HRI + RA report

Edoardo Colonna 1697124
Matteo Zaramella 2025806

February 2024

Abstract

Applications of assistive robotics have increased in recent times due to the advances that have been made in the various disciplines on which this field is based. Not only new technical and technological capabilities of systems but also advances in research subjects such as Human-Robot Interaction or Reasoning Agents have made it possible to develop robots that can act in human environments and perform collaborative tasks. In this project, the implementation of a receptionist robot, DIAGhino, that can answer questions and accompany DIAG students and visitors to various classrooms is proposed. To do this it relies both on implementation details aimed at improving its social skills, making it more interesting and interactive for users, and on its reasoning skills by exploiting the knowledge it has of the environment and schedules.

The work carried out for this project was conceived, developed and edited by Edoardo Colonna and Matteo Zaramella. Both contributed equally to all aspects of the project

Contents

1 Introduction

In recent years we are witnessing a massive increase in the use of robots around the world. This is obviously due to the incredible developments that robotics and artificial intelligence have made and continue to make month after month. In addition to the increases in robotic units in industrial or research sectors, AI and robotics are becoming topics of news and public debate in this period. Just think of the great media coverage that generative models such as GPT-3 [1] or Dall-E[2] have had with regards to AI or the outcry that the videos [3] of the performances of the Boston Dynamics robots unleash. This means that in addition to the substantial increase that will occur for industrial process automation applications, current projections of the future trends [4] are showing positive toward a substantial increase in home and support robotics as well. In fact, there are many fields where it will be possible to use intelligent robotic agents in ways that will migrate the experience and lives of human users: hospital robots [5] [6], employed in food service [7] [8] or assistive social robots in elderly care [9]. These are just a few of the most popular applications to date involving robots operating and collaborating among humans. The implications of this process of developing robots capable of acting in spaces previously designed for humans are many. They range from very different fields, from ethics to computer engineering. In this sea of disciplines that, some more than others, go into collaborative and cooperative robotics, of particular note are Human-Robot Interaction, abbreviated as **HRI** and Reasoning and Planning Agents or **RA** (just Reasoning Agents). The first is crucial in that the behaviour of humans toward robots is complex and articulated, and if you want to be able to develop robots with which humans actually want to act you have to take these aspects into account. Regarding the reasoning ability of automata, it goes without saying that employing them in dynamic and changing environments brings additional challenges compared to industrial robots used in production lines, where the spaces and actions they can perform are limited. An agent capable of reasoning and planning based on novel elements with respect to its knowledge is critical if it is to be used in applications similar to those mentioned above. This project has developed the implementation of an assistive robot to be employed in a receptionist task in a university setting. In particular, the robot's purpose is to greet visitors and students at the entrance and ask them what their destination within the building is. Based on the answers it receives and a list of information it has about class schedules, the classrooms where they are held, and the professors who teach them, it should be able to correctly point to the classroom the user is looking for and in case the user wants it to, it should lead the user to the classroom in question. To carry out its task, the automaton must use reasoning techniques to understand what to do.It must also be designed to follow those HRI guidelines that make the experience with the human user as pleasant and natural as possible. the implementation of the project was developed with Pepper, a robot well known and used for similar tasks, equipped with sensors that allow it to act and perceive the environment in which it finds itself. Pepper was only used in a simulated environment, without testing its actual behaviour in reality. Despite this, the objective of this project is to propose a streamlined and effective architecture and implementation in which the most important aspect is to be able to understand what the user wants to know about and how to help him with this. Implementations for tasks of this type are in our opinion one of the most effective uses that can be made of collaborative and assistive robotics today. Already being able to count on the advanced motor and calculation capabilities of robots already ready for large-scale distribution as in the case of Pepper, all those jobs for which human beings may not derive great satisfaction due to the repetitiveness of requests and actions can be studied to be carried out by robots. In the next sections, we will first propose an analysis of related works in the field of HRI and RA regarding projects similar to the one proposed in this work, in order to extrapolate insights and paths already followed that lead to valid results. Subsequently, the task that our agent, DIAGhino, has to perform is illustrated in detail, together with the implementation choices and the architecture. Simulations that were carried out to evaluate the validity of the proposed solution are then described and analyzed, underlining which particular aspects of the HRI and RA are used by the robot. Finally, the conclusions and some considerations on the project carried out are presented, together with the possible developments and improvements that can be made.

2 Related Work

In HRI, interaction refers to the process of working together to achieve a goal. Motivations for discussions in HRI are the increased interest in applications involving this kind of interaction with non-expert users and in general deployment of robots in public spaces. As for any radically new technology, to make it really useful for humans, something that creates familiarity with it is needed. Based on different application fields, robots could have higher or lower social skills requirements. An agent designed for firefighting tasks would not necessitate the same charisma and empathy that is requested for a home companion robot. The characteristics on which the familiarity that a human being feels towards a robot depends are pretty much the same as those on which familiarity with other human beings depends: the exchange of information is fundamental to understanding what "the other" intends to do or is thinking. The study of social signal processing is a fundamental aspect in the field of HRI as autonomous agents must be able to understand the behavior of humans around them in order to act accordingly. Face detection, person tracking or speech recognition are just some of the technologies that are coming into play to make robots able to understand the emotions or facial and physical expressions of human beings. At the same time, this need to interpret emotions or facial expressions is also necessary on the part of the human being towards the automaton. In this sense, it is interesting to briefly analyze the work proposed by Carpinella et al. [10] "the Robotic Social Attributes Scale (RoSAS)": they developed a standardized measurement tool to determine how perceived attributes affect the quality of interaction with robots. Social psychology has established two universal dimensions of person perception that are warmth and competence [11]. Existing research indicates social categorization processes of person perception can be generalised to robots [12] [13]. In order to measure to measure human and robotic interaction they developed a scale based on 18 items that comprised the 3 factors that are **Competence**, **Warmth** and **Discomfort**. The RoSAS scale is a widely used tool that has been cited in a lot of other studies. One example of its use can be taken from the tests proposed in the paper: they used RoSAS to validate whether the judgment of robots varies as a function of appearance, in particular with gender. They presented a set of 9 faces (fig.1) which vary under gender and machineness. Participants had to rate all the faces on the RoSAS scale. In this particular experiment, they discovered that human-like robots were seen as more competent than machinelike ones. At the same time, female and humanlike robots are perceived as warmer and more competent while male and machinelike robots are rated the highest on discomfort. These results prove that human evaluation of a robot can effectively vary based on its aesthetic characteristics and that some robots can generate a higher sense of discomfort with respect to others. However, physical appearance is not the only thing that matters when we talk about interactions between humans and robots. Let's imagine getting an agent who is so well-groomed and advanced that it is particularly pleasant to interact with him, perhaps one who manages to overcome the Uncanny Valley [14]. We could expect to be the perfect assistant for social interactions. But then we discover that. His behaviour and actions are slow, unintuitive and therefore annoying. Such an agent would end up not attracting the slightest interest of human beings after the first few interactions. This simple example justifies the direction taken by some research that aimed to create more interactive and performing agents that could somehow connect strongly with the users. With this in mind, we give some brief notes regarding the work proposed by Mason and Lopes in [15]. Their idea is to develop a Robotic system that interacts with users, and through repeated actions, it adapts to them. The aim is to create a robot that can execute orders and anticipate needings of the users. So a robot that based on training, can learn. The approach consists of a user who says to the robot verbally how to do a task, and the robot executes it, so the user gives feedback. In this way, a user profile can be created and the robot can start to do tasks without the used command. Instead, it anticipates his needs. The complex implementation needs environments as a vector in feature space and a dialog system that allows robot-user interaction, and it's not limited to fixed sentences. Then a classifier is used to train the robot to generalize the information. This approach scored great results in experiments in which people who interacted with the robot left feedback on the interaction. There is still plenty of work to do to make such an implementation Usable in everyday life but these results are very promising. In this work we leveraged concepts from both this paper and the one of Carpinella et al. as a ground base to design the implementation of our assistive agent: from [10] we take intuitions about what helps an agent to be considered more sociable and to create less discomfort, especially in the choice of interactions and gestures, given that they are the only factors we can modify in Pepper, the robot we work with, while we take inspiration from [15] to develop the action-logic and behaviour of Pepper in order to make the experience of using our agents much pleasant and stimulating as possible. From here on, we complete this section with some study cases and interesting studies of both HRI and RA fields on which we based our work

2.1 HRI

In this section, we are interested in proposing a brief review of case studies of robotics applications in collaboration and interaction tasks with humans. In this way, you can highlight problems and ideas already developed for projects similar to this one. This can help you have a more solid starting point for subsequent implementation. As already mentioned previously, the characteristics required by autonomous agents operating in environments shared with humans mainly

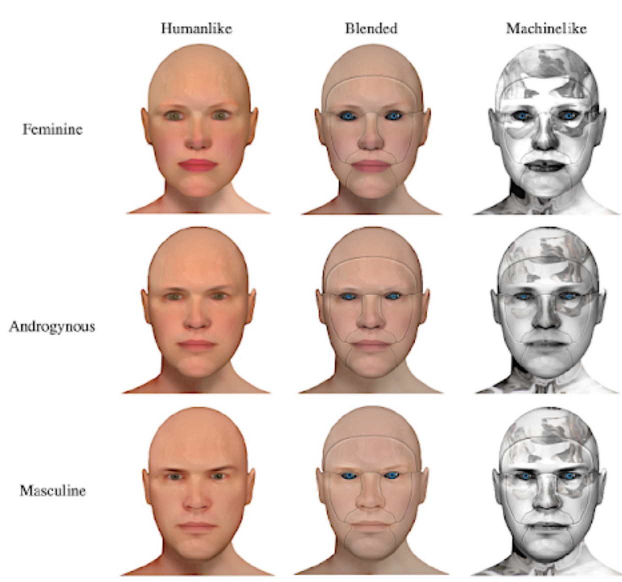

Figure 1: a set of 9 faces which vary under gender and machinelikenes

depend on the purpose of the agent's application. In addition to structural and behavioural efficiency, which we will analyze better in the next paragraph, in application areas like ours, in which the robot must carry out a task that falls within the hospitality sector, aesthetic and social characteristics are also necessary to generate greater harmony with human users. An interesting work has been carried out by Lopez et al.[16]: they proposed an automatic hotel assistant system based on a series of mobile platforms that interact with guests and service personnel to help them with different tasks. These tasks include bringing small items to customers, showing them different points of interest in the hotel, accompanying the guests to their rooms and providing them with general information. In a sense, their project is pretty similar to the one that we are interested in implementing. Their system is also more advanced since they also add as a feature the possibility of performing tasks based on events triggered by the building's automation system. Among the most interesting aspects of this research as far as we are concerned, it is reported that their robots, a representation of which is shown in fig.2, are oriented towards close interaction with guests and service personnel. They have social interaction capabilities, including touch sensors, touchscreens, and voice recognition systems, as can be seen in fig.3. They reported that two different kinds of robots with different capabilities of interaction have been used: BellBot robot interacts with users using touch sensors on its cheeks and hands, as well as a touchscreen and a laser sensor. The Sacarino robot interacts with users using a touchscreen and a voice recognition system. It also includes a chatbot for more flexible recognition and provides useful information to guests. They also shaped the appearance of their automata by basing their choices on the Uncanny Valley theorem. For this reason, they decided to give them a more animated toy-looking look to avoid making them unpleasant. for two monthsSacarino has been working, mainly on providing information to users and validating the performance of the dialogue system. So with this work, they have demonstrated effective use of assistive robots. To date, they have not reported the results of a possible survey filled out by customers, which would be very useful for understanding where to direct subsequent changes to systems of this kind.

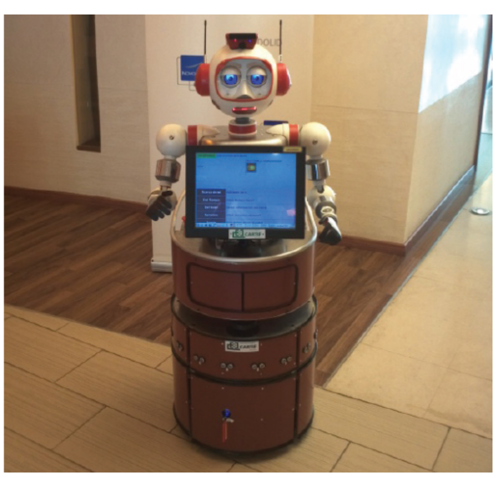

Figure 2: Sacarino at the hotel.

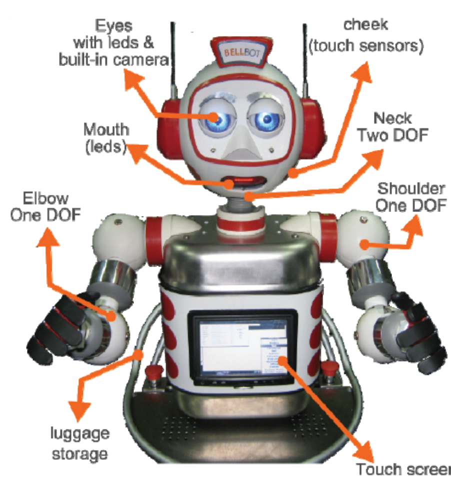

Figure 3: Bellbot body description.

To have a clearer vision of human judgment regarding these types of robot applications, it is useful to analyze the work proposed in the article of Seo and Lee of 2021 [17]. In this study, they investigated consumer behaviours and acceptance of service robots at restaurants. They integrated trust, perceived risk, and satisfaction with the Technology Acceptance Model (TAM) to understand the factors influencing consumer responses towards robot restaurant visitors. The results showed that perceived usefulness (PU) had a direct positive impact on consumers' intention to revisit robot restaurants. However, perceived ease of use (PEOU) did not have a significant direct impact on behavioural intention. Trust was found to have a significant positive impact on both PU and PEOU, indicating that trust plays a crucial role in the acceptance of robot service at restaurants. Trust also decreased perceived risk and increased satisfaction with robot service. On the other hand, perceived risk had a negative impact on satisfaction and behavioural intention, while satisfaction had a positive impact on behavioural intention. The implications of these findings with regard to hospitality and assistive robot applications are pretty interesting. Marketers and operators should focus on enhancing consumers' perception of trust in robot service by promoting the reliability and safety of the robot system. Another important step is to reduce perceived risk by ensuring error-free and perfect robot performance. Providing information and guarantees can help alleviate consumers' concerns and increase their satisfaction with robot service. Another well-known study on Human-Robot interaction in hospitality services, parallel to the one that has been just seen, is the one of Kervenoael et al. [18]. They performed an analogous study by conducting a survey of 443 Singaporean visitors in contact with robots within their daily hospitality service environment and also interviewed five hospitality managers. Parts of their findings agree with [17], stating that visitors' intentions to use social robots are influenced by factors such as perceived usefulness, perceived ease of use, service assurance, personal engagement, tangibles, empathy, perceived value, and information sharing. But they also report interesting findings about the robot's aspect. The study also found that the physical appearance of the robots and the level of empathy and information sharing exhibited by the robots play a significant role in visitors' intentions to use them. This is not only to be intended as the simple look of the robot, that can be observed also when it is still. Instead is more about the combination with the level of empathy, which can be increased by working on the robot's gestures, movements, naturalness of the artificial voice and so on. The results of the study indicate that when visitors perceive that social robots offer benefits that outweigh the costs, they are more likely to have a positive intention to use them. This is another fundamental concept that has to be kept in mind when developing such systems. In the case of this project, the user must be encouraged to exploit the automaton, otherwise, following a simple optimization of the mental effort required, he will be more inclined to find the way or the information on his own rather than requesting assistance from the android. This aspect was strongly taken into consideration for the implementation which will be illustrated in the next paragraphs. There are still other case studies that have been considered in this project, such as "'LUCAS': The Library Assistant Robot" [19] or "Pearl: A Mobile Robotic Assistant for the Elderly" [20], and also surveys studies ([21] [22] [23]). As regards the relevant aspects of our interest, the considerations drawn from these other studies are the same as those already explained in the previous lines. What most distinguishes these works are the implementation choices aimed at obtaining robots with different behaviors and empathy. These choices fall into the other research sector, which we illustrate in the next paragraph.

2.2 RA

What we have seen until here regards the outputs that the robots are able to generate and the reactions of humans to them. Humans are more interested in interacting with them if, among other things, they perceive a high usefulness, ease of use and empathy. In order to develop these characteristics in our system, is also important to refer to **planning** and **reasoning** study areas. The task in which we are interested in this work is pretty complex with respect to other application fields. For example, trying to make the robot less cumbersome or more intuitive is something that is not required in industrial applications. Again

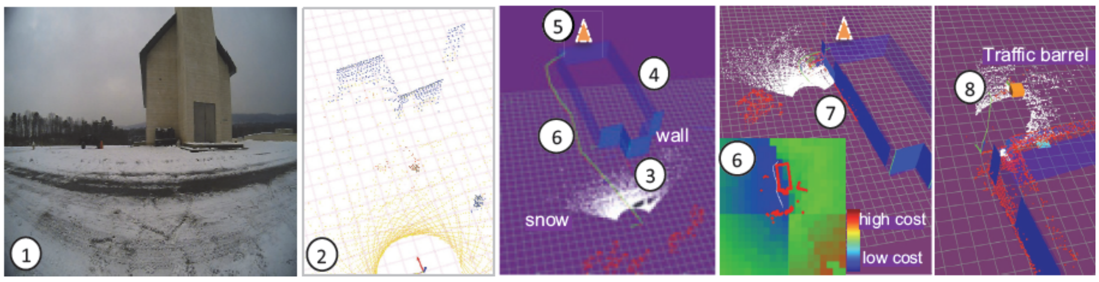

Figure 4: Images of the end-to-end process for executing a command task outdoors for an autonomous robot. Space objects are detected and then labelled. From these, Tactical Behaviour Computations language is used to compute inference.

with reference to the latter, agents tend to be able to operate in a confined space and perform limited actions that will rarely go outside the established plan. Developing the logic for these types of agents is obviously more straightforward. In contrast, when it comes to assistive robots that must work in contact and/or collaboration with humans, one enters an already less deterministic and more unknown environment. Under these conditions, it is impossible for the robot to already have total knowledge of the world around it. Suffice it to say that the human presence makes it completely unpredictable since each person can act following free will that cannot be contained within a database. In this section, we are going to present interesting studies related to solutions adopted for this type of application. In "Toward Mobile Robots Reasoning Like Humans" [24] paper the authors discuss the development of an intelligence architecture for mobile robots to reason and navigate like humans. The goal is to enable robots to understand natural language commands, recognize objects in the environment, and perform high-level reasoning tasks. The architecture combines cognitive components, semantic perception, and natural language understanding. They used a semi-structured language that specifies high-level cognitive tasks to develop tactical behaviour. The document presents the results of outdoor experiments conducted in a military training facility. One example is reported in fig.4. The system successfully carried out various commands, such as navigating to a specific location or object. The results show that the system can effectively reason and navigate in real-world environments, even in the presence of uncertainty and incomplete information. Also in Martin and Lopes's work [15] that has been presented before, an interesting approach for interpreting and applying human commands has been proposed. The interesting aspect of this study is that the system uses learning to allow the robot to adapt known tasks to new situations or learn new tasks. the authors propose using social guidance, where a human reward signal is combined with environmental rewards to program the robot to solve new tasks and adapt to the user's desires and action preferences. This is only an example of how articulated the understanding and reasoning architecture can be. Not only the ability to understand what to do but also how this is done is fundamental. Recalling that humans are interested in safety when dealing with robots, the movements of the agents and how they are performed should be thoroughly implemented. Human-following or human-accompanying behaviour has been studied in works like the one of Ohya [25]. As a workaround, the robot uses a camera to detect a light-emitting device carried by the human. By analyzing the position and distance of the device, the robot can follow the human and adjust its path to avoid obstacles. Nowadays implementation can rely on big steps ahead in computer vision and perception technologies. At the same time, the control strategies remain the same. They should generate control decisions through an appropriate reasoning mechanism based on what they perceive. Changes in the

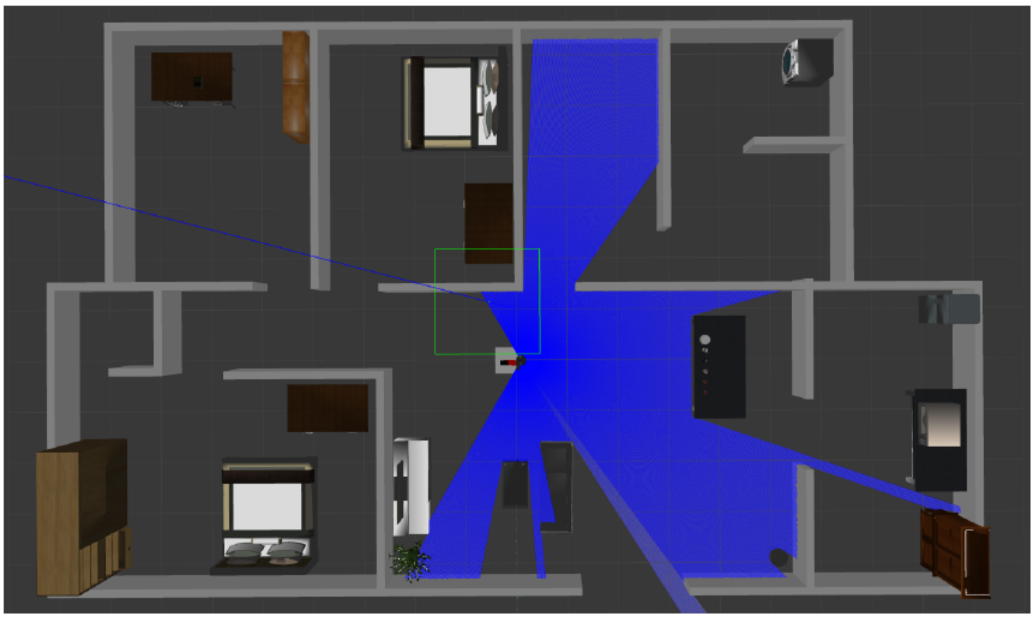

Figure 5: domestic simulated environment from [27], which includes two bedrooms, one kitchen, one living room, and one study

scenario can occur at any moment and the controller must robustly handle each possible occurrence while maintaining a safe behaviour [26]. Just to analyze an example of solutions for such problems, we can refer to the work of Cui, Shuai and Chen[27] of 2021. They proposed a planning system for a general-purpose service robot that aims to behave more intelligently in domains with incomplete information, under-specified goals, and dynamic changes. The system is based on semantic reasoning and combines natural language understanding, task-oriented knowledge acquisition, and semantic-based automated task planning. During the execution of the task, the robot continuously senses the environment, obtaining useful information. This information is used to adapt to dynamic changes and handle faults. The planner retrieves preconditions and effects for each action from domain knowledge to detect and handle environmental changes that conflict with the actions as early as possible. Their implementation is really advanced and for this reason so effective, as they showed in their results. Their robot is able to perform a long list of functions, from autonomous mapping to Natural Language Understanding. There would still be many more studies to report, but with these examples that have been analyzed in depth we have already given a fairly clear idea of the complexity and components required for a robot to perform a collaborative task in an open environment with the presence of humans. Although in our project we do not reach the levels of complexity seen in some of the papers presented, we nevertheless take great cues from them to best solve the problems our agent may encounter.

3 Solution

We are going to explain the implementation that was developed for the specific task of this project. The various aspects of the project are analyzed below, from the conditions and definition of the task, the methodologies used in the solution we propose, details regarding the tools used in the development, and an analysis of the results obtained.

3.1 Task definition

As a scenario where to employ our receptionist robot, which from now on we will call DIAGhino in honour of DIAG, we decided to use a university building. Taking a cue from the floor plan of the DIAG ("La Sapienza" University of Rome) building, the agent has the role of helping students and visitors enter the building. A schematic of the imaginary floor plan of the environment in which the task takes place is shown in fig.6. This in particular consists of 5 classrooms, called A1, A2, B1, B2 and the library, respectively. As will be better explained later, this visualization is for illustrative purposes only, since in the implementation the automaton is only aware of the distances of these classrooms from its starting point, without actually having a floor plan available. The operating environment is straightforward and streamlined. The addition of environmental elements such as clutter, stairs or furniture has been avoided; although they would make the environment more realistic, this would increase the complexity of the task and consequently of the implementation. DIAGhino is provided with the day's timesheet based on which he will be able to develop his reasoning to give the most appropriate response to the person who questions him. An example of a timesheet of a lesson day is given in tab.1. As can be seen, there are different time slots and classes, and in the various boxes appear the names of the classes being taught. Empty cells in the table mean that at that particular time, the classroom is not being used for lessons, and therefore can be used by students. The library is open throughout the opening hours of the day, which runs from 8:00 am until 4:00 pm. The other component of knowledge that DIAGhino has is the list of which subjects are associated with the professors who teach them. The display of this list is shown in tab.2. DIAGhino's job is to receive incoming visitors and students to DIAG and help them by providing information consistent with what they need. DIAGhino, as previously mentioned, is a Pepper robot, which will be described in more detail later. As such it can rely on all of Pepper's original sensors, hardware and software units, including the NLP module that allows it to interpret directions given to it vocally by people and respond in turn with an artificial voice. In addition, DIAGhino can move freely in 2D space to go to points of interest and can use arm and head joints while talking. This last function has a purpose only for the interactive and social aspects of DIAGhino. In our implementation the limbs cannot be used to perform specific manoeuvres such as picking up objects, opening doors, or touching people. There are two types of interactions possible with DIAGhino: in the first, the user asks for information and receives the answer. The information DIAGhino can respond to questions about what is happening in classrooms and where to find specific courses, lessons or professors. Typical questions that he can be asked are for example, " I need to go to the classroom where the Deep Learning class is being held," and " At what time does Professor Navigli's lesson start today?" " Where can I find a place to study ?". The second type of interaction always includes an answer regarding the questions seen before but also includes the possibility of being escorted to the classroom of interest. In the case where the person agrees to be accompanied, DIAGhino moves toward the entrance of the room by having the user follow him and once the objective is reached, DIAGhino takes leave of the person and returns to the starting station, ready to meet the next user. It

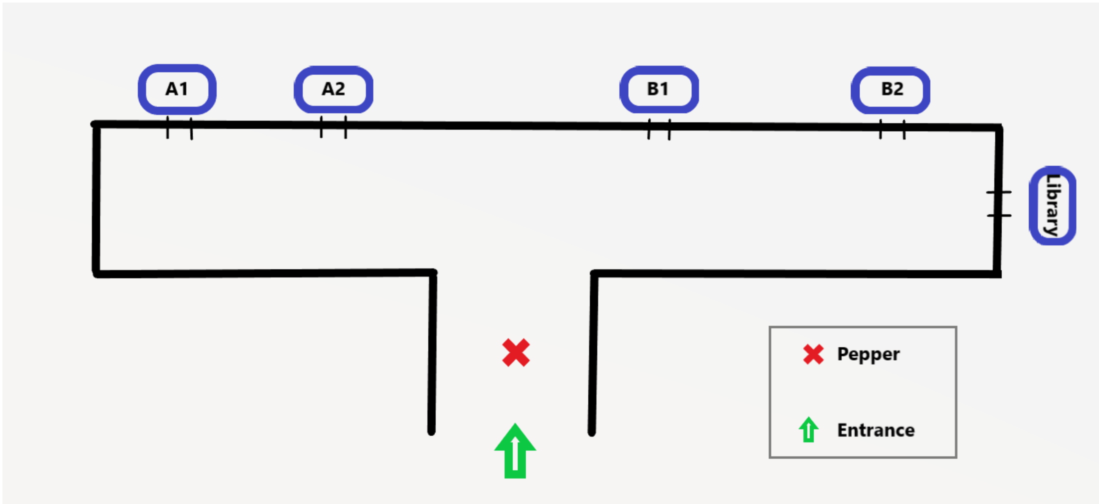

Figure 6: Illustration of the environment in which DIAGhino performs his duties as a receptionist. The entrance is where people come from, and the Red Cross is where he shows up whenever he has finished a trip with the previous guest. The map shows the 5 classrooms to which people can be taken, including the library.

Time	A1	A2	B1	B2	Library
08:00 - 10:00	NeuralNetworks	-	-	Deep Learning	Open
10:00 - 12:00	Robotics	Machine Learning	Artificial Intelligence	-	Open
12:00 - 14:00	-	Natural Language Processing	Social Networks	Neuroengineering	Open
14:00 - 16:00	Computer Vision	-	Reinforcement Learning	-	Open

Table 1: Class schedule of a typical day. DIAGhino relies on this knowledge to formulate its reasoning. Shown in the different schedules of the day are the lessons of the subjects taught in the diverse classrooms. Empty blocks mean that that classroom is empty and free in that slot

is also able to reason based on the current time. If he is asked to go to a class to attend a lesson that has now ended, however, DIAGhino can reason about it and respond to the user that it is now too late. If, on the other hand, the class has already started but is still going on, DIAGhino does not ask the user if he or she would like to be accompanied but instead directly starts walking, telling the person to follow it so as not to waste any more time. The proposed implementation involves agent activation the moment the agent senses the presence of motion in front of him. In addition, he can reset his state and return to the basic one if, despite having perceived someone and started talking to him, he receives no response or the person goes straight to ignore him. This allows him to have greater behavioural flexibility and to handle ambiguous cases. Another aspect to consider is that in this task the tablet present on the robot's belly is not used. The interactions all take place by voice. We believe that the use of a touch screen in an interaction that takes place while the user is standing and that involves a task in which the robot moves is neither invited nor congenial. As has already been stated several times, the use of both Reasoning and Planning (RA) applications and accoutrements and implementation decisions based on Human-Robot Interaction (HRI) field studies is essential for this project.

As for the HRI-related aspects, in this project, we tried to make the most of the robot's features that we could modify to maximize its social capabilities. Referring to the studies done for RoSAS [11], we worked with the idea in mind to maximize both warmth and competence and to reduce as long as possible any element that could generate discomfort. As for DIAGhino's appearance, being based on an existing automaton, Pepper, we could not work on his appearance or voice. We do note, however, that Pepper has appearances that we would call neutral, without being specifically male or female. This, according to studies

Subject	Professor
NeuralNetworks	Prof. Scardapane
Deep Learning	Prof. Silvestri
Robotics	Prof. De Luca
Machine Learning	Prof. Iocchi
Artificial Intelligence	Prof. Nardi
Natural Language Processing	Prof. Navigli
Social Networks	Prof. Anagnostopolus
Neuroengineering	Prof. Astolfi
Computer Vision	Prof. Amerini
Reinforcement Learning	Prof. Capobianco

Table 2: List of the day's subjects associated with the professors who teach those classes.

on the RoSAS scale, should place DIAGhino on a median human rating level according to the aspects mentioned above. What we could work on instead to improve his social performance concerns his interactions with users. Since his intended purpose is an informal but professional task, we tried to give him a warm but not funny personality. DIAGhino expresses this personality through his body movements and what he says. Regarding the former, we made DIAGhino gesticulate while talking. His gesturing is not casual but specific to the different phases of his speech. When he has to greet and offer his help to the person in front of him, he will move his arms and head in a saute-like manner. On the other hand, when he considers the response to be given to the user based on the requests, he will assume a pensive pose. When he takes the user to the room he has requested, while he is being followed he waves his arms while speaking, and when he takes his leave he makes the gesture of removing his hat while hinting at a bow. We believe that these types of movements can help to avoid conveying a feeling of coldness on the part of the automaton. We also believe, based on what is reported in the literature, that this may even entice the user to continue the interaction by capturing his attention and intriguing him. We believe that a robot that gesticulates very little induces the person interacting with it to immediately understand that he or she should not expect new movements and to consequently focus only on what the robot says eliminating a possible communication channel such as body language. Regarding what DIAGhino says, we decided to provide him with a simple and straightforward set of answers, considering that the person interacting with him has the goal of receiving information about a place he needs to go. So it is assumed that he has neither time nor desire to chat. Despite his streamlined language, DIAGhino provides clear and comprehensive answers. Its ability to understand what the user is referring to regardless of the information provided, whether it is the professor's name or a time and a classroom, allows for a very natural interaction with DIAGhino. Other robotic assistive systems ask them a series of questions and based on your answers they arrive at the node of information you are looking for. You are forced to follow their entire data acquisition process, though. Whereas DIAGhino is immediately ready to respond after the user's first request, which can be freely expressed. Its language system is designed to avoid getting into annoying interaction loops in which, not understanding the request, the robot keeps formulating the same request. If DIAGhino realizes that what is being asked is not part of his knowledge, he suggests that perhaps it was the wrong building and invites the user to go to the faculty site to find the answers he is looking for, thus interrupting the conversation himself before entering in loops. Another function we have implemented is a phase of voice interactions during the accompanying phase of the guest to the classroom. Following a robot silently without knowing how far the arrival point is is very different from the actual interaction that would occur in the case of a human being accompanying. Indeed, one would tend to fill the time with conversation, thus keeping the dialogue alive and decreasing any awkwardness. Since it may be unlikely to expect the user to initiate a conversation with DIAGhino while being escorted, we have arranged for DIAGhino himself to talk to the user. While being followed to the classroom, the robot gives some general cultural information about the building, talking about its history, or talking about the DIAG department and its foundation. The duration of this conversation is not proportional to the length of the walk but generally succeeds in giving the idea of a chat.

The other main elment DIAGhino relies on to perform its task is its reasoning and planning (RA) module. The environment in which the agent must operate is nondeterministic, in that interactions with the user cannot be known a priori and do not fall into a possible finite list. The other nondeterministic element would be, in a real implementation, the environment in which the robot must move. Of the corridors of a university are walked by many people performing nondeterministic movements and gestures. This would imply modules that can handle this, such as a path planner and an online implementation that would regularly check if the execution of the plan is happening correctly, and if not, reshape it. Instead in our implementation, it is not dynamic but static, i.e., when DIAGhino moves through the corridors to reach the classrooms, no unexpected external elements can intervene. This implies that in our case the module dedicated to defining movements is offline while only the one dedicated to communication is online. The plan to be executed is generated by a nondeterministic off-line planner since it intervenes after the communication part with the user and thus once it has acquired the details of the problem to be solved, i.e., the information in which the user is interested. The plan is generated using a breadth-first search algorithm, a pseudo code of which is given in 17.

Algorithm 1 Breadth-First Search

```
function BFS(start, end)
    visited ← empty set
    queue ← empty deque with [(start, [start])]
    while queue is not empty do
        node, path ← queue.popLeft()
        if node == end then
            return path
        end if
        if node ∉ visited then
            visited.add(node)
            for neighbor in self.mappa[node] do
                queue.append((neighbor, path + [neighbor]))
            end for
        end if
    end while
    return None
end function
```

What is returned is a list of states ranging from the current one the agent is in, to the final one in which the task is resolved. An example is reported in fig.7, where a screenshot of a terminal where the plan found by the planner is printed, describing step by step the connections between states. These states are represented by nodes, shown in the map in fig.6: in addition to the rooms and corridors, intermediate states (SI,S2,.—) are included. These nodes, represented by a graph, allow connection to all possible locations the robot can reach. The planner can handle other maps provided to it in this graph format, remaining generic with respect to the particular environment in which it is used. However,

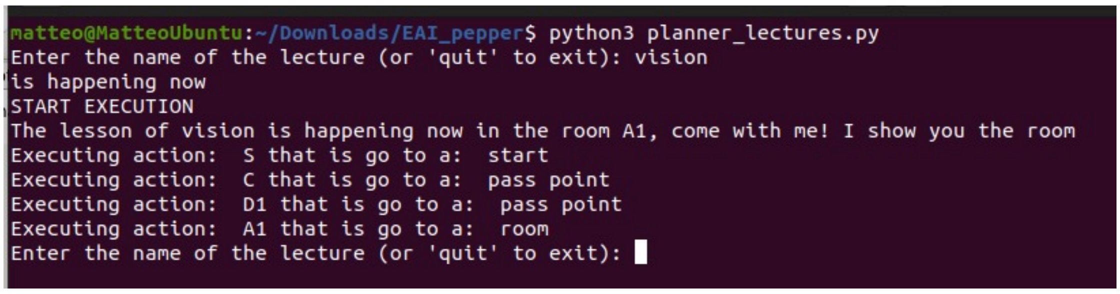

Figure 7: Screenshoot of a terminal where the plan generated by the planner for a specific request of the user has been printed.

in our implementation, we have opted for a formal, a priori definition of the movements the robot must make in order to move from one node to another. This is due to the fact that lacking a path planning module and not doing an online check of where it is actually going, DIAGhino must be provided with the spatial details of the movements it must make, i.e. how many metres it must move, in which direction and how far it must turn. The complete pipeline describing the various stages of an interaction can thus be summarised as follows: initially, a user verbally approaches DIAGhino, which understands, through its native NLP module, whether the request is pertinent to the services it can provide or not. In the affirmative, the NLP module terminates the keywords to find out what the user is interested in, which may be the classroom, the subject, the professor or spaces where to study. These key terms are fed to the planner which uses them to generate a plan that satisfies what the user has requested and which is then executed. Execution takes place by means of logic blocks implemented in Choregraph, which transition the robot between different states. DIAGhino's executor and actions system is based on a simple structure implemented in Choregraphe simulation environment. Its main advantage is dimensional, or breadth scalability. In fact, it is linearly adaptable to an increase in the amount of data, such as by adding schedules for all days of the week or by increasing the number of classrooms or even the number of floors, assuming it can even move you from floor to floor. Instead, it is hard to expand the complexity of the depth of its interactions. DIAGhino is able to perform two kinds of interactions, where the most complex one is the one in which it accompanies the visitor. Because of the way the logic module is implemented, the depth of interactions is represented by the number of consecutive action blocks, such as those that can be seen in Fig. 9. The more branches the iteration steps with the user can take, the more consecutive actions are expected before returning to the initial state, and the more the amount of blocks that must be created grows exponentially. Should an agent be designed to support longer interactions, another system would have to be used for implementation. Nonetheless, for the functions that our robot has to perform in this design, the reasoning module turns out to be efficient and well-functioning, as is also shown in the results.

3.2 Pepper

Pepper, developed by SoftBank Robotics, is an advanced humanoid robot engineered to engage with humans across a spectrum of environments, like customer service, education, or entertainment realms. Standing at a height of approximately 4 feet, Pepper embodies a striking resemblance to a human figure, complete with a head, arms, and a torso. Its design is not just aesthetically pleasing but also highly functional, boasting an array of sensors that empower it to perceive and interact with its surroundings with impressive precision. At the heart of Pepper's sensory apparatus lie depth sensors, ingeniously integrated to provide the robot with a nuanced understanding of the spatial dimensions of objects in its vicinity. These sensors play a pivotal role in facilitating Pepper's navigation ensuring seamless obstacle avoidance, and enhancing its operational efficiency. Moreover, Pepper includes also infrared sensors, meticulously crafted to discern the presence of objects and individuals nearby. This capability serves a dual purpose: maintaining a safe distance from obstacles and fostering meaningful interactions with humans, underlining Pepper's commitment to safety and engagement alike. Complementing its tactile awareness are embedded touch sensors, positioned throughout Pepper's body. This feature enables the robot to detect and respond to human touch, whether it's a gentle tap on its head or an enthusiastic high-five, consequently fostering a sense of connection and interactivity. Furthermore, Pepper is equipped with an impressive auditory apparatus, comprising microphones and audio sensors that enable it to perceive and comprehend spoken commands and conversations. This auditory apparatus empowers Pepper to engage in verbal communication and respond promptly to voice prompts, enhancing its versatility in various applications. The ability to see, Pepper's integrated cameras, typically housed within its head, enables the robot to recognize faces, interpret gestures, and identify objects within its environment. This visual data serves as a vital navigational aid, allowing Pepper to manoeuvre with precision and interact seamlessly with its surroundings. Not to be outdone, ultrasonic sensors augment Pepper's awareness by detecting nearby objects and obstacles through the emission and reception of high-frequency sound waves. This sensory feedback further enhances Pepper's spatial cognition, reinforcing its ability to navigate complex environments with ease. Rounding out its sensory suite are gyroscopes and accelerometers, instrumental in maintaining Pepper's balance and posture during dynamic interactions. Whether it's walking or engaging in intricate gestures, these sensors ensure that Pepper remains poised and stable, enhancing the overall quality of human-robot interaction. Pepper's advanced sensors help it see and understand the world around it well. Whether it's finding its way through a busy area or having a chat, Pepper shows how human-robot interaction can make a big difference in many areas of life.

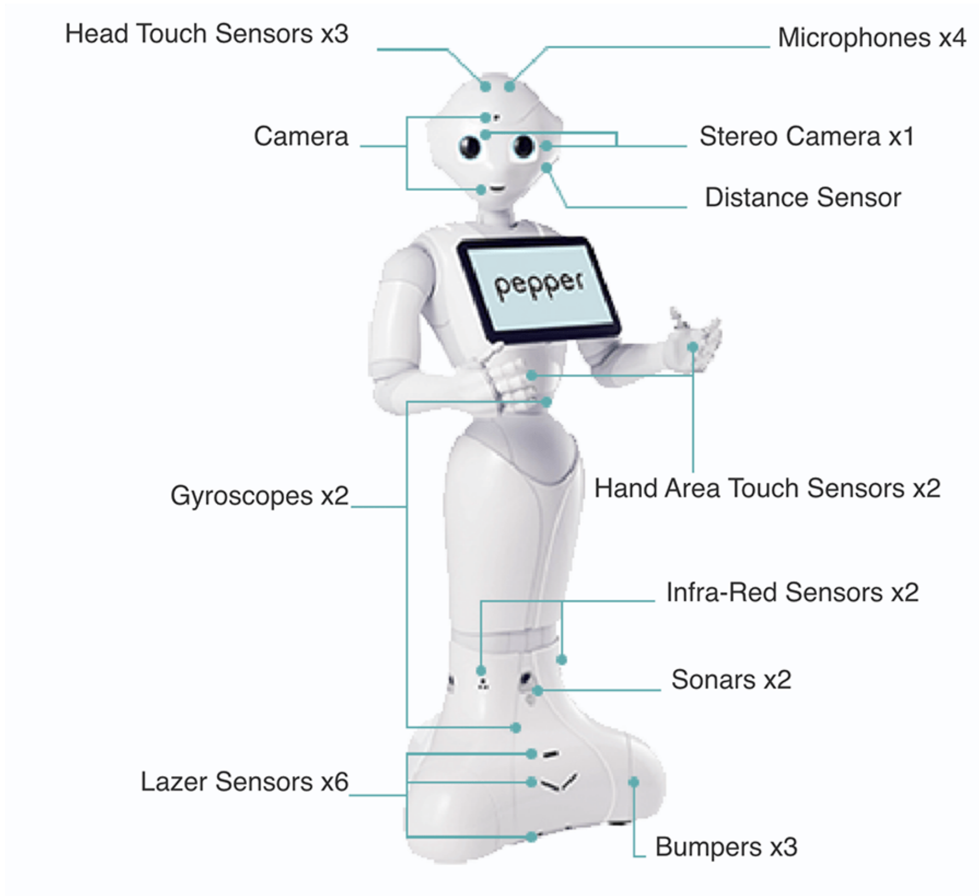

Figure 8: Thanks to its articulated multi-modal system, it can best percept the environment, interact with humans, and move algimentally in dynamic and interactive environments.

4 Implementation

We utilized Choregraphe for our project implementation. Choregraphe, developed by SoftBank Robotics, serves as a software tool specifically designed for programming and managing Pepper, a humanoid robot. It boasts a user-friendly graphical interface, enabling seamless creation, testing, and deployment of behaviours for Pepper. Through Choregraphe, users can visually craft intricate behaviours by simply dragging and dropping modules to define actions, conditions, and transitions. Additionally, the software offers animation and motion editing tools, empowering users to meticulously adjust the robot's movements and gestures. It encompasses features for creating and organizing dialogues and simplifying speech interactions with the robot. Furthermore, Choregraphe incorporates a simulation mode, facilitating the testing of programs within a virtual environment. In essence, Choregraphe presents a comprehensive platform for programming Pepper robots, catering to diverse applications such as education, research, and entertainment. We employed Choregraphe to simulate Pepper and code the programming aspect of our project using the box scheme method. This method revolves around individual boxes, each facilitating a specific action such as walking, speaking, or moving. Among these boxes are predefined ones like "say" or "go to," which enable the robot to perform basic tasks. Additionally, there are boxes like the "Python script" box, allowing customization by writing code tailored to specific requirements. Each box features at least one input field to trigger its function or receive information and an output field to either activate another box or transmit information. These inputs/outputs can take on four types: Bang, number, string, or dynamic. A "Bang" is a simple signal to initiate the box's execution, while numbers and strings represent values that the box can utilize. The "dynamic" type can encompass any of the preceding types. This ver-

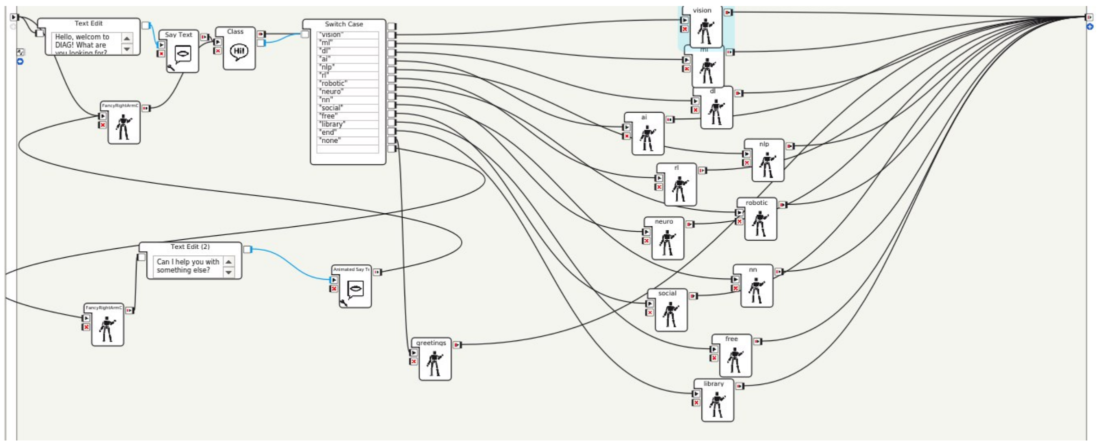

Figure 9: A screenshot of Choregraphe showing a visualization of logical blocks scheme. DIAGhino actions are concatenated in a consequential fashion, based on what it perceives from the environment.

satility is crucial as it facilitates sequencing the box executions and passing along values or information acquired through interactions, which remain useful for future actions. The execution process begins with an initial start and concludes with a general end to the program, facilitating the simultaneous passage of inputs to multiple boxes and enabling the concurrent execution of multiple actions. For testing and simulation purposes, Choregraphe incorporates a simulated map, a blank space where Pepper can freely navigate, utilize its tablet, and engage in dialogue or interaction with the user via a dedicated panel at the bottom of the screen, akin to a chat interface. This greatly simplifies the testing process and makes it easier to understand and correct the robot's behaviour when interacting with users. During the simulation phase, you have the flexibility to either execute all the code or selectively run specific components, thereby augmenting debugging capabilities and facilitating fine-tuning of the code. Furthermore, the virtual simulation enables a thorough understanding and correction of Pepper's behaviour in human interaction scenarios, thanks to its precise and accurate simulation program. Although Pepper may not be the latest robot on the market, Choregraphe stands out as state-of-the-art software for programming this robot, despite occasional instances of sluggishness or the need for rebooting. These challenges stem from the complexity of the simulation software and the constraints of the available CPU resources.

5 Results

Simulations were carried out with Choregraphe to evaluate the correct performance of the task and the behaviour of DIAGhino. The following images, fig.11, fig.10 and fig.12 show some screens taken from the videos that were produced attached to this report to show how it works. The three proposed videos are explained below. In "not existing class.mp4" video is an example of a first-level interaction with DIAGhino. The video shows the visualization of the simulation, in which at the top is the robot and at the bottom are displayed the sentences that the user and DIAGhino exchange. In fig.10 a screenshot of the described video is reported. The virtual environment in which the robot is represented is bare of furnishings and structures, but the robot has in mind a representation of the environment like that shown in fig.6. In this example, the user asks for directions to find out in which class the lecture regarding Large Language Models is being taught, which is not on the timesheet available to DIAGhino, which then, having reasoned about it, tells the user that he probably has the wrong building and suggests that he checks the university's Web site. DIAGHino then begins a new series of commands asking the user if he needs anything else. Receiving a negative response, DIAGHino salutes by making his hat gesture and returns ready to serve the next person. An example of second-level interaction with DIAGhino can be seen in "Computer vision class.mp4." The person interacting with the robot explains that he is looking for Professor Amerini's class. DIAGhino reasons about the request, finds the right information and reports back to the user all the data regarding the class, i.e., the time, classroom, and subject name. Knowing the current time, DIAGhino understands that the lesson being discussed has already started, and so to make the user not waste any more time, he directly tells the user to follow him and begins his journey to the classroom. As he accompanies him, being followed, DIAGhino tells about the building and Antonio Ruberti. When they arrive in front of the classroom, the automaton turns around, shows the classroom entrance, and greets the late student. In fig.11 There is a screenshot of the video in which DIAGhino can be seen walking toward the classroom while talking to the user while gesturing. In the last video, "free rooms to study.mp4", a less deterministic interaction is shown because the possible correct answers to be given to the user are multiple. In fact, it can be seen that what DIAGhino is asked for is " ...a free room to study." Based on its knowledge, the robot knows that the classrooms considered suitable for studying are either the library at any time it is open or classrooms where no classes are currently being taught. Understanding that it has been asked for a place to study and not a specific place, DIAGhino could indifferently answer one of the free places between the library or classrooms. Instead, the answer it provides to the user is that there is definitely the library as a congenial place to study but it also provides him with the list of currently vacant classrooms. It also asks the user if he would like to be escorted to one of these free classrooms, then receiving as a response the name of a classroom among the ones DIAGhino has listed, it starts escorting him in that direction, from which follows actions similar to those described in the previous video. As in all other simulations, once DIAGHino is dismissed, he returns to his starting position.

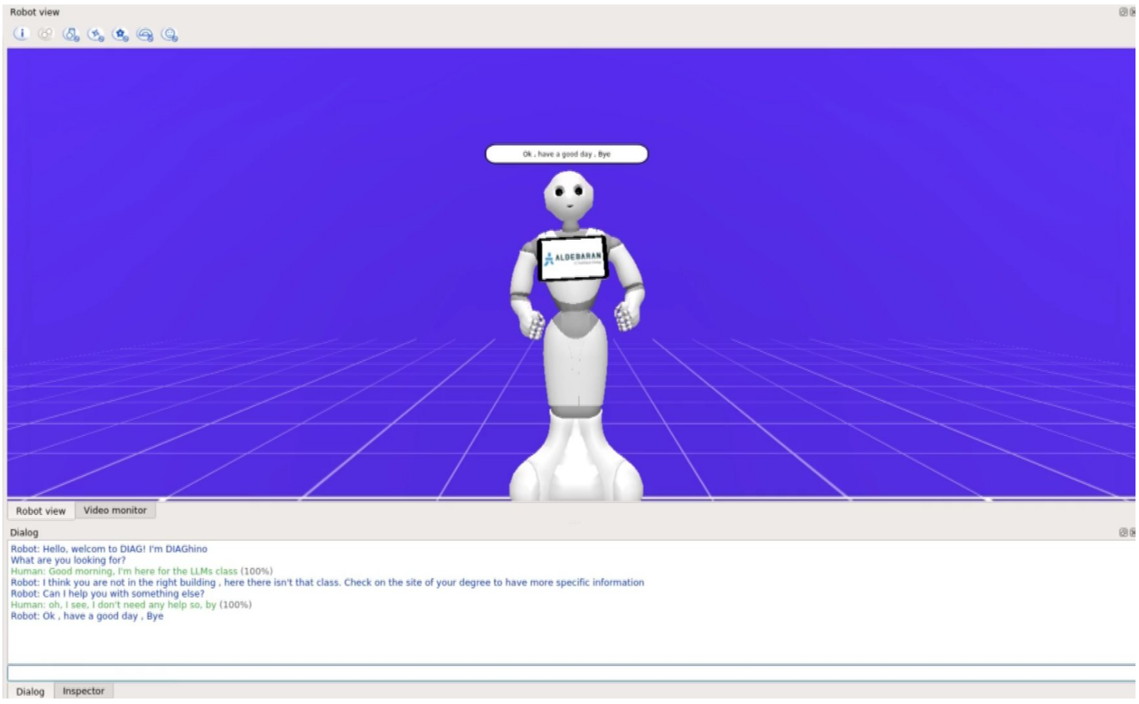

Figure 10: An example of an interaction in which the robot, receiving a request for information for a class that is not in the knowledge, suggests that the user go check on the university website

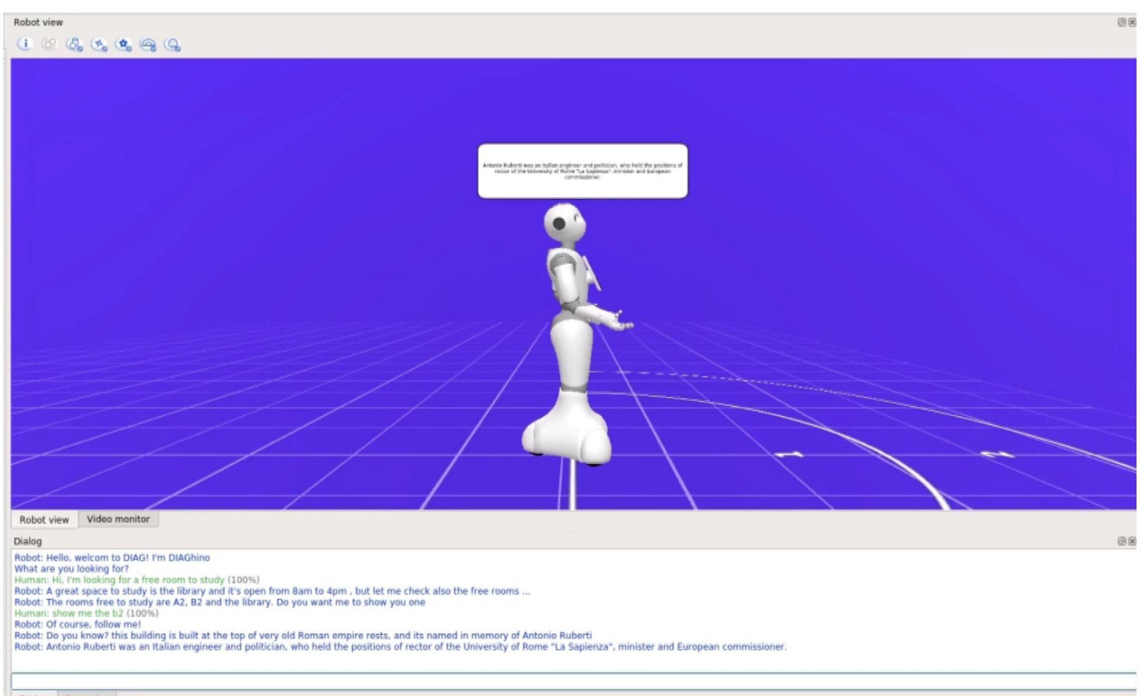

Figure 11: An example of an interaction in which the robot is escorting a user to a classroom. On the way, DIAGhino gestures and tells the story of the building and the DIAG department.

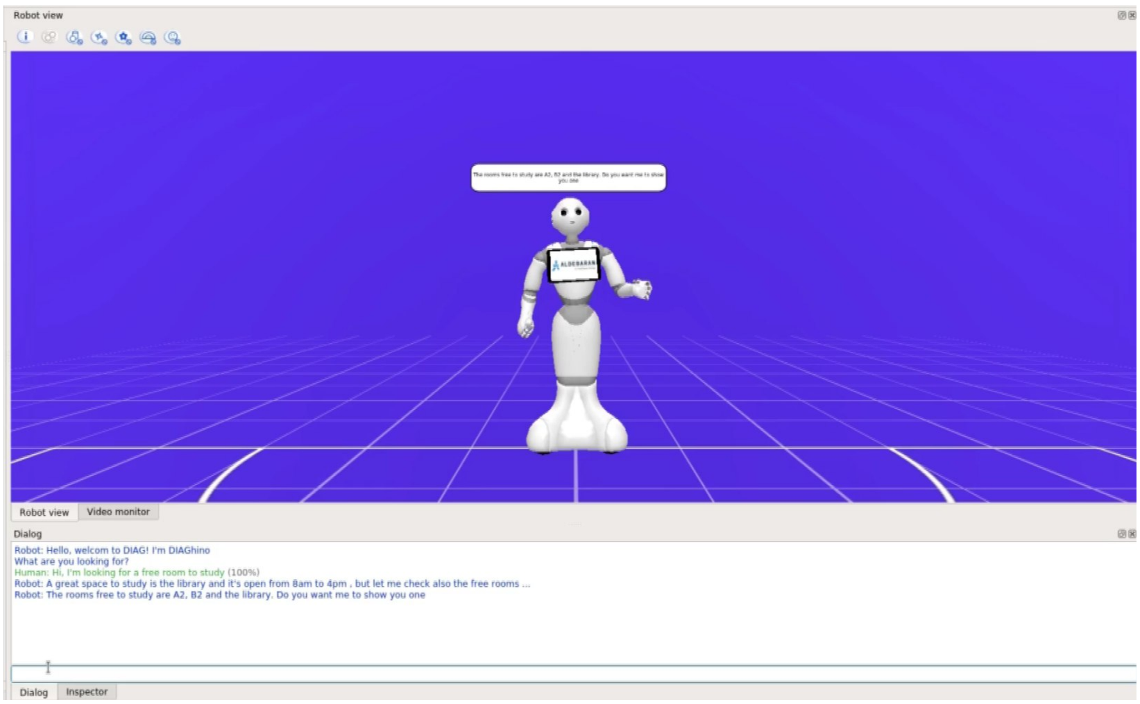

Figure 12: An example of an interaction in which the robot tries to figure out what the user wants to do since there are multiple alternatives to choose from.

6 Conclusions

This paper presented the implementation of a robot receptionist, DIAGhino, capable of answering questions from visitors to a university and escorting them inside the building. The relevant literature was first reviewed to go over the main aspects and complications that have been encountered in applications similar to this one. Approaches were studied with regard to both the HRI and RA domains. Then the problem and its setting were explained in depth. DIAGhino reasoning capabilities were explained and implementation choices were defined at the level of dialogues, gestures, and body language that fall under the aspect of its social capabilities. Finally, videos of actual simulations proposed to show the actual capabilities of the robotic agent were described. This implementation is based on many assumptions and simplifications. Starting from the environment in which the robot performs its tasks to the absence of dynamic and disruptive elements while the task is being performed. In addition, simplifications were introduced about the complexity of knowledge on which DIAGhino was to develop its reasoning. Despite this, our agent is able to perform its tasks correctly within a simulated virtual environment, acting as we expected in each different interaction proposed to it. This work is a basic structure that we have created to solve the task in the simplest and most optimal conditions. Several possible steps forward can be made. One of the possible new functions that could be introduced is that of managing multiple people at the same time. If two people showed up together at DIAGhino asking to go to two different classrooms, a reasoning module could be developed that takes into account the map of the building and the distances of the different places. At this point DIAGghino should plan the optimal route to take the two people to their two destinations while minimizing the distance travelled. This would make its capabilities more robust to more realistic operability conditions. Another aspect to implement is his motor behavior in case other people move in the same environment as him. As seen in the papers covered in the Related Works, the movement choices made by the robot are also important to convey safety, reliability and competence. Based on what it perceives with its sensors, movement choices should be introduced, such as stopping and letting pass or just trying to divert other people. These capabilities, in addition have to be supported by an implementation capable of managing possible unexpected events during the movement of the robot, but it must be studied from a social aspect based on studies done on the subject. The more complex and complex the environment becomes, the more modules are needed in the implementation to maintain robust behavior like the one demonstrated in our simulations. Another big step forward would be made by carrying out tests in a real environment on Pepper models. We are sure that dozens more implementation needs would arise when testing the robots, making the process of developing and improving the project much easier. From the results obtained it is much easier to evaluate the actual capabilities of the reasoning module rather than the social and relational qualities. All the choices we made in developing DIAGhino's personality are based solely on ideas from the other studies analyzed and on our beliefs regarding what a robot of this type should be like to be pleasant and encourage its use. However, our beliefs are not enough to evaluate DIAGhino's actual social skills. For this reason, a further step forward would be to be able to test our implementation in a real environment with real users and then ask them to fill out a survey, so as to be able to define a statistical and scientific evaluation of the HRI module. In developing this project we had to address questions regarding the behavior of robots that we took for granted or to which we did not associate particular relevance. In having to decide how to develop DIAGhino's personality, we realized how complex it is to generate an object that must have characteristics that are unique to humans. We stopped and questioned what makes human behaviour "human" and then we understood that transmitting our ideas through means of communication that are limited in number, such as only the voice and gestures, compared to the entire set on which it can rely a human being makes it even more difficult. The ad hoc definition of phrases and movements to make DIAGhino socially valid becomes increasingly difficult as the different behaviours and tasks that he may have to carry out increase. We also understood that to develop more complex systems we need to use completely different architectures from those used for this project, which as seen before are partly limited and difficult to scale. The recent and innovative applications in the field of Natural Language Processing are already demonstrating that they can increase the capabilities of the reasoning and behavioural systems of robots, opening the doors to new possible future scenarios in this field. Furthermore, the most cutting-edge humanoid robot models that companies have unveiled in recent years allow us to imagine even more specific and robust collaborative robotics applications, where the strong movement capabilities of these automata can help a lot in developing feelings of acceptance, trust and interest in the humans who interact with them, to improve and expand the use and applications of these technologies.

References

[1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.

[2] A. Ramesh, M. Goyal, F. Chollet, M. Paluri, M. Mathieu, T. Aila, and A. Radford, "Zero-shot text-to-image generation," *arXiv preprint arXiv:2102.12092*, 2021.

[3] Boston Dynamics, "Atlas Gets a Grip — Boston Dynamics," 2022.

[4] N. Karnik, U. Bora, K. Bhadri, P. Kadambi, and P. Dhatrak, "A comprehensive study on current and future trends towards the characteristics and enablers of industry 4.0," *Journal of Industrial Information Integration*, vol. 27, p. 100294, 2022.

[5] M. Kyrarini, F. Lygerakis, A. Rajavenkatanarayanan, C. Sevastopoulos, H. R. Nambiappan, K. K. Chaitanya, A. R. Babu, J. Mathew, and F. Makedon, "A survey of robots in healthcare," *Technologies*, vol. 9, no. 1, p. 8, 2021.

[6] C. S. González-González, V. Violant-Holz, and R. M. Gil-Iranzo, "Social robots in hospitals: a systematic review," *Applied Sciences*, vol. 11, no. 13, p. 5976, 2021.

[7] K. Berezina, O. Ciftci, and C. Cobanoglu, "Robots, artificial intelligence, and service automation in restaurants," in *Robots, artificial intelligence, and service automation in travel, tourism and hospitality*, pp. 185–219, Emerald Publishing Limited, 2019.

[8] J. M. Garcia-Haro, E. D. Oña, J. Hernandez-Vicen, S. Martinez, and C. Balaguer, "Service robots in catering applications: A review and future challenges," *Electronics*, vol. 10, no. 1, p. 47, 2020.

[9] J. Broekens, M. Heerink, H. Rosendal, *et al.*, "Assistive social robots in elderly care: a review," *Gerontechnology*, vol. 8, no. 2, pp. 94–103, 2009.

[10] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas): Development and validation," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pp. 254–262, 2017.

[11] S. T. Fiske, A. J. Cuddy, and P. Glick, "Universal dimensions of social cognition: warmth and competence," *Trends in Cognitive Sciences*, vol. 11, no. 2, pp. 77–83, 2007.

[12] J. Carpenter, J. Davis, N. Erwin-Stewart, T. Lee, J. Bransford, and N. Vye, "Gender representation and humanoid robots designed for domestic use," *I. J. Social Robotics*, vol. 1, pp. 261–265, 08 2009.

[13] F. Eyssel and D. Kuchenbrandt, "Social categorization of social robots: Anthropomorphism as a function of robot group membership," *The British journal of social psychology / the British Psychological Society*, vol. 51, 11 2011.

[14] M. Mori, K. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics Automation Magazine*, vol. 19, pp. 98–100, 06 2012.

[15] M. Mason and M. Lopes, "Robot self-initiative and personalization by learning through repeated interactions," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 433–440, 2011.

[16] J. López, D. Pérez, E. Zalama, and J. Gómez-García-Bermejo, "Bellbot - a hotel assistant system using mobile robots," *International Journal of Advanced Robotic Systems*, vol. 10, no. 1, p. 40, 2013.

[17] K. H. Seo and J. H. Lee, "The emergence of service robots at restaurants: Integrating trust, perceived risk, and satisfaction," *Sustainability*, vol. 13, no. 8, 2021.

[18] R. de Kervenoael, R. Hasan, A. Schwob, and E. Goh, "Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors' intentions to use social robots," *Tourism Management*, vol. 78, p. 104042, 2020.

[19] J. Behan and D. O'Keeffe, "Lucas: The library assistant robot, implementation and localisation," in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, vol. 1, pp. 1140–1146, 2005.

[20] M. Pollack, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, J. Matthews, J. Dunbar-Jacob, C. Mccarthy, M. Montemerlo, J. Pineau, and N. Roy, "Pearl: A mobile robotic assistant for the elderly," 06 2002.

[21] Y. Chen, Y. Luo, C. Yang, M. Yerebakan, S. Hao, N. Grimaldi, S. Li, R. Hayes, and B. Hu, "Human mobile robot interaction in the retail environment," *Scientific Data*, vol. 9, 11 2022.

[22] H. Admoni and B. Scassellati, "Social eye gaze in human-robot interaction: A review," *Journal of Human-Robot Interaction*, vol. 6, p. 25, 03 2017.

[23] A. Goudey and G. Bonnin, "Must smart objects look human? study of the impact of anthropomorphism on the acceptance of companion robots," *Recherche et Applications en Marketing (English Edition)*, vol. 31, 05 2016.

[24] J. Oh, A. Suppé, F. Duvallet, A. Boularias, L. Navarro-Serment, M. Hebert, A. Stentz, J. Vinokurov, O. Romero, C. Lebiere, and R. Dean, "Toward mobile robots reasoning like humans," 01 2015.

[25] A. Ohya, "Human robot interaction in mobile robot applications," in *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication*, pp. 5–10, 2002.

[26] T. Tanaka, J. Ohwi, L. V. Litvintseva, K. Yamafuji, S. V. Ulyanov, and I. Kurawaki, "A mobile robot for service use: behaviour simulation system and intelligent control," in *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS'97*, vol. 1, pp. 366–371, IEEE, 1997.

[27] G. Cui, W. Shuai, and X. Chen, "Semantic task planning for service robots in open worlds," *Future Internet*, vol. 13, no. 2, p. 49, 2021.