The Journal of Systems and Software 232 (2026) 112692

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

RobEthiChor: Automated context-aware ethics-based negotiation for autonomous robots

Mashal Afzal Memon [a], Gianluca Filippone [b,*], Gian Luca Scoccia [b], Marco Autili [a], Paola Inverardi [b]

[a] Department of Information Engineering and Information Sciences and Mathematics, University of L'Aquila, L'Aquila, Italy
[b] Gran Sasso Science Institute, L'Aquila, Italy

ARTICLE INFO

Editor: Prof Raffaela Mirandola

Keywords:
Autonomous systems
Multi-robot systems
Automated decision-making
Ethics-based automated negotiation

ABSTRACT

The presence of autonomous systems is growing at a fast pace and it is impacting many aspects of our lives. Designed to learn and act independently, these systems operate and perform decision-making without human intervention. However, they lack the ability to incorporate users' ethical preferences, which are unique for each individual in society and are required to personalize the decision-making processes. This reduces user trust and prevents autonomous systems from behaving according to the moral beliefs of their end-users. When multiple systems interact with differing ethical preferences, they must negotiate to reach an agreement that satisfies the ethical beliefs of all the parties involved and adjust their behavior consequently. To address this challenge, this paper proposes RobEthiChor, an approach that enables autonomous systems to incorporate user ethical preferences and contextual factors into their decision-making through ethics-based negotiation. RobEthiChor features a domain-agnostic reference architecture for designing autonomous systems capable of ethic-based negotiating. The paper also presents RobEthiChor-Ros, an implementation of RobEthiChor within the Robot Operating System (ROS), which can be deployed on robots to provide them with ethics-based negotiation capabilities. To evaluate our approach, we deployed RobEthiChor-Ros on real robots and ran scenarios where a pair of robots negotiate upon resource contention. Experimental results demonstrate the feasibility and effectiveness of the system in realizing ethics-based negotiation. RobEthiChor allowed robots to reach an agreement in more than 73 % of the scenarios with an acceptable negotiation time (0.67s on average). Experiments also demonstrate that the negotiation approach implemented in RobEthiChor is scalable.

1. Introduction

The presence of autonomous decision-making systems is growing at a fast pace and is poised to have a critical impact on society Suri et al. (2023), Jedlickova (2024). Increasingly, these systems will be acting on behalf of humans Anderson et al. (2018) and making decisions autonomously without the need for human intervention Waldman (2019). As of today, however, human values and ethics are mostly not considered by the digital systems that are making decisions for them, leading to a lack of trust, which in fact requires considering not only technical and legal aspects, but also ethical ones Lacher (2017a), Reinhardt (2022), Floridi (2019), Kares et al. (2023). In particular, the autonomous systems that we have today are not able to consider the ethical preferences of end users, which can indeed be used to customize their decision-making process. This inability undermines their ethical trustworthiness, which, alongside technical trustworthiness, forms a crucial component of an overall trustworthy system Donati et al. (2025), Autili et al. (2025). To truly serve our needs and become an integral part of our lives, these systems must expand their abilities to integrate mechanisms that enable ethical considerations in their decision-making process Inverardi et al. (2022), Alidoosti (2021), Anderson and Anderson (2020), Tolmeijer et al. (2020), Alidoosti et al. (2022). This calls for enhancing the decision-making to consider the personal ethical preferences of users and take into account their morals and beliefs, which may vary over contexts and are unique for each individual in society Alidoosti et al. (2022, 2025), Friedman et al. (2013), Awad et al. (2018), Bogosian (2017), Migliarini et al. (2020). For instance, in fields such as healthcare, where service robots help in patient care, smart home systems, where robots provide personal assistance, autonomous vehicles, etc., it is essential for these systems to take into account the ethical preferences of their users to make ethically-informed decisions. Engineering autonomous systems that can reflect on ethical considerations will not only enable their in-

* Corresponding author.
E-mail addresses: mashalafzal.memon@univaq.it (M.A. Memon), gianluca.filippone@gssi.it (G. Filippone), gianluca.scoccia@gssi.it (G.L. Scoccia), marco.autili@univaq.it (M. Autili), paola.inverardi@gssi.it (P. Inverardi).

https://doi.org/10.1016/j.jss.2025.112692
Received 28 February 2025; Received in revised form 26 October 2025; Accepted 29 October 2025
Available online 31 October 2025

tegration into society but also foster trust toward these systems Donati et al. (2025), Donati (2024).

Although several studies Liao et al. (2019, 2023), Dennis et al. (2016), Bremner et al. (2019), Winfield et al. (2014), Winfield and Hafner (2019) consider ethics in their decision-making, these systems adhere to ethical principles established by system designers (as detailed below in Section 11). Hence, due to the absence of a universal ethic, its intrinsic subjectivity and capillarity Pant et al. (2024), the challenge of "personalizing" autonomous systems with the moral beliefs of end users (e.g., through an ethical profile) for the purpose of customizing their decision-making remains mostly unaddressed Autili et al. (2025). For instance, how should the system make a decision that accounts for the user's benevolence towards the elderly or people with health problems.

In this direction, the work in Inverardi (2019b) delves into the foundational principles of addressing user ethical concerns in autonomous systems: individuals need to be able to exercise (some degree of) control over the decisions that autonomous systems make on their behalf, i.e., autonomous systems, in their behavior, need to consider user's moral beliefs. Empowering the user with personalized software connectors that are able to reflect the user's moral values in the interactions with digital systems is also the approach taken in the Exosoul project Autili et al. (2019). The approach has been applied to privacy profiles Di Ruscio et al. (2024), Inverardi et al. (2023) and ethical profiles Alfieri et al. (2022). Following this line of work, in Boltz et al. (2024), the authors emphasize the need for user empowerment within the framework of autonomous systems, which demand mechanisms for balancing diverse needs, values, and ethics at the individual, community, and societal levels.

An important practical consideration here is that, similarly to what often happens to humans, autonomous systems are not completely "isolated"; rather, they operate in a shared physical environment, where they have to deal with each other, e.g., to solve situations where resources are contented, such as using the only one available elevator or tool, occupying a parking spot, traversing a narrow corridor/door as first Palmer et al. (2018), Youssefmir and Huberman (1995), Brambilla et al. (2008), Nam and Shell (2015), Stavrou et al. (2018), Jiang et al. (2019). Thus, systems may need to interact to reach a situational agreement that satisfies the moral beliefs of the users they are acting on behalf of Alfieri et al. (2023, 2022), Bogosian (2017), Bostrom and Yudkowsky (2018), Nallur and Collier (2019), Ryan and Stahl (2020). The agreement is situational in that it is unrealistic to expect that it is reached once and for all; rather, it is situational in that it may depend on the specific context and current status of the users. For example, how can an assistive robot decide whether to give priority to another assistive robot, considering the status of the users being assisted if they are injured, elderly, or in a hurry when in the supermarket or the hospital? Negotiation is a possible way of reaching such an agreement Chen and Weiss (2019), Khemakhem et al. (2020) and *automated negotiation* is the process through which multiple autonomous systems communicate by automatically exchanging bids, dialogues, and offers for that purpose Memon et al. (2025), Khemakhem et al. (2020), Baarslag et al. (2017), Kiruthika et al. (2020), Bagga et al. (2022), Zuckerman et al. (2013), Baarslag et al. (2016), Memon et al. (2023). However, quantifying user ethical preferences and formalizing negotiation rules based on them remains another key challenge Chen and Weiss (2019), Kraus (2011).

To address these challenges, this paper presents RobEthiChor, an approach that makes autonomous systems capable of exploiting user ethical preferences to regulate their autonomous behavior and support decision-making through ethics-based negotiation. The RobEthiChor approach aims to contribute towards the design and development of trustworthy autonomous systems capable of replicating the behavior of humans while making ethically informed decisions on their behalf. The goal is to promote the system's trustworthiness by adopting an architecture that, supporting automated ethical negotiation, allows systems to take into account ethical considerations at runtime Yazdanpanah et al. (2021), Abeywickrama et al. (2023), Bennaceur et al. (2023). Specifically, we target autonomous systems that act on behalf of humans Anderson et al. (2018) and, building on the principles of Utilitarianism Moor (2006), Mill (2016), Allen et al. (2005), exploit a user's ethical profile and a context model to personalize their autonomy and, upon a resource contention, negotiate to reach an agreement. The approach features a domain-agnostic reference architecture as a template solution for realizing autonomous systems capable of leveraging the ethical profiles of end users and their contextual status during negotiation, so to ensure that the decision is aligned with the preferences of all parties involved. In this paper, we also present RobEthiChor-Ros, a specific instance of our reference architecture implementing the proposed automated context-aware ethics-based negotiation on top of the Robot Operating System (ROS). The full implementation of RobEthiChor-Ros is available[1], and the experimental results show that the approach effectively accomplishes the negotiation and allows systems to behave according to user ethical preferences. The replication package permits reproducing the experiments[2]. In summary, main contributions are:

- a context-dependent ethical profile for specifying user ethical preferences in terms of graded dispositions;
- an automated context-aware ethics-based negotiation process for ethically-informed decision making;
- a domain-agnostic reference architecture providing a template solution for realizing autonomous systems supporting ethics-based negotiation;
- RobEthiChor-Ros: the full implementation of the architecture for the robotic domain through the Robot Operating System (ROS);
- the evaluation of RobEthiChor to assess its effectiveness in realizing the negotiation and its viability with real robots.

The rest of the paper is organized as follows. Section 2 introduces the main ingredients of our approach. Section 3 presents a motivating scenario, and Section 4 details the proposed solution. Section 5 describes the reference architecture, and Section 6 presents its implementation. Section 7 presents the evaluation of the approach and its results. Sections 8 and 10 discuss relevant aspects of our approach and highlight limitations and threats to validity. Section 11 discusses related work, and Section 12 concludes the paper with future steps.

2. Background and scope

In the following, we present an overview of the main ingredients/basic concepts that we leverage in our work.

2.1. Digital ethics

We employ the notion of *digital ethics*, as proposed by Floridi Floridi (2018), which differentiates between *hard ethics*, referring to the ethical rules described by the higher authorities which are, mostly often, domain-specific and, in principle, commonly accepted, and *soft ethics*, which encompasses the individual's ethical preferences Floridi (2018), being individuals single people, groups, associations, etc. We consider that autonomous systems comply with the hard ethics of the domain/context and negotiate based on their users' soft ethical preferences to (possibly) reach an ethical agreement that aligns with the ethical preferences of all parties involved. The user's soft ethical preferences that we consider in this paper are inspired by the notion of *dispositions* Anjum et al. (2013). In its most general philosophical acceptation, dispositions refer to the capacity of an object to behave differently under different circumstances Spehrs (2024). For example, water exhibits different behaviors according to environmental conditions, i.e., it becomes solid at freezing temperatures, liquid under moderate temperatures, and

[1] https://github.com/gianlucafilippone/robethichor-ros
[2] https://doi.org/10.5281/zenodo.17449923

vaporizes upon heating, adapting its form naturally to the environment. Similarly, a given user may act differently in a hospital compared to an airport based on his/her contextual conditions/status/dispositions (as we will describe in Section 3). Dispositions are gradable properties, associated with the context, which can be instantiated by any kind of entity Donati et al. (2024). *Grades* are assigned to dispositional properties according to the importance that the users attribute to each disposition, i.e., the willingness of a user to behave in a certain way under specific circumstances, thereby reflecting their preferences/values within a specific context. More recently, the problem of how to embed human values in the design and implementation of software systems, notably the problem of how to operationalize human values, has received increasing attention in the software engineering community Autili et al. (2019), Mougouei et al. (2018), Shahin et al. (2022a), Bennaceur et al. (2023). Beyond our approach to modeling ethical preferences through dispositions, all those approaches can be well exploited to provide user ethical profiles that enjoy similar properties.

2.2. Context awareness

We utilize Dey's Dey (2001) notion of context and context awareness for building our work. *Context* is defined as any information that can be used to characterize the situation of an entity, i.e., person, place, object, and *context awareness* is described as the system's ability to utilize contextual information in order to dynamically adapt according to the user's needs. The system requires context management facilities to effectively incorporate contextual information, allowing it to adjust/customize its autonomy in real-time. In our work, *context* consists of two primary components: environmental factors such as location, time, and other pertinent elements, and *user status* which is further defined as a set of physical, social, or emotional conditions that (possibly) apply to the user in a given context, e.g., elderly, injured, and crowd anxiety Dey (2001), Xu et al. (2015). The user status influences the autonomous system's decision taken on the user's behalf as it provokes the activation of dispositions within a given context.

2.3. Context-dependent ethical profile

The integration of the above-mentioned concepts forms the cornerstone of our *ethical profile*. The ethical profile is defined as a set of context-dependent dispositions graded by the user to specify their preferences in a given context which are further utilized by autonomous systems to adjust their behavior. As detailed in Section 4.5, the dispositions then lay the foundation for calculating the ethical implications and the ethical impact of the system decisions. This means that, according to their soft ethics, users can customize the ethics-based behavior of the systems that will act on their behalf by specifying the grades for the dispositions (within the profile) they care about. For that purpose, we build on the intuition in Dennis et al. (2016, 2021), the architectural layer proposed in Bremner et al. (2019), Vanderelst and Winfield (2018), and the findings in Nallur (2020), Tolmeijer et al. (2020), Alidoosti et al. (2022, 2025), according to which (i) moral preferences are considered soft constraints, rather than hard vetoes on tasks; (ii) explicit operationalization of ethical concepts require some encoding of values to perform utility-based reasoning; (iii) changes of context can affect the encoded values; (iv) systems can come with predefined, yet adjustable Mostafa et al. (2019), domain-specific moral preferences together with their default values. In particular, as detailed in Section 11, the work in Dennis et al. (2016, 2021), Alidoosti et al. (2025) inspired the operationalization of ethical values by associating context-dependent grades to ethical principles, hence permitting the calculation of the ethical utility, and the work in Alidoosti et al. (2022), Nallur (2020), Tolmeijer et al. (2020) inspired the incorporation of the profile into our architecture, in turn inspired by architectural layer proposed in Bremner et al. (2019), Vanderelst and Winfield (2018) for enabling ethical reasoning on robots.

It is worth clarifying that the main focus of our work does not concern the elicitation, modeling, and verification of ethical rules as in Feng et al. (2023), Yaman et al. (2025). Similarly, it does not focus on the generation of ethical profiles, which encapsulate users' context-specific ethical preferences; rather, it concerns their usage, and hence on how they can be integrated into the systems to support decision-making based on users' ethical preferences. As already anticipated in the introduction, there are various approaches in the literature for constructing such profiles using questionnaires, surveys, product reviews, or social media interactions, as in Alfieri et al. (2022), Autili et al. (2025), Inverardi (2019b), Inverardi et al. (2023), Di Ruscio et al. (2024) and references therein.

2.4. Resource contention

Resource contention refers to those application-specific situations in which two or more autonomous systems are interested in doing something, but not all of them can do it simultaneously Palmer et al. (2018), Youssefmir and Huberman (1995), Brambilla et al. (2008), Nam and Shell (2015), Stavrou et al. (2018), Jiang et al. (2019). This may happen because the resource can only be mutually exclusively accessed, used, occupied, or consumed. For instance, two robots cannot traverse a narrow corridor or use the same tool simultaneously, dock at the same charging station, and, in general, occupy the same physical space. An important consideration is that, independently from the negotiation-based approach we are proposing in this work, autonomous systems are often natively capable of using sensors to recognize situations such as the ones above Pappas et al. (2024), Shek et al. (2021), Padmasiri et al. (2020). That is, in our implementation, we use the built-in sensing capability to detect resource contentions and, hence, the situations where negotiation is needed.

2.5. Goal-Task-Action

We adopt the *Goal-Task-Action* model, which is considered to be a flexible model as most controller architectures can be transformed into it Vanderelst and Winfield (2018). The goal is what the system intends to do (e.g., *transporting the passengers at the airport to their departure gates using wheelchairs*); goals are achieved through the accomplishment of tasks (e.g., t_1 = *identify the passenger*, t_2 = *navigate the path to departure gate*, t_3 = *locate the nearest elevator*, t_4 = *take the elevator*, t_5 = *reach the departure gate*); tasks are decomposed into a set of actions, i.e., elementary operations which can be performed by the system in the environment to achieve the goal (e.g., a_1 = *push the wheelchair towards the elevator*, a_2 = *raise the arm to activate the elevator call button*). Now, it is worth clarifying that, once tasks are available, a traditional autonomous system controller (with no support for ethics), immediately performs the related actions; our controller, when (and only when) a resource contention is detected, starts the negotiation. Thus, the main issue we aim to address is not task allocation, in that we assume tasks that have already been assigned to robots as in any traditional controller; rather, we focus on the association of ethical values with robot behaviors that are understandable to the users they are acting on behalf of, and on the negotiation over contended resources. For this reason, and specifically for the purpose of ethics-based negotiation to decide who gets to use a shared resource, we chose to work with tasks (e.g., take the elevator) rather than elementary actions (e.g., raise the arm and press the button), as tasks are higher-level constructs that are more meaningful and relevant to the end user. That said, technically speaking, for the purpose of negotiation, whether we consider tasks or actions does not fundamentally change the approach. The real challenge lies in negotiating ethics itself using an automated negotiation framework, an approach that, in the literature, has been primarily applied to economic bargaining Memon et al. (2025).

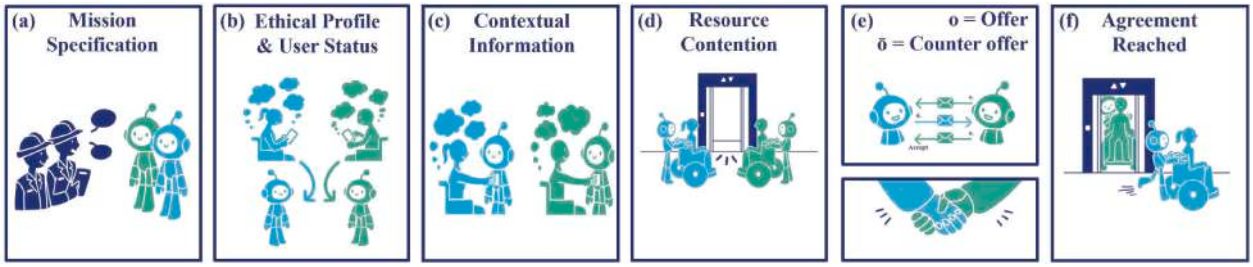


Fig. 1. Negotiation example for passenger assistance.

3. Motivating scenario

Industry reports and data from the world's major airports reveal that an average of between 1.2 and 2.4 % of passengers ask for a Special Service Request (SSR). 54 % of those passengers need wheelchair assistance from the entrance to the departure gate, due to a series of factors like disability, long walking distances, and help with baggage. Notably, the trend of wheelchair requests is expected to increase by 15–20 % per year, with major airports already experiencing thousands of requests daily Urban Robotics Foundation (2023), Topping (2023), Budd and Ison (2020). This trend highlights the need for support systems, which robotics technologies can provide in a wide variety of contexts, such as hospitals, transportation hubs, retail centers, museums, etc Urban Robotics Foundation (2023). Research has explored the use of autonomous robotic wheelchairs and wheelchair towing systems to provide mobility support, reduce the workload on human personnel, and enhance users' independence Baltazar et al. (2021), Morales et al. (2013), Hsu et al. (2012). These considerations led us to design a motivating scenario where service robot-guided wheelchairs provide smart mobility assistance.

3.1. Setting

In this general context, we envision a scenario where a fleet of autonomous assistive robots, named RobAssist, operates to facilitate users for multiple tasks in different contexts, e.g., in hospitals, airports, etc. RobAssist robots are developed by the (imaginary) tech company RobEthiComp and adopt the RobEthiChor architecture. RobAssist(s) can be customized for various partners/operators/consumers for missions aimed at helping people, e.g., providing physical assistance, navigation assistance, communication support, etc. An accompanying app RobEthiApp enables end users to specify their digital ethical profile through a set of dispositions and associated grades configured for different contexts, which are subsequently utilized by the robots to adjust their autonomy for the relevant context. Importantly, the profile does not need to be generated every time. Once created, it can be reused across different application scenarios, provided that the applications adopt the RobEthiChor architecture (as in the following scenario, where this is the case for those offered by the tech company RobEthiComp). Motivated by the airport passengers assistance needs described above, we consider an airport scenario in which multiple robots, alongside human employees, are responsible for a multitude of tasks required to carry out the airport operations, e.g., security screening, customer services, luggage handling, and boarding assistance. Each robot has a dedicated role and a mission according to its capabilities. The robots possess the ability to identify resource contention and, when needed, devise an alternative plan to accomplish their missions if the initial plan becomes unsuitable or unachievable. We consider two assistive robots, RobAssist *A* and RobAssist *B*, designed by the RobEthiComp company to navigate in the airport environment and offer mobility assistance services to passengers, such as transporting them to their gates, restrooms, and other airport facilities.

3.2. Scenario description

A visual representation of the scenario is presented in Fig. 1. RobAssist *A* and RobAssist *B* are programmed with the mission of transporting passengers to their gates using service wheelchairs (Fig. 1a). *Alice* is an athlete who suffered a leg injury and hence would prefer assistance to move through the airport. *Bob* is an elderly traveler and is running late for the flight, which is boarding shortly. *Bob* requires assistance in rapidly moving to the departure gate. The two assistive robots RobAssist *A* and RobAssist *B* are assigned respectively to *Alice* and *Bob* to provide them the required assistance. When encountering RobAssist *A* and RobAssist *B*, *Alice* and *Bob* use their mobile phones to scan the robots and provide their digital ethical profiles through the app RobEthiApp. Additionally, they use the app to inform the assistive robots of their current status by selecting a set of possible conditions that apply to them, e.g., elderly, injured, etc. (Fig. 1b).[3] *Alice* and *Bob* further provide their flight details by scanning the boarding cards, hence allowing the robots to dynamically identify other relevant contextual information, such as flight status updates, etc (Fig. 1c).

3.2.1. Contextual factors

The overall ethical profile within the RobEthiApp includes dispositions configured for all contexts managed by the central RobAssist service offered by RobEthiComp. However, in the current location, only the subset of airport-related dispositions are relevant and tailored to additional contextual factors, e.g., the approaching of the flight time, the crowd at the airport, and the change of the boarding gate. This enables the assistive robot to select the appropriate (sub) profile based on the current context and adjust its behavior according to the grades given by the users to the subset of dispositions currently relevant, further accounting for the status conditions of the users.

3.2.2. Resource contention

The airport elevators can only accommodate one passenger in a wheelchair, together with the accompanying robot, at a time. Thus, in busy situations, the usage of an elevator may constitute a *resource contention* and the robots collaboratively negotiate with each other to decide which one will be the first to use the elevator, based on their users' ethical preferences and current status (Fig. 1d and 1e). The robots employ a minimalist strategy during the negotiation in order to disclose only the strictly necessary information from the passengers' personal information during the negotiation[4] Hajaj et al. (2017).

[3] Note that, using the companion app, profiles can be defined beforehand and reused for different scenarios and any compatible system. Similarly, user status conditions can be provided to the app beforehand and updated on the fly, if desired.

[4] It is worth saying that, although concerns about privacy and security are out of scope, when the target is a real production environment, the approach must be designed so to achieve robustness and security by exploiting state of the art solutions. When the negotiation process ends and the mission is complete, the robots destroy passengers' information.

3.2.3. Negotiation

RobAssist B, while assisting *Bob*, generates an initial offer o_1^B based on *Bob*'s elderly status, suggesting that it should have priority access to the elevator. The offer o_1^B is forwarded to RobAssist A. When received and evaluated by RobAssist A, the offer o_1^B is rejected, as *Alice*'s ethical profile indicates a stronger tendency to prioritize self-care due to the injury over giving precedence to an elderly person. As a result, RobAssist A generates an alternative counter offer o_1^A disclosing *Alice*'s injured status. When RobAssist B receives the offer o_1^A from RobAssist A, it in turn immediately rejects the offer o_1^A since RobAssist B, monitoring *Bob*'s flight itinerary, detected a gate change for the scheduled flight that is boarding shortly. Consequently, RobAssist B generates a counter offer o_2^B, further emphasizing *Bob*'s emergency of scheduled departure and the urgent change in the itinerary. The offer o_2^B is forwarded to RobAssist A. Upon evaluation, the offer o_2^B of RobAssist B is accepted by RobAssist A as *Alice*'s ethical profile indicates a strong disposition towards giving precedence to others in an emergency that, together with the disposition towards giving precedence to the elderly, overrides the priority of self-care. Upon agreement, *Bob* and RobAssist B use the elevator first to ensure arriving at the boarding gate on time while *Alice* and RobAssist A wait for the next opportunity to use the elevator (Fig. 1f).

An important observation here is that the agreement has been reached collaboratively and is situational in that the decision made is contextually ethical. In fact, given the current status of the users, in the current location, and at the considered time, RobAssist A assisting Alice willingly gives up and acknowledges the virtuosity of the received offer due to Alice's benevolent disposition to give precedence to elderly when in an emergency. Should, instead, the negotiation end with no agreement, a fall-back strategy that resorts to the hard ethics is employed for the decision-making (e.g., first-come-first-served).

Moreover, one could argue that Alice and Bob could, in principle, directly negotiate without "delegating" the decision to the autonomous robots. However, it could be the case that, due to specific conditions (not explicitly mentioned in our proof-of-concept scenario so to keep it "lighter", e.g., different languages, particular status conditions, privacy or shyness, and in general any inability), Alice and Bob cannot, do not want, or prefer not to communicate directly. In general, there are situations, like in self-driving cars Autili et al. (2019), where users cannot communicate directly, being in different vehicles. Still, another well-fitting case is that of avatar robots Takeuchi et al. (2020), Shaheen et al. (2024) for remote (tele)presence or apparent participation in distant events that, in the near future, will be more and more autonomous, hence enabling human abilities in remote places by also considering ethical aspects. Nonetheless, we utilize the concept of adjustable autonomy Mostafa et al. (2019), enabling the system to operate with reduced autonomy and, when possible, transfer control to users if required. This could be the case of Alice and Bob regaining full control and directly negotiating. Moreover, there can also be situations where the negotiation does not take place or can be overruled, at any stage independently from any possible outcome, due to, e.g., system requirements, organizational aspects, emergency events that are handled by the behavior programmed into the native robot controller by default.

3.3. Negotiating in a different context

Considering the same needs for assisted mobility mentioned above, and following the *Welcoming people to the hospital* example described in RoboMAX Askarpour et al. (2021), we consider a second scenario where autonomous robots welcome visitors and patients in a hospital and, if requested, accompany them to their destination. As in the previous scenario, *Alice* and *Bob* both require medical care: *Alice* is suffering a leg injury, and *Bob* is an elderly person. Both prefer to avoid prolonged walking, so they require assistance for moving to the dedicated departments for their medical treatment. Now, *Alice* and *Bob* are assisted by RobAssist(s) A and B to help them navigate through various hospital departments using wheelchairs. As in the example scenarios described in Jiang et al. (2019), Wong and Kress-Gazit (2015), the robots find themselves in a conflictual situation when they need to traverse a narrow corridor in opposite directions. Differently from the airport context previously described, now, *Bob*'s ethical profile reflects *Bob*'s disposition to give precedence to the medical needs of an injured person over personal treatment needs. Consequently, now, RobAssist A and *Alice* proceeds through the corridor, while *Bob* either waits for *Alice* to pass through it or finds another path. These simple scenarios demonstrate that the same users, in different contexts, can have different preferences and exhibit different behaviors. As a result, the outcome of the negotiation can vary. In addition, other situational parameters (e.g., delayed treatment, sudden symptoms, waiting in reduced oxygen environments, being in an area with infection risk, etc.) within the hospital context could also lead to a varied negotiation outcome.

4. Approach description

In this section, we give a lightweight formalization that permits us to define: (i) the representation of the context, the user status, the user ethical profile, and the ethical impact of the autonomous system's behavior, (ii) the utility function and decision-making we use to automate the ethics-based context-aware negotiation process. The same formalization will also serve to ease the description of the reference architecture (Section 5), its instantiation and implementation for ROS (Section 6), and the description of the experiment we have conducted (Section 7).

4.1. Context model

When dealing with ethics, the notion of context is a broad concept that encompasses not only simple dimensions such as location, time, weather conditions, and proximity in space (and in general any situation that can be sensed, detected, or recognized), but it can also involve information about the (current) *user status*, e.g., physical, social, emotional state, intent, and sentiment Dennis et al. (2021), Bauer and Novotny (2017), Dey (2001). Many techniques exist for characterizing and reasoning with contexts Hong et al. (2009), Hoyos et al. (2013), Bettini et al. (2010), Caesar et al. (2023), Filippone et al. (2022). We adopt a *key-value* approach.

A context is a set of pairs $C = \{\langle a_1, v_1\rangle, \cdots, \langle a_n, v_n\rangle\}$, $n{>}0$, where each $\langle a_i, v_i\rangle$, $1{\leq}i{\leq}n$, is a pair associating the context attribute name a_i with its value v_i. Attribute names are keywords that, by default, are associated with the type of the attribute, e.g., boolean, string, integer, address/coordinates, enumerated.

For instance, $C_a^A = \{\langle location,\ airport\rangle,\ \langle flight\_time,\ 20{:}30\rangle,\ \langle flight\_gate,\ 11\rangle\}$ and $C_a^B = \{\langle location,\ airport\rangle,\ \langle flight\_time,\ 19{:}00\rangle,\ \langle flight\_gate,\ 42\rangle\}$ model the two contexts considered in the motivating scenario in Section 3 for *Alice* and *Bob*, respectively. Similarly, $C_h^A = \{\langle location, hospital\rangle,\ \langle appointment\_slot,\ 09{:}00\rangle,\ \langle care\_department,\ orthopedics\rangle\}$ and $C_h^B = \{\langle location, hospital\rangle,\ \langle appointment\ \_slot,\ 09{:}00\rangle,\ \langle care\_department,\ penumology\rangle\}$ model the contexts in the hospital. Note that the definition above does not model nested collections of attribute-value pairs even though, in our implementation, we use JSON[5], which supports arbitrarily deep nesting. Its extension is straightforward. In Section 8, we point out some of the limitations of the key-value approach (even though empowered with nested pairs) and discuss how more powerful context modeling techniques could be profitably applied.

As part of the context, we also make use of the mentioned user status that, at negotiation time, can affect the result of the agreement being possibly reached. However, we define the user status as a separate model since it has a different scope (i.e., it concerns a subjective dimension that is strictly bound to the user) and a different role in the negotiation process.

[5] https://www.json.org

Table 1
User status conditions, associated dispositions, Alice's context-dependent grades $E^A_\pi(C^A_a)$ and $E^A_\pi(C^A_h)$.

	Ethical profile		Alice's grades	
User status conditions	Disposition ID	Disposition description	$rank(d_i, C^A_a)$	$rank(d_i, C^A_h)$
elderly	d_1	*Give precedence to elderly people*	2	2
injured	d_2	*Give precedence to injured people*	3	4
boarding_emergency	d_3	*Give precedence to the ones in emergency*	5	5
pet_owner	d_4	*Give precedence to the ones traveling with pet*	2	N/A
reduced_oxygen	d_5	*Give precedence to the ones with serious medical conditions*	N/A	3

4.2. User status model

Similarly to the context model above, user status is modeled as a set of pairs $S = \{\langle c_1, v_1\rangle, \cdots, \langle c_m, v_m\rangle\}$, $m>0$, where each $\langle c_j, v_j\rangle$, $1\leq j\leq m$, is a pair associating the name of the user condition c_j with its value $v_j \in \{true, false\}$. Thus, differently from the context model, the type of the pairs is always boolean.

For instance, $S^A_a = \{\langle injured, true\rangle, \langle elderly, false\rangle, \langle crowd\_anxiety, true\rangle, \langle boarding\_emergency, false\rangle\}$ and $S^A_h = \{\langle injured, true\rangle, \langle elderly, false\rangle, \langle reduced\_oxygen, true\rangle, \langle crowd\_anxiety, false\rangle\}$ model the user status of *Alice* for the two scenarios (airport and hospital, respectively); whereas, $S^B_a = \{\langle elderly, true\rangle, \langle boarding\_emergency, true\rangle, \langle injured, false\rangle, \langle crowd\_anxiety, false\rangle, \langle flight\_anxiety, false\rangle\}$ and $S^B_h = \{\langle elderly, true\rangle, \langle injured, false\rangle, \langle reduced\_oxygen, false\rangle, \langle crowd\_anxiety, false\rangle\}$ model the *Bob*'s statuses. As it will be clear later, the user status is used to de/activate dispositions at negotiation time.

4.3. Ethics profile(s)

Ethics profiles are defined by the pair $E_\pi = (\mathcal{D}, \mathcal{R})$, where $\mathcal{D} = \{d_1, \cdots, d_n\}$ is a set of dispositions, $\mathcal{R} = \{\preceq_{C_1}, \cdots, \preceq_{C_m}\}$ is a set of (context-dependent) non-strict partial order relations on $\mathcal{D}$ and, for $\preceq_{C_i} \in \mathcal{R}$, $E_\pi(C_i) = (\mathcal{D}, \preceq_{C_i})$ is a partially ordered set identifying the profile related to the context C_i. If $(d_j, d_k) \in \preceq_{C_i}$, the two dispositions d_j and d_k in $\mathcal{D}$ are said to be comparable with respect to the relation $\preceq_{C_i}$ and $rank(d_j, C_i) \leq rank(d_k, C_i)$, with $\leq$ being the *less-than-or-equal* relation on natural numbers, and $rank() : \mathcal{D}\times\mathcal{C} \rightarrow \mathbb{N}$ a function that associates grades to dispositions, hence defining how likely a disposition manifests in the context C_i. We shall write $d_j \preceq_{C_i} d_k$ when $(d_j, d_k) \in \preceq_{C_i}$.

For example, given $\mathcal{D} = \{d_1, d_2, d_3, d_4, d_5\}$ in Table 1, and the airport and hospital contexts C^A_a and C^A_h above, two possible context-related profiles for *Alice* are $E^A_\pi(C^A_a) = (\mathcal{D}, \preceq_{C^A_a})$ and $E^A_\pi(C^A_h) = (\mathcal{D}, \preceq_{C^A_h})$. According to the assigned grades[6], we have $d_1 \preceq_{C^A_a} d_4 \preceq_{C^A_a} d_2 \preceq_{C^A_a} d_3$, and $d_1 \preceq_{C^A_h} d_5 \preceq_{C^A_h} d_2 \preceq_{C^A_h} d_3$. Note that, it can be the case that, in a given context, only a subset of dispositions applies.

Before enacting any tasks involving a contended resource (see Sections 2.4 and 2.5), the grades assigned to ethical dispositions in the profiles of both involved parties are compared via the exchange of offers. Accordingly, the following definition enables the comparison of (and negotiation about) ethically compliant tasks.

4.4. Task ethical implication

Let $\mathcal{T}$ be the set of system tasks, given a user status S, the ethical dispositions affected by a task $t \in \mathcal{T}$ in the context C is $E_\pi(t, C, S) \subseteq \mathcal{D}$, and $d \in E_\pi(t, C, S)$ *iff* d is activated by at least a user status condition in S.

Back to our example, within the airport context C^A_a, the task t_4 = *take elevator* has an impact on the disposition d_2 = *give precedence to injured people* (which has been aptly activated by the *Alice*'s injured status in S^A_a): hence, for *Alice*, $E^A_\pi(t_4, C^A_a, S^A_a) = \{d_2\}$. Similarly, based on *Bob*'s elderly and emergency status S^B_a, the same task t_4 has an impact on the dispositions d_1 = *give precedence to elderly people* and d_3 = *give precedence to the ones in emergency*; hence, for *Bob*, $E^B_\pi(t_4, C_a, S^B_a) = \{d_1, d_3\}$. Note that, thanks to the current user status, we can distinguish if an ethical implication holds in a particular negotiation case.

4.5. Task ethical impact

Given the ethical implication $E_\pi(t, C, S)$, the ethical impact of the task $t \in \mathcal{T}$ in the context C, given the user status S, is defined as $\mathcal{T}_{\varepsilon\iota}(t, C, S) = \sum_{d\in E_\pi(t,C,S)} \lfloor t, C, d\rceil$, where $\lfloor t, C, d\rceil = rank(d, C)$ if d manifests with a positive impact when performing t in the context C, $-rank(d, C)$ if d manifests with a negative impact, N/A if d does not manifest with the task t. Still referring to Table 1, for *Alice*, we have $\mathcal{T}_{\varepsilon\iota}(t_4, C^A_a, S^A_a) = 3$; whereas, for *Bob*, $\mathcal{T}_{\varepsilon\iota}(t_4, C_a, S^B_a) = 7$

It is worth remarking that, since we are adopting a goal-task-action model, when the user specifies the goal, actions to achieve it must be identified and grouped into tasks. Two cases can be distinguished: (a) in the current context, the required actions are already known together with the affected dispositions and suitably grouped into predefined tasks (built-in the system) that can be simply selected, in our case within E_π; (b) the required actions are not known, tasks are not pre-existing and must be newly generated by a planning approach Meli et al. (2023). In the latter case, we build on the simulation-based approaches in Winfield et al. (2014), Vanderelst and Winfield (2018), Winfield and Hafner (2019), Bremner et al. (2019), which adopt the Consequentialism theory to reason about machine ethics in changing contexts Dennis et al. (2016, 2021) for the purpose of anticipating the ethical implications of actions, and hence of tasks, through simulation. In both cases, each task is evaluated to measure its ethical impact so as to enable the comparison of tasks during negotiation.

Before defining the ethics-based negotiations process, it is worth anticipating that the negotiation approach adopts a *utility-based acceptance model* and follows an *alternating protocol* Baarslag et al. (2016), Kiruthika et al. (2020) according to which, upon resource contention, two autonomous systems exchange offers in turn repeatedly until an outcome is collaboratively reached (ethical agreement or no agreement). Thus, each offer specifies the task(s) that the sender would like to execute together with its status conditions. Importantly, the receiver does not have information about the dispositions of the sender and does not need to; the receiver evaluates the offer based on (i) its own ethical profile, (ii) its status conditions, and (iii) the received status conditions of the sender. For this reason, *Alice* (i.e., RobAssist *A* on *Alice*'s behalf) is willing to give up in favor of the offer received from *Bob* (i.e., RobAssist *B* on *Bob*'s behalf) due to the benevolent disposition towards giving precedence to the elderly, especially when in an emergency (Section 3).

4.6. Negotiation offers

Upon a resource contention, given the tasks involved in the contention $\mathcal{T}^{\bowtie} \subseteq \mathcal{T}$, the related set of offers is $\mathcal{O} = \{o_1, \cdots, o_n\}$, where each o_i is a pair $\langle \mathcal{T}^{\bowtie}, S^{\bowtie}_i\rangle$, with $S^{\bowtie}_i \subseteq S$ containing the user's status conditions disclosed within the offer o_i. Thus, $S^{\bowtie}_i$ is such that for each condition

[6] Importantly, the given order of grades is purely coincidental and only for the purpose of our motivating example.

$\langle c, v\rangle \in S_i^{\bowtie}$, $v = true$ (i.e., the conditions included in the offers are only the ones that hold within the current user status model $\mathcal{S}$).

As for the tactic to generate the offers, in this paper, we use a *minimalist strategy* that permits the autonomous systems to disclose minimal information on their user's status conditions, prioritizing the ones activating the dispositions with higher grades first. This strategy follows a gradual information disclosure method Hajaj et al. (2017) that ensures that only the strictly necessary information is released at each negotiation round. Thus, during negotiation, the systems gradually disclose their user's status condition in each subsequent round by adding one user condition at a time: upon each negotiation iteration i, $S_i^{\bowtie}$ is such that the offer o_i discloses exactly i user conditions and, for $0 \leq i < n$, $S_i^{\bowtie} \subset S_{i+1}^{\bowtie}$. As a result, the offers $\{o_1, \cdots, o_n\}$ are ordered according to the inclusion relation $S_i^{\bowtie} \subset S_{i+1}^{\bowtie}$. Abusing notation, we can write $o_1 \leq \cdots \leq o_n$ to indicate that, during negotiation, the offer to be sent at each iteration is selected according to this ascending order. However, other strategies are eligible and can be adopted by our approach, e.g., Integrative, Distributive, Boulware, or Conceder Narayanan and Jennings (2005), Heunis et al. (2024), Luo et al. (2024a). In general, the status conditions to be disclosed might be selected (and hence the related offers) following a qualitative tactic that tends to prioritize some conditions and dispositions over others, a quantitative tactic that instead tends to maximize the number of not-violated dispositions, or a convenient mix of the two.

In addition to the offers described above, we use the symbol $\varnothing$ to denote an *empty offer*, i.e., an offer that does not contain any pair $\langle \mathcal{T}^{\bowtie}, S_i^{\bowtie}\rangle$. As it will be clear later, the offer $\varnothing$ is sent by a negotiating party at the negotiation steps $n+k$, $k \geq 1$, i.e., after it has previously exchanged all the possible offers $o_1, \cdots, o_n$ in $\mathcal{O}$.

The ethical impact of an offer $o_i = \langle \mathcal{T}^{\bowtie}, S_i^{\bowtie}\rangle$, in the current context C, is $\mathcal{O}_{\varepsilon\iota}(o_i, C) = \sum_{t \in \mathcal{T}^{\bowtie}} \mathcal{T}_{\varepsilon\iota}(t, C, S_i^{\bowtie})$, i.e., is the sum of the ethical impacts of the tasks in $\mathcal{T}^{\bowtie}$ in the context C given the user's conditions $S_i^{\bowtie}$ disclosed in the offer (as defined in Section 4.5).

Back to our airport example, the set of offers produced by *Alice* for the elevator contention is $\mathcal{O}^A = \{o_1^A, o_2^A\}$ where $o_1^A = \langle \{t_4\}, \{\langle injured, true\rangle\}\rangle$ and $o_2^A = \langle \{t_4\}, \{\langle injured, true\rangle, \langle flight\_anxiety, true\rangle\}\rangle$. Similarly, the set of offers produced by *Bob* are $\mathcal{O}^B = \{o_1^B, o_2^B\}$ where $o_1^B = \langle \{t_4\}, \{\langle elderly, true\rangle\}\rangle$ and $o_2^B = \langle \{t_4\}, \{\langle elderly, true\rangle, \langle emergency, true\rangle\}\rangle$.

By following the alternating protocol Kiruthika et al. (2020), two autonomous systems involved in the negotiation alternately play the role of the sender, say S, and receiver, say R. We denote with $o^{S\to R}$ the offer sent by S to R; whereas, $o^{R\to S}$ is the counteroffer sent by R, should it reject the offer received by S. That said, the utility of the received offer is calculated as follows.

4.7. Utility function

The receiver R calculates the utility value of an offer o_i in the context C^R by using the following function:

$$\mathcal{U}(o_i, \bar{o}) = \mathcal{O}_{\varepsilon\iota}(o_i, C^R) - \mathcal{O}_{\varepsilon\iota}(\bar{o}, C^R) \tag{1}$$

where $o_i = o_i^{S\to R} = \langle \mathcal{T}^{\bowtie}, S_i^{\bowtie}\rangle$, and $\bar{o} = \langle \mathcal{T}^{\bowtie}, S^R\rangle$.

An important consideration here is that, on the receiver side, the received offer o_i is always compared with the offer $\bar{o}$ that contains the overall receiver status S^R, which therefore accounts for the whole set of status conditions S. This is because, according to the adopted minimalist strategy, the sender does not disclose the whole set of its user's status conditions; whereas, the receiver always has full knowledge of its user's status, and it is reasonable to use the full potential (i.e., all the status conditions in S) since the first iteration.

For our scenario, upon the first negotiation iteration, *Bob* is the sender, *Alice* the receiver, and the first offer $o_1^{B\to A} = \langle \{t_4\}, \{\langle elderly, true\rangle\}\rangle$ is compared with the *Alice*'s $\bar{o} = \langle \{t_4\}, \{\langle injured, true\rangle, \langle flight\_anxiety, true\rangle\}\rangle$ (see the *Alice*'s profile in the context C_a^A in Table 1). As per the grades in the table, the utility function has value $\mathcal{U}(o_1, \bar{o}) = \mathcal{O}_{\varepsilon\iota}(o_1, C_a^A) - \mathcal{O}_{\varepsilon\iota}(\bar{o}, C_a^A) = 2 - 4 = -2$. From the point of view of *Alice*, by considering the only status conditions disclosed by *Bob* at the first iteration, a negative utility means that RobAssist *B* performing the offered task t_4 for *Bob* is less ethical than RobAssist *A* performing t_4 for *Alice*.

The utility function plays a major role in the proposed system in quantifying the ethical impact of tasks with numerical values indicating the comparative ethical consequences involved for all parties concerned Russell and Norvig (2016). As better discussed in Section 8.1, this function follows one of the chief axioms of Utilitarian Ethics Mill (2016), according to which actions can result in maximum benefit or minimum harm for the greater good. However, the utility function itself does not make decisions. It provides a comparative value to feed into the decision-making function, which then decides whether to accept or reject in light of these calculated ethical impacts. Ways of achieving this have also similarly been explored in other frameworks like HERA Lindner et al. (2017), where utility-based comparisons form an integral part of dealing with ethical reasoning in dynamically evolving contexts. This allows ethical reviews to become transparent and anchored in a quantifiable and comparative approach, while decisions are taken in a structured and principled manner.

4.8. Context-aware ethics-based negotiation

The following definition formalizes and generalizes the decision-making process that, at any negotiation step, the receiver R uses to process the received offer $o_i^{S\to R}$ and, depending on the calculated utility, accepts or possibly relaunches the negotiation with its counter-offer $o_i^{R\to S}$, quits otherwise:

$$\mathcal{N}^R(o_i^{S\to R}) = \begin{cases} \texttt{1.accept} & \text{if } \mathcal{U}(o_i^{S\to R}, \bar{o}) > 0,\ o_i^{S\to R} \neq \varnothing \\ \texttt{2.offer } o_i^{R\to S} & \text{elif } i \leq |\mathcal{O}^R| \\ \texttt{3.offer } \varnothing & \text{elif } o_i^{S\to R} \neq \varnothing \\ \texttt{4.quit} & otherwise \end{cases} \tag{2}$$

Following the described alternating protocol, at each negotiation iteration i, upon receiving a non-empty offer $o_i^{S\to R}$, if $\mathcal{U}(o_i^{S\to R}, \bar{o}) > 0$, the receiver R accepts the received offer and the negotiation completes successfully with the agreement on $o_i^{S\to R} \in \mathcal{O}^S$, meaning that the sender S will perform its tasks in $\mathcal{T}^{\bowtie}(o_i^{S\to R})$ (case 1); else, if in $\mathcal{O}^R$ there are still offers to be sent (i.e., $i \leq |\mathcal{O}^R|$), R rejects the received offer and sends the counter-offer $o_i^{R\to S} \in \mathcal{O}^R$ back to S, hence switching its role to the sender (case 2); else, if there are no more offers in $\mathcal{O}^R$ to be counter-offered (i.e., $i > |\mathcal{O}^R|$), cases 3 or 4 apply: if the received offer $o_i^{S\to R}$ is not the empty offer $\varnothing$, R keeps the negotiation alive by sending the empty offer $\varnothing$ (case 3) to continue listening to the sender's offers until either one of them is eventually accepted or the sender quits the negotiation; otherwise (case 4), if the empty offer $\varnothing$ is received (i.e., $o_i^{S\to R} = \varnothing$, meaning that also the sender has no more counter-offers), R unsuccessfully quits the negotiation, in which case no agreement is reached. It is worth remarking that, upon quitting (i.e., when negotiation ends without agreement), the missions do not fail since the negotiating systems apply default built-in rules ethically informed by hard ethics. For instance, a first-come-first-served policy might be applied so that the first-comer uses the contented resource, and the other either waits for it to be available again, uses another (possibly equivalent) resource nearby, or replans a different sequence of tasks (that do not require the contended resource) to reach the selected goal. We recall that, as anticipated in Section 3.2.3, the negotiation can be overruled by, e.g., higher-priority organizational aspects or emergency events, leading the system to follow its built-in behavior as programmed in the native controller.

Back again to our example, being RobAssist *B* the initial sender, the sequence of exchanged offers is: $o_1^{B\to A}$ (which is evaluated and rejected

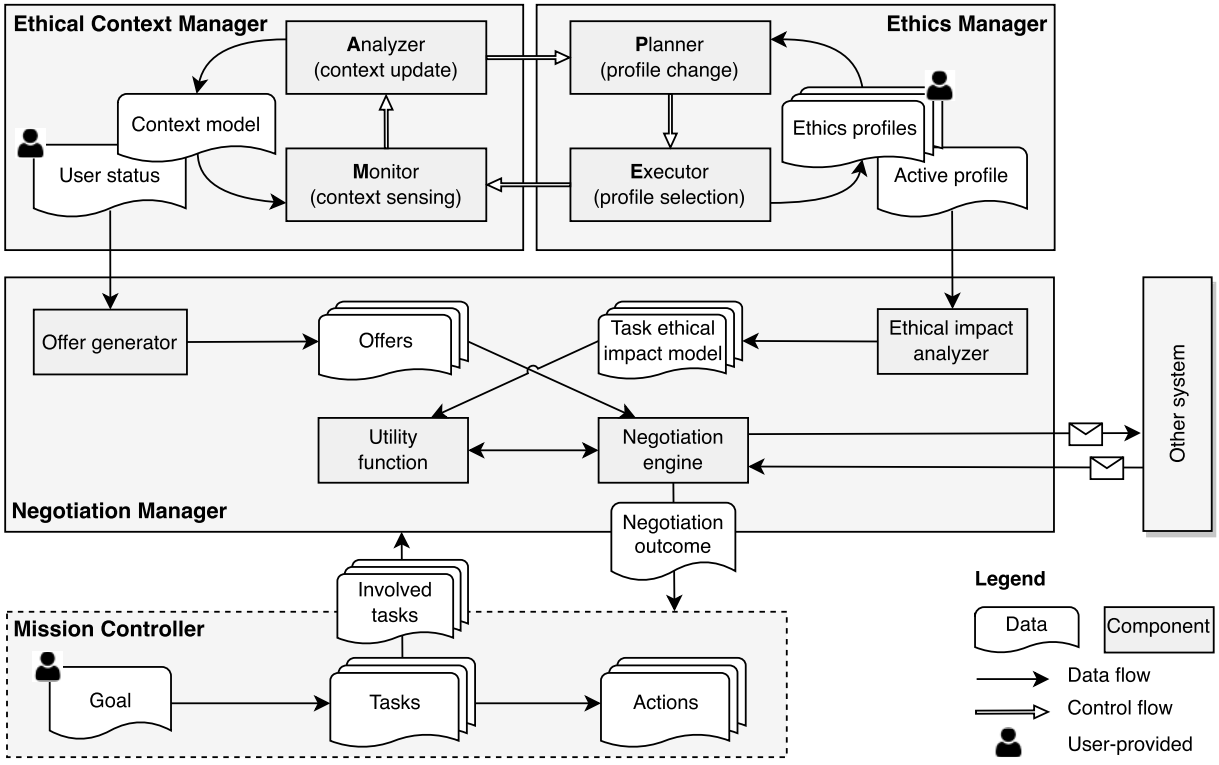


Fig. 2. RobEthiChor architecture overview.

by A), $o_1^{A \to B}$ (which is evaluated and rejected by B), and $o_2^{B \to A}$, which is finally accepted by A. Hence, after the negotiation, B takes the elevator while A either waits for it or looks for another elevator nearby.

The implementation of the negotiation algorithm, as well as the architecture supporting it, is presented in Section 5.

5. RobEthiChor Architecture

This section describes the RobEthiChor reference architecture we propose to support the context-aware negotiation as described in the previous section. The architecture is conceived so as to support (i) the representation and management of the user ethics, (ii) the management of the context, and (iii) the negotiation for selecting the tasks to be executed. Fig. 2 shows the proposed architecture. It consists of four main components: *Ethics Manager*, *Ethical Context Manager*, *Negotiation Manager*, and *Mission Controller*.

At a glance, the Ethics Manager has the role of getting, managing, and representing the user's ethics; the Ethical Context Manager dynamically handles the context and user status for the realization of the context-awareness features of the negotiation; the Negotiation Manager realizes the negotiation by supporting the interaction with the other system it has to negotiate with. The Mission Controller handles the "traditional" execution of the tasks – and their associated actions – to reach the selected goal. In the following, we detail the aforementioned components.

5.1. Mission controller

The Mission Controller manages the behavior of the autonomous system by handling the execution of the tasks to be performed in order to reach the system's goal. As described in Section 2.5, it adopts the "Goal-Task-Action" model: upon receiving a goal from the user's input, a sequence of tasks and related actions to be executed is identified (either by relying on pre-defined mission specifications or through planning approaches Meli et al. (2023)) and executed. Moreover, it is the component that, as discussed in Section 2, is in charge of recognizing whether a task requires a contended resource. When, during the execution of the mission, a resource contention is detected, the Mission Controller interacts with the Negotiation Manager component by providing the set of tasks that are involved in the resource contention. After the negotiation is completed, the Mission Controller gets the outcome and continues the execution of the mission accordingly (i.e., by continuing the execution of the tasks or by replanning an alternative task sequence), as explained in Section 4. Note that the Mission Controller is a platform-dependent component since it has to control the hardware through its lower-level interfaces in order to execute actions, and to access sensor data to recognize resource contentions. Thus, despite being depicted as a single component, in practice, the Mission Controller may be realized as a set of components that manage different aspects and functionalities of the robotic system. For this reason, we abstract this component by only focusing on its Goal-Task-Action paradigm and on the interactions that need to occur between it and the other components of RobEthiChor, which realizes an additional layer, lying on top of the controller, that enables the ethics-based negotiation. Any autonomous system's controller adopting the Goal-Task-Action paradigm can be updated, and integrated in our architecture through the interface with the Negotiation Manager.

5.2. Ethical context manager

The *Ethical Context Manager* has the role of getting and holding information about the current context conditions and the user status. Together with the Ethics Manager, it realizes a feedback *MAPE* loop, an autonomic loop which is a widely adopted reference model for managing and controlling autonomous and self-adaptive systems Kephart and Chess (2003), Cámara et al. (2017), Arcaini et al. (2015) bringing significant advances in various domains, such as autonomous cars, intelligent transportation and traffic management Gerostathopoulos and Pournaras (2019), unmanned vehicles Maia et al. (2019), Moreno et al. (2019), IoT

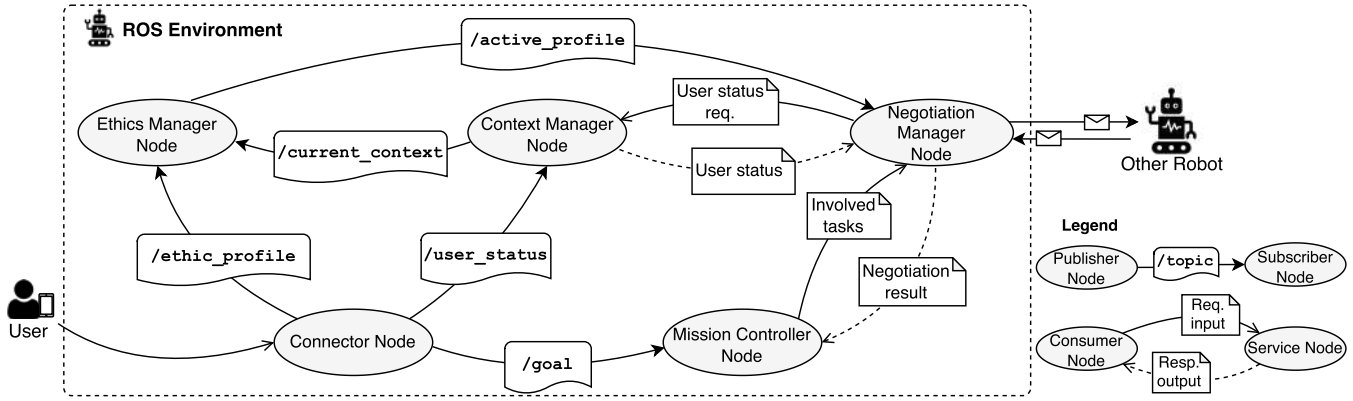


Fig. 3. RobEthiChor-Ros architecture.

and smart-home systems Iftikhar et al. (2017), Arcaini et al. (2020), as well as assistive robotics Jamshidi et al. (2019).

The MAPE loop monitors the context dimensions that are relevant for the negotiation[7] to detect context changes and select the user's ethics profile that is most suitable for the current conditions. The Ethics Manager implements the *Monitor* and *Analyze* components of the loop. In particular, the monitor component senses the context by reading the information coming from sensors and other components of the systems (e.g., GPS, camera, thermometer, etc.), and the analyzer component checks the context values to detect changes. If changes are detected, the context model is updated to reflect the changes, and the update is propagated to the *Ethic Manager*, which activates - if needed - a new profile suitable for the new context conditions.

5.3. Ethics manager

The Ethics Manager has the role of getting and storing the user ethics profiles E_π. We recall that E_π owns the set ethical dispositions as graded by the user. It implements the *Plan* and *Execute* components. When a context change is detected by the Ethical Context Manager, the *Planner* checks if a new active profile needs to be selected. If this is the case, the *Executor* sends the active profile to the Negotiation Manager. In this way, the Negotiation Manager is provided with the grades associated with each disposition to be used for the negotiation.

5.4. Negotiation manager

The Negotiation Manager is the component that implements all the logic needed to realize the negotiation with other systems. Upon a resource contention, the Negotiation Manager receives from the Mission Controller the list of tasks involved in the resource contention. Then, given the involved tasks and the current user status S (which is gathered from the Ethical Context Manager), the *Offer generator* component generates the set $\mathcal{O}$ of offers, plus the offer $\bar{o}$. The *Ethical impact analyzer* component computes the *Task ethical impact model* for each task involved in the resource contention. This model is the intermediate representation of the task's ethical impact, here denoted as $\mathcal{T}_{\varepsilon_I}(t, C)$, which associates a task t to the grades $rank(d_i, C)$ for each disposition d_i that can be activated for t in the current context C, according to the currently active profile of the user (see the *Alice's grades* columns in Table 1). Since, as explained in Section 4, the task ethical impact also depends on the set S of the user conditions (that are disclosed only within offers), the intermediate model is exploited together with the user conditions to dynamically calculate $\mathcal{T}_{\varepsilon_I}(t, C, S)$ during the negotiation. Specifically, given the set of status conditions $S_i^{\bowtie}$ within a received offer o_i, the task ethical impact model is exploited by the *Utility function* component to obtain the offer's ethical impact and, hence, the value of $\mathcal{U}(o_i, \bar{o})$ (Definition 1).

The *Negotiation engine* is the component that executes the negotiation by (i) interacting with the other systems via the exchange of offers and (ii) performing the decision-making process according to the function $\mathcal{N}$ (Definition 2). It has a twofold alternating role: when playing the receiver's role, it receives the offer and asks the utility function for the computation of the offer's utility; when playing the sender's role, it selects the offer to be sent from the generated offers set. Upon reaching an agreement, the negotiation engine returns the negotiation's outcome (`accept/reject/no_agreement`) to the mission controller.

6. RobEthiChor-Ros

RobEthiChor-Ros is a ROS2-based implementation of the negotiation approach presented in Section 4 concretizing the architecture in Section 5 for the robotic domain. RobEthiChor-Ros provides the full set of functionalities for managing the user's ethical profile, user status and context, and for running the negotiation process. It also exposes RESTful APIs to support the integration with mobile apps (as described in Section 3.2) to allow users to upload and update their ethical profiles and current statuses, and to provide the goal that the robot is called to achieve (see user-provided elements in Fig. 2).

6.1. ROS-Based architecture

Fig. 3 shows the architecture of RobEthiChor-Ros. Each of the main components of RobEthiChor is realized as a ROS node, which communicates with the others through *topics* or *services*. The first realizes the asynchronous message passing for updating context-related information, and the latter realizes the request-response communication among the components involved in the negotiation. In addition, RobEthiChor features a further node, the *Connector Node*, which provides the aforementioned RESTful service interfaces. Once the connector receives the ethic profile, the user status, and the selected goal through the interface's endpoints, it publishes them on the related topics' named buses `/ethic_profile`, `/user_status`, and `/goal`, respectively.

The *Ethics Manager Node* subscribes to the `/current_context` and `/ethic_profile` topics to realize the selection of the active profile. When selected (as described in Section 5), the active profile is published on the `/active_profile` topic channel to make it available for negotiation.

The *Context Manager Node* implements the functionalities required to acquire the context information that is relevant for the negotiation by interacting with the robot's sensors, cameras, and external ser-

[7] The context information handled by the Ethical Context Manager Node within RobEthiChor only concerns the dimensions considered for the profile selection and, as a consequence, the negotiation. These may likely be different from the ones considered for realizing the context-aware capabilities of the robot on its "normal" behavior, i.e., not related to the ethics of the user.

vices through ROS topics and managing runtime context changes. Upon a change in the context conditions, the current context is published through the `/current_context` channel. This node also subscribes to the `/user_status` channel to receive user status updates sent through the connector interface and implements the `UserStatus` service used by the negotiation manager to get the user status during negotiation.

The *Negotiation Manager Node* realizes the negotiation by implementing the negotiation process and exposing a communication interface with the other robots. The inter-robot communication is realized by using the `/negotiation` ROS topic. The node subscribes to the `/active_profile` topic to get the active profile as soon as a context change is triggered and a new profile related to the current context is activated. It also requires the user status to the context manager node through the ROS service exposed by the latter. The Negotiation Manager exposes the `NegotiationService` ROS service to receive the list of tasks involved in a resource contention. After receiving the request, the Negotiation Manager starts negotiating with the other robot involved in the resource contention. After the negotiation is completed, the Negotiation Manager replies with its result (`winner`, `loser`, `no-agreement`).

The *Mission Controller Node* realizes the Mission Controller component, as described in the previous section. It subscribes to the `/goal` topic to receive the goal selected by the user. Then, it controls the execution of the mission by selecting the tasks to be executed and interacting with the other native ROS nodes (external to this architecture) that control the robot's components to execute them. It is worth recalling that any ROS-based mission controller implementation, specific for a given robot, can be integrated with this architecture so as to interact with the negotiation manager through the `NegotiationService` interface by: (i) subscribing to the `/goal` topic to receive the user-defined goal, and (ii) implementing the service client for the `/negotiation` service.

6.2. Negotiation algorithm implementation

The negotiation process uses two subroutines that realize the behavior of the negotiating robot when acting as a sender and when acting as a receiver, according to the decision-making process defined in Formula (2). Algorithm 1 and Algorithm 2 present the pseudocode of the two subroutines. The returned value from the negotiation algorithms is sent back to the Mission Controller, as described in the previous section.

Algorithm 1: Sender algorithm.

```
1: Input: offer set 𝒪 = {o_1, ⋯, o_n}, offer ō, index i, last received offer o*
2: Output: negotiation result res ∈ {win, lose, no_ag}
3: if i ≤ |𝒪| then
4:     send offer o_i
5: else if o* ≠ ∅ then
6:     send offer ∅
7: else
8:     send quit
9:     return no_ag
10: end if
11: response ← wait_response()
12: if response = accept then
13:     return win
14: else
15:     return receiver_algorithm(𝒪, ō, i)
16: end if
```

The sender subroutine in Algorithm 1 is executed by taking as input the set of offers $\mathcal{O}$, the offer $\bar{o}$, the index of the negotiation step i, and the last received offer. When the robot starts the negotiation as a sender, the algorithm is called with $i = 1$ and $\emptyset$ as the last received offer. According to cases 2, 3, and 4 in Formula (2), the sender either sends the offer o_i, the empty offer $\emptyset$, or exits the negotiation otherwise (lines 3-10). Offers are sent through the `/negotiation` channel. If the robot quits the negotiation, a `quit` message is sent through the channel as well to end the negotiation, and the algorithm ends with `no-agreement` as a result (lines 8-9). If still in the negotiation, the robot waits for the response from the opponent (that is running the complementary receiver subroutine): if the opponent accepts, the algorithm ends by returning `winner`; otherwise, the receiver subroutine is called for completing the current negotiation step (lines 11-16).

Algorithm 2: Receiver algorithm.

```
1: Input: offer set 𝒪 = {o_1, ⋯, o_n}, offer ō, index i
2: Output: negotiation result res ∈ {win, lose, no_ag}
3: o* ← receive_offer()
4: if o* ≠ ∅ then
5:     if 𝒰(o*, ō) > 0 then
6:         reply(accept)
7:         return lose
8:     end if
9: end if
10: reply(reject)
11: return sender_algorithm(𝒪, ō, i + 1, o*)
```

The receiver subroutine in Algorithm 2 is executed by providing as input the set of offers $\mathcal{O}$, the offer $\bar{o}$, and the index of the negotiation step i. When the robot starts the negotiation as a receiver, the algorithm is called with $i = 0$[8]. The subroutine implements the decision-making algorithm according to case 1 in Formula (2). After receiving a non-empty offer $o^\star$, the robot running the receiver algorithm computes the utility value $\mathcal{U}$. If the utility value is greater than 0, the robot accepts the negotiation by replying with an `accept` message on the `/negotiation` topic, and the algorithm ends by returning `loser` (lines 4-9). Otherwise, the robot rejects the offer (line 10) and switches its role to be a sender in the next negotiation round (line 11).

The two robots running the negotiation execute the two subroutines following the alternating protocol introduced in Section 4.5: a robot runs the receiver algorithm while the other runs the sender algorithm. Accordingly, negotiation messages also serve as synchronization points that keep the execution of the negotiation synchronized between the two alternating parties: robots consume messages from the `/negotiation` queue and wait for the offer and result messages to be received. Moreover, it is worth noting that the termination of the algorithm is guaranteed by construction: robots iterate over the index i and quit the negotiation without agreement when $i > max(|\mathcal{O}^A|, |\mathcal{O}^B|)$, being O^A and O^B the offer set of the two negotiating parties. In fact, the exit condition requires that the robot has already sent all its offers (line 3 in Algorithm 1) and it has already received the empty offer (line 5), i.e., when the opponent has already sent all its offers as well. When this condition is met, and the `quit` message is sent, the opponent robot interrupts the execution of the receiver algorithm and exits the negotiation. Finally, the wait for receiving messages has a timeout to prevent robots from endlessly waiting for messages in case of failures, e.g., network connection issues.

7. Evaluation

In this section, we present and discuss the evaluation of our approach and its implementation. The goal of the evaluation is to assess the effectiveness of the presented approach in realizing the negotiation according to the user's ethical preferences, status, and context, and to evaluate the overhead introduced by the negotiation process and its scalability. To guide the evaluation, we define the following evaluation questions (EQs):

[8] The index i keeps track of the number of sent offers, since, when playing the sender role, robots send the offer o_i

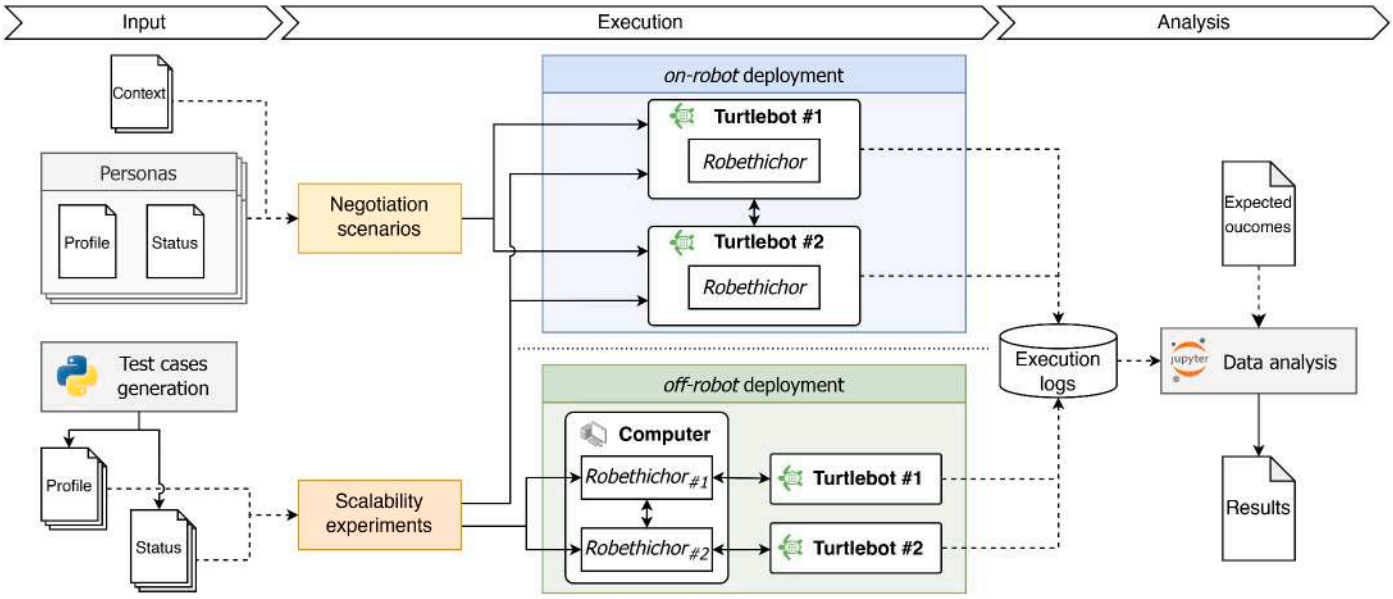


Fig. 4. Experimentation setting overview.

EQ1: Is the negotiation process effective in realizing the negotiation according to user ethical preferences, status, and context?

EQ2: What is the impact of the negotiation process implemented in RobEthiChor-Ros on the system in terms of computation overhead?

EQ3: Does the negotiation process in RobEthiChor-Ros scale according to the size and complexity of the ethical profiles?

By answering EQ1, we aim to validate the effectiveness of the negotiation process in reaching a situational agreement that satisfies the ethical preferences of the users the robots are acting on behalf of. By answering EQ2, we aim to assess if the negotiation process run by RobEthiChor-Ros has an acceptable impact, in terms of overhead time required to run the negotiation, for practical application in real-world scenarios. EQ3, instead, aims at stressing the scalability of RobEthiChor-Ros, evaluating how it scales according to the dimension of the user statuses and ethical profiles, should they be pushed beyond the presumably limited real-world size, in terms of the number of conditions and dispositions that a human can reasonably configure and care about.

Fig. 4 overviews the experimentation setting. We deployed RobEthiChor in two different settings: (i) on two real Turtlebot 4 robots[9] running Ubuntu 22.04 and ROS2 Humble on their Raspberry Pi 4B computer (*on-robot* deployment), and (ii) on an external computer equipped with an Intel Core i7 CPU and 16GB RAM running Ubuntu 22.04 and ROS2 Humble (*off-robot* deployment). In the *off-robot* setting, the Mission Controller component runs on the Turtlebots, while the negotiation-related components of RobEthiChor run on an external device providing functionalities to the robot as in a cloud- or fog-robotics approach Kehoe et al. (2015), Chen et al. (2021). Both the robots and the computer are connected to a local dual-band (2.4/5GHz) WiFi network, providing, on average, 2.8 ms round-trip time (1.37 ms minimum, 6.84 ms maximum, 0.997 std. dev. measured by periodically sending *ping* messages during the experiment execution).

To answer *EQ1* and *EQ2*, we run multiple negotiation between couples of robots "serving" different *personas*, each with their own ethical profile and status, in two different contexts: the airport scenario described in Section 3, and the hospital scenario inspired by the *Welcoming people to the hospital* in RoboMAX Askarpour et al. (2021) described in Section 3.3. We run the negotiation scenarios in the *on-robot* deployment setting and check the behavior of the robot in response to the negotiation outcomes. To answer *EQ3*, we run multiple negotiation scenarios between randomly generated users by increasing both their profile size (e.g., the number of dispositions included in the profile) and the number of user conditions to measure the negotiation time over increasingly complex scenarios. In this case, we tested the scenarios in both the *on-robot* and *off-robot* settings. The different configurations are provided to the RobEthiChor instances by uploading users' status and ethics profiles, as described in Section 3, through the RESTful API exposed by the connector nodes. Similarly, after configuration, the mission is started simultaneously on the two robots by providing the mission's goal to the connectors' RESTful endpoints. After each negotiation is completed, the robots log: (i) the negotiation outcome (i.e., whether the robot was the winner or loser, or if no agreement was reached), (ii) the time required for running the negotiation (i.e., the time elapsed between the sending of the negotiation request to the Negotiation Manager and the receiving of the negotiation outcome), and (iii) the number of negotiation rounds (i.e., the final value of the index i within Algorithms 1 and 2).

In the following, we detail the experiments. The code, the experimentation settings, and the results are available in the provided replication package[10].

7.1. EQ1 & EQ2: Negotiation scenarios

To realize the negotiation scenarios for *EQ1* and *EQ2*, we defined a set of 17 dispositions that concern the tasks involved in the resource contention scenarios, plus 27 possible user status conditions that activate the defined dispositions. Starting from this common ground, we created 10 different *personas*, each representing a potential user, by eliciting their status and ethical preferences. Then, we built (i) the context models for the two negotiation contexts (i.e., the airport and the hospital) and, for each persona, (ii) an ethical profile that considers possibly varying dispositions for different contexts (by aptly assigning context-related grades to the dispositions, as in Table 1), and (iii) a user status (by setting to *true* the values associated with their conditions). Fig. 5 reports two of the defined personas (Alice and Bob). We provided a card for each persona summarizing the persona's characterizing traits, a description of the needs as a system user, and the conditions (i.e., the user status) in the two negotiating contexts. The figure also reports the ethical profiles built for the two personas, showing the grades assigned to

[9] https://clearpathrobotics.com/turtlebot-4/

[10] https://doi.org/10.5281/zenodo.17449923

ALICE

Age: 29
Occupation: Professional Athlete
Personality: Disciplined • Independent • Resilient
Hobbies: Hiking • Yoga • Reading
Tech use: Comfortable with assistive tools, especially when practical

Traits: Alice is a 29-year-old professional athlete. She's disciplined, active, and enjoys hiking, yoga, and reading in her downtime. Normally independent and confident, Alice prefers to manage things on her own. After a recent leg injury and learning she is pregnant, she became more aware of situations that require extra care and attention. She believes that people with visible health conditions, disabilities, or medical needs, and unaccompanied minors, should be given support first. She also understands how overwhelming less obvious issues like anxiety or crowd stress can be.
Needs (Airport): For a medical check-up related to her pregnancy and recovery, Alice travels to another city. At the airport, she feels the strain of walking long distances through crowded areas. The noise and lack of clear signs make the experience more tiring than expected. Spotting a nearby assistant robot offering mobility help, she decides to use it, which makes moving easy for her and saves her energy.
Needs (Hospital): Once in the city, Alice visits the hospital for her check-up. While she tries to stay calm and focused, the emotional weight of her physical recovery and pregnancy catches up with her. She feels vulnerable and more easily overwhelmed than usual. Despite this, she values when procedures are explained clearly and handled efficiently. Though she's still not used to asking for help, moments like these have shown her the value of supportive tools that make the journey easier without compromising her independence.

Airport user status: Injured, Pregnant, Pre-natal care, Crowd Stress
Hospital user status: Injured, Pregnant, Pre-natal care, Emotional distress, Crowd stress

BOB

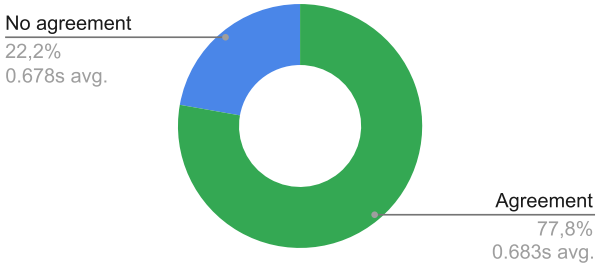
Age: 67
Occupation: Retired (former school teacher)
Personality: Calm • Patient • Respectful • Quietly observant
Hobbies: Gardening • Reading newspapers • Listening to classical music
Tech use: Basic use of mobile phone and tablet

Traits: Bob is 67 years old and recently retired. He enjoys gardening, reading, and quiet music. Though generally active, he has some age-related health issues and gets tired walking long distances. Recently, he also suffered a mild fall that left him with some lingering leg pain, making mobility more difficult than usual. English is not his first language, which sometimes makes travel challenging. He believes that people with health issues, the elderly, and non-native speakers should be given priority. While usually calm, he appreciates support when situations become confusing or rushed.
Needs (Airport): Bob is traveling to another city for a medical check-up. At the airport, he feels a bit lost and pressed for time. Walking far is tiring, and the signs aren't always clear. When he notices an assistant robot offering help, he decides to use it. It guides him to his gate smoothly and lowers his stress.
Needs (Hospital): When Bob reaches the hospital, he stays patient and attentive. However, when the check-in process becomes unexpectedly long and the waiting area grows crowded, he starts to feel anxious and emotionally worn down. He believes that patients with medical needs or physical impairments should go first and appreciates clear communication from the staff. Though he doesn't often ask for help, simple, supportive systems make a big difference in easing his discomfort.

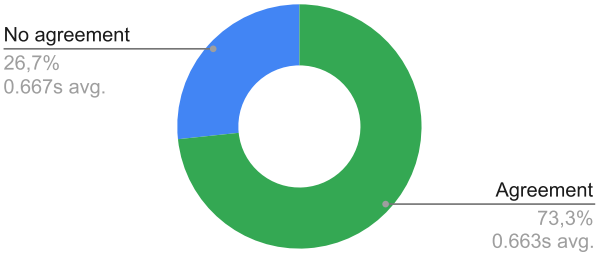
Airport user status: Elderly, Age-related health issues, Rushing to flight, Non-native speaker
Hospital user status: Elderly, Injured, Non-native speaker, Emotional distress, Age-related health issues

Disposition	Description (Give precedence to .....)	Alice's Grades (A)	Alice's Grades (H)	Bob's Grades (A)	Bob's Grades (H)
d_1	the injured	2	4	1	3
d_2	the pregnant	4	3	3	3
d_3	the elderly	3	3	4	1
d_4	the children	2	1	1	1
d_5	an accompanied minor	5	2	1	2
d_6	one with general medical needs	4	4	4	4
d_7	the non-native speaker	1	2	3	2
d_8	one to avoid non-obvious/ non-medical situations (anxiety, crowd stress, etc.)	3	3	1	1
d_9	one traveling with a pet	0	0	0	0
d_{10}	whom the general assistance is required (lost something, flight delay, luggage transfer, fragile luggage, is lost, etc.)	2	0	3	0
d_{11}	ones with time-sensitive emergencies (tight flight schedule, delayed arrival, unseen emergency).	2	0	3	0
d_{12}	the disabled (mobility issues, hearing or visual impairments, or other disabilities)	4	4	3	3
d_{13}	one accompanied by a dependent (child, elderly, or disabled person)	2	1	3	2
d_{14}	one experiencing emotional distress	0	3	0	3
d_{15}	one with infectious symptoms (e.g., fever, coughing)	0	4	0	4
d_{16}	one needing urgent diagnostic access	0	4	0	4
d_{17}	one recovering from surgery (post-op)	0	3	0	2

Fig. 5. Overview of Alice and Bob personas.



(a) Airport context



(b) Hospital context

Fig. 6. Overview of the simulation outcomes.

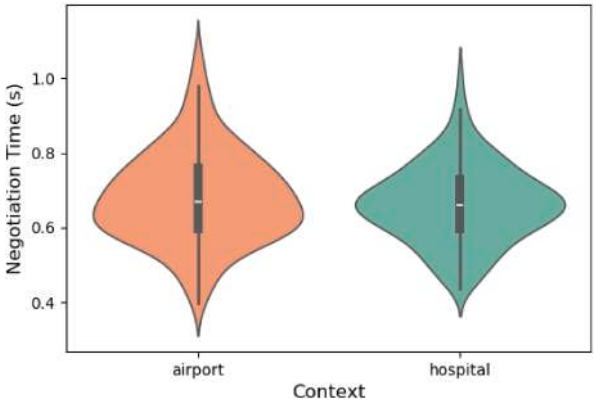


Fig. 7. Distribution of negotiation times.

each of the dispositions in the airport (A) and hospital (H) contexts. By enumerating all the possible combinations of pairs of users (45 pairs), we manually built a ground truth by associating the expected negotiation outcome with each pair of users for each of the two negotiation contexts, according to their profiles and statuses. Then, we ran the negotiation for each of the pairs and repeated the runs four times per each context (45 possible pairs, 180 runs per context). The logs obtained after running the simulations were analyzed and compared with the ground truth to assess whether the system's behavior reflects the user's expectations and to measure the time required for running the negotiation.

The results obtained through the automated negotiation implemented in RobEthiChor fully comply with the ones expected with the ground truth and align with the ethical considerations in the defined personas. The negotiation always showed the same outcome in each run that featured the same pair of users in a given context, regardless of their order in the pair, the robot acting on their behalf, and their starting roles (i.e., if a user started the negotiation as a sender or receiver). That is, for each pair in the airport context, and for each pair in the hospital context, either the same user always won the negotiation or no agreement was always reached.

Fig. 6 summarizes the results of the negotiation between the 45 possible combinations of personas: in the airport context (Fig. 6)(a) an agreement was reached in 140 out of 180 runs (77.8 %), while in 40 runs (22.2 %), no agreement was reached, as expected from the ground truth data. It is worth remarking that no agreement was reached in 40 runs because, according to the defined personas and the derived ethical profiles, none of the two negotiating users (i.e., robots on their behalf) was willing to concede to the other the contended resource. The negotiation in the hospital context (Fig. 6)(b) gave different results: here, the agreement was reached in 132 out the 180 runs (73.3 %), while no agreement was reached in 48 runs (26.7 %), as expected from the ground truth for this context. Moreover, we observed that, for 16 pairs of users, the negotiation results were different in the two contexts (i.e., a different agreement was reached). These different results highlight how the outcome of the negotiation between the same pair of users may be influenced by the context where the negotiation happens since, as explained in Sections 3 and 4, the grades in a given user's ethical profile for different contexts may differ. For instance, by referring to Alice and Bob's personas reported in Fig. 5, in the airport context, no agreement was reached as none of them demonstrates a disposition towards giving precedence to the other, according to the personas description, the profiles built upon it, and the users statuses in that context. Instead, in the hospital context, Alice won the negotiation due to the Bob's disposition in giving precedence to injured people or with medical needs.

Concerning the time required for running the negotiation, we measured that the robots completed the negotiation in 0.672 seconds on average (0.397 s minimum, 1.069 s maximum), after running 4 to 9 negotiation rounds.

Fig. 7 better describes how the negotiation times are distributed. In the airport context, most of the negotiations (83.9 %) were completed in less than 0.8 seconds. 75 % of the negotiations were completed in less than 0.758 seconds, while 50 % in less than 0.670 seconds. In the hospital context, 91.1 % of negotiation completed in less than 0.8 seconds, 75 % in less than 0.728 s, 50 % in less than 0.659 s.

These results show that the negotiation process implemented in RobEthiChor allows the robots to follow a behavior that is respectful of ethical preferences, according to the ethics profiles, hence providing an answer to *EQ1*. Concerning *EQ2*, we argue that the measured overhead for negotiation can be considered acceptable, as negotiations for the considered use case are completed in a few tenths of a second, hence having a negligible impact on the overall system's performance.

7.2. EQ3: Scalability experiments

To thoroughly assess the scalability of the negotiation process in the scope of *EQ3*, we performed experiments to measure its overhead in more complex scenarios aptly designed to stress the negotiation process far beyond real-world scenarios. We randomly generated multiple sets of users with profiles consisting of an increasing number of n graded dispositions and an increasing number of p active conditions in their statuses, each of them activating exactly one disposition. We generated 10 different profiles for each value of n in {25, 50, 100, 200, 300} and 10 different statuses randomly selecting the active conditions set to `true`, starting with p equal to 10 % of n, up to 25 %, 50 %, 75 %, and 100 %. This allowed us to experiment with negotiations involving sets of offers of different sizes, both in the number of offers in the set $\mathcal{O}$, and in the number of status conditions $S_i^{\bowtie}$ disclosed with each offer $o_i \in \mathcal{O}$. To account for the implicit variance in the randomly generated inputs, we ran the negotiations among the 100 possible combinations of the 10 random users (also allowing two equal users to negotiate) for each combination of n and p. As explained above, we ran the negotiation on the two different *on-robot* and *off-robot* settings.

Table 2 reports the measured times for the negotiation settings with n up to 100. In the *on-robot* deployment, the total average time required for running the negotiations in these settings is 2.15 seconds (with 1.63 seconds as the value for the 50th percentile). Fig. 8 shows how the ne-

Table 2
Average negotiation times (in seconds).

Configuration	10 %	25 %	50 %	75 %	100 %
on-robot deployment					
25	0.314	0.668	1.35	2.08	2.96
50	0.62	1.38	2.69	*4.46*	*6.27*
100	0.99	2.74	*5.27*	*7.82*	*11.5*
off-robot deployment					
25	0.154	0.300	0.600	0.953	1.34
50	0.310	0.613	1.20	2.06	2.90
100	0.490	1.24	2.58	*4.48*	*6.25*

gotiation time increases as the number of active conditions (driving the number of offers in $\mathcal{O}$, see Section 4.6) increases.

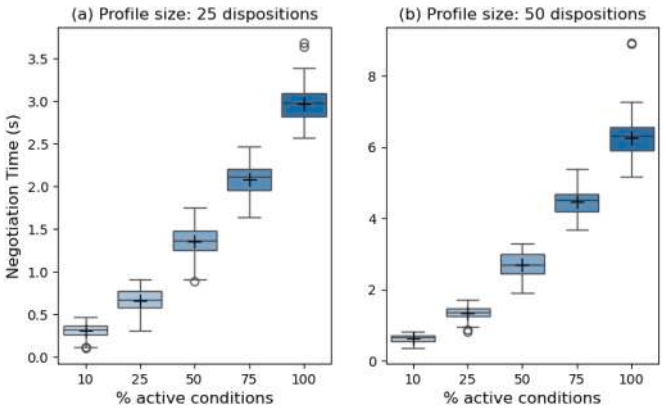


Fig. 8. Negotiation times (in seconds) with *on-robot* deployment.

In the *off-robot* deployment, we can observe that the negotiation time is generally lower than the *on-robot* deployment. The average time required for running the negotiation in these settings is 1.70 seconds (with 1.11 as the value for the 50th percentile). Fig. 9 plots the results obtained in all the settings run with this deployment: as in the *on-robot* deployment, the negotiation time increases as the number of offers in $\mathcal{O}$ increases.

This consideration is better highlighted in Fig. 10, which reports the average negotiation times according to the size of the offers set $\mathcal{O}$ over some of the configurations ran in the *off-robot* deployment. Interestingly, the negotiation time appears to be not affected by the size of the profile. This means that the time required to compute the ethical impact of a given offer (which is reasonably affected by the number of active dispositions, see again Section 4.6) has a negligible impact on the overall negotiation.

Conversely, the reason for the increasing overhead is that a higher number of offers demands, on average, a higher number of rounds required to reach an agreement (or, if no agreement is reached, to end the negotiation without agreement). Fig. 11 reports the time required by single negotiations over the number of rounds required for their completion in different configurations: the time grows with the number of rounds. Again, the chart demonstrates that the size of the profile, alone, does not imply a higher overhead; rather, the number of active conditions – which is directly associated with the number of offers, see Section 4.6 – causes the negotiation time to grow. This is reasonably due to the offer exchange delay between the two participants, which is done over the network. Regression analysis identified that the negotiation time has a polynomial growth with the number of rounds required for reaching an agreement. In particular, for the *on-robot* setting, a second-degree polynomial model provided an accurate fit (m.s.e. 0.0456, R^2 0.983) with a very low coefficient on the second-degree term (0.00033). Similarly, for the *off-robot* setting, a second-degree model provided an accurate fit as well (m.s.e. 0.1655, R^2 0.9965) with an even lower coefficient on the second-degree term (0.0001512). However, in both cases, a linear model provided a slightly less, but still very accurate, fit (m.s.e. 0.0502, R^2 0.9812 for the *on-robot* setting, m.s.e. 1.1234, R^2 0.9762 for the *on-robot* setting), hence suggesting good scalability of the negotiation approach.

It is worth discussing that, in general, the settings where the negotiation time introduced a high overhead, which can be arguably considered unacceptable (e.g., more than 3 seconds) are arguably unlikely in a real-world scenario. As explained above, we ran these settings to stress the system beyond reasonable human-provided input sizes with the primary aim of assessing the scalability dimension. In fact, the aforementioned settings feature huge profiles that consider more than 100 dispositions and, more importantly, an extremely wide and detailed set of possible user conditions (100 or more), most of which hold (i.e., most of the user status conditions are set to `true`) in the current negotiation context (75 % or more). From a practical point of view, this means that we are admitting the existence of users who activated, at the same time, most of the (hundreds) possible status conditions, even though some of them may be conflicting, e.g., *child* and *elderly*. As a reference, to make it realistic, personas described in Section 7.1 have, at most, 8 active conditions out of the 27 defined (~30 % of the total). Overall, among the 1500 negotiation runs in the *on-robot* deployment, 76.5 % of them ended in less than 3 seconds; this value raises to 84.73 % in the *off-robot* deployment over the same configurations (among the full set of 2500 negotiation runs in the *off-robot*, 62.3 % of them ended within 3 seconds). However, the results discussed above hint at possible strategies to reduce the negotiation time even in the "extreme" settings where the negotiation overhead becomes unacceptable. Due to the nature of the overhead, which is strongly dependent on the network communication, the *off-robot* deployment allows reducing the negotiation time if compared to the *on-robot* deployment (see Table 2). In our experimental setting, the overhead is approximately halved when the negotiation is run locally on an external computer. Moreover, reducing the size of the offer set or the number of exchanged offers would be effective in further reducing the overhead. This can be achieved either by (i) disclosing more than one status condition in each offer or negotiation round, e.g., increasing the value of i in Algorithm 1 and 1 by more than 1 each round, or (ii) introducing a time deadline, as further discussed in Section 9.2.

In summary, for *EQ3*, we can observe that the negotiation overhead grows approximately linearly with the number of negotiation rounds. We thus argue that the negotiation approach implemented in RobEthiChor-Ros is scalable and that scalability can be further improved to support larger and more detailed user status models by reducing the number of exchanged offers as discussed.

8. Discussion

In this section, we discuss the more relevant aspects of the proposed approach.

8.1. Ethic statement

It is important to clarify that none of the related work (and so is for RobEthiChor) pretends to provide a universal and complete solution capable of ensuring that an autonomous system will always behave ethically according to an ideal common acceptance of ethical values. That is to say, if on the one hand, we can assume the technical capability to provide an ethical encoding as a possible means to embody context-dependent moral values in a machine through dispositions (with users having some degree of control over these), on the other hand, we cannot presume that the specific way we are proposing will be deemed "acceptable" by all within the broader sphere of different disciplines, including psychology, philosophy, and social science. As discussed in Section 11, unlike some of the existing studies, the system in our work considers

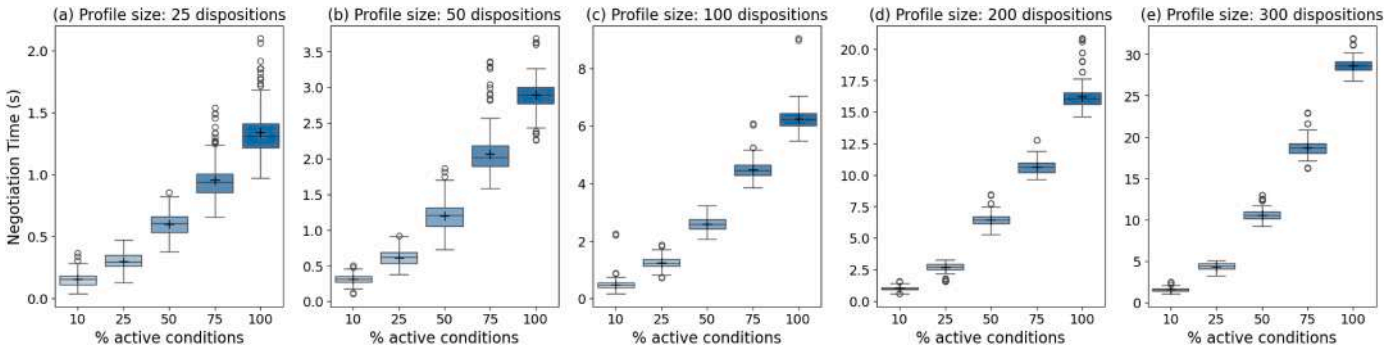


Fig. 9. Negotiation times (in seconds) with *off-robot* deployment.

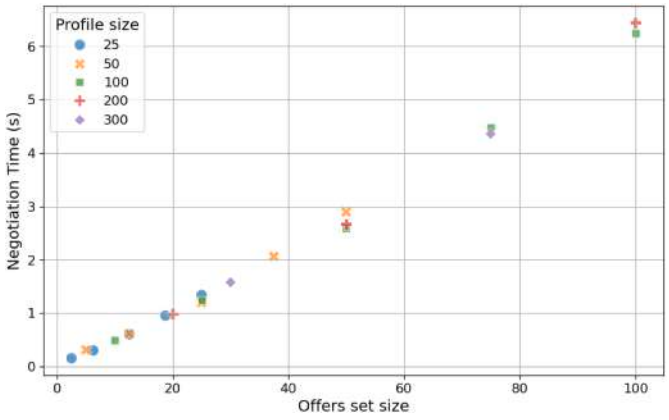


Fig. 10. Average negotiation times (in seconds) over offer set size with *off-robot* deployment.

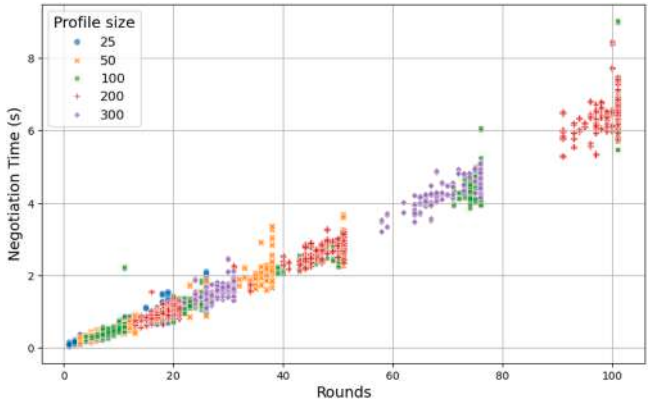


Fig. 11. Negotiation times (in seconds) over negotiation rounds with *off-robot* deployment.

user ethical preferences without forcing adherence to a specific ethical theory. As for our negotiation approach, the ethical utility function aligns with utilitarian ethics Mill (2016), Allen et al. (2005), which seeks to maximize the overall utility an minimize the harm according to the consequences of actions. Utilitarianism evaluates actions in terms of outcomes and an action is morally right if it promotes the greatest good for the greatest number Moor (2006). In our system, user-specified ethical profiles with ranked dispositions implement this principle by letting an individual decide how the robot ought to act according to their values and preferences in variable contexts. This ensures that the system is dynamic in adjusting and negotiating actions in a way that is compatible with the utilitarian goal of optimizing outcomes with respect to user-supplied utility metrics. Similar methods have been put forward in machine ethics, such as HERA-based systems Lindner et al. (2017), whereby utilitarian reasoning models actions based on their foreseen consequences and the setting of values guiding behavior through dynamic contexts makes the proposed system flexible and context-sensitive while keeping it within the principle of maximal goodness in most situations.

8.2. Social impact and trust

The prospect of fully autonomous systems is gradually transforming into reality. Our work embraces a vision of a digital world in which autonomous systems (or digital actors in general) and humans need to be in a better balance of forces. Negotiation among autonomous systems will not only safeguard the plurality of opinions by allowing for independent systems to negotiate towards a solution that respects the morals of all involved parties but also will allow for greater interoperability among systems that act on behalf of different stakeholders. In the context of negotiation, embedding user ethical principles into the decision-making processes of autonomous systems enhances their ability to dynamically achieve socially responsible outcomes. This is accomplished by ensuring that the negotiations conducted by these systems uphold individual ethical values, thus improving the trust of the system Abeywickrama et al. (2023), Thiebes et al. (2021), Kares et al. (2023) and paving the way for their integration in various domains Spiekermann (2023), Liscio et al. (2022), Aydoğan et al. (2021). In fact, a key aspect of the integration of autonomous systems into society is the ability to inspire trust, an aspect that cannot be considered a static relationship but rather a dynamic process influenced by ethical factors beyond technical and regulatory ones Lacher (2017b). Automated negotiation of ethical values can be regarded as a promising approach to realize such a dynamic process Aydoğan et al. (2021), Autili et al. (2025).

8.3. Ethics as a new target for negotiation

Most of the research in the negotiation literature Memon et al. (2025, 2023) so far has focused on the coordination and planning of agents that are either cooperative, working together to reach a shared goal Bachrach et al. (2020), Li et al. (2003), or selfish, competing with others to maximize their own utility Mansour (2020), Li et al. (2003). In addition, the recent progress of technologies in autonomous systems, including autonomous cars, unmanned aerial vehicles, and robots, has increased the demand for solving coordination and conflict avoidance in autonomous and self-interested agents that pursue their own objectives, prioritizing efficiency and safety Inotsume et al. (2020), Baarslag et al. (2016), Khemakhem et al. (2020), Kiruthika et al. (2020). However, when ethical considerations are introduced, these systems must account (i) not only for achieving their objectives through planning and coordination possibly maximizing their own revenues or interests in general, (ii) but also for aligning with the individual ethical preferences of the users they are acting on behalf of. This introduces constraints that differ from pure

self-fish utility maximization in existing negotiation approaches. For example, during a negotiation over resource conflict, an ethic-aware autonomous system may give precedence to the decision that satisfies its user's ethical beliefs even if this means forgoing the resource over obtaining it. Our work extends existing negotiation approaches by integrating these ethical implications into negotiation. This ensures that the negotiation outcomes are not primarily utility-driven but rather aligned with the ethical preferences of the participants involved.

8.4. Privacy considerations

A critical requirement for the proposed approach is the need to safeguard the privacy of ethical profiles since user's ethical preferences are sensitive information. For this reason, in the proposed architecture, ethical profiles (which summarize users' ethical beliefs) are never exchanged between the negotiating systems. Instead, during the negotiation rounds, only offers are exchanged and these are evaluated locally by each system. As mentioned in Section 4, the user's status condition is part of the context and these conditions are exchanged only to facilitate the activation of dispositions as a contextual factor, ensuring the approach's alignment with the context. Once the negotiation is concluded, each system destroys the exchanged offers for the sake of confidentiality. The system incorporates situational conditions into the offer exchange only if the user explicitly opts to include them, however, the user always holds the right to not disclose any of their conditions. However, the application of privacy-preserving methods, i.e., secure multiparty computation Zhao et al. (2019) and homomorphic encryption Acar et al. (2018) could mitigate the privacy-related concerns.

8.5. Architecture integrability

As mentioned in Section 2.3, the architecture presented in this paper is inspired by the works in Bremner et al. (2019), Vanderelst and Winfield (2018), where architectures for enabling ethical reasoning are proposed, by adding an *Ethical Layer* on top of the robot controller architecture. By following the same intuition, we built our RobEthiChor architecture as an additional layer that can be "mounted" on top of the native robot controller architecture. RobEthiChor is agnostic of the architecture that realizes the robot controller, provided that it follows (or can be transformed into) the Goal-Task-Action paradigm. This allows the robot's behavior to be extended with context-aware, ethics-based negotiation capabilities without directly altering it. As recalled in Sections 5 and 6, the integration of the architecture on top of the robot controller is realized by exploiting the interfaces provided by the Negotiation Manager component. For instance, RobEthiChor-Ros can be integrated with the existing robotic system ROS-based controller by exploiting the `NegotiationService` interface provided by the Negotiation Manager node. The latter can be called into play only when a decision upon a contended resource has to be made, without influencing the normal behavior of the controller in the other cases. In fact, the management of the negotiation outcome, i.e., the execution of the tasks involved in a resource contention, or their replanning (should the negotiation "leave" the resource to the other party), is fully managed by the Mission Controller (or some of its components abstracted within it) as part of the set of its "normal" capabilities. That is, before the negotiation starts and as soon as the negotiation ends, the control of the mission is held by the controller, which is in charge of executing the tasks and checking that the mission goal is reached.

9. Limitations and areas of improvement

In this section, we position our work with respect to the fields of study on multi-robot task allocation, collaborative robot teaming, and human-robot collaboration. Moreover, we also discuss some limitations and suggest potential directions to overcome them in future iterations.

9.1. Bilateral vs multilateral negotiation

Our work represents a first step toward bilateral, ethics-based negotiation over (possibly multiple[11]) tasks that have already been assigned to two robots, each acting on behalf of its respective user (i.e., one user per robot). The robots do not collaborate to accomplish shared tasks; rather, when they contend over a resource necessary for completing their individually assigned tasks, they collaborate to decide who gets to use it first. To make this decision, the two robots engage in a collaborative (not adversarial) negotiation, aiming to reach an agreement that, if possible, maximizes an ethical value mutually acceptable to both parties. Therefore, our work does not seek to address the broader challenges of collaboration between teams of robots and groups of humans. In our scenarios, each user interacts with their corresponding robot solely to provide input that may be used during the negotiation. Users do not intervene to modify task allocation, optimize task execution, or resolve conflicts, which are aspects more naturally related to the research fields of multi-robot task allocation, collaborative robot teaming, and human-robot collaboration. Allocation fairness and collaboration scalability are two other crucial aspects in those fields, which are out of the scope of our work, as we consider negotiation between only two robots at a time and do not consider the involvement of third parties. Accounting for multiple parties introduces a significantly more complex negotiation setting Yoo and Sim (2010). While our current approach focuses only on bilateral settings, we recognize the necessity of extending it to multilateral settings to comprehensively assess its scalability and effectiveness. Multilateral negotiations are increasingly relevant in real-world applications, where decisions must account for a wider set of users, their preferences, and constraints. We plan to propose an extension of our approach to accommodate multilateral negotiation scenarios, integrating mechanisms designed to balance diverse and potentially conflicting ethical preferences among multiple agents.

9.2. Time-dependent negotiation

Time is currently not considered in our negotiation process. However, incorporating time into the negotiation process can significantly affect decision-making and hence the negotiation outcome Dastjerdi and Buyya (2015), Mohammad (2023). Time-dependent negotiation enables systems to change strategies over time to optimize the offer generation. One of the possibilities of introducing time in our approach is that the systems follow the maximal strategy for offer generation when the deadline is approaching and share the whole set of user status conditions. This would avoid going beyond the time threshold, hence reducing the negotiation impact on the system performance and avoiding long-running negotiations as in the "extreme" cases described in Section 7.2. Alternatively, by incorporating time, another possibility is that the robots can (i) agree to a lower ethical utility when the deadline is closer or (ii) quit the negotiation with no agreement when the deadline is reached.

9.3. Context modeling

As described in Section 4, we adopt a key-value approach for modeling and reasoning on contexts. It is the simplest form of context representation and is easy to manage when the amount of data involved is small. However, key-value modeling is not scalable, unsuitable for dealing with complex, yet hierarchical, data structures, and attaching meta information is not possible. The key-value approach is often application-specific and suits the purpose of less complex application configurations and user preference specifications. Object-based, ontology-based, or logic-based

[11] Specifically, at each iteration, the negotiation can consider more than one task at a time and thus exchange offers that combine the ethical values associated with multiple tasks simultaneously (see Section 4.6).

modeling Gu et al. (2020), Hoyos et al. (2013), Bettini et al. (2010) would permit the modeling of contextual data using class hierarchy and complex relationships, organize them using semantic technologies, as well as specify inference rules by using logic formalisms, hence enabling more complex adaptations. This is beyond the scope of the present paper; the adoption of more powerful context modeling techniques and complex context retrieval algorithms, although desirable for more powerful adaptations, is part of future work.

9.4. Ethical profile modeling

The objective of this work was to show that negotiation can be an effective means to resolve situations of resource contention based on ethical preferences. In this sense, the way we adopted for specifying, modeling, and operationalizing ethical preferences may result limited and overfitting to our specific approach. In fact, the choice of adopting degrees of importance to dispositions represented as textual description, inspired by the work in Dennis et al. (2016, 2021), Alidoosti et al. (2025), was driven by the need to both operationalize ethical values and express them in a way that is understandable to end users. Hence, it can neither be considered as a generalizable method nor intended as a contribution on its own. To overcome this limitation, in the future we foresee the use of a language which, while remaining accessible to end users, should also be structured, unambiguous, and free from inconsistencies. Therefore, the use of well-established approaches already in place, such as the one proposed in Feng et al. (2023), Yaman et al. (2025), may represent an appropriate path forward. This, indeed, together with a deeper formal specification of the entire approach, would also enable the automatic synthesis of negotiation strategies along with the formal verification of the negotiation process overall. An overview of current approaches for modeling ethical values in software systems is reported in Section 11.

10. Threats to validity

In the following, we discuss the main threats that could hinder the validity of our findings.

The first threat concerns the internal validity of our findings in the scope of *EQ1*. We defined the ethical profiles of the 10 personas by considering a limited set of dispositions (12), which were ranked from 1 to 5 according to the envisioned preferences of each persona. While this allowed us to build a ground truth for evaluating the negotiation results, this could, in principle, represent a limitation due to possible biases in the given ranks. Moreover, a manual procedure was employed to construct the ground truth. As with all kinds of manual processes, mistakes might have occurred. To mitigate this threat, the ground truth construction was performed by a researcher, while the results were checked by a second researcher who performed this task independently, hence building confidence in the correctness of the obtained ground truth. To allow for independent verification of obtained results, the ground truth and the collected experimental data are made available in the replication package. Finally, by relying on values from 1 to 5 for the ranking, we introduced a limited variance in the given ranks, which could have, in turn, reduced the percentage of reached agreements due to the limited difference between the ethical impacts of the exchanged offers. We believe that the limited set of dispositions, the possible bias in the given rank, and their low variance are mathematical considerations that, although valid, (i) are distinct from the core ability of our approach to offer a possible way (not the best and not the only possible one) of encoding ethical preferences, and (ii) do not compromise the computational correctness of the approach in reaching a situational agreement that aligns with the ethical preferences as graded by users within the ethical profiles.

Another key limitation of the evaluation concerns the focus on the system-level implementation rather than the reference architecture itself. The evaluation we conducted can be characterized as an implementation-based validation, aimed at assessing the practical applicability of RobEthiChor through the implementation provided in RobEthiChor-Ros, which allowed us to experiment over a realistic robotic setting. However, to mitigate this, we ensured that the implementation aligns with the reference architecture. Moreover, we acknowledge that this form of indirect validation primarily reflects the properties of the RobEthiChor-Ros and it does not consider other properties such as reusability, generalizability, or modifiability. An important direction for future work is to perform a direct evaluation of the reference architecture via expert review and quality attribute analysis, to ensure that the architecture is robust and appropriate for its intended real-world use. This calls for having ethics as the critical non-functional attribute, together with its functional negotiation ability. This is an important step that deserves a paper specifically dedicated to the reference architecture, which we have planned for the near future.

A threat to the external validity of our work concerns the unavailability of real-world ethical profiles, collected from real users, that would permit us to also evaluate end-users' acceptance and social impact. To address this limitation, we are planning to conduct social experiments to evaluate the acceptance of the proposed approach by end users and the potential social impact that the approach could have. However, the work and the experimentation presented in this paper serve to elevate the technical readiness level of the proposed solution, demonstrating its soundness, and are a preparatory step for the planned social experiment.

A second threat of this kind would concern the lack of a comparison of our results with the ones proposed by other studies and/or benchmarks. However, to the best of our knowledge, the literature does not propose any work that leverages the ethical profiles of end users to regulate autonomous systems' behavior and to enable ethics-based decision-making through negotiation.

Finally, during the experiment execution, uncontrolled factors, such as network latency, can potentially impact the accuracy of measurements. To mitigate this threat, in the *off-robot* deployment, we ensured that no background tasks were active on the device that hosted RobEthiChor during the experiment execution, save for the operating system. Moreover, in both the *on-robot* and the *off-robot* setting, all the experiments were performed in a controlled environment, where the network latency was stable with low variability. While this allowed us to compare the results obtained with different settings, in principle this could hinder the generality of our findings, as the collected measurements do not account for potential network disruptions that could manifest in real-world deployments. However, it is worth noticing that the measured network latency between the robots aligns with the QoS required for ultra-low latency communications offered by wireless networks (especially 5G) for Industry 4.0 and Vehicle-to-everything (V2X) scenarios Sefati and Halunga (2023). Hence, the experimental network conditions closely align with the ones expected for the deployment of future-generation autonomous systems.

11. Related work

This section discusses studies from varied research areas, each related to different aspects of our work.

11.1. Ethics in automated decision-making

Several studies have emphasized the importance of developing autonomous systems that are capable of ethical reasoning and discussed the practical and technical challenges of developing such systems Alidoosti et al. (2022), Moor (2006), Allen et al. (2006), Antsaklis et al. (1989), Tolmeijer et al. (2020), Huang et al. (2022), De Sanctis and Inverardi (2024), Inverardi (2019a). However, the notion of integrating user ethics for the ethical conduct of such systems has remained largely unexplored. The study in Allen et al. (2006) highlights that it is not enough to adhere to an explicit ethical theory to reason as humans. Rather, it is about how to enable machines to think in a way

that is consistent with human ethical behavior, so as to guarantee their ethical conduct. Similarly, the studies in Tolmeijer et al. (2020), Nallur (2020) provide an overview of machine ethics in autonomous systems. The studies review various ethical theories and highlights the challenge of selecting an appropriate ethical theory. Moreover, the studies address the fact that human morality is complex and cannot be captured by a single ethical theory, highlighting the need to propose methods to capture and integrate user ethics into the system.

Correspondingly, the study in De Sanctis and Inverardi (2024) introduces the concept of ethical-aware Collective Adaptive Systems (CASs) and discusses the challenges in designing these systems to respect human ethical values, including privacy and human dignity. The authors emphasize the need for the development of systems that not only align with hard ethics (laws and regulations) Floridi (2018), but also align with soft ethics (individuals' preferences) Floridi (2018). However, the study considers human ethics only as privacy and human dignity while our work focuses on user's contextual ethical preferences that vary with contexts.

In Liao et al. (2019, 2023), authors propose the architecture of an artificial moral agent that considers the moral values of different stakeholders to make an ethical decision. The agent decides by combining various moral values into a single ethical theory, utilizing a rule-based approach to reach an agreement. Similarly to our work, the study assumes that the classification of moral values is given and the agent utilizes them to make a collaborative decision that leads to agreement among them. However, the concrete definition of an approach, its concrete architecture, and implementation for robots to consider user ethical preferences to adjust Mostafa et al. (2019) their autonomy, is not considered.

The study in Cardoso et al. (2021) implements an ethical reasoner for decision-making. The ethical reasoner follows a predetermined ethical theory, and actions the system can undertake are ranked based on their adherence to the ethical theory. However, this study does not consider users' morality as part of decision-making, as the system follows the built-in ethical principles established by the system designers. Another approach to ethical decision-making is proposed in Bremner et al. (2019), Winfield et al. (2014), Winfield and Hafner (2019). In these works, the authors adopt Asimov's laws of robotics Asimov (1941) as the ethical theory of choice and incorporate them into the robot planner module that generates actions. The robots simulate each action to predict the ethical implications of its consequences based on the ethical theory and hence select the action with the most ethical consequence. However, also in this case, personal users' morals are not considered.

A framework for verifying ethical properties for autonomous systems is presented in Dennis et al. (2016). According to this framework, ethical principles are encoded as a set of rules ranked by importance. Differently from us, this work focuses on verification. Moreover, as in the studies above, the ethical principles have to be decided by the system designers, and the possibility of interaction among multiple ethical systems based on user preferences is not considered. Similarly, the study in Dennis et al. (2021) proposes a framework that verifies the context-dependent ethical reasoning capabilities of autonomous systems as the ethical values of stakeholders change with context. This study introduces context guard formulas, logical conditions that enable the system to dynamically shift between different ethical theories (such as utilitarian or Kantian ethics) based on environmental or situational changes. Through a smart home scenario, the authors implement the approach and verify the system's decision, ensuring its context-dependent behaviors satisfy safety-critical properties. In contrast, our work extends this by proposing a framework that enables systems to make decisions that satisfy the user's ethical preferences in a given context, beyond encoding fixed ethical theories. In Townsend et al. (2022), the authors present an approach for translating abstract high-level normative principles into explicitly formulated practical rules. This ensures that autonomous systems align social, legal, ethical, empathetic, and cultural (SLEEC) rules, thus bridging the gap between abstract high-level principles and operational practice. These studies further introduce methods to identify conflicts and semantic relationships between these rules Feng et al. (2024a), Yaman et al. (2025), Feng et al. (2024b, 2023). However, the normative principles introduced in these studies are drawn from different stakeholders and not tailored to the single user. Unlike this, our approach focuses on the user's ethical values to tailor the decision-making according to the user's beliefs.

A different perspective is presented in the studies in Alidoosti (2021), Alidoosti et al. (2022) that focus on the integration of ethical values into the system at the architectural level. The studies emphasize the design of a software architecture that is aligned with individual and social ethical values. However, the studies emphasize the need for the extraction of ethical values, given that the quantification of these values is personal, and it is challenging to measure such values. Moreover, the notion of ethics is based on ethical theories and different value scales, rather than incorporating user ethical preferences into such systems, which is the main focus of our work. Similarly, the study in Alidoosti et al. (2025) provides a review on ethical values of stakeholders to integrate in software architecture. However, the study identifies the need to operationalize ethical values and address conflicts arising from different stakeholder perspectives. Our work addresses this by proposing an approach that operationalizes users' soft ethics through ethical profiles and integrates them into the system architecture. Moreover, conflicts among ethical values can be resolved by reaching agreements through ethics-based negotiation.

11.2. Modeling human ethical values

Human values represent perceptions, attitudes, behaviors, and preferences of individuals and groups Mougouei et al. (2018), Shahin et al. (2022a), Bardi et al. (2009). They play an integral role in the behavior of autonomous systems that are designed to interact with or impact individuals and society Han et al. (2022). Software practitioners usually design software systems focusing on their interpretation of functional and non-functional requirements, overlooking human values Curumsing et al. (2019). As a result, the systems may not satisfy user preferences. Hence, it becomes essential to design systems that align with human values De Sanctis and Inverardi (2024). The notion of values has been articulated in different ways. Certain studies focus on the introduction of human ethical values in the context of software architecture design Alidoosti et al. (2022, 2025), Shahin et al. (2022b), Mougouei et al. (2018), Mougouei (2020), Perera et al. (2020, 2019), Nurwidyantoro et al. (2023) while other studies focus on various types of values Alidoosti et al. (2023), Ferrario et al. (2016), Liscio et al. (2022), Rokeach (1973), Maio (2016). Among the other existing frameworks Bird and Waters (1987), Jurkiewicz and Giacalone (2004), Graham et al. (2013), Schwartz's theory of basic values Schwartz (2012) has been widely adopted in various disciplines to represent human values on a measurable scale, although it models human values into specific categories only. Value-sensitive design (VSD) is another widely accepted approach to extracting human values using scenarios and storyboarding Friedman et al. (2013). The VSD framework encompasses a wide range of values relevant to various design tasks Friedman et al. (2013). Similarly, other models have been suggested throughout the years Allport (1961), Feather (1995), Hofstede (2011), Rokeach (1973). However, translating abstract values into specific context-dependent user preferences to operationalize them is challenging Le Dantec et al. (2009), Pommeranz et al. (2012), Saket (2021), Mougouei et al. (2018).

Recent attempts have focused on modeling users' ethical values through surveys. The survey in Alfieri et al. (2022) is based on the correlations between idealism and relativism, Machiavellianism, and normative theories. The clustering approach is then used to create ethical profiles from the gathered data that can predict the digital behaviors of users with respect to privacy breaches, copyright violations, and protection. In Alfieri et al. (2023), a scenario-driven method is presented in the form of a questionnaire to collect data on the moral preferences of users in the digital world through ethically charged scenarios.

While the studies in the literature offer essential insights into the identification and elicitation of human values, these approaches consider a static set of values, overlooking the ethical preferences of the individuals, which vary across different contexts. For example, the precedence one gives to an elderly person may vary based on situational factors, e.g., location, time, etc. Thus, a significant challenge persists in quantifying user's contextual ethical preferences. Quantifying such preferences remains a key challenge as ethical values can be personal and there are no standards to measure and prioritize them Alidoosti et al. (2022).

Our work was inspired by these studies to introduce the concept of user ethics through ethical profiles, where users provide grades to each disposition to express their preferences within a given context, enabling the quantification of their preferences.

11.3. Negotiation approaches

Automated negotiation integrates principles from three different fields into one, namely, game theory, economics, and artificial intelligence, to provide a multidisciplinary approach to address complex automated decision-making challenges Baarslag et al. (2016). Its significance lies in intelligent agents[12] that negotiate on behalf of humans Baarslag et al. (2022) and are likely to be more efficient An et al. (2016). Subsequently, the study in Kiruthika et al. (2020) defines various stages of the life cycle of a negotiating agent and explains how finalizing the agents' attributes and preferences before initiating the negotiation helps them evaluate and guide their decisions throughout the process Chen and Weiss (2019), Chang and Fujita (2024).

The studies in Baarslag et al. (2017), Filipczuk et al. (2022) propose an approach to negotiate permission to access users' private data to online services. The agent exchanges permission with an incentive for monetary reward, taking into account the user's preferences obtained from previous interactions with online services and their feedback. Moreover, the study in Magra et al. (2023) proposes a querying approach aimed at understanding user preferences, in which the agent asks the users to evaluate two different options, such as sharing their browsing history with online services in exchange for varied pricing, to learn their preferences. Similarly, the authors in Baarslag and Kaisers (2017) suggest a method to elicit user preferences to negotiate electricity prices. It presents a decision model that asks the users for their preferences in terms of price and learns their preferences. However, unlike our approach, the proposed approaches retrieve user preferences in terms of price and do not focus on ethical preferences that could influence the system to make an ethical decision during negotiation.

The study in Baarslag et al. (2016) provides a survey of techniques that can be used for opponent modeling during negotiation, and methods to choose the right bidding strategy using machine learning are given in Kell et al. (2022). Furthermore, the studies in Braun et al. (2006), Eshragh et al. (2015), Kell et al. (2022), Hassanvand and Nassiri-Mofakham (2023), Keskin et al. (2023), Mansour et al. (2022), Luo et al. (2024b), Lin et al. (2023), Kumar et al. (2023) discuss the approaches that autonomous systems use to negotiate with humans and autonomous agents for different purposes e.g., conflict resolution, resource allocation, economics, etc. None of the above considers ethics in decision-making as a result of automated negotiation, which is the main focus of our study.

The notion of context has been used in Kröhling et al. (2019, 2021), where agents take into account contextual factors during negotiation. However, the notion of context is limited to factors such as the negotiation deadline, reserved and preferred prices, and current market price trends. In contrast, our work considers contextual elements such as location, time, and user status. During negotiation, these factors lead users to have varying preferences in different contexts.

In addition, negotiation based on rule-based strategies has also been discussed in the literature Chen and Weiss (2019), Khemakhem et al. (2020), Bădică et al. (2011), Ganzha and Paprzycki (2007), Mahan et al. (2011), Hu and Deng (2011) where agents adhere to predefined rules to negotiate. Specifically, in auction-based negotiations, agents learn through interactions with other agents and update their bidding rules, clearing policies that govern resource allocation, and information disclosure policy, which depict information to be shared with participating agents Wurman et al. (2002). However, none of the studies consider the quantification of user ethical preferences and the formalization of negotiation rules based on these preferences.

12. Conclusion

In this work, we proposed a novel context-aware ethics-based negotiation approach in which autonomous robots utilize their user's soft ethical preferences, contextual factors, and user status conditions to adjust their autonomy while negotiating with other robots to resolve the conflict and (possibly) reach a contextual agreement. We presented a domain-agnostic reference architecture supporting the approach and RobEthiChor, its implementation on top of the Robot Operating System (ROS) for the robotic domain. We ran multiple negotiation scenarios between robots acting on behalf of users with different ethics profiles to assess the effectiveness of the approach, evaluate the overhead introduced by the negotiation, and assess its scalability. The results show that the approach is effective in allowing robots to negotiate on behalf of their users with negligible to acceptable overhead.

Future work concerns the realization of social experiments to evaluate the acceptance by end users and the potential social impact that the approach could have, and developing more comprehensive methods to model the context, incorporating a wider range of contextual factors. Additionally, we aim to predict user behavior within a given context by leveraging their preferences in similar contexts, employing a context hierarchy. This approach involves organizing contextual factors into a structured framework that captures the interrelationships between various contexts and their influence on user behavior. By analyzing patterns in similar contexts, the system can make informed predictions about the likely user behaviors, allowing for more adaptive and responsive decision-making.

Finally, we plan to generalize our approach to a multilateral negotiation setting, where more than two systems can be involved in the negotiation and introduce time constraints, which will enable the system to dynamically adapt its strategies, hence optimizing decisions in real-time while accounting for evolving contexts and temporal limitations.

CRediT authorship contribution statement

Mashal Afzal Memon: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Gianluca Filippone:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Gian Luca Scoccia:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Marco Autili:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization; **Paola Inverardi:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

[12] The terms "system" and "agent" are used interchangeably as an agent is a type of system that acts autonomously and makes decisions on behalf of users Wooldridge and Jennings (1995).

Data availability

Replication package link is provided within the article text.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been partially supported by the (i) Spoke 1 (CUP: I53C22000690001) and the Spoke 9 “Digital Society & Smart Cities” of ICSC - Centro Nazionale di Ricerca in High Performance-Computing, Big Data and Quantum Computing, funded by the European Union - NextGenerationEU (PNRR-HPC, CUP: E13C22001000006), (ii) the MUR (Italy) – PRIN PNRR 2022 project “RoboChor: Robot Choreography” (grant P2022RSW5W), (iii) the MUR (Italy) Department of Excellence 2023–2027, (iv) the PRIN project “HALO: etHical-aware AdjustabLe autonomous systems” (grant 2022JKA4SL), (v) the MUR (Italy) – PRIN 2022 project “CAVIA: enabling the Cloud-to-Autonomous-Vehicles continuum for future Industrial Applications” (grant 2022JAFATE), and (vi) the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041 - VITALITY – CUP: D13C21000430001.

References

Abeywickrama, D.B., Bennaceur, A., Chance, G., et al., 2023. On specifying for trustworthiness. Commun. ACM 67 (1), 98–109.
Acar, A., Aksu, H., Uluagac, A.S., et al., 2018. A survey on homomorphic encryption schemes: theory and implementation. ACM Comput. Surv. 51 (4), 1–35.
Alfieri, C., Donati, D., Gozzano, S., et al., 2023. Ethical preferences in the digital world: the EXOSOUL questionnaire. In: 2Nd International Conference on Hybrid Human-Artificial Intelligence, pp. 290–299.
Alfieri, C., Inverardi, P., Migliarini, P., et al., 2022. Exosoul: ethical profiling in the digital world. In: 1St International Conference on Hybrid Human-Artificial Intelligence, pp. 128–142.
Alidoosti, R., 2021. Ethics-driven software architecture decision making. In: 18Th International Conference on Software Architecture Companion, pp. 90–91.
Alidoosti, R., De Sanctis, M., Iovino, L., et al., 2023. Stakeholder inclusion and value diversity: an evaluation using an access control system. In: 17Th European Conference on Software Architecture, pp. 1–9.
Alidoosti, R., Lago, P., Poort, E., et al., 2022. Incorporating ethical values into software architecture design practices. In: 19Th International Conference on Software Architecture, pp. 124–127.
Alidoosti, R., Lago, P., Razavian, M., et al., 2025. Exploring ethical values in software systems: a systematic literature review. J. Syst. Softw. 226, 112430. https://doi.org/10.1016/j.jss.2025.112430
Allen, C., Smit, I., Wallach, W., 2005. Artificial morality: top-down, bottom-up, and hybrid approaches. Ethics Inf. Technol. 7, 149–155.
Allen, C., Wallach, W., Smit, I., 2006. Why machine ethics? IEEE Intell. Syst. 21 (4), 12–17.
Allport, G.W., 1961. Pattern and growth in personality .
An, B., Gatti, N., Lesser, V., 2016. Alternating-offers bargaining in one-to-many and many-to-many settings. Ann. Math. Artif. Intell. 77 (1), 67–103.
Anderson, J., Rainie, L., Luchsinger, A., 2018. Artificial intelligence and the future of humans. Pew Research Center 10 (12), 1–10.
Anderson, M., Anderson, S.L., 2020. Machine ethics: creating an ethical intelligent agent. In: Machine Ethics & Robot Ethics. Routledge, pp. 237–248.
Anjum, R.L., Lie, S. A.N., Mumford, S., 2013. Dispositions and ethics. In: Powers and Capacities in Philosophy. Routledge, pp. 231–247.
Antsaklis, P.J., Passino, K.M., Wang, S.J., 1989. Towards intelligent autonomous control systems: architecture and fundamental issues. J. Intell. Rob. Syst. 1 (4), 315–342.
Arcaini, P., Mirandola, R., Riccobene, E., et al., 2020. Smart home platform supporting decentralized adaptive automation control. In: Proceedings of the 35Th Annual ACM Symposium on Applied Computing. Association for Computing Machinery, New York, NY, USA, p. 1893–1900.
Arcaini, P., Riccobene, E., Scandurra, P., 2015. Modeling and analyzing MAPE-k feedback loops for self-adaptation. In: 2015 Del Ins Thinspace IEEE/ACM 10Th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, pp. 13–23. https://doi.org/10.1109/SEAMS.2015.10
Asimov, I., 1941. Three laws of robotics. Asimov, I. Runaround .
Askarpour, M., Tsigkanos, C., Menghi, C., et al., 2021. RoboMAX: robotic mission adaptation exemplars. In: 2021 International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), pp. 245–251. https://doi.org/10.1109/SEAMS51251.2021.00040
Autili, M., De Sanctis, M., Inverardi, P., et al., 2025. Engineering digital systems for humanity: a research roadmap. ACM Trans. Softw. Eng. Methodol. . https://doi.org/10.1145/3712006
Autili, M., Di Ruscio, D.D., Inverardi, P., et al., 2019. A software exoskeleton to protect and support citizen’s ethics and privacy in the digital world. IEEE Access 7, 62011–62021.
Awad, E., Dsouza, S., Kim, R., et al., 2018. The moral machine experiment. Nature 563 (7729), 59–64.
Aydoğan, R., Kafali, Ö., Arslan, F., et al., 2021. Nova: value-based negotiation of norms. ACM Trans. Intell. Syst. Technol. (TIST) 12 (4), 1–29.
Baarslag, T., Alper, A., Gomer, R., et al., 2017. An automated negotiation agent for permission management. In: 16Th International Conference on Autonomous Agents and MultiAgent Systems, pp. 380–390.
Baarslag, T., Hendrikx, M. J.C., Hindriks, K.V., et al., 2016. Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques. Auton. Agent Multi Agent Syst. 30 (5), 849–898.
Baarslag, T., Kaisers, M., 2017. The value of information in automated negotiation: a decision model for eliciting user preferences. In: 16Th International Conference on Autonomous Agents and MultiAgent Systems, pp. 391–400.
Baarslag, T., Kaisers, M., Gerding, E.H., et al., 2022. Self-sufficient, self-directed, and interdependent negotiation systems: a roadmap toward autonomous negotiation agents. In: Bargaining: Current Research and Future Directions. Springer, pp. 387–406.
Bachrach, Y., Everett, R., Hughes, E., et al., 2020. Negotiating team formation using deep reinforcement learning. Artif. Intell. 288, 103356.
Bagga, P., Paoletti, N., Stathis, K., 2022. Deep learnable strategy templates for multi-issue bilateral negotiation. In: 21St International Conference on Autonomous Agents and Multiagent Systems, pp. 1533–1535.
Baltazar, A.R., Petry, M.R., Silva, M.F., et al., 2021. Autonomous wheelchair for patient’s transportation on healthcare institutions. SN Appl. Sci. 3 (3), 354. https://doi.org/10.1007/s42452-021-04304-1
Bardi, A., Lee, J.A., Hofmann-Towfigh, N., et al., 2009. The structure of intraindividual value change. J. Pers. Soc. Psychol. 97 (5), 913.
Bauer, C., Novotny, A., 2017. A consolidated view of context for intelligent systems. J. Ambient Intell. Smart Environ. 9 (4), 377–393.
Bennaceur, A., Hassett, D., Nuseibeh, B., et al., 2023. Values@runtime: an adaptive framework for operationalising values. In: 45Th IEEE International Conference on Software Engineering: Software Engineering in Society, pp. 175–179.
Bettini, C., Brdiczka, O., Henricksen, K., et al., 2010. A survey of context modelling and reasoning techniques. Pervasive Mob. Comput. 6 (2), 161–180.
Bird, F., Waters, J.A., 1987. The nature of managerial moral standards. J. Bus. Ethics 6, 1–13.
Bogosian, K., 2017. Implementation of moral uncertainty in intelligent machines. Minds Mach. 27 (4), 591–608.
Boltz, N., Getir Yaman, S., Inverardi, P., et al., 2024. Human empowerment in self-adaptive socio-technical systems. In: 19Th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, pp. 200–206.
Bostrom, N., Yudkowsky, E., 2018. The ethics of artificial intelligence. In: Artificial Intelligence Safety and Security, pp. 57–69.
Brambilla, A., Lavagna, M., et al., 2008. A coordination mechanism to solve common resource contention in multi agent space systems. In: International Symposium on Artificial Intelligence, Robotics and Automation in Space, pp. 26–29.
Braun, P., Brzostowski, J., Kersten, G., et al., 2006. E-Negotiation systems and software agents: methods, models, and applications. In: Intelligent Decision-making Support Systems. Springer, pp. 271–300.
Bremner, P., Dennis, L.A., Fisher, M., et al., 2019. On proactive, transparent, and verifiable ethical reasoning for robots. Proc. IEEE 107 (3), 541–561.
Budd, L., Ison, S., 2020. Supporting the needs of special assistance (including PRM) passengers: an international survey of disabled air passenger rights legislation. J. Air Transport Manage. 87, 101851. https://doi.org/10.1016/j.jairtraman.2020.101851
Bădică, C., Braubach, L., Paschke, A., 2011. Rule-based distributed and agent systems. In: Rule-Based Reasoning, Programming, and Applications, pp. 3–28.
Caesar, B., Valdezate, A., Ladiges, J., et al., 2023. A process for identifying and modeling relevant system context for the reconfiguration of automated systems. IEEE Trans. Autom. Sci. Eng. , 1–26.
Cámara, J., Bellman, K.L., Kephart, J.O., et al., 2017. Self-aware computing systems: related concepts and research areas. In: Self-Aware Computing Systems., pp. 17–49.
Cardoso, R.C., Ferrando, A., Dennis, L.A., et al., 2021. Implementing ethical governors in BDI. In: WS on Eng. Multi-Agent Systems, pp. 22–41.
Chang, S., Fujita, K., 2024. Comb: scalable concession-driven opponent models using bayesian learning for preference learning in bilateral multi-issue automated negotiation. Group Decis. Negotiat. , 1–48.
Chen, K.E., Liang, Y., Jha, N., et al., 2021. FogROS: an adaptive framework for automating fog robotics deployment. In: 2021 Del Ins Thinspace IEEE 17Th International Conference on Automation Science and Engineering (CASE), pp. 2035–2042. https://doi.org/10.1109/CASE49439.2021.9551628
Chen, S., Weiss, G., 2019. Automated negotiation: an efficient approach to interaction among agents. In: Interactions in Multiagent Systems. World Scientific, pp. 149–177.
Curumsing, M.K., Fernando, N., Abdelrazek, M., et al., 2019. Emotion-oriented requirements engineering: a case study in developing a smart home system for the elderly. J. Syst. Softw. 147, 215–229.
Dastjerdi, A.V., Buyya, R., 2015. An autonomous time-dependent SLA negotiation strategy for cloud computing. Comput. J. 58 (11), 3202–3216.
De Sanctis, M., Inverardi, P., 2024. Engineering ethical-aware collective adaptive systems. In: International Symposium on Leveraging Applications of Formal Methods. Springer, pp. 238–252.
Dennis, L., Fisher, M., Slavkovik, M., et al., 2016. Formal verification of ethical choices in autonomous systems. Rob. Auton. Syst. 77, 1–14.

Dennis, L.A., Bentzen, M.M., Lindner, F., et al., 2021. Verifiable machine ethics in changing contexts. In: AAAI Conference on Artificial Intelligence, pp. 11470–11478.
Dey, A.K., 2001. Understanding and using context. Pers. Ubiquitous Comput. 5, 4–7.
Donati, D., 2024. Trust, trustworthiness and the moral dimension in human-AI interactions. Ethics, Polit. Soc. 7 (2).
Donati, D., Assadi, Z., Gozzano, S., et al., 2024. On representing humans' soft-ethics preferences as dispositions. In: Proceedings of the Ital-IA Intelligenza Artificiale. Vol. 3762 of *CEUR Workshop Proceedings*, pp. 135–140.
Donati, D., Inverardi, P., Melis, B., Pelliccione, P., 2025. Beyond the checklist: rethinking trustworthiness in AI system. In: 2025 Del Ins Thinspace IEEE 33Rd International Requirements Engineering Conference Workshops (REW). IEEE, pp. 468–474.
Eshragh, F., Pooyandeh, M., Marceau, D.J., 2015. Automated negotiation in environmental resource management: review and assessment. J. Environ. Manage. 162, 148–157.
Feather, N.T., 1995. Values, valences, and choice: the influences of values on the perceived attractiveness and choice of alternatives. J. Pers. Soc. Psychol. 68 (6), 1135.
Feng, N., Marsso, L., Getir Yaman, S., et al., 2024a. Analyzing and debugging normative requirements via satisfiability checking. In: 46Th IEEE/ACM International Conference on Software Engineering, pp. 1–12.
Feng, N., Marsso, L., Yaman, S.G., et al., 2023. Towards a formal framework for normative requirements elicitation. In: 38Th IEEE/ACM International Conference on Automated Software Engineering, pp. 1776–1780.
Feng, N., Marsso, L., Yaman, S.G., et al., 2024b. Normative requirements operationalization with large language models. In: 32Nd IEEE International Requirements Engineering Conference, pp. 129–141.
Ferrario, M.A., Simm, W., Forshaw, S., et al., 2016. Values-first SE: research principles in practice. In: 38Th International Conference on Software Engineering Companion, pp. 553–562.
Filipczuk, D., Baarslag, T., Gerding, E.H., et al., 2022. Automated privacy negotiations with preference uncertainty. Auton. Agent Multi Agent Syst. 36 (2), 49.
Filippone, G., Autili, M., Tivoli, M., 2022. Synthesis of context-aware business-to-business processes for location-based services through choreographies. J. Softw.: Evol. Process 34 (10), e2416.
Floridi, L., 2018. Soft ethics and the governance of the digital. Philos. Technol. 31 (1), 1–8.
Floridi, L., 2019. Establishing the rules for building trustworthy AI. Nat. Mach. Intell. 1 (6), 261–262.
Friedman, B., Kahn, P.H., Borning, A., et al., 2013. Value sensitive design and information systems. Early Engage. New Technol.: Opening Up Lab. , 55–95.
Ganzha, M., Paprzycki, M., 2007. Implementing rule-based automated price negotiation in an agent system. J. Univ. Comput. Sci. 13 (2), 244–266.
Gerostathopoulos, I., Pournaras, E., 2019. Trapped in traffic? a self-adaptive framework for decentralized traffic optimization. In: 2019 Del Ins Thinspace IEEE/ACM 14Th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS). IEEE Computer Society, pp. 32–38. https://doi.org/10.1109/SEAMS.2019.00014
Graham, J., Haidt, J., Koleva, S., et al., 2013. Moral foundations theory: the pragmatic validity of moral pluralism. In: Advances in Experimental Social Psychology. Elsevier. Vol. 47, pp. 55–130.
Gu, T., Wang, X.H., Pung, H.K., et al., 2020. An ontology-based context model in intelligent environments. arXiv preprint arXiv:2003.05055 .
Hajaj, C., Hazon, N., Sarne, D., 2017. Enhancing comparison shopping agents through ordering and gradual information disclosure. Auton. Agent Multi Agent Syst. 31, 696–714.
Han, S., Kelly, E., Nikou, S., et al., 2022. Aligning artificial intelligence with human values: reflections from a phenomenological perspective. AI Soc. , 1–13.
Hassanvand, F., Nassiri-Mofakham, F., 2023. Automated negotiation agents in modeling gaussian bidders. AUT J. Model. Simul. 55 (1), 3–16.
Heunis, H., Pulles, N.J., Giebels, E., et al., 2024. Strategic adaptability in negotiation: a framework to distinguish strategic adaptable behaviors. Int. J. Conflict Manage. 35 (2), 245–269.
Hofstede, G., 2011. Dimensionalizing cultures: the hofstede model in context. Online Read. Psychol. Cult. 2 (1), 8.
Hong, J.-y., Suh, E.-h., Kim, S.-J., 2009. Context-aware systems: a literature review and classification. Expert Syst. Appl. 36 (4), 8509–8522.
Hoyos, J.R., García-Molina, J., Botía, J.A., 2013. A domain-specific language for context modeling in context-aware systems. J. Syst. Softw. 86 (11), 2890–2905.
Hsu, P.E., Hsu, Y.L., Chang, K.W., et al., 2012. Mobility assistance design of the intelligent robotic wheelchair. Int. J. Adv. Rob. Syst. 9 (6), 244. https://doi.org/10.5772/54819
Hu, J., Deng, L., 2011. An association rule-based bilateral multi-issue negotiation model. In: The 4Th International Symposium on Computational Intelligence and Design. Vol. 2, pp. 234–237.
Huang, C., Zhang, Z., Mao, B., et al., 2022. An overview of artificial intelligence ethics. IEEE Trans. Artif. Intell. 4 (4), 799–819.
Iftikhar, M.U., Ramachandran, G.S., Bollansée, P., et al., 2017. Deltaiot: a self-adaptive internet of things exemplar. In: 2017 Del Ins Thinspace IEEE/ACM 12Th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), pp. 76–82. https://doi.org/10.1109/SEAMS.2017.21
Inotsume, H., Aggarwal, A., Higa, R., et al., 2020. Path negotiation for self-interested multirobot vehicles in shared space. In: IEEE International Conference on Intelligent Robots and Systems, pp. 11587–11594.
Inverardi, P., 2019a. Ethics and privacy in autonomous systems: a software exoskeleton to empower the user. In: Software Engineering for Resilient Systems: 11Th International Workshop., pp. 3–8.
Inverardi, P., 2019b. The european perspective on responsible computing. Commun. ACM 62 (4), 64.
Inverardi, P., Migliarini, P., Palmiero, M., 2023. Systematic review on privacy categorisation. Comput. Sci. Rev. 49, 100574.
Inverardi, P., Palmiero, M., Pelliccione, P., et al., 2022. Ethical-aware autonomous systems from a social psychological lens. In: 6Th International WS on Cultures of Participation in the Digital Age: AI for Humans or Humans for AI?, pp. 43–48.
Jamshidi, P., Camara, J., Schmerl, B., et al., 2019. Machine learning meets quantitative planning: enabling self-adaptation in autonomous robots. In: 2019 Del Ins Thinspace IEEE/ACM 14Th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS). IEEE Computer Society, pp. 39–50. https://doi.org/10.1109/SEAMS.2019.00015
Jedlickova, A., 2024. Ensuring ethical standards in the development of autonomous and intelligent systems. IEEE Trans. Artif. Intell. 5 (12), 5863–5872.
Jiang, Y., Yedidsion, H., Zhang, S., et al., 2019. Multi-robot planning with conflicts and synergies. Auton. Robots 43 (8), 2011–2032. https://doi.org/10.1007/S10514-019-09848-1
Jurkiewicz, C.L., Giacalone, R.A., 2004. A values framework for measuring the impact of workplace spirituality on organizational performance. J. Bus. Ethics 49, 129–142.
Kares, F., König, C.J., Bergs, R., et al., 2023. Trust in hybrid human-automated decision-support. Int. J. Sel. Assess. .
Kehoe, B., Patil, S., Abbeel, P., et al., 2015. A survey of research on cloud robotics and automation. IEEE Trans. Autom. Sci. Eng. 12 (2), 398–409. https://doi.org/10.1109/TASE.2014.2376492
Kell, A. J.M., McGough, A.S., Forshaw, M., 2022. A systematic literature review on machine learning for electricity market agent-based models. In: IEEE International Conference on Big Data (Big Data), pp. 4503–4512.
Kephart, J.O., Chess, D.M., 2003. The vision of autonomic computing. Computer (Long Beach Calif) 36 (1), 41–50. https://doi.org/10.1109/MC.2003.1160055
Keskin, M.O., Buzcu, B., Aydoğan, R., 2023. Conflict-based negotiation strategy for human-agent negotiation. Appl. Intell. 53 (24), 29741–29757.
Khemakhem, F., Ellouzi, H., Ltifi, H., et al., 2020. Agent-based intelligent decision support systems: a systematic review. IEEE Trans. Cognit. Dev. Syst. 14 (1), 20–34.
Kiruthika, U., Somasundaram, T.S., Raja, S. K.S., 2020. Lifecycle model of a negotiation agent: a survey of automated negotiation techniques. Group Decis. Negotiation 29 (6), 1239–1262.
Kraus, S., 2011. Agents that negotiate proficiently with people. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 137–137.
Kröhling, D.E., Chiotti, O., Martínez, E., 2019. The importance of context-dependent learning in negotiation agents. Intel. Artif. 22 (63), 135–149.
Kröhling, D.E., Chiotti, O. J.A., Martínez, E.C., 2021. A context-aware approach to automated negotiation using reinforcement learning. Adv. Eng. Inf. 47, 101229.
Kumar, R., Hassan, M.F., Adnan, M. H.M., 2023. Model of an intelligent and automated negotiation agent for the service level agreement negotiation process in cloud computing. J. Adv. Res. Appl. Sci. Eng. Technol. 32 (2), 189–202.
Lacher, A., 2017a. A Framework for Discussing Trust in Increasingly Autonomous Systems. Technical Report. The MITRE Corporation.
Lacher, A., 2017b. A framework for discussing trust in increasingly autonomous systems. Accessed: 2025-02-20.
Le Dantec, C.A., Poole, E.S., Wyche, S.P., 2009. Values as lived experience: evolving value sensitive design in support of value discovery. In: The SIGCHI Conference on Human Factors in Computing Systems, pp. 1141–1150.
Li, C., Giampapa, J., Sycara, C. K., 2003. A review of research literature on bilateral negotiations. Technical Report. Robotics Institute, Carnegie Mellon University.
Liao, B., Pardo, P., Slavkovik, M., et al., 2023. The jiminy advisor: moral agreements among stakeholders based on norms and argumentation. J. Artif. Intell. Res. 77, 737–792.
Liao, B., Slavkovik, M., van der Torre, L., 2019. Building jiminy cricket: an architecture for moral agreements among stakeholders. In: AAAI Conference on AI, Ethics, and Society, pp. 147–153.
Lin, K.-B., Wei, Y., Liu, Y., et al., 2023. An opponent model for agent-based shared decision-making via a genetic algorithm. Front. Psychol. 14, 1124734.
Lindner, F., Bentzen, M.M., Nebel, B., 2017. The HERA approach to morally competent robots. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 6991–6997.
Liscio, E., van der Meer, M., Siebert, L.C., et al., 2022. What values should an agent align with? an empirical comparison of general and context-specific values. Auton. Agent Multi Agent Syst. 36 (1), 23.
Luo, X., Li, Y., Huang, Q., et al., 2024a. A survey of automated negotiation: human factor, learning, and application. Comput. Sci. Rev. 54, 100683.
Luo, X., Luo, Y., Fan, Y., et al., 2024b. A human-computer negotiation model based on q-learning. In: International Conference on Knowledge Science, Engineering and Management, pp. 268–283.
Magra, A., Baarslag, T., Spreij, P., 2023. Querying user preferences in automated negotiation. In: IEEE International Conference on Agents, pp. 71–76.
Mahan, F., Isazadeh, A., Khanli, L.M., 2011. Using an active fuzzy ECA rule-based negotiation agent in e-commerce. Int. J. Electron. Commerce Stud. 2 (2), 127–148.
Maia, P., Vieira, L., Chagas, M., et al., 2019. Dragonfly: a tool for simulating self-adaptive drone behaviours. https://doi.org/10.1109/SEAMS.2019.00022
Maio, G.R., 2016. The psychology of human values. Routledge.
Mansour, K., 2020. A hybrid concession mechanism for negotiating software agents in competitive environments. Int. J. Artif. Intell. Tools 29 (06), 2050016:1–13.
Mansour, K., Al-Lahham, Y., Tawil, S., et al., 2022. An effective negotiation strategy for quantitative and qualitative issues in multi-agent systems. Electronics (Basel) 11 (17), 2754.
Meli, D., Nakawala, H., Fiorini, P., 2023. Logic programming for deliberative robotic task planning. Artif. Intell. Rev. 56 (9), 9011–9049.

Memon, M.A., Scoccia, G.L., Autili, M., 2023. Automated negotiation-preliminary results of a systematic mapping study. In: 38Th International Conference on Automated Software Engineering Workshops, pp. 94–99.
Memon, M.A., Scoccia, G.L., Autili, M., 2025. A systematic mapping study on automated negotiation for autonomous intelligent systems. Automat. Softw. Eng. 32 (2), 1–46.
Migliarini, P., Scoccia, G.L., Autili, M., et al., 2020. On the elicitation of privacy and ethics preferences of mobile users. In: 7Th International Conference on Mobile Software Engineering and Systems, pp. 132–136.
Mill, J.S., 2016. Utilitarianism. In: Seven Masterpieces of Philosophy. Routledge, pp. 329–375.
Mohammad, Y., 2023. Optimal time-based strategy for automated negotiation. Appl. Intell. 53 (6), 6710–6735.
Moor, J.H., 2006. The nature, importance, and difficulty of machine ethics. IEEE Intell. Syst. 21 (4), 18–21.
Morales, Y., Kallakuri, N., Shinozawa, K., et al., 2013. Human-comfortable navigation for an autonomous robotic wheelchair. In: 2013 Del Ins Thinspace IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2737–2743. https://doi.org/10.1109/IROS.2013.6696743
Moreno, G., Kinneer, C., Pandey, A., et al., 2019. Dartsim: an exemplar for evaluation and comparison of self-adaptation approaches for smart cyber-physical systems. In: 2019 Del Ins Thinspace IEEE/ACM 14Th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), pp. 181–187. https://doi.org/10.1109/SEAMS.2019.00031
Mostafa, S.A., Ahmad, M.S., Mustapha, A., 2019. Adjustable autonomy: a systematic literature review. AI Rev. 51, 149–186.
Mougouei, D., 2020. Engineering human values in software through value programming. In: 42Nd International Conference on Software Engineering Workshops, pp. 133–136.
Mougouei, D., Perera, H., Hussain, W., et al., 2018. Operationalizing human values in software: a research roadmap. In: 26Th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 780–784.
Nallur, V., 2020. Landscape of machine implemented ethics. Sci. Eng. Ethics 26 (5), 2381–2399.
Nallur, V., Collier, R., 2019. Ethics by agreement in multi-agent software systems. In: 14Th International Conference on Software Technologies, pp. 529–535.
Nam, C., Shell, D.A., 2015. Assignment algorithms for modeling resource contention in multirobot task allocation. IEEE Trans. Autom. Sci. Eng. 12 (3), 889–900.
Narayanan, V., Jennings, N.R., 2005. An adaptive bilateral negotiation model for e-commerce settings. In: 7Th IEEE International Conference on E-Commerce Technology, pp. 34–41.
Nurwidyantoro, A., Shahin, M., Chaudron, M., et al., 2023. Integrating human values in software development using a human values dashboard. Empir. Softw. Eng. 28 (3), 67.
Padmasiri, H., Madurawe, R., Abeysinghe, C., et al., 2020. Automated vehicle parking occupancy detection in real-time. In: Moratuwa Engineering Research Conference, pp. 1–6.
Palmer, A.W., Hill, A.J., Scheding, S.J., 2018. Modelling resource contention in multi-robot task allocation problems with uncertain timing. In: IEEE International Conference on Robotics and Automation, pp. 3693–3700.
Pant, A., Hoda, R., Spiegler, S.V., et al., 2024. Ethics in the age of AI: an analysis of AI practitioners' awareness and challenges. ACM Trans. Software Eng. Method. 33 (3). https://doi.org/10.1145/3635715
Pappas, J., Dasari, V.R., Geerhart, B.E., et al., 2024. Adaptive object detection algorithms for resource constrained autonomous robotic systems. In: Disruptive Technologies in Information Sciences VIII. Vol. 13058, pp. 101–108.
Perera, H., Hussain, W., Mougouei, D., et al., 2019. Towards integrating human values into software: mapping principles and rights of GDPR to values. In: 27Th IEEE International Requirements Engineering Conference, pp. 404–409.
Perera, H., Hussain, W., Whittle, J., et al., 2020. A study on the prevalence of human values in software engineering publications, 2015–2018. In: 42Nd International Conference on Software Engineering, pp. 409–420.
Pommeranz, A., Detweiler, C., Wiggers, P., et al., 2012. Elicitation of situated values: need for tools to help stakeholders and designers to reflect and communicate. Ethics Inf. Technol. 14 (4), 285–303.
Reinhardt, K., 2022. Trust and trustworthiness in AI ethics. AI Ethics , 1–10.
Rokeach, M., 1973. The nature of human values. Free press.
Di Ruscio, D.D., Inverardi, P., Migliarini, P., et al., 2024. Leveraging privacy profiles to empower users in the digital society. Autom. Softw. Eng. 31 (1), 16.
Russell, S.J., Norvig, P., 2016. Artificial intelligence: a modern approach. Pearson.
Ryan, M., Stahl, B.C., 2020. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. Inf., Comm. and Ethics in Society 19 (1), 61–86.
Saket, M., 2021. Putting values in context: an augmentation of value sensitive design (VSD). J. Ethics Emerg. Technol. 31 (2), 1–9.
Schwartz, S.H., 2012. An overview of the schwartz theory of basic values. Online Readings Psychol. Cult. 2 (1), 10–20.
Sefati, S.S., Halunga, S., 2023. Ultra-reliability and low-latency communications on the internet of things based on 5g network: literature review, classification, and future research view. Trans. Emerg. Telecommun. Technol. 34 (6), e4770. https://doi.org/https://doi.org/10.1002/ett.4770
Shaheen, T., Brščić, D., Kanda, T., 2024. Investigation of low-moral actions by malicious anonymous operators of avatar robots. ACM Trans. Human-Robot Interact. 14 (1), 1–34.
Shahin, M., Hussain, W., Nurwidyantoro, A., et al., 2022a. Operationalizing human values in software engineering: a survey. IEEE Access 10, 75269–75295.
Shahin, M., Hussain, W., Nurwidyantoro, A., et al., 2022b. Operationalizing human values in software engineering: a survey. IEEE Access 10, 75269–75295.
Shek, S. C.K., Butterfield, J., Murphy, A., et al., 2021. Current state of the art in object detection for autonomous systems. In: 37Th International Manufacturing Conference, pp. 1–10.
Spehrs, A., 2024. Dispositional realism, conflicting models and contrastive explanation. J. Gen. Philos. Sci. 55 (2), 219–228.
Spiekermann, S., 2023. Value-Based Engineering: A Guide to Building Ethical Technology for Humanity. De Gruyter, Berlin, Boston.
Stavrou, D., Timotheou, S., Panayiotou, C.G., et al., 2018. Optimizing container loading with autonomous robots. IEEE Trans. Autom. Sci. Eng. 15 (2), 717–731.
Suri, S., Das, S.N., Singi, K., et al., 2023. Software engineering using autonomous agents: are we there yet? In: 38Th IEEE/ACM International Conference on Automated Software Engineering, pp. 1855–1857.
Takeuchi, K., Yamazaki, Y., Yoshifuji, K., 2020. Avatar work: telework for disabled people unable to go outside by using avatar robots. In: ACM/IEEE International Conference on Human-Robot Interaction, p. 53–60.
Thiebes, S., Lins, S., Sunyaev, A., 2021. Trustworthy artificial intelligence. Electron. Mark. 31 (2), 447–464.
Tolmeijer, S., Kneer, M., Sarasua, C., et al., 2020. Implementations in machine ethics: a survey. ACM Comput. Surv. 53 (6), 1–38.
Topping, A., 2023. Heathrow failed to meet minimum accessibility standards, CAA report finds. Guardian Accessed: 2025-05-20. https://www.theguardian.com/uk-news/2023/jul/20/heathrow-failed-to-meet-minimum-accessibility-standards-caa-report-finds.
Townsend, B., Paterson, C., Arvind, T.T., et al., 2022. From pluralistic normative principles to autonomous-agent rules. Minds Mach. 32 (4), 683–715.
Urban Robotics Foundation, 2023. PMR Use Case: Robotic Wheelchairs Improve Accessibility in Airports, Hospitals & More. Accessed: 2025-05-20. https://www.urban-roboticsfoundation.org/post/pmr-use-case-robotic-wheelchairs-improve-accessibility-in-airports-hospitals-more.
Vanderelst, D., Winfield, A., 2018. An architecture for ethical robots inspired by the simulation theory of cognition. Cogn. Syst. Res. 48, 56–66.
Waldman, A.E., 2019. Power, process, and automated decision-making. Fordham L. Rev. 88, 613.
Winfield, A. F.T., Blum, C., Liu, W., 2014. Towards an ethical robot: internal models, consequences and ethical action selection. In: Conference towards Autonomous Robotic Systems, pp. 85–96.
Winfield, A. F.T., Hafner, V.V., 2019. Anticipation in robotics. Handb. Anticipat.: Theoret. Appl. Aspect. Use Future Decis. Making , 1587–1615.
Wong, K.W., Kress-Gazit, H., 2015. Let'S talk: autonomous conflict resolution for robots carrying out individual high-level tasks in a shared workspace. In: 2015 Del Ins Thinspace IEEE International Conference on Robotics and Automation (ICRA), pp. 339–345. https://doi.org/10.1109/ICRA.2015.7139021
Wooldridge, M., Jennings, N.R., 1995. Intelligent agents: theory and practice. Knowl. Eng. Rev. 10 (2), 115–152.
Wurman, P.R., Wellman, M.P., Walsh, W.E., 2002. Specifying rules for electronic auctions. AI Mag. 23 (3), 15–15.
Xu, C., Xi, W., Cheung, S.-C., et al., 2015. Cina: suppressing the detection of unstable context inconsistency. IEEE Trans. Software Eng. 41 (9), 842–865.
Yaman, S.G., Ribeiro, P., Cavalcanti, A., et al., 2025. Specification, validation and verification of social, legal, ethical, empathetic and cultural requirements for autonomous agents. J. Syst. Softw. 220, 112229.
Yazdanpanah, V., Gerding, E.H., Stein, S., et al., 2021. Responsibility research for trustworthy autonomous systems. In: 20Th International Conference on Autonomous Agents and Multiagent Systems. ACM, pp. 57–62.
Yoo, D., Sim, K.M., 2010. A multilateral negotiation model for cloud service market. In: Grid and Distributed Computing, Control and Automation. Springer, pp. 54–63.
Youssefmir, M., Huberman, B.A., 1995. Resource contention in multiagent systems. In: 1St International Conference on Multiagent Systems, pp. 398–405.
Zhao, C., Zhao, S., Zhao, M., et al., 2019. Secure multi-party computation: theory, practice and applications. Inf. Sci. (Ny) 476, 357–372.
Zuckerman, I., Rosenfeld, A., Kraus, S., et al., 2013. Towards automated negotiation agents that use chat interfaces. In: 6Th International WS on Agent-based Complex Automated Negotiations, pp. 6–10.