BRILLO: Personalised HRI with a Bartender Robot

Alessandra Rossi[1*], Silvia Rossi[1*], Maria Di Maro[1] and Antonio Origlia[1]

[1*]Department of Electrical Engineering and Information Technologies, University of Naples Federico II, Napoli, Italy.

*Corresponding author(s). E-mail(s): alessandra.rossi@unina.it; silvia.rossi@unina.it;
Contributing authors: {maria.dimaro2, antonio.origlia}@unina.it;

Abstract

The BRILLO (Bartending Robot for Interactive Long-Lasting Operations) project aims to create an autonomous robotic bartender that can interact with customers while accomplishing its bartending tasks. In such a scenario, people's novelty effect connected to the use of an attractive technology is destined to wear off and, consequently, negatively affects the success of the service robotics application. For this reason, providing personalised natural interaction while people access its services is fundamental for increasing users' engagement and, consequently, their loyalty. In this paper, we present a novel robotic system that is able to not only provide a recommanded service (from the ordering to the preparation of a drink), but also personalise the verbal and non-verbal interaction. In particular, we described the developed three-layer ROS architecture integrating a perception layer managing the processing of different social signals, a decision-making layer for handling multi-party interactions, and an execution layer controlling the behaviour of a complex robot composed of arms and a face. Finally, user modelling through a beliefs layer allows for personalized interaction. We also present the results of both people's interaction, experience and performances in a real user case. The user study involved 116 participants and showed that BRILLO is considered an easy-to-use and attractive system by the users.

Keywords: Personalised HRI, service robotics, multi-modal interaction, long-lasting collaborations, robot bartender

1 Introduction

In recent years, service robots have been employed in a variety of contexts that involve direct interactions with multi-users in public environments. A particularly challenging one is the bartending domain, which combines the complexity of efficiently manipulating objects and the need to keep users engaged for a long-lasting interaction. This, particularly, reflects the approaches of modern businesses that aim to achieve customer satisfaction and retail by presenting an equally high-quality product and service [25]. Several projects [15, 18] explored the robotic bartending domain by developing automatic serving robots that are able to serve multiple users. Long-lasting interactions, however, can be established when the service robot is capable of showing social intelligence through personalised interactions [33].

Current literature has identified several aspects that affect people's perception of social intelligence in robots. For example, a robot with facial features can be perceived as more intelligent than one without any [40], and a robot that is able to move naturally can enhance people's acceptance of the robot and convey a sense of security [14], and a robot that is able to model human behaviours and express appropriate emotions can positively affect the interactions [41]. Among those, the possibility of providing personalised services can increase users' interaction on a long timescale and their involvement with the robot [38]. Socially intelligent bartender robots that are able to mix task execution, dialogue, and social interaction in response to customers' states and intentions were more efficient than non-social ones [7]. A robotic system that integrates a "system of record" (e.g., analysing and storing past interactions, and preferences to optimise sales for specific users) and a "system of engagement" (e.g., aiming at facilitating and enhancing the experience via a personalised and natural interaction) can play a crucial role in customers relationship management.

To achieve this goal, a complex human-robot interaction (HRI) and control architecture have to be designed to comprehend different software components allowing for efficient and simultaneous execution of multiple tasks and for providing essential capabilities, such as storing past events [21], constructing models of others' actions, beliefs, and intentions [13], modelling the domain knowledge, selecting actions and behaviours, and planning [17]. For example, in the JAMES project [7], a bartender robot was able to engage participants in conversation by producing facial expressions and lip-synchronising speech. The iCub robot in Tanevska et al.'s study [37] adapted its gaze and body to convey comfort and discomfort according to the engagement of the participants. The CORTEX cognitive architecture [23] was used to allow a salesman robot to convince potential customers to follow the robot towards a selling boot. The robot was able to identify the customers, understand people's willingness to follow it and answer some specific questions.

While cognitive architectures have been investigated for a long time, real social and service robot implementations in complex scenarios, such as the bartending service, have only recently been developed [24]. Moreover, there is an inconsistency between current robots' ability to generate verbal and non-verbal expressive behaviours (such as spoken language, gestures, and emotions), and their capability to understand the situational context and engage the users in natural dialogues according to the users' intentions and desires [20]. The development of such interaction and personalization capabilities in real-world settings is still an open challenge for the HRI community due to the complexity of interaction pipelines and the lack of maturity of some of the underlying detection and processing algorithms [16]. To address these challenges, the BRILLO project aims to create a robotic platform that is able to accomplish both the expected management of a bar counter, such as drinks manipulation, and the socially intelligent interaction, context-awareness, and personalisation of the robot's behaviours. In this work, we present the cognitive architecture and functionalities developed for satisfying the multiple agents' needs involved in the interaction, in terms of preferences, moods, and differences, and at the same time increasing the customers' retention.

Robots have been used as attraction in the food and beverage industries [2], even if they did not have direct interaction with the customers. However, a robot that is not able to interact naturally may create anxiety in humans [19]. In this work, we also present the results of a user study for the evaluation of the BRILLO robot's attractiveness and people's perception of the robot during direct and bystander interactions with the system.

2 A Multi-modal and Multi-user Scenario

A typical use case scenario of a BRILLO service point includes multiple human users, a bartender robot for preparing and serving drinks, and a totem kiosk to register, recognise the users, and manage orders. We modelled the interactions considering the following aspects: a) degree of interaction modalities (e.g. formal vs colloquial) between the bartender and a customer; b) degree of knowledge shared between the bartender and a customer; c) representation of the customers and needs through personas. In order to define these requirements for our system, we collected data via phone-interviews with human bartenders, and we analysed several online videos of humans' interactions in bars registered by live cameras. From the analysis of the data, different phases of the interactions (see Section 2.2) and different personas (i.e. workers during a break, group of colleagues, family, group of friends, regular costumer and curious person) were identified. Users' first interactions are with the totem kiosk that welcomes them and allows them to register in case it is their first visit. If the users are returning customers, they are recognised via biometric facial recognition. Then, the customers are able to order a drink from the menu displayed on the kiosk, or they can decide to make their order sitting at the counter by directly interacting with the bartender robot. At the bar station, the bartender robot recognises the registered users by their faces, takes orders (in case they want to modify them or to have suggestions from the robot), and serves them. While it is preparing the drink, the bartender robot interacts with the users according to their profile which is built upon typical customers' personas (such as workers on a lunch break, groups of friends or family members, and regular users) and needs, such as previous orders, knowledge about the user's general interests, observing their engagement levels, processing their moods based on the sentiment analysis of their dialogues and facial expressions [30, 36]. The bartender robot is also able to manage multiple orders and users, by opportunely scheduling and adapting its behaviours.

The richness of the one-to-one and many-to-one human-robot interactions in the above-mentioned scenarios requires the development of a complex and sophisticated HRI architecture that allows the robot to show a social comprehension of the context and other agents involved in the interaction, and, at the same time, generates matching verbal and non-verbal socially acceptable behaviours. The bartender robot, for example, needs to intelligently adapt its dialogues, pose and gestures, according to the user's needs, in terms of situational context (drink orders, group dynamics, etc.).

2.1 The BRILLO Bartender Robot

We adopted a minimalist anthropomorphic structure for the bartender robot (see Figure 2). The robot has two Kuka[1] LBR iiwa 14 R820 robotic arms (each of 7 DoF and gripper), attached to a fixed-torso (only one arm has been integrated into the current version of the robot), and a Furhat Robotics head[2] (called Furhat, 3 DoF). The bartender robot is equipped with a variety of external sensors to improve and support its capability to perceive and assess the environment, the users, and the activities of the other agents involved in the interaction. In particular, a 4x2MP IR 180° Multi-sensor Panoramic Network Bullet camera[3] is mounted under the Furhat head and two microphone arrays[4] are placed on the right and on the left of the robot to perform source separation and noise reduction to isolate the customers' voices. The BRILLO robot is shown in Figure 1.

Currently, this robot is able to prepare smoothies, cocktails and variations of the two types of drinks. The choice to adopt an anthropomorphic structure for the robot comes from the evidence that people are social entities, and they are more comfortable interacting with agents that can show social behaviours [28].

2.2 Interaction Patterns and Dialogue

The interaction pattern with the bartender robot is composed of the following phases: 1) greetings or greetings and wait; 2) recommendation; 3) orders and changes request; 4) order confirmation; 5) personalised casual interaction; 6) complimentary close. In the presence of a single or two users, the robot welcomes them ("greetings") and, then, recommends drinks, while in the presence of multiple users, the robot welcomes them, and invites them to wait their turn to be served ("greetings and wait"). After the recommendation of the drinks, the customers can place their order and/or ask for order customisation ("orders and changes request"). If the users do not ask for any other change, the order can be confirmed to the robot ("order confirmation"). During the drink preparation phase, the robot can dialogue with two users at a time to entertain them. The robot engages the customers in casual dialogues based on data collected in previous interactions if they are recurring customers, or by building interactions at run-time. According to the social feedback received from past interactions or current turns, appropriate topics of conversation are selected. The robot

[1] Kuka Robotics https://www.kuka.com

[2] Furhat Robotics https://furhatrobotics.com

[3] https://us.dahuasecurity.com/?product=4x2mp-ir-180-multi-sensor-panoramic-network-bullet

[4] PureAudio USB Array Microphone

Fig. 1: The BRILLO robot.

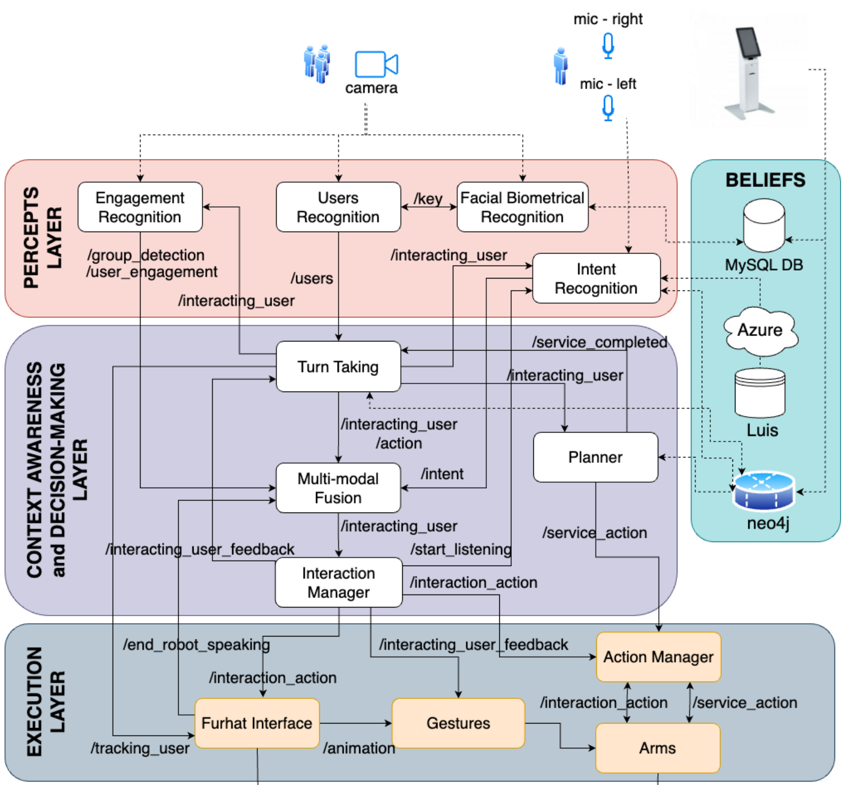

Fig. 2: Overview of the BRILLO project's ROS architecture and topics. It is composed of a component to build the Beliefs and three principal layers: 1) Percepts, 2) Context-awareness and Decision Making, and 3) Execution.

closes the interaction after it prepared the drink ("complimentary close").

3 ROS Architecture

In a recent survey on cognitive architectures, Kotseruba and Tsotsos [8] identified seven core cognitive abilities: perception, attention mechanisms (used in multi-user interaction), action selection, memory, preference learning, reasoning, and metareasoning. We started from such core cognitive abilities (excluding metareasoning) to develop the functionalities for enabling the BRILLO system to perceive the users, build dynamic and efficient interactions while adapting the robots' behaviours to the users' needs and preferences.

The global architecture adopted for the BRILLO project is shown in Figure 2. Our system is based on a three-layer control architecture that follows a user-centred approach considering possible expected interactions types and customers' needs as highlighted by previous interviews with human bartenders and observations of humans' interactions and behaviours in real bars.

From a software engineering perspective, the developed modules operate asynchronously by means of ROS nodes and communication via topic subscriptions.

3.1 Percepts Layer

The *Percepts layer* manages the robots' multimodal perception capabilities which consist of the information obtained by the modules for processing the visual and speech inputs.

3.1.1 Engagement Recognition Node

This node uses the inputs from the camera and microphones to evaluate the level of engagement and the affective responses of the users during in the interaction.

User, Group, and Engagement

Authentication mechanisms are widely used both in online and mobile systems. Researchers have been conducting extensive efforts to create and improve authentication systems, such as biometric-based, that can be faster and easier to manage than password account management [34]. The biometric-based authentication systems may use data such as voice, iris, fingerprint, palm print, and face. In a bartending scenario, multiple user access and complex interactions are expected, therefore, the recognition of the user needs to be developed considering real-time constraints, such as disturbances caused by noisy and vast areas, the presence of many users, and the need for fast service to avoid crowds. The most used and appropriate technique for recognising users in similar scenarios is face recognition [34].

The identification, detection, and tracking of the users in a BRILLO scenario are carried out using two cameras (an RGB camera at the totem, and a panoramic camera at the bar station) that allow the agent to collect data in real-time. The processed visual information is used for the following reasons: 1) face biometric data to identify a registered customer; 2) customers' engagement via their body pose; 3) user tracking while at the counter; and 4) group recognition.

The data from the camera are processed by YOLO[5] for object detection, while the face recognition is done through the OpenFace library[6]. The combination of these two techniques allows accurate results even with different lights, lower quality of the frames and not fully frontal faces. The user pose is estimated using a Skeleton-Based approach through the OpenPose library[7]. The system estimates whether a user belongs to a group of other people using a Multilayer Perceptron classifier trained on the dataset called Ego-Group which is one of the few public datasets having a robot egocentric view of the surrounding space [35]. Moreover, a Multilayer Perceptron model based on the user's pose and group information, and trained on the same dataset, classifies the user's engagement with the robot. Screenshots from the used dataset with the results of the Engagement recognition module are shown in Figure 3. On 5-fold cross-validation, an average accuracy of 94.33% was achieved for engagement prediction and 97.12% for group identification.

Evaluation of the Affective Response

Engagement and group detection information are here assessed also with emotion recognition from facial expressions that is carried out using the software called Affectiva[8]. The classification of

[5]YOLO library https://pjreddie.com/darknet/yolo/

[6]OpenFace library https://github.com/TadasBaltrusaitis/OpenFace

[7]OpenPose library https://github.com/CMU-Perceptual-Computing-Lab/openpose

[8]Affectiva software https://www.affectiva.com/

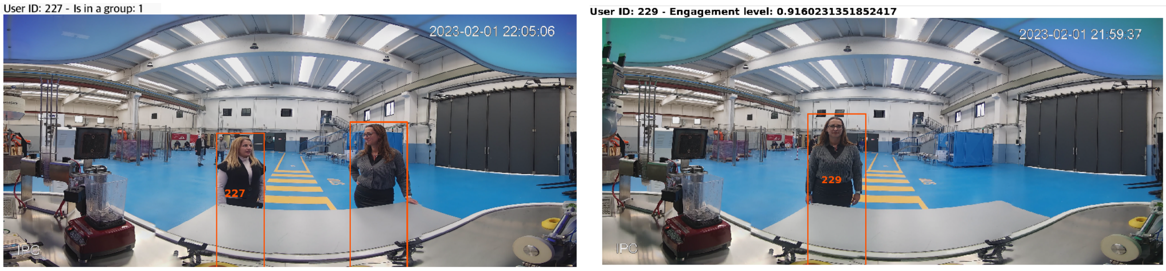

Fig. 3: Screenshot of results from our engagement and group recognition modules: (a) User 227 in the figure on the left is recognised as in a group; (b) User 229 in the figure on the right is recognised as engaged with the robot.

people's emotions via Affectiva allows a classification of facial expressions according to seven main emotions (anger, contempt, disgust, fear, joy, sadness, and surprise). It also allows measuring the positive or negative valence for measuring the experience, every few hundred milliseconds, and with high accuracy.

The analysis of the voice, both simultaneous speech and high entropy, is implemented using the Python library called My-Voice-Analysis[9]. It allows for evaluating people's mood (neutral, calm, or pacey), gender (female or male), speech rate, energy, frequency, average speech interval duration, and speaking duration. Moreover, to build a natural and fluid interaction, a social robot needs to be able to understand sentiments hidden in the content of the user's speech [1]. Therefore, the user's voice is also further analysed to classify the emotions in the text. Currently, the sentiment analysis is carried out using the Azure Cognitive Services Text Analytics libraries[10] which label the speech-to-text as positive, negative, or neutral.

3.1.2 Intent Recognition Node

A bartending stand is typically found in noisy areas where a lot of people chat and, possibly, where music is playing. For this reason, to enable speech-based interaction, it is necessary to equip the stand with adequate hardware to isolate the customers' voices. This module makes use of the following different sub-processes: 1) automatic speech recognition (ASR) system to obtain the speech transcription, and 2) machine learning approaches for Natural Language Understanding (NLU).

The BRILLO bartender stand is equipped with two microphone arrays to perform source separation and noise reduction to isolate the customers' voices. The audio stream is processed remotely using an instance of the Azure ASR service[11], which produces the utterances' transcriptions.

Concerning the NLU module, the following intents are modelled using the Microsoft Azure Service LUIS[12]:

- AnswerGreeting: this answers to the greeting phase initialised by the robot; this phase can also include information concerning the user's state.
- Order: with this intent, users can place a drink order.
- OrderConfirm: with this intent, an order is confirmed.
- OrderReject: with this intent, an order is conversely disconfirmed.
- NewsConfirm: the user confirms to want to listen to proposed news during the drink preparation phase.
- NewsReject: the news proposed during the preparation phase is not well received, and therefore rejected.
- NewsStop: the user asks to stop with the news phase;
- Evaluation: with this intent, a user can provide an evaluation of a proposed drink.
- GetFlavour: with this intent, the flavour and/or the ingredient contained in the drink are retrieved.

[9] My-Voice-Analysis library https://github.com/Shahabks/my-voice-analysis
[10] Azure services https://azure.microsoft.com/
[11] Azure services https://azure.microsoft.com/en-gb/
[12] https://www.luis.ai/

Table 1: Intent Recognition Performances

Intent	Precision	Recall	F-score
AnswerGreeting	1	1	1
OrderConfirm	0.83	0.83	0.83
OrderReject	0.75	0.75	0.75
Order	1	0.81	0.9
NewsConfirm	0.86	1	0.92
NewsReject	0.88	1	0.94
NewsStop	1	0.75	0.86
Evaluation	1	1	1
GetFlavour	0.80	0.80	0.80

- None: if the intent is not recognised, the robot asks the user to repeat their sentence.

As far as orders are concerned, the user intents are modelled for the application domain, considering the type of product, possible modifications, and cancellations.

Intent performances were computed by dividing the examples collected in the data set by subdividing them into a training set (80%) and a test set (20%). The results are shown in Table 1. On average, the intent recognition module performed well with an average $F - score = 0.89$.

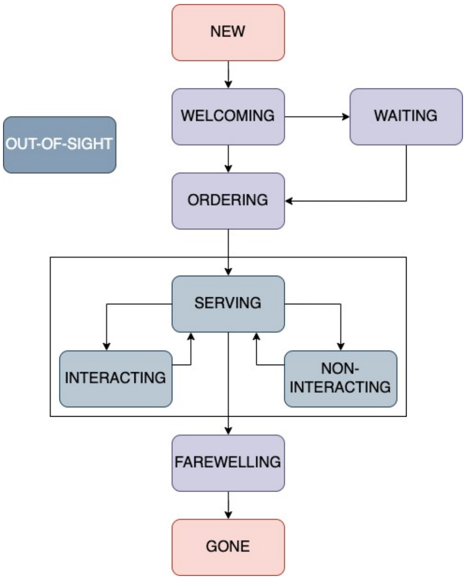

Fig. 4: Possible interaction states for each user. Transitions between states are achieved by means of service and interaction actions. Users can transition from any state (except for "gone") to an out-of-sight state, which will be re-assigned once the robot recognises them again.

3.2 Users' Beliefs

According to Kotseruba and Tsotsos [8], different types of memory can be identified in a cognitive architecture. In our architecture, a short-term memory stores the information related to the current users and situation states (e.g., current customers' engagement state, users' interaction states, see Figure 3).

A working memory keeps track of the global state of the systems in terms of the list of orders to be served, that are represented in terms of goals to be reached, the current active intentions to be executed (drinks currently in preparation), and their plans of execution that are composed of both service and interactive actions.

Long-term memory stores information about users' static information (i.e., personal data), preferences on drinks and topics of conversation, and previous interactions' history. The user's personal data are stored in a MySQL database, while the length of the interaction, topics found of interest for the conversation, and interaction preferences, such as the type of the ordering (at the totem or the bar), and an average estimation of the engagement are stored with the drinks orders.

Finally, semantic memory is used to store semantic information (ontology-like structure) relying on the use of the Neo4J[13] graph database platform. The relationships between orders, cocktails and drinks, ingredients, and flavours are connected in a semantic graph that also includes the flavorDB[14] database. The user's past interactions stored in the long-term memory are also associated with the semantic graphs.

3.3 Context Awareness and Decision Making Layer

This layer stores and processes the proper context-awareness information the robot needs to produce socially acceptable and natural behaviour and prepare the drinks. Moreover, such information is used to decide the user to be served/interacted with and the action plan.

[13] https://neo4j.com/
[14] https://cosylab.iiitd.edu.in/flavordb/

3.3.1 Turn-taking Node

The BRILLO system instantiates a desire for interaction and drinks order for each recognised user (*/users* topic). To accomplish such a goal, we defined a set of possible interaction states for the user, starting from the initial greetings towards the serving and farewell state (see Figure 4). Each active user is currently in one state and, to transition from one state to another, the proper set of actions (both service and interactive) have to be planned. At each interaction cycle, the Turn-taking module selects one active user at a time. It instantiates an intention to transition from one state to the following. After such a transition is obtained the system begins with a new cycle.

At each cycle, the robot decides on the active user by considering:

- The arrival order;
- The users' associated personas in terms of the willingness to interact with the robot;
- The engagement level. The robot wants to maximize overall engagement;
- Whether the user is in a group or not. Users within are expected to be interacting more with their peers than the robot;
- The time passed from the last interaction with the user.

These are implemented through simple empirical rules. Nevertheless, the robot tries to involve all users, albeit to a different extent, to increase engagement [3]. The user and their current interaction state are shared on the */interacting_user* topic. States and transitions as represented in Figure 4.

3.3.2 Multi-modal Fusion Node

The multi-modal fusion module for the human assessment is deployed to keep track of the user's emotional response by merging the data (sentiment from text semantics, voice and facial expression) as elaborated by the engagement module. This information is passed to the Interaction Manager module to decide which appropriate interactive action should be executed and how such action should be executed.

3.3.3 Robots' Actions Planner

In our system, we distinguished two types of actions:

- *Service actions*, that are the bartending actions necessary to prepare and serve a drink. It may involve one or both arms of the robot;
- *Interactive actions*, that are the robot's actions for interacting and entertaining the user while they are at the bar station. These actions may be verbal utterances, gestures (when one of the robot's arms is not engaged in any service action), and facial expressions.

3.3.4 Planner Node

Orders (service actions) are processed by the robot as soon as the active user is in the serving state. ROS plan libraries for AI planning[15] are used to schedule service actions according to the scheduled orders for each user. The Planner module creates, therefore, a sequence of basic actions for each arm and for each ordered drink to be executed by the robot. The trajectories necessary to achieve each basic action to serve a drink are pre-recorded and basic actions are expressed in terms of preconditions to be checked before the execution (e.g., the mixer is empty) and the time to execute each action.

3.3.5 Interaction Manager Node

The selection of the interactive actions to be used for the selected user relies on Influence Diagrams[16], which integrate the probabilistic estimates of engagement coming from the users' interaction modes (i.e., sentiment from speech, facial expressions, semantics) with a utility estimation linked to the possible speech acts or gesture. Bayesian networks allow handling probabilistic input coming from the NLU module, among others, to take into account confidence measures when selecting the next action. The utility of each possible machine move (Action) is a function of the chosen Action and of the actual intent of the user. At decision time, the system calculates an estimate of the actual intent base on the probability

[15] ROSPlan's source code and documentation https://github.com/KCL-Planning/ROSPlan

[16] In BRILLO, Influence diagrams are implemented using the AGRUM library[4] https://agrum.gitlab.io/

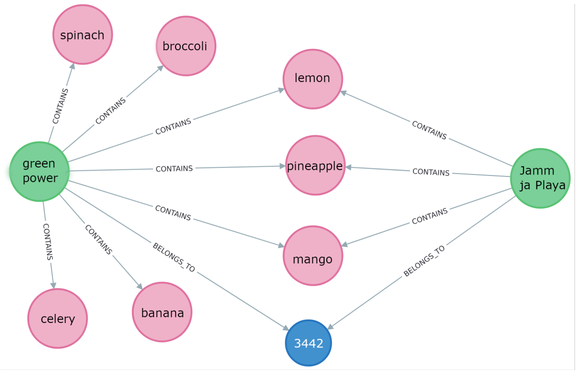

Fig. 5: Graph showing the similarity between the last ordered DRINK node, *Green Power*, and another DRINK node (in green) belonging to the same category SMOOTHIE (in blue). The similarity is here computed on the highest number of FOOD_INGREDIENT nodes (in pink) in common.

that each intent is correct, and assigns a utility value to the dialogue move which consists of a repetition request (AskRepeat). The utility of each possible Action is computed based on the probability distribution over all intents. Finally, the system selects the Action with the highest utility.

Interaction customisation is applied in this phase when the user is known to the system. Pragmatic models of interaction are investigated and included in the architecture. For instance, clarification requests (CRs) expressing counter-expectations for past ordering actions are adapted to engage the user in the dialogue and add further details to the user model. For order confirmation, CRs with a confirmation function are employed, i.e., when the CR initiator has some kind of hypothesis [22]. Specifically, this is useful when low confidence scores occur due to background noise or multiple orders.

To support recommendations, previous interactions with the users are stored in the Neo4j database and used to build a profile. For past interactions involving drinks that were not previously selected, the robot asks for explicit feedback, in the form of a rating on a scale of 1 to 5. Using the graph structure represented in Neo4j and, in particular, the data coming from FlavorDB (see Figure 5), the system is able to compute drinks' similarity in terms of shared ingredients. On this basis, different recommendation strategies can be selected according to different parameters: a) type of persona, a not mandatory piece of information asked at the totem kiosk during the registration phase; b) degree of knowledge of the user based on the number of interactions (known user: at least one previous interaction; unknown user: first interaction); c) evaluation of past consumed drinks (positive evaluation: ≥ 3; negative evaluation: ≤ 2); d) acceptance of the recommendation (if the recommendation is not accepted another strategy is selected) (see Table 2).

For entertainment purposes, the robot is also able to present news extracted daily from different sources (i.e., Italian Twitter page for news, called Ansa[17], and Italian Twitter Comic page, called Lercio[18]) and categories (e.g., politics, science, sports, culture) [32]. Explicit feedback is also used in this case to support user profiling and to select topics from the most appropriate categories. In particular, the robot initially suggests a topic of interaction, based on similarities among users and/or personas; once a topic is accepted, the corresponding news is selected from a serious or entertaining source; afterwards, the user is asked to give feedback concerning their interest about the news. If the feedback is positive, another news belonging to the same category is proposed, otherwise the category is changed. News items can be provided as long as available, or the customer is still willing to listen to them (i.e., they can stop the robot by telling it - for example, they could say something like "Enough news"). Other social signals can also cause the robot to stop entertaining the current client.

Moreover, social recovery strategies are considered to account for interaction failures [26]. For this reason, the BRILLO robot uses social interacting behaviours, such as paying attention during a conversation, being transparent on the process of thinking and elaborating a response, to be perceived with higher anthropomorphism and animacy. Moreover, since verbal irony may help to dismiss controversial and incongruous issues [9], the BRILLO robot uses it to produce sarcasm and irony when there are misunderstandings of the contextual situations, such as when the users' facial, voice, and content of the text produce ambiguous results [31].

[17] Ansa website https://www.ansa.it
[18] Lercio website https://www.lercio.it/

Table 2: Recommendation Strategies for drinks

Persona	Recommendation Strategies		
worker	known user	a) client's preferred drink	
		b) users' most ordered drink	
		c) ask what the client wants to order	
	new user	a) users' most ordered drink	
		b) ask what the client wants to order	
other (specified or not)	known user	last drink was positively evaluated	a) similar drink from the same category (smoothie or cocktail)
			b) similar drink from another category
			c) users' most ordered drink
			d) ask what the client wants to order
		last drink was negatively evaluated	a) users' most ordered drink
			b) ask what the client wants to order
	new user	a) users' most ordered drink	
		b) ask what the client wants to order	

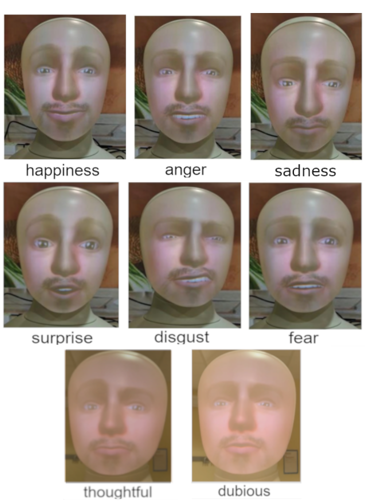

Fig. 6: Affective expressions created for the Furhat robot.

3.4 Execution Layer

The *Execution layer* manages the orders and the interactions of the robots with the users according to the knowledge and beliefs of the agents, and the situational context.

3.4.1 Action Manager Node

The Action Manager module acts as a scheduling mechanism to improve the efficiency of making the drinks by using one or two arms in parallel. It schedules the interactive actions, related to the robot's gestures, and the service actions by orchestrating the robot's arm movements. The next actions are chosen to balance the expected contribution of the action towards the goal of preparing a drink, and the robot's actions needed to maintain the engagement and entertainment of the users to their social expectations. For example, the robot might engage the users in a casual conversation while preparing their drink to foster a more endearing interaction if its arms are not busy in the drink preparation.

3.4.2 Arms Node

This module is a wrapper for favouring the communication between the BRILLO system and the two KUKA arms' Programmable Logic Controller. The controller has been implemented using the KUKA SunriseOS system to securely manage the arms, the grippers and each movement in the stand. BRILLO has seven different movements to simulate liveness. The designed gestures are: waving to salute people, raising and lowering its arm to simulate breathing, and randomly extending one of its arms while it is talking with a user, a typical Neapolitan gesture to show that it did not understand what the person replied to the robot, yawing and stretching its arms to catch people's attention when it is not interacting with them.

3.4.3 Furhat Interface Node

To make the interaction more engaging, the robot is endowed with a module in charge of guiding the Turn-taking through gaze and speech pauses [32]. The system, in particular, manages the behaviours of this robot by considering the presence and the absence of the user's attention. When a low user's engagement is perceived, the robot produces facial expressions and vocal sounds to catch the user's attention (e.g., clearing its throat). During a dialogue, the robot adapts its facial expressions and vocal sounds whether it is listening to the user's speaking, understanding or not understanding the speech, and expressing an emotion about the topic of conversation. The facial and vocal sounds are based on the Facial Action Coding System (FACS). In particular, we used Action Units that compose the FACS for endowing the Furhat robot with two sets of facial emotions: 1) we defined Ekman's 6 basic emotions [5], which have been implemented using Python and Furhat Robotics' APIs, and 2) we generated a greater and more personalised variety of synthesised human-like expressions using a two-tier Generative Adversarial Network, as described in [10]. An example of the six emotions created for the Furhat robot is shown in Figure 6.

4 In-the-field Scenario

To carry out an "in the field" evaluation of our system as a whole, with a focus on the interaction phase in a non-controlled environment, we tested it during a national faire called Maker Faire, that was held in Rome. To this extent, we collected people's responses to evaluate the quality and efficiency of both services and the interaction. In particular, we evaluated the execution times for the preparations of the drinks and interacting gestures, the perceived usability, the dialogue length necessary to achieve the domain task and the related communicative goal, and the acceptance of the recommendation provided.

4.1 Approach

We used a between-subjects design in which we manipulated people's participation in the interaction with BRILLO by asking them to either actively interact with the robot (**A-HRI** condition - active) or evaluate the system by observing other people interacting with the robot (**O-HRI** condition - observing). During the **A-HRI** condition, participants were in front of the robot, and they directly talked with the robot by placing an order and talking with it according to their news preferences; while in **O-HRI** participants' engagement with the robot was limited to just observing the robot, how this served the participants who were actively interacting with the robot, and other people's interaction with BRILLO. In this latter condition, participants were behind or at the side of the people who were interacting with BRILLO under **A-HRI** condition. In this condition, we wanted to evaluate whether the BRILLO robot attracts bystanders' attention and positively affect them even if they are not currently engaged. Based on our findings in [32], participants were served two at the time.

To measure the system's performances and people's dialogues with BRILLO and their perception of the interaction and the robot, we collected subjective and objective measures. In particular, as objective observations, we collected drink preparation times, response times, percentage of speech and intent recognition, length of dialogues, etc. To evaluate participants' perception, in both conditions, we asked them to complete two sets of questionnaires: User Experience Questionnaire (UEQ) [11] and Unified Theory of Acceptance and Use of Technology (UTAUT) [39]. Participants completed the questionnaires at the end of the (active and observed) interactions.

4.2 Participants

We collected responses from 116 people. We excluded some participants because they did not answer all the items of the questionnaires. Each participant was assigned to one condition. For the UEQ, participants were overall distributed among the two experimental conditions as follows: 1) 61 participants in the **A-HRI** condition; 2) 49 participants in the **O-HRI** condition. From the initial sample of participants, the resulting group of participants completing the UTAUT questionnaire is composed of 59 people for the A-HRI condition and 48 people for the O-HRI condition.

4.3 Results

4.3.1 Objective Measures

We used objective measurements to evaluate the performances of our system, such as the length of the dialogues and accuracy of the speech recognition in an *in-the-wild* setting, and the timing for the drink preparation.

Dialogues

The dialogues had an average duration of 3.76 minutes (min 2, max 8, st.dv. 1.06). We also observed that the intents were not correctly recognised with an average value of 0.76 per interaction user session, and the system asked users to repeat with an average value of 1.10 per interactive user session. Correlations were also calculated between dialogue length and repetition request, and between dialogue length and intent failure, but no correlation was found. We used a Pearson's correlation for evaluating the relation between dialogue length and requests for repetition and intent failures. We observed a moderate positive correlation ($r = 0.41$, $p < 0.001$) between the dialogue lengths (avg. 3.76, st.dv. 1.06) and people's asking the robot to repeat (avg. 1.10, st.dv. 1.18). We did not find any correlation neither between the intent failures (avg. 0.73, st.dv. 0.89) and dialogue length ($p = 0.67$), and repetition requests and intent failures ($p = 0.77$). The failures and repetitions were often due to the experimental environment, which was in an open space with several people crowded to observe the robot and the interactions. In such situations, people were either distracted and, therefore, did not respond to the robot, or their responses were provided with a too-low voice.

Drink Preparation

The system was also evaluated considering the preparation times of the different drinks (i.e., smoothies, smoothies with added alcohol, and cocktails) using both the right and left arm. The robot is able to prepare some drinks with just one or both robotic arms. For example, the Red Passion smoothie can be prepared only with the left robotic arm, while the Gin Lemon is prepared with the collaboration of both arms. During the preparation of the drinks, the robotic arms were used at a velocity of 60% over the maximum allowed to make sure the movements were not discomforting for the users and to avoid spilling the drink content.

Table 3 shows the average time (in seconds) used by the robot for the preparation of the drinks. The average drink times refer to the preparation of the robot with one or both arms, about the positioning of the customer. For example, if the customer is on the left, the robot uses its left arm to position the glass closer to them (i.e., on the left). Similarly, the robot uses its right arm to place the glass on its right side if the customer is on the right side. Regardless of the glass's final position, the average preparation time for smoothies was 173.04s, that of alcoholic smoothies was 178.57s, and the average preparation time for cocktails was 67.06s. It should be noted, however, that the preparation of alcoholic smoothies involves a higher number of movements (i.e., operations and trajectories) than those of a simple smoothie. These two drinks share the same fruit base, but the use of the two arms is able to optimize the preparation times of alcoholic smoothies, which would otherwise be almost doubled. Similarly, the average preparation time for smoothies and alcoholic smoothies requires longer time than the preparation for cocktails. Indeed, the first two types of drinks require the use of the fruit base which it is stored in a fridge and blender, which are not used for the preparation of the cocktails. The robot uses one or more alcoholic beverages and a shaker for the preparation of the cocktails.

Table 3: Preparations time for drinks using left, right or both robotic arms.

Drink	Arm	Served	Average Time (s)
Smoothie	Left	On the right	181.58
		On the left	164.50
Alcoholic Smoothies	Both	On the right	187.09
		On the left	170.05
Cocktail	Right	On the right	108.04
		On the left	26.09

Gesture Movements

During the users' interactions with BRILLO, the robot makes gestures both with the robotic head and arms. In particular, the robot makes arm gestures when it is not busy preparing a drink. As mentioned, some gestures are performed with one or both arms. For example, the simulation of the breathing is done by the robot using both robotic arms, while the greeting and goodbye gestures are done using either one or two arms. On average, the robot spends around 11.6 seconds for each gesture.

4.3.2 Evaluation of the Interaction

Participants used two scales to rate each item of the UEQ and UTAUT questionnaires where values were from a negative to a positive connotation (from 1 to 7, and from 1 to 5, respectively).

Evaluation of the User Experience

The User Experience Questionnaire consisted of 26 different items for evaluating usability and user experience aspects, including the following aspects: attractiveness, learnability, efficiency, dependability, stimulation, and novelty. These factors can be grouped to measure the attractiveness, pragmatic quality (i.e., efficiency, perspicuity and dependability) and hedonic quality (i.e., stimulation and novelty) of the interaction with the robot. The grouped dimensions respectively allow evaluation of pure valence, pragmatic quality aspects (i.e., goal-directed) and hedonic quality aspects (i.e., non-goal-directed).

We measured the reliability of the responses across the single factors and the grouped dimensions by using Cronbach's α for both conditions. In condition A-HRI, the sub-scales had a medium-high level of internal consistency as determined by Cronbach's α of 0.81, 0.57, 0.60, 0.77 and 0.54, respectively for attractiveness, learnability, efficiency, stimulation, and novelty. A low internal consistency was instead found by Cronbach's α of 0.29 for dependability. In the condition O-HRI, Cronbach's α were 0.87, 0.69, 0.70, 0.42, 0.77, and 0.80 respectively for attractiveness, learnability, efficiency, dependability, stimulation, and novelty factors. Since the values for dependability in both conditions were medium-low, we decided to follow Laugwitz et al.'s [12] suggestions, and analyse the single scores for each item of the dependability factor.

The ratings of the UEQ scale are values between -3 (horribly bad) and +3 (extremely good). Therefore, the values for the subscales between -0.8 and 0.8 are considered to have a more or less neutral evaluation of the corresponding scale, values > 0.8 represent a positive evaluation and values < -0.8 represent a negative evaluation.

We, firstly, analysed participants' overall perception of BRILLO for the different constructs of User Experience Questionnaire. As we can observe in Table 4, participants' overall experience with the robot was not considered negative (means are all above 0). In particular, they perceived BRILLO as novel, attractive and easy to be learned (i.e., novelty, attractiveness, learnability). In contrast, they perceived more neutrally the robot's ability to be fast and solve its tasks without effort, and inspire fun and motivating to use it (i.e., efficiency, stimulation).

Table 4: Descriptive for participants' scores to the UEQ questionnaire.

Construct	Min	Max	Mean	Std. D.
Novelty	-3	3	1.46	1.125
Attractiveness	-1	3	1.39	1.076
Learnability	-1	3	1.19	1.108
Stimulation	-2	3	.93	1.223
Efficiency	-3	3	.70	1.135

Figure 7 shows the participants' mean ratings of the UEQ subscales. The robot was rated positively for attractiveness, ease of use and understanding, and originality in both experimental conditions. Efficiency was rated as neutral. As we can observe, the robot received lower scores in the interacting condition (A-HRI) compared to the non-interacting condition (O-HRI). However, this difference was not statistically significant. We believe that these results are due to people's expectations during the drink preparation phase. The preparation of the drinks lasted around 2 minutes, which may be considered too slow for a robot.

Figure 8 shows the mean values of the participants' responses for the items that represent the dependability factor. Participants perceived the robot as more supportive and up to their expectations while they directly interacted with it. We can

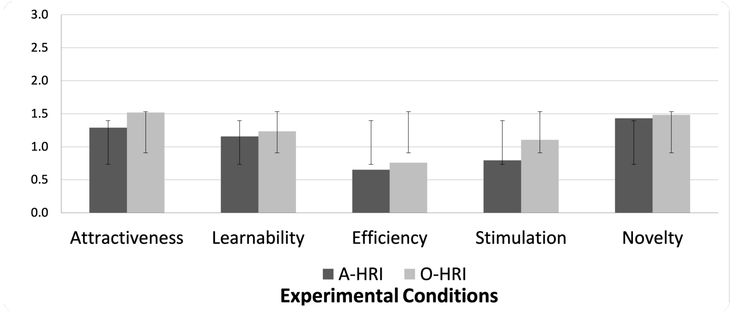

Fig. 7: Mean value collected from the responses of the factors of the UEQ. On the left, are the mean values of the responses of the participants in A-HRI condition, and on the right are those of the participants in O-HRI condition.

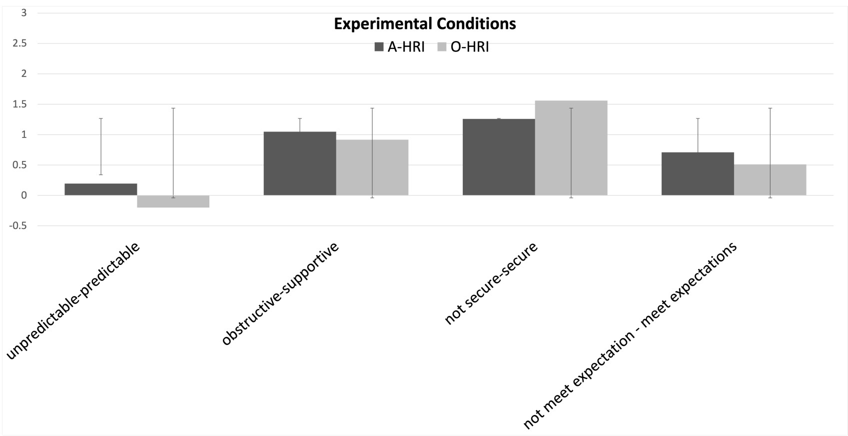

Fig. 8: Mean value collected from the responses of the dependability items of the User Experience Questionnaire. On the left, are the mean values of the responses of the participants in A-HRI condition, and on the right are those of the participants in O-HRI condition.

instead observe that they felt more secure while they were observing BRILLO. We hypothesise that since they were standing farther away from the robot, they were less negatively impressed by the dimensions and movements of the robot. Moreover, participants perceived BRILLO as more unpredictable in O-HRI than A-HRI condition. We believe that they did not feel in control of the interaction because they were only observers of someone else's interaction.

As we can also observe with visual inspection, Mann-Whitney U tests determined that there were no statistical differences in the above-mentioned UEQ subscales across the two conditions ($p > 0.3$ for all comparisons).

Finally, we can observe the overall mean scores of the grouped constructs are positive (values > 0.8) for both the experimental conditions (see Figure 9). These results show that participants were equally satisfied by the pragmatic qualities of the robot (i.e., perspicuity, efficiency and dependability). On the contrary, their overall impressions and excitement of BRILLO (i.e., attractiveness and pragmatic quality) were higher in A-HRI compared to O-HRI. We believe that these results may have been affected by a possible greeter sense of anxiety and robots' capability of communications [19], as it is shown by further analysis presented in the next section.

User Acceptance and Perception

We used the Unified Theory of Acceptance and Use of Technology model to evaluate the likelihood

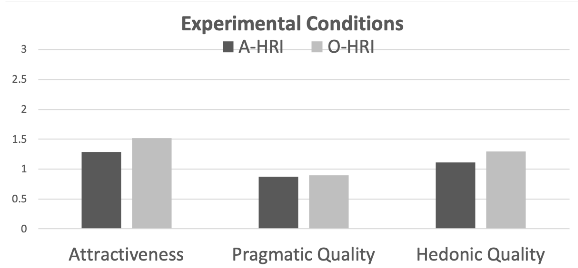

Fig. 9: Mean values for the grouped measure of attractiveness, pragmatic and hedonic qualities of the User Experience Questionnaire.

of adopting the BRILLO system by the participants. The questionnaire, in particular, measures the perceived performance expectancy, effort expectancy, social influence, and facilitating conditions. The questionnaire is composed of 36 items that build the following constructs: ANX, Anxiety; ATT, Attitude; FC, Facilitating Conditions; ITU, Intention to Use; PAD, Perceived Adaptability; PENJ, Perceived Enjoyment; PEOU, Perceived Ease of Use; PS, Perceived Sociability; PU, Perceived Usefulness; SI, Social Influence; SP, Social Presence; TR, Trust. Participants' ratings were collected using a 5-point scale from 1 (Strongly disagree) to 5 (Strongly agree).

As a first step, we measured the reliability of the responses across the different domains using Cronbach's α for all the items in both conditions. The constructs with a Cronbach's α greater than 0.7 were considered medium-high, while those with a Cronbach's α lower than 0.7 were considered too low and, therefore, we removed the questions/answers affecting the reliability of the construct. In particular, we removed the first two questions of ANX in A-HRI condition, the first question of PAD in A-HRI and O-HRI conditions, the last question of PENJ in O-HRI conditions, the first three questions of PEOU in O-HRI and A-HRI conditions, the first question of PU in O-HRI condition. The resulting Cronbach's α are above 0.7.

We then evaluated the overall participants' perception of BRILLO based on the different constructs. As observed in Table 5, participants positively rated most of the UTAUT's constructs (i.e., ratings with average greater than 2.5). In particular, they perceived BRILLO as a system that could help the work of someone who uses it (i.e., PAD, FC, PU), that can create comfort, pleasant feelings and is very useful (i.e., ANX[19], PENJ, ATT, PEOU), can exert social influence on one's behaviour (i.e., SI, PS). While they neither agree nor disagree on whether they would like to use the system (i.e., ITU), BRILLO had a strong social presence (i.e., SP), and they would trust BRILLO (i.e., TRUST).

Table 5: Descriptive for participants' scores to the UTAUT questionnaire.

Construct	Min	Max	Mean	Std. D.
Anxiety	1.00	5.00	4.1121	.90741
Perceived Enjoyment	1.60	5.00	3.8991	.87021
Perceived Adaptability	1.50	5.00	3.7103	.80399
Attitude	1.00	5.00	3.6480	1.00087
Perceived Ease of Use	1.00	5.00	3.6121	1.05801
Perceived Usefulness	1.00	5.00	3.4393	1.08216
Social Influence	1.00	5.00	3.1776	1.10371
Facilitating Conditions	1.00	5.00	3.1729	.97396
Perceived Sociability	1.00	5.00	3.0748	1.03318
Intention of Use	1.00	5.00	2.4673	1.06877
Trust	1.00	5.00	2.4673	1.12910
Social Presence	1.00	4.60	2.3290	.89916

A Mann-Whitney U test was run to determine if there were differences in the scores of the different constructs across the conditions. Distributions of the UTAUT scores when the user actively interacted and passively observed the interaction with BRILLO were similar, as assessed by visual inspection. We used an exact sampling distribution for U, unless otherwise stated. Differences in SP scores were found statistically significant higher for A-HRI than O-HRI, $U = 1750.5$ and $p = 0.036$. This implies that people's perception of BRILLO as a social entity (SP) increased while they directly interacted with the robot. We observed that the ANX and PEOU scores were statistically significantly lower for A-HRI than O-HRI, respectively with $U = 886.5$ and $p < 0.001$, and $U = 1090.0$ and $p = 0.038$. These results are particularly interesting, showing that people felt to be able to easily use the system, even if they were more anxious of doing it. There are several factors associated to people's attitude towards robots, in particular when these have social and anthropomorphic characteristics.

[19]Note that higher scores of Anxiety (ANX) correspond to lower levels of anxiety.

Table 6: Descriptive statistics (minimum, maximum, mean values, standard deviation) of the responses at the different constructs of UTAUT for the users non-interacting with the robot (on the left) and those interacting with the robot (on the right).

	O-HRI				A-HRI			
	Minimum	**Maximum**	**Mean**	**Std.D.**	**Minimum**	**Maximum**	**Mean**	**Std.D.**
Anxiety	2	5	4.35	0.85	1	5	3.82	0.90
Attitude	1	5	3.78	0.99	1.33	5	3.48	0.99
Facilitating Conditions	1	5	3.33	0.94	1	5	2.98	0.99
Intention to Use	1	5	2.50	1.11	1	5	2.42	1.02
Perceived Adaptability	1.5	5	3.72	0.85	2	5	3.69	0.76
Perceived Enjoyment	1.8	5	3.92	0.77	1.6	5	3.87	0.99
Perceived Easy of Use	1	5	3.79	0.99	1	5	3.38	1.11
Perceived Sociability	1	5	3.01	1.08	1.25	5	3.15	1.06
Perceived Usefulness	1	5	3.57	0.86	1	5	3.27	0.92
Social Influence	1	5	3.19	1.09	1	5	3.15	1.12
Social Presence	1	4.4	2.17	0.86	1	4.6	2.52	0.91
Trust	1	5	2.49	1.15	1	4.5	2.44	1.11

Notably, as highlighted by Erebak and Turgut [6], people may develop different types of anxiety, from robot anxiety to social and interaction anxiety. In the latter, the communication capabilities of the robots produce higher anxiety during the interaction with a robot as it happens while interacting with another person.

Median scores for A-HRI and O-HRI conditions were not statistically significantly different for the following constructs: ATT with $U = 1148.0$ and $p = 0.090$, FC with $U = 1146.0$ $p = 0.087$, ITU with $U = 1336.5$ and $p = 0.615$, PAD with $U = 1188.0$ and $p = 0.857$, PS with $U = 1490.0$ and $p = 0.642$, PENJ with $U = 1432.5$ and $p = 0.920$, PU with $U = 1131.5$ and $p = 0.071$, SI with $U = 1384.0$ and $p = 0.839$, and TRUST with $U = 1405.5$ and $p = 0.946$.

Table 6 shows the descriptive statistics (minimum, maximum, mean values and standard deviations) of the responses at the different constructs of UTAUT for the users non-interacting with the robot (O-HRI) and those interacting with the robot (A-HRI). These results show that participants had overall very positive perception and interaction with BRILLO. We can observe that for all the construct the evaluation resulted in a positive score (> 2.5) for both the conditions. In particular, very good scores (> 3.5) were obtained in the case of Anxiety (ANX, whereas higher scores correspond to lower levels of anxiety), Attitude (ATT), Perceived Adaptation (PAD), Perceived Enjoyment (PENJ) and Perceived Utility (PU). These results highlight that users, both in the case of direct interaction with the robot, but also as observers, were able to understand the abilities of the robot to adapt its behaviour to the user, and also that its behaviour is enjoing and useful. Median scores were obtained in the case of Intention to Use (ITU), TRUST and Social Presence (SP).

From our observations of the interactions, we believe that longer and long-term interactions would lead to even more positive scores for some constructs, as for example the ITU. Similarly, since the participants at the fair interacted only once with the robot. Hence, while the users were able to recognize the adaptation capabilities of the robot, the personalised recommendation system was not able to generate the interaction recommendation based on their previous preferences, but only based on-the-fly preferences. Further research we made showed that personalised interactions increase people's interest in the conversation with the BRILLO robot [32]. Furthermore, trust is a very complex construct that can depend on and be influenced by several factors, including reliability and efficiency [27]. We believe, indeed, that the limitations of drinking the beverage prepared negatively affected the perception of trust in the robot.

Finally, people provided some spontaneous comments that led us to rethink the robot's design. The robot currently has a humanoid appearance and a human face, however, it is equipped with two industrial robotic arms, which may have negatively influenced their perception of the robot, and so affecting the evaluation of SP. For example, people mentioned that the arms were too big compared to the rest of the robot's face and less reassuring. For these reasons, a future industrialization of the solution will have to take more into account the design of the robot.

5 Conclusions

The use of robots for the automation of the supply of food and beverages is a commercially attractive and modern application of robotic technologies, and it is used as a strategy to renew the image of the service and thus stimulate people's curiosity. However, the maintenance, in the long term, of the degree of interest in a commercial proposal linked to leisure can be only effectively achieved by implementing strategies for personalising the experience according to previous interactions and adapting the robot's behaviours.

In this work, we presented a ROS Architecture for a bartending robot aiming at providing an efficient service while keeping engaged the human customers in a dynamic and personalised interaction. The system has been tested in an ecological environment evaluating the effectiveness of the developed modules and of the whole architecture in terms of personalization of the experience and adaptation of behaviours, including content and frequency of dialogues, movements and poses to the customers' moods and preferences. Software modules were designed for the management of dialogues and the perception of social signals, such as, for example, a biometric-based algorithm for the recognition of social signals linked to the dynamics of interaction with the robot and Bayesian networks for the modelling of the interaction and the estimation of the probability of success of new proposals based on previous interactions.

We ran a user study during a national fair, and the system performed very well both in terms of service and social interactions whether people were observing or directly interacting with the system. However, long-lasting interactions need to be further investigated with paying customers to fully demonstrate that BRILLO is able to provide a service that is fast but also personalised. This will require the involvement of regular users over time. The use of personalised and recommended systems in public environments raises several issues, in particular connected to the users' privacy, as we already started to investigate in [29]. However, we believe that further investigations are also needed for testing the effects of other factors, such as collection and disclosure of private information in public (e.g., facial biometrics, drink preferences and habits), payments methods for purchasing drinks, variety of conversational topics.

BRILLO can present an added value to the commercial proposal, taking it beyond the simple experience associated with the presence of new technology and increasing customers' loyalty.

Acknowledgments. This work has been supported by Italian PON I&C 2014-2020 within the BRILLO research project "Bartending Robot for Interactive Long-Lasting Operations", no. F/190066/01-02/X44, and Italian PON R&I 2014-2020 - REACT-EU Azione IV.4 (CUP E65F21002920003).

Declarations

- Conflict of interest: The authors declare that they have no conflict of interest.
- Ethics approval and consent to participate: Participants were handed an information sheet for the study, and they signed consent forms before participating in the study. Anonymised data were collected.
- Data are available upon request.

References

[1] Atzeni M, Reforgiato Recupero D (2018) Deep learning and sentiment analysis for human-robot interaction. In: The Semantic Web: ESWC 2018 Satellite Events. Springer, Cham, pp 14–18

[2] Berezina K, Ciftci O, Cobanoglu C (2019) Robots, Artificial Intelligence, and Service Automation in Restaurants, pp 185–219. https://doi.org/10.1108/978-1-78756-687-320191010

[3] Bruce A, Nourbakhsh I, Simmons R (2002) The role of expressiveness and attention

in human-robot interaction. In: Proceedings 2002 IEEE ICRA, pp 4138–4142

[4] Ducamp G, Bonnard P, De Sainte Marie C, et al (2020) agrum/pyagrum : a toolbox to build models and algorithms for probabilistic graphical models in python. In: 10th International Conference on Probabilistic Graphical Models, pp 173–184

[5] Ekman P, Freisen WV, Ancoli S (1980) Facial signs of emotional experience. Journal of personality and social psychology 39(6):1125

[6] Erebak S, Turgut T (2020) The mediator role of robot anxiety on the relationship between social anxiety and the attitude toward interaction with robots. AI Soc 35(4):1047–1053. https://doi.org/10.1007/s00146-019-00933-8

[7] Foster ME, Gaschler A, Giuliani M (2013) How can i help you: Comparing engagement classification strategies for a robot bartender. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction. Association for Computing Machinery, New York, NY, USA, ICMI '13, p 255–262

[8] Kotseruba I, Tsotsos JK (2020) 40 years of cognitive architectures: core cognitive abilities and practical applications. Artificial Intelligence Review 53(1):17–94

[9] Kwon W, Clarke I, Vaara E, et al (2020) Using verbal irony to move on with controversial issues. Organization Science 31(4):865–886. https://doi.org/10.1287/orsc.2019.1333, https://arxiv.org/abs/10.1287/orsc.2019.1333

[10] Lambiase PD, Rossi A, Rossi S (2023) A two-tier gan architecture for conditioned expressions synthesis on categorical emotions. International Journal of Social Robotics pp 1–17. https://doi.org/10.1007/s12369-023-00973-7

[11] Laugwitz B, Held T, Schrepp M (2008) Construction and evaluation of a user experience questionnaire. In: Holzinger A (ed) HCI and Usability for Education and Work. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 63–76

[12] Laugwitz B, Held T, Schrepp M (2008) Construction and evaluation of a user experience questionnaire. In: Symposium of the Austrian HCI and usability engineering group, Springer, pp 63–76

[13] Lemaignan S, Warnier M, Sisbot EA, et al (2017) Artificial cognition for social human–robot interaction: An implementation. Artificial Intelligence 247:45 – 69. https://doi.org/10.1016/j.artint.2016.07.002, special Issue on AI and Robotics

[14] Lichtenthaler C, Lorenzy T, Kirsch A (2012) Influence of legibility on perceived safety in a virtual human-robot path crossing task. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, pp 676–681

[15] Masuda T, Misaki D (2005) Development of japanese green tea serving robot "t-bartender". In: IEEE International Conference Mechatronics and Automation, 2005, pp 1069–1074 Vol. 2, https://doi.org/10.1109/ICMA.2005.1626700

[16] Mohamed Y, Lemaignan S (2021) Ros for human-robot interaction. In: Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems

[17] Moulin-Frier C, Fischer T, Petit M, et al (2017) Dac-h3: a proactive robot cognitive architecture to acquire and express knowledge about the world and the self. IEEE Transactions on Cognitive and Developmental Systems 10(4):1005–1022

[18] Ngo HQT, Nguyen TP, Nguyen H (2018) Investigation on barbot to serve human in public space. In: 2018 4th International Conf. on Green Technology and Sustainable Development, pp 300–305, https://doi.org/10.1109/GTSD.2018.8595543

[19] Nomura T, Kawakami K (2011) Relationships between robot's self-disclosures and human's anxiety toward robots. In: 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp 66–69, https://doi.org/10.1109/WI-IAT.2011.17

[20] Prescott TJ, Robillard JM (2021) Are friends electric? the benefits and risks of human-robot relationships. iScience 24(1):101,993. https://doi.org/10.1016/j.isci.2020.101993

[21] Prescott TJ, Camilleri D, Martinez-Hernandez U, et al (2019) Memory and mental time travel in humans and social robots. Philosophical Transactions of the Royal Society B 374(1771):20180,025

[22] Rieser V, Moore JD (2005) Implications for generating clarification requests in task-oriented dialogues. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), pp 239–246

[23] Romero-Garcés A, Calderita LV, Martínez-Gómez J, et al (2015) Testing a fully autonomous robotic salesman in real scenarios. In: 2015 IEEE International Conference on Autonomous Robot Systems and Competitions, pp 124–130, https://doi.org/10.1109/ICARSC.2015.19

[24] Romero-Garcés, Adrián, Luis V, Calderita L, et al (2015) The cognitive architecture of a robotic salesman. In: Conferencia de la Asociación Española para la Inteligencia Artificial CAEPIA'15, pp 16–24

[25] Rossi A, Rossi S (2021) Engaged by a Bartender Robot: Recommendation and Personalisation in Human-Robot Interaction, Association for Computing Machinery, New York, NY, USA, p 115–119. URL https://doi.org/10.1145/3450614.3463423

[26] Rossi A, Dautenhahn K, Koay KL, et al (2017) How the timing and magnitude of robot errors influence peoples' trust of robots in an emergency scenario. In: Kheddar A, Yoshida E, Ge SS, et al (eds) Social Robotics. Springer International Publishing, Cham, pp 42–52

[27] Rossi A, Dautenhahn K, Koay KL, et al (2018) The impact of peoples' personal dispositions and personalities on their trust of robots in an emergency scenario. Paladyn Journal of Behavioral Robotics 9. https://doi.org/https://doi.org/10.1515/pjbr-2018-0010

[28] Rossi A, Garcia F, Cruz Maya A, et al (2019) Investigating the effects of social interactive behaviours of a robot on people's trust during a navigation task. In: Towards Autonomous Robotic Systems (TAROS 2019), Lecture Notes in Computer Science. Springer International Publishing, Cham, pp 349–361

[29] Rossi A, Perugia G, Rossi S (2021) Investigating customers' perceived sensitivity of information shared with a robot bartender. In: Li H, Ge SS, Wu Y, et al (eds) Social Robotics. Springer International Publishing, Cham, pp 119–129

[30] Rossi A, Caputo A, Scafora A, et al (2022) Investigating customers' preferences of robot's serving styles. In: 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp 1017–1020, https://doi.org/10.1109/HRI53351.2022.9889629

[31] Rossi A, Elizabeth John N, Taglialatela G, et al (2022) Generating emotional gestures for handling social failures in hri. In: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp 1399–1404, https://doi.org/10.1109/RO-MAN53752.2022.9900637

[32] Rossi A, Menna C, Giordano E, et al (2024) Evaluating customers' engagement preferences for multi-party interaction with a robot bartender. In: Ali AA, Cabibihan JJ, Meskin N, et al (eds) Social Robotics. Springer Nature Singapore, Singapore, pp 371–381

[33] Rossi S, Rossi A, Dautenhahn K (2020) The secret life of robots: Perspectives and challenges for robot's behaviours during non-interactive tasks. International Journal of Social Robotics 12:1265–1278

[34] Rui Z, Yan Z (2019) A survey on biometric authentication: Toward secure and privacy-preserving identification. IEEE Access 7:5994–6009. https://doi.org/10.1109/ACCESS.2018.2889996

[35] Song S, Chandrasekhar V, Cheung NM, et al (2015) Activity recognition in egocentric life-logging videos. In: Jawahar CV, Shan S (eds) Computer Vision - ACCV 2014 Workshops. Springer International Publishing, Cham, pp 445–458

[36] Staffa M, Rossi A, Bucci B, et al (2021) Shall i be like you? investigating robot's personalities and occupational roles for personalised hri. In: Li H, Ge SS, Wu Y, et al (eds) Social Robotics. Springer International Publishing, Cham, pp 718–728

[37] Tanevska A, Rea F, Sandini G, et al (2019) A cognitive architecture for socially adaptable robots. In: 2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), pp 195–200

[38] Tung V, Law C (2017) The potential for tourism and hospitality experience research in human-robot interactions. International Journal of Contemporary Hospitality Management 29(10):2498–2513

[39] Venkatesh V, Morris MG, Davis GB, et al (2003) User acceptance of information technology: Toward a unified view. MIS quarterly pp 425–478

[40] Walters ML, Koay KL, Syrdal DS, et al (2009) Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. In: in Procs of New Frontiers in Human-Robot Interaction: Symposium at AISB09 Convention, pp 136–143

[41] Wykowska A, Chellali R, Al-Amin MM, et al (2014) Implications of robot actions for human perception. how do we represent actions of the observed robots? International Journal of Social Robotics 6(3):357–366