-ordinates the Logic and Formal Methods Group ( $\lambda$ -ForM) of AALL. He is a Assistant Professor (tenure track) at the School of Applied Mathematical and Physical Sciences at NTUA. Dr. Petros Stefaneas has done extensive work on the application of formal verification and specification techniques to engineering problems. His current research includes formal methodologies for information privacy, legal documents, and open data policies. Other research interests include abstract model theory, algebraic specifications, computational creativity, and semantics and ethics of computer science. Orcid: 0000-0002-2096-9914*

*Panayiotis V. Frangos was born in Thessaloniki, Greece, in 1959. He received the Bachelor Degree from the NTUA, Greece, in 1983, and the Master’s and Ph. D. Degrees from the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, USA, in 1985 and 1986 respectively, all in Electrical Engineering. Since 1989, after his military service in Greece, he has been with the Faculty of the Department of Electrical and Computer Engineering, NTUA, first as a Lecturer (1989 - 92), and subsequently as an Assistant Professor (1992 - 96), Associate Professor (1996 - 2000), and Full Professor (2000 - today). Orcid: 0000s-0002-8607-5737*