Technical Methods for Social Robots in Museum Settings: An Overview of the Literature

Mehdi Hellou1,2 · JongYoon Lim1 · Norina Gasteiger1,3 · Minsu Jang4 · Ho Seok Ahn1

Accepted: 16 June 2022 / Published online: 9 August 2022 © The Author(s) 2022

Abstract
Robots have become increasingly present in our daily lives, and we see several settings wherein this technology might positively impact people. One of these settings, and the focus of this paper, is the museum. This paper provides an overview of the literature on how robots can help people within the museum context and explores the state-of-the-art methods or technologies employed by researchers to allow robots to interact within this environment. Five key features were found to be essential to integrate into social museum robots: social navigation (including safe and successful motions); perception (how robots can use visual skills to understand the environment); speech (verbal communication with visitors); gestures (non-verbal cues) and behavior generation (how the robot combines the four skills to adopt different behaviors). The paper subsequently analyzes the integration of these features into robotic systems for a complete and appropriate collection of behaviors in the museum setting.

Keywords Social robots · Robotics · Design · Artificial intelligence · Museum

1 Introduction

Studies concerning social robots are increasingly being conducted and their employment in everyday life has become a real need. Many researchers and roboticists believe we can use this technology in public spaces to assist people [1, 2]. Different sectors are affected by the employment of social robotics, such as restaurants [3–5], shopping malls [6], healthcare [7, 8], exoskeletons [9] and collaborative robots [10, 11]. Each sector deals with general issues we may encounter when developing robotic services for humans. One of the main issues remains the safety of users, considering how developers need to design their robots to accomplish their tasks while remaining safe. For example, this includes making the machine aware of human motions and gestures not to disrupt their activity or collide with them [11], supporting the user’s work, and contributing to their health, e.g., helping them to avoid lifting heavy objects [9, 11] or manipulating dangerous products [10]. When designing robots for human–robot collaboration or interaction, safety is not the only consideration. Robotic services also need intuitive interfaces and services to be useful. These could include a virtual simulation to visualize steps needed for the robot and the human operator to assemble a product [10] or an exoskeleton arm for the upper-limb rehabilitation [7]. Either way, the device must include an element of interactivity. This becomes increasingly important when robots and humans work collaboratively, or when social interactions between robots and humans are required. More work has also focused on developing personalized robots for users that can adapt their behaviors to user’s personalities, needs and preferences or to different cultures [12, 13]. For example, a robot serving drinks to users according to their preferences and the time of the day [14]. This notion of personalization has become an increasingly necessary feature for social robots that try to establish interactive exchanges with users. It is also an element that needs to be considered when developing social robots for museums.
In alignment with the previous applications of social robots, museums could benefit from social robotics by developing robots as guides, educators, entertainers, or a mix of these services [15]. More specifically, social robots might guide people to see different exhibits [16–21] provide information on the building and employees [22], or help them visit the museum regardless of their ability to do it by themselves [23]. Social robots can also assist people with daily and repetitive tasks, thus permitting them to dedicate their spare time to learn new skills at work [17]. Indeed, social robots are not intended to replace current employees in a museum but instead have been developed to collaborate with them to achieve their tasks. These tasks may include guiding visitors toward points of interest or pointing out a human expert who could provide more information [22]. This could be beneficial when museums are confronted with increasing numbers of visitors [16] or allow employees to know their most popular exhibits [24]. However, facilitating easier access to culture, regardless of an individual’s social background or physical abilities [16, 23], and encourage more people, including youth, to visit museums [19, 25].

Social robots: commonly reported and essential to integrate into museum robotic systems. This ability requires functional modules to enable the robot to verbally communicate with people by employing speech. Lastly, we explore non-verbal cues (i.e., gestures) which emphasize the robot’s speech and convey emotions simultaneously. The paper concludes with a focus on how the different approaches may be used, to implement these abilities and create complete behaviors for museum robots.

2 Methodology

This review takes a general and non-systematic approach. Majority of the literature was identified from Google Scholar and Microsoft Academic Scholar, with literature from social science and engineering databases. These included IEEE Xplore, Springer Link, ScienceDirect, ACM Digital Library and MIT Press Direct. To search the databases, we used the following keywords and phrases across multiple searches: "museum robots", "social robots in museum settings", "receptionist social robots", "social navigation", and "guiding service robots". The keywords were selected as they captured the topic of interest and reflected the categories used to present the findings. Other relevant literature was identified through snowballing and suggestions by networks/colleagues. We focused on papers published within the last 10 years to cover recent methodologies and state-of-the-art methods that would be relevant to social robots in museum environments. However, others submitted before 2010 we also considered, if they were deemed relevant for the review and helped to provide context to the information. For example, one paper was included as it was one of the few that proposed a real-world experiment in a museum and not a laboratory-based experiment. Papers were only included in the review if they were peer-reviewed, featured details on technical methods and were published in English (Table 1).

One author (MH) made initial decisions on which papers to include or exclude. The inclusion of the final selection of literature was then determined by discussion and relevance to the topic, among the other authors. We analyzed the papers by determining which factors each paper used to build their robots for museums, including the tasks achieved by the robot and its social features (Table 2), the robot’s abilities (Table 3) and the technical methods employed (Table 4). An overview of the papers on social museum robots included in the paper is depicted in the Appendices. We will detail these features in the following sections, as different social skills that are relevant to the museum context. Other relevant literature is drawn on to provide more context and understanding of the technical methods.

Across the papers, five features were found to be most commonly reported and essential to integrate into museum robots: social navigation (including safe and successful motions); perception (how robots can use visual skills to adapt to the environment); speech (verbal communication); gestures (non-verbal cues) and behavior generation (how the robot combines the five skills to adopt different behaviors). These features were discussed before and during the analysis of the different articles. They were determined to be essential characteristics that describe a museum guide robot. These will be discussed in turn below. Figure 1 presents a summary of the five features essential for a museum guide robot.

3 Social Navigation

As tour guides, museum robots should guide visitors through exhibits. This ability requires functional modules to enable the robot to navigate inside the museum, avoid obstacles and
accurately follow precomputed paths. It also needs to respect conventional rules or social rules [26] that respect people’s privacy around the robot, usually referred to as social distance [18, 27] or proxemics [13]. The next section presents some methods highlighting how social navigation might be applied to museum robots.

### 3.1 Localization, Mapping, and Path-Planning Methods

Robust navigation needs effective sensors that enable the robot to see and evaluate its surroundings, such as laser scanners [16, 17, 19, 22, 23] combined with appropriate software that can employ these sensors to ensure appropriate navigation. For example, the Robotics Operating System (ROS) is a flexible framework for writing robot software with a collection of tools, libraries, and conventions that aim to simplify the creation of complex and robust behavior across a wide variety of robotic platforms. These libraries include tools that enable localization, mapping and path-planning with the well-known ROS navigation stack [16, 20, 23, 26]—a package that includes a set of algorithms to permit a tour-guide robot to move freely or follow preplanned paths while avoiding obstacles. Path planning is performed by a planning module that provides both local and global planners. The program employs information from odometry and sensor streams and outputs velocity commands to send to a mobile base. Developers can also directly change one or several features of the algorithm to fit their needs. An overview of the package is depicted in the Fig. 2. In [26], the authors followed this process to determine a new and better way to avoid people by following Pandey’s work [28].

Some other researchers have preferred to adapt their robot’s locomotion methods manually [17, 18, 22]. For example, the authors in [17] integrated a navigation system that performs the mapping, localization, and path-planning of a tour-guide robot called Doris. For localization, the robot employed the sensor fusion procedure by combining a laser range finder (LRF) with an omnidirectional camera to detect visual markers. The mapping is characterized by an XML...
| References                        | Robot                               | Main purpose and social features                                                                 |
|----------------------------------|-------------------------------------|--------------------------------------------------------------------------------------------------|
| Shiomi et al. [24]               | Robovie and Robovie-M               | Guide and explain exhibits to visitors                                                             |
|                                  |                                     | Social navigation, speech and gestures                                                                 |
| Sakamoto et al. [30]             | Omni Wheel Mobile Robot             | Approach visitors                                                                                   |
|                                  |                                     | Social navigation and perceptions                                                                   |
| Huang et al. [44]                | ASIM0                               | Approach and explain exhibits to visitors                                                            |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Kanda et al. [35]                | N/S                                 | Recognize group of visitors                                                                          |
|                                  |                                     | Perceptions                                                                                          |
| Del Duchetto et al. [16]         | Scitos G5                           | Guide visitors                                                                                      |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Andrist et al. 2013 [45]         | N/S                                 | Explain exhibits                                                                                     |
|                                  |                                     | Speech                                                                                               |
| Takamasa Lio et al. [27]         | ASIM0                               | Approach and explain exhibits to visitors                                                            |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Miguel Kaouk Ng et al. [23]      | N/S                                 | Visit the museums remotely                                                                            |
|                                  |                                     | Social navigation                                                                                     |
| Vásquez et al. [17]              | Doris robot                         | Guide and explain exhibits to visitors                                                                |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Yousuf et al. [36]               | Robovie-R Ver.3                    | Engage and guide visitors                                                                             |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Graterol et al. [48]             | N/S                                 | Provide personalized comments and explanations                                                       |
|                                  |                                     | Speech                                                                                               |
| Zhang et al. [18]                | PIONEER3-DX                         | Guide visitors                                                                                       |
|                                  |                                     | Social navigation                                                                                     |
| Stricker et al. 2012 [22]        | Konrad and Suse mobile robot        | Guide and provide information to visitors                                                             |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Alvarez-Santos et al. [37]       | Pioneer 3-DX                        | Guide and explain exhibits to visitors                                                                |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Bellarbi et al. [26]             | B2lr                                | Guide visitors                                                                                       |
|                                  |                                     | Social navigation and perceptions                                                                    |
| Pang et al. [19]                 | MAVEN EDGAR                         | Guide and explain exhibits to visitors                                                                |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Moualla et al. [38]              | Berenson                            | Develop artwork appreciations from the visitors’ feedbacks                                            |
|                                  |                                     | Social navigation and perceptions                                                                    |
| Cheng et al. [20]                | DuckieBot                           | Guide and explain exhibits to visitors                                                                |
|                                  |                                     | Perceptions, speech and gestures                                                                     |
| Okano et al. [21]                | Multi-robot                         | Multi-robots work in collaboration                                                                    |
|                                  |                                     | Social navigation, perceptions and speech                                                              |
| Mondou et al. [25]               | Nao                                 | Explain and play interactively with visitors                                                           |
|                                  |                                     | Perceptions, speech and gestures                                                                     |
| Patompak et al. [33]             | N/S                                 | Social navigation in crowded environments                                                             |
|                                  |                                     | Social navigatin and perceptions                                                                    |
| Ratsamee et al. [32]             | Enon robot                          | Social navigation in crowded environments                                                             |
|                                  |                                     | Social navigation and perceptions                                                                    |
| Varitimiadis et al. [47]         | N/S                                 | Explain autonomously exhibits to visitors                                                             |
|                                  |                                     | Speech                                                                                               |
## Table 3 Abilities of the different museum robots

| References               | Robot                     | Abilities                                                                 |
|--------------------------|---------------------------|---------------------------------------------------------------------------|
| Shiomi et al. [24]       | Robovie and Robovie-M     | Identify people<br>Speak and explain exhibitions to visitors<br>Move to different exhibits<br>Greet people by handshaking, hugging people and playing rock, paper, and scissors<br>Suggests exhibits to visitors, according to their identity |
| Sakamoto et al. [30]     | Omni Wheel Mobile Robot   | Generate a personalized approach or avoiding behaviors for people<br>Take into account the internal state of the robot and the spatial relationship with the target |
| Huang et al. [44]        | ASIMO                     | Greet visitors<br>Generate non-verbal friendly behaviors<br>Explain exhibits<br>Approach and move away from visitors |
| Kanda et al. [35]        | N/S                       | Recognize a group of people in a walking or static position<br>Detect when a person is leaving or joining the group |
| Del Duchetto et al. [16] | Scitos G5                 | Guide and provide information on exhibits to visitors<br>Detect and track people<br>Autonomous navigation around the museum<br>Autonomous docking |
| Andrist et al. [45]      | N/S                       | Provide information concerning the buildings and history of a city<br>Change speech according to rhetorical and knowledge ability<br>Rhetorical ability changes the organization of the dialogue, use of additional verbal cues such as metaphors, the timing of speech |
| Takamasa Lio et al. [27] | ASIMO                     | Social features such as gaze, and speech<br>Explain exhibits<br>Greet visitors by their names and express a friendlier attitude to repeat visitors<br>Personalized approach and retreat behavior called speak-and-retreat |
| Miguel Kaouk Ng et al. [23] | N/S                       | Move remotely into the museum<br>Autonomous navigation |
| Vásquez et al. [17]      | Doris robot               | Guide visitors in museums<br>Navigation systems: mapping, localization and path planning<br>Expression system to detect emotions<br>Generate expressive responses and facial gesture |
| Yousuf et al. [36]       | Robovie-R Ver.3           | Identify people interested in a specific exhibit or in the robot<br>Engage interested visitors<br>Greet people and take them to the exhibit |
| Graterol et al. [48]     | N/S                       | Detect emotions and store this information in a semantic repository<br>Link emotions of users to objects (e.g., a painting or a plate) for personalized comments |
| Zhang et al. [18]        | PIONEER3-DX               | Predict the user’s moving speed and adapt to it to maintain the social distance<br>Navigate and guide a user or a group of users to a target point<br>Wait and follow a user who deviates from the guided path<br>Restart to the main guiding tasks when this one was interrupted |
| References          | Robot                        | Abilities                                                                                                                                           |
|---------------------|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Stricker et al. [22]| Konrad and Suse mobile robot | Provide information about the building (employees, labs, meeting rooms and offices)                                                              |
|                     |                              | Guide visitors to these points of interest                                                                                                        |
| Alvarez-Santos et al. [37] | Pioneer 3-DX | Follow an instructor                                                                                                                              |
|                     |                              | Record routes when it is requested → reproduce the routes when it is asked                                                                         |
|                     |                              | Recognize gestures to achieve a specific task                                                                                                     |
|                     |                              | Record voice from an instructor to have explanation on a exhibit                                                                               |
|                     |                              | Navigate and avoid obstacles                                                                                                                     |
| Bellarbi et al. [26] | B21r                         | Navigate in indoor environments                                                                                                                  |
|                     |                              | Avoid people in respecting their social space                                                                                                    |
|                     |                              | Detect humans, their pose and body orientation                                                                                                   |
| Pang et al. [19]    | MAVEN EDGAR                  | Navigate, avoid obstacles and guide visitors to see the different exhibits                                                                       |
|                     |                              | Detecting and tracking humans                                                                                                                     |
|                     |                              | Greet and narrate to explain exhibits to visitors with gaze attention and gestures to emphasize the speech                                         |
|                     |                              | Tell a story by using a Microsoft PowerPoint for language heritage                                                                               |
| Moualla et al. [38] | Berenson                     | Navigate randomly in a museum by avoiding obstacles at the same time                                                                           |
|                     |                              | Detect and recognize objects                                                                                                                     |
|                     |                              | Learn artwork appreciation from visitors’ instruction                                                                                            |
|                     |                              | Move in a preferred way by going to artworks it appreciates                                                                                      |
| Cheng et al. [20]   | DuckieBot                     | Follow pre planned routes and guide visitors to see different artworks on the path                                                             |
|                     |                              | Provide detailed explanation on the exhibits                                                                                                      |
|                     |                              | Recognize gestures in order to interact with users                                                                                              |
| Okano et al. [21]   | Multi-robot                  | Receptionist robot: greets visitors, counts people, and estimates traffic flow                                                                  |
|                     |                              | Photography service robot: shoot photographs for visitors and print pictures                                                                       |
|                     |                              | Questionnaire robot: give visitors a questionnaire on a tablet screen                                                                            |
|                     |                              | Exhibition guide robot: Follow visitors and guide them through the exhibits using voice explanations                                            |
|                     |                              | Patrol robot: move freely by remote control and observe the surroundings                                                                         |
| Mondou et al. [25]  | Nao                          | Detect and recognize users who already interacted with it                                                                                         |
|                     |                              | Resume or start a quiz                                                                                                                           |
| Patompak et al. [33] | N/S                          | Navigation and obstacles avoidance                                                                                                               |
|                     |                              | Generate human-like path by taking into account people’s features: gender, relationship with the robot and physical distance                      |
|                     |                              | Determine human state                                                                                                                          |
| Ratsamee et al. [32]| Enon robot                   | Adapt robot’s navigation according to the human’s intention                                                                                    |
| Varitimiadis et al. [47] | N/S              | Chatbot for visitors in a museum                                                                                                                 |
|                     |                              | Respond to questions about the museums and exhibits                                                                                               |
|                     |                              | Received questions and answers them by voice or text                                                                                              |
|                     |                              | Acquire new knowledge about the museum and exhibits                                                                                              |
| References         | Robot                          | Methods                                                                 |
|-------------------|-------------------------------|-------------------------------------------------------------------------|
| Shiomi et al. [24]| Robovie and Robovie-M        | RFID tags to detect users and analyze the time they spent for each exhibit to personalize their behaviours for each user RBS for decision-making about the dialogues, motion and navigations, and interactive behaviors |
| Sakamoto et al. [30]| Omni Wheel Mobile Robot       | Utility function that generates aggressive or passive involvement for the agent toward the human target Two variables control the internal state of the robot for the motion: control and acceptance |
| Huang et al. [44]| ASIMO                         | Non-verbal friendly behaviors according to four aspects: Response time, approach speed, individual distance and attentiveness Use collected data from human–human interaction to generate these behaviors |
| Kanda et al. [35]| N/S                           | Sensor pole equipped with two laser range finders to compute the body orientation and position of people in a group Allow to analyze velocities of different person |
| Del Duchetto et al. [16]| Scitos G5 | Petri net plans (PNP) to control the execution of different tasks (when and how) Execution Rule (ER) to manage unexpected events ROS framework and package from STRAND project used to main modules such as navigation, people tracking, task scheduling and data collection |
| Andrist et al. [45]| N/S                           | RBS systems and scripted dialogues Virtual visit of the city during the experiment with robots that explains each building aside from the participants Four behavioral conditions according to knowledge (low and high) and rhetorical (low and high) ability |
| Takamasa Lio et al. [27]| ASIMO | Landmarks methods for localization RFID tags to identify people Range camera sensors to detect and track people Use body orientation and position to get the visitor’s attention RBS with a behavior selector that controls utterances, gestures and locomotion according to the situation and a behavior executor use these arguments to execute the right action according to different templates |
| Miguel Kaouk Ng et al. [23]| N/S | ROS framework for navigation, sensors abstraction and process, and transmission of this information to a web interface Web user interface as the GUI for users |
| Vásquez et al. [17]| Doris robot                   | Stack up architecture to allow the communication between the different layer Localization by sensors fusion: laser range finder + visual marker detection + omnidirectional camera fused with an extended Kalman filter Mapping with XML files that represent the world (objects, path, landmarks, etc.) Path planning with brand and bound algorithm for exterior and A* for interior’s environments Path following with two fuzzy controllers XML files for robot’s utterances and fuzzy controller to control emotional responses and facial gestures Control actions of the robots with a new programming language: the combination of C, Pascal, and JavaScript |
| Yousuf et al. [36]| Robovie-R Ver.3              | Sensor pole with two laser sensors: one at the top to track position and body orientation of visitors and another one at the ankle level to track the position and orientation of the guide robot |
| References                   | Robot                  | Methods                                                                 |
|------------------------------|------------------------|-------------------------------------------------------------------------|
| Graterol et al. [48]         | N/S                    | EMONTO (EMonto ONTOlogy) an ontology for emotions interpretations        |
|                              |                        | Speech-to-text via Google API and Python Library SpeechRecognition       |
|                              |                        | Neural networks to detect emotions through text                         |
|                              |                        | Adaptable software that is able to add new features, such as face recognition or gestures |
| Zhang et al. [18]            | PIONEER3-DX            | Artificial potential field method to enable adaptive motion control      |
|                              |                        | Dijkstra algorithm for path generation                                  |
| Stricker et al. [22]         | Konrad and Suse mobile robot | Server with a webserver and a database in MySQL                        |
|                              |                        | Each robot’s server and software based on MIRA middleware to share data between distributed modules (e.g., data from sensors) |
|                              |                        | Topological map for localization and navigation                         |
|                              |                        | DWA as a motion planner to choose actions                               |
|                              |                        | Probabilistic multi-hypothesis people tracking system based on a 6D Kalman Filter that tracks the position and velocity of people |
| Alvarez-Santos et al. [37]  | Pioneer 3-DX           | Human detection with range camera and laser scanner                     |
|                              |                        | Finite State Machine to recognize the gestures                          |
|                              |                        | Augmented GUI wherein the user can interact by using their hands        |
| Bellarbi et al. [26]         | B21r                   | Pandey’s work to set up the method for avoiding a person               |
|                              |                        | Kinect and their software to compute human pose and body orientation with Euler angle |
|                              |                        | ROS navigation stack for path planning and navigation                   |
| Pang et al. [19]             | MAVEN                  | MAVEN: Virtual Human Toolkit for a virtual character on a flat 2D screen |
|                              | EDGAR                  | Kinect RGB-D camera to detect and track users                           |
|                              |                        | Planner module to execute each task according to a plan                |
|                              |                        | RBS for speech and narration                                            |
|                              |                        | EDGAR: Speech and animation module possible thanks to PowerPoint API    |
|                              |                        | Automated dialog system with Kinect sensor                              |
|                              |                        | Kinect sensor to detect people from voices and facial stimuli           |
| Moualla et al. [38]          | Berenson               | Perceptual algorithm to learn negatives or positives emotions for objects |
|                              |                        | Dynamical neutral field to control robot’s navigation                  |
|                              |                        | Tracking system to detect objects                                       |
| Cheng et al. [20]            | DuckieBot              | ROS for autonomous driving                                              |
|                              |                        | AprilTag to identify artworks and provide explanations to visitors     |
|                              |                        | Lane detection and OpenCV to follow different coloured predefined paths |
|                              |                        | Fisheye camera to convert 2D images in 3D to determine robot current coordination in the image |
|                              |                        | Hybrid machine learning algorithm (k-mean and k-nearest neighbors) to learn and recognize hand gestures |
| Okano et al. [21]            | Multi-robot            | RT middleware and robot service network protocol (RSNP) for communication |
|                              |                        | Database to save information                                            |
|                              |                        | Laser Range Finder to count the number of people                       |
|                              |                        | Beacon that guide visitors through the exhibits using voice explanations |
|                              |                        | Remote control for patrolling service                                   |
| Mondou et al. [25]           | Nao                    | Linear programming method with Choregraphe                              |
|                              |                        | Dynamic supervision: XML files and Python Software Development Kit     |
|                              |                        | Database: contents for scenario, scripted dialogue, arguments (e.g., gestures) and parameters values specified by a curator and experts |
|                              |                        | Finite State Machine for modelization                                  |
|                              |                        | Supervision tool generate actions for NAO                              |
|                              |                        | Time automaton to represent each context                                |
Table 4 (continued)

| References                  | Robot       | Methods                                                                 |
|-----------------------------|-------------|-------------------------------------------------------------------------|
| Patompak et al. [33]        | N/S         | Use Social Force Model’s algorithm to generate the robot’s path         |
|                             |             | Fuzzy logic to define a metric for the three features (gender, social and |
|                             |             | physical distance)                                                      |
|                             |             | Human state is defined by the human’s velocity and discomfort zone      |
| Ratsamee et al. [32]        | Enon robot  | Predict human’s motion by using the methods from the SFM, and           |
|                             |             | additional arguments, such as face orientation and the human’s         |
|                             |             | predicted path                                                         |
| Varitimiadis et al. [47]    | N/S         | Dialogflow for the baseline model of the chatbot                       |
|                             |             | Semantic Web technology to analyze the questions of visitors            |

![Diagram](image)

Fig. 1 Five features essential for a museum guide robot. The figure includes the main values of each feature that are detailed in the following sections.

file representing the robot’s environment (e.g., tables, chairs, doors, chargers, objects), routes to follow, and landmarks, while the path-planning follows the concepts of Branch and Bound and A* algorithms. In another example, authors used a topological map of the environment (university buildings) for their mobile guide robot [22]. The general Dijkstra algorithm [29] is employed to search the shortest path by using the topological map, and for localization, they used a particle filter-based Monte Carlo localization. Their robots Suse and Konrad were equipped with sonars, laser, and tactile collision sensors, to ensure safe navigation during the guide.

3.2 Social Rules

In addition to effective and safe navigation systems, museum robots must deal with crowded public spaces wherein several visitors will move around them. Specifically, the robot needs to avoid people in its path. As previously mentioned,
authors in [26] introduce a method to avoid people by following Pandey’s work. This includes the robot passing a person from their left side or passing behind them, when the robot faces the user from the profile side. Navigation can also be personalized by adding different approaches or avoiding behaviors. For example, in [30] the authors integrated different internal states to the robot represented by both criteria: control as an orientation to aggressive involvement with a target and acceptance as an orientation to passive involvement with a target. In other words, the control manages the robot’s motion toward a target, whereas acceptance manages the robot’s direction according to the target. In doing so, a utility function is defined as a representation for the agent’s preferred spatial relation based on its internal state.

Navigation robots may also be defined according to human motion and intention. Researchers have tried to generalize and explain human behaviors when navigating into environments with other actors (humans and objects) in static or dynamic motions. At the moment, the best model has been proposed by Helbing et al. [31] with their Social Force Model (SFM), a mathematical model employed to predict the motion of pedestrians. It was supposed that pedestrians followed some ‘social forces’ that describe the internal motivation of individuals to perform actions, and authors defined them through three terms. First, each person had a specific destination they desired to reach; hence they adjusted their velocity and direction to attain their goals. Secondly, a pedestrian wanted to keep certain distances to other people and objects in their surroundings. The third term described the potential attraction to other people (e.g., friends) or objects the pedestrian can encounter. The authors established equations and methods to depict these behaviors with these definitions and some developers tried to adapt these notions to mobile robots. For example, Ratsamee et al. [32] proposed a social navigation model for mobile robots respecting social rules in different crowded situations. More specifically, they adapt the robot’s motions when it encounters a human. They supposed that the human had different intentions toward the robot, and they tried to adapt its motions according to these situations: they want to avoid, maintain their course, or approach the robot. As a result, the authors defined a social navigation model (SNM) inspired by [31] wherein different attractive and repulsive forces with other social signals such as face orientation are supposed to predict the human's motions. They also included hypothetical areas around the robot to illustrate the human’s intentions or probability of approaching, avoiding, or avoiding the robot.

Patompak et al. [33] also established their navigation model for mobile robots from the SFM’s mathematical model. They called the Social Relationship Model (SRM), and similar to [32], defined it by different elements. First, the authors used the internal human state to establish their model. This state relied on the velocity of the human and their discomfort zone that the robot should not cross. Second, the authors included social characteristics that would adapt the shape and range of the discomfort zone of humans. Three elements were considered: gender (Male/Female), familiarity concerning the robot (Familiar/Stranger), and the physical distance between the human and the robot (Near/Far). A fuzzy logic approach is employed to define a metric of these factors. The different values are attributed with the help of different studies on social navigation in human society [34].
3.3 Guiding a Group of People

Guiding one person is already a complex task for social robots, but this difficulty increases when handling a group of users. Indeed, this situation is expected in museums since many visits are made in groups. In this case, robots need to adapt their motions according to a group of people. Kanda et al. [35] present a method that can recognize a group of visitors and detect when a person joins or leaves the group. This method is validated in both static and dynamic situations, and the authors employed a sensor pole system to recognize the group of people. Further details are provided in the next section.

Another example is provided by Zhang et al. [18] who set up a method generating virtual potential artificial fields for adaptive motion control. These potential fields work as attractive or repulsive forces that enable the robot to guide a user or a group of users while maintaining a social distance according to their movement speed. It also allows the robot to wait and follow users when they deviate from the guided path temporarily and return to the original task afterward. In doing so, the robot plan a path by using Dijkstra algorithm [29] and control its motion by using the artificial potential field method that based its methodology on the combination of attractive and repulsive forces. These forces are generated as virtual artificial potential fields, and the motion of the robot is controlled by the gradient of the potentials. The robot moves towards the position with a lower potential. Along the generated path are subgoal points created in order to bring the robot closer to the main goal. When the robot approaches an obstacle, the potential becomes greater, and a virtual repulsive force is generated to keep the robot away from it.

4 Perceptions

4.1 Human Recognition, Engagement and Identification

Tour-guide robots also need appropriate perception systems to adapt to their environments. By using internal or external cameras, perception may bring new information to the robot, such as identifying users or their emotions. Overall, a large number of papers use this system to detect and track human visitors [16, 18, 19, 22, 26, 27, 35–37] which allows robots to start interacting with new visitors and track them during their guiding tasks [18, 19, 22]. It is also a way to engage them when they seem to be interested in artwork [27, 36]. To make this possible, researchers have employed various sensors including cameras [20, 21, 26, 37, 38] or LRF and sonar sensors [22]. For example, in [35], the authors employed a sensor pole system to determine the position and orientation of visitors and assume if a pair of people belong to the same group by analyzing different features. The features include the distance between the two people, the absolute difference of velocities, the absolute difference in orientations of velocity and the absolute difference between an orientation of relative position and orientation of an average of differences of velocities between the two people.

Another example of human detection is depicted in [37], wherein the authors developed their software, combining information from a range camera and a laser scanner. First, the collected depth images from the range camera and data from the laser scanner are segmented into different blobs using the distances between adjacent points. A filtering process is then provided to discard irrelevant blobs by using several shape properties, such as the object’s size and height/width ratio, combined with a laser leg detection algorithm [39]. A human tracker is then used with a probabilistic algorithm that compares the probabilities of a detected human to be one of the humans that the robot is already tracking. In doing so, each human is distinguished by a specific a color and texture.

Guide robots may also use visual sensors to recognize and identify people. Few papers in museum settings use this process due to the lack of convenient technologies or methodologies to recognize a person’s identity in crowded places. Indeed, methods for identifying people (e.g., facial recognition algorithms) are complex and almost impossible to set up in this context. Some authors employed alternative techniques, such as RFID tags [24, 27], to read people’s identity. In [24], the authors used RFID tags to identify visitors, greet them by their names, and have access to additional information, such as the time they spent on each exhibit. This helped the authors discover which exhibits were preferred.

Recognizing the attention of visitors toward an exhibit or the robot itself might also give additional feedback to personalize a robot’s behavior. For example, Yousuf et al. [36] studied different strategies for the mobile robot Robovie-R to engage interested visitors. These strategies consisted of engaging visitors who are interested in the robot or in a specific exhibit. More specifically, two different situations are analyzed: (1) when the body and face of visitors are orientated toward an exhibit for five seconds, the robot considers visitors to be interested in the exhibit and proposes some explanations; (2) when the body and face of visitors are orientated toward the robot for five seconds, it considers visitors to be interested in it and suggests showing them the closest exhibit. The authors used a sensor pole wherein two laser sensors had been attached: a laser on the top to track the position and body orientation of visitors, and another at the ankle level to track the position and orientation of the guide robot. This technique permits them to define the human’s target of attention and determine their targets of interest (exhibits or robots). In [27], they considered a different approach for visitors who look...
at exhibits called *speak-and-retreat*. The method included a phase wherein the robot approached the visitors and says one segment of the utterances for an explanation (speak phase). After that, the robot immediately retreats without yielding a conversational turn and waits for an opportunity to provide its next explanation (retreat phase).

### 4.2 Probabilistic Methods for Visual Perception

When analyzing visual methods and algorithms in the current literature, we generally refer to autonomous methods with Machine Learning (ML) models that show promising results in current computer vision issues, including object detection. It is, though, rarely employed by autonomous robots in general, due to sometimes unpredictable outputs. There are still, though, some papers using these approaches for tour-guide robots. For instance, in [38], the authors combined a perceptual algorithm in a humanoid robot to learn artificial taste in artwork pieces using feedback from visitors. In doing so, the authors integrated a neural network (NN) to define local views in an artwork, wherein each local view corresponds to a neuron in the NN. Subsequently, according to the visitor’s impression, the robot assigned a positive or negative emotion to the local views.

We also saw the deployment of probabilistic methods to recognize crowd emotion [40], particularly bio-inspired probabilistic models, which used information from camera sensors to recognize positive or negative emotion from individuals. The authors depicted three algorithms based on Dynamic Bayesian networks to collect local features [41] and clustered by using Self-Organized maps (SOM) [42]. The classification and learning are then realized using the Ortony, Clore and Collins (OCC) model [43] to define the cause of emotions. The only difference with all three algorithms is the manner they partition the area to detect emotions. Among the three, the algorithm using DBN to study dynamic motions was far quicker than the other two. The model focused its detection by analyzing the individual’s motivation to reach a destination using parameters such as position or time. For instance, it determined how much time a person spent reaching a point or spent on a specific zone.

### 5 Speech

In addition to guiding visitors through the different exhibits, museum tour guide robots should explain the exhibits. They should also employ dialogue with visitors, to respond to their needs.

#### 5.1 Rules-Based Systems

Many of the reviewed papers employ scripted dialogue prepared by experts with knowledge about the different exhibits. In this case, a Rules-based System (RBS) is used to control and organize the robot’s utterances [16, 17, 19–22, 24, 25, 27, 36, 44, 45]. The RBS’s architecture differs in structure, parameters, and depth for each paper, and they are unparalleled for each system. This may occur due to the environment, the dialogue’s complexity and the other modules (e.g., perceptions and navigations modules). For example, Pang et al. [19] presented two social robots, MAVEN (Mobile Avatar for Virtual Engagement) virtual human character mounted on a mobile robotic platform for guiding services in a museum and EDGAR (Expressions Display and Gesturing Avatar Robot Pang) a humanoid robot which produces speech in English and Mandarin for culture learning and education. In order to provide responses and information about the museum, both robots have a text-to-speech service with a Kinect sensor that acquires the audio data and transmits to a speech recognizer, which converts spoken words to textual strings. Moreover, the robot EDGAR employs a Microsoft PowerPoint application to tell a story collaboratively with a human on stage in front of children for culture and languages heritage purposes. In addition to illustrating the story, the slide allows storing the robot’s script and animation clues to enable the robot to make gestures. The robot is composed of an automated dialogue system for social interaction, which searches a database to generate an appropriate response.

During guided tours, the robot’s explanations can also adapt its rhetorical ability for more effective speech. As demonstrated by Andrist et al. [45], the dialogue of an expert in a museum can be summarized using two elements: rhetorical ability and practical knowledge. Practical knowledge refers to the level of information the expert has concerning the museum and its artworks. At the same time, the rhetorical ability is defined as the ability to communicate the knowledge effectively and persuasively to visitors [46]. The authors proposed using these definitions to integrate them into a robot with different levels (high or low). Experimenters confronted participants with robots that differed in their amount of practical knowledge and rhetorical ability, e.g., a robot could be implemented with low knowledge and high rhetorical level, while another possessed high knowledge and a low rhetorical level. During the experiments, the robot presented a virtual city composed of different landmarks and buildings, and a questionnaire was given to participants at the end to assess the robots’ performances.
5.2 Dialogue and Speech Based on Machine or Deep Learning Methods

The models depicted above are generally based on scripting dialogue or RBS methods to build the robot's speech. Another type of model, named 'chatbot', is increasingly employed for autonomous systems mixing current state-of-the-art techniques in Natural Language Processing (NLP), Natural Language Generator (NLG) and ML. In the lexicon, the chatbot is defined as 'A computer program designed to simulate conversation with human users, especially over the Internet' [3]; hence, these models need to understand what humans say and respond appropriately to their requests. Their advantages reside in the fact that we can use them in any domain, including museums, as soon as we determine what information the chatbot needs to learn to provide convenient services to people. For example, a framework proposed by Google called 'Dialogflow' allows developers to set up their conversational bot easily. Many examples exist in the literature, such as the one presented by Varitimiadi et al. [47], who implemented a chatbot for a museum called MuBot to provide simple, interactive and human-friendly apps for visitors. The chatbot integrated several technologies with ML, NLP/NLG and a Semantic Web to research information using the Dialogflow framework to link the different modules for an effective program. Moreover, a Knowledge Base (KB) that relies on a Semantic Technology (RDF, SPARQL query language, and OWL ontologies) is integrated for data extraction, representation, linking, reasoning and querying. The authors successfully developed an unfinished version of the Cretan MuBot developed for the Archaeological Museum of Heraklion and allowed users to query the chatbot with a famous museum exhibit, the ‘Snake Goddess’ figurine. Even if the first version needs to be tested in experiments, the paper proposes explanations and methodologies that could be implemented into social museum robots.

Emotion detection through speech may also help to adjust the behaviors of tour-guide robots. In [48], the authors developed a semantic repository based on ontology called EMONTO (an Emotion ONTOlogy) to allow social robots to detect emotions through multimodal cues. They adapted their model on emotion detection in text and implemented a case study of tour-guide museum robots. The model relies on different technologies to convert audio data to text-based data, using the Google Application Programming Interface (API), a neural network (NN) to label the emotions in texts based on NLP transformers, and the ontology EMONTO for museums to register the emotions that artworks produce in visitors. The particularity of the ontology enables the robot to detect emotions through different verbal and non-verbal cues (e.g., gestures, face, voice, text). Even though the framework operates as expected with text, further upgrades and experiments need to confirm this ontology with other elements.

The previous paragraph gives us a clear idea of how to set up the robot’s speech. However, it is essential in human–robot interaction to have appropriate devices and software to recognize speech and transcribe the Text to Speech (TTS). Most of the robots integrated their speech synthesizers as part of an individual module. However, some devices and software is available for the public, such as Kinect sensors that possess their speech recognition systems [19] or the Google API [48] that provide an efficient speech recognizer with intuitive tutorials. Moreover, some robots have their TTS ready to be used without configuration or implementation of other systems. This is the case for humanoid robots NAO from SoftBank Robotics Europe; they contain a speech recognizer and synthesizer, although some papers recommend using other programs (e.g., Google Speech-to-Text1) to have a performant speech recognition module [3].

6 Gestures

In previous sections, we have covered essential abilities for robots in a museum. Indeed, by combining the navigation with appropriate perception and speech skills, the robot can ensure primary functionalities in a museum, such as guiding visitors, explaining exhibits and tracking visitors. However, additional features may reinforce the robot’s behavior and improve its interaction with visitors. Gestures support this statement by enabling the social robot to animate its speech [19, 24, 25, 27] and convey motions [17, 38]. Often, researchers script a robot’s gestures at the same time as they create its behaviors [19, 24, 25, 27].

6.1 Gestures Recognition

Gestures can also be part of the communication between the robot and the user. The user may use this information to express some needs to the robots as established in [20, 37]. For example, a mobile guide robot called DuckieBots in [20] invites visitors to follow it on a preplanned route to show different artworks in a museum. The visitors can interact with the robot by using hand gestures with a wearable device, including altering the moving speed of the robot or controlling the voice function during the robot’s explanations (e.g., replay, pause, move forward). In doing so, a hybrid ML algorithm is used to detect and recognize hand gestures. It combines a k-means algorithm [49] to remove noise from the input sensing data in the preprocess stage and select the samples used for the learning. The k-nearest neighboring (KNN) algorithm is employed to learn to recognize hand gestures on this trained database. In [37], the users can also interact with a mobile robot by using hand gestures. The objective is to

---

1 https://cloud.google.com/speech-to-text
implement a method that allows a robot to follow an instructor to record routes, paths and explanations for exhibits. For that purpose, the authors employed a Finite State Machine (FSM) to model gestures, whose inputs are a series of points that defines the trajectory made by the hand.

Gesture recognition may also enable human activity recognition (HAR). Wen Qi et al. [50] implemented a HAR in real-time using smartphones for telemedicine and long-term treatment by introducing an unsupervised online learning algorithm called Ada-HAR. It could update the classifier when the algorithm detected new activities and included them as new classes. In the article, the activities comprised dynamic exercises such as jogging and jumping, static postures (e.g., standing and sitting), and transitions activities (e.g., from sitting to standing). The idea was to use the Ada-HAR framework for healthcare guidance by only using smartphones. It might be interesting to integrate this technology in museum environments, since smartphones are widely present in our daily lives.

On the other hand, gesture recognition needs to face issues that are difficult to resolve, such as occlusion. Ovur et al. [51] studied this matter by introducing an adaptive multisensor fusion methodology for hand pose estimation with two Leap Motions. The authors integrated two separate Kalman filters for adaptive sensor fusion of palm position and orientation in their model. They showed that their method could perform stable and continuous hand pose estimation in real-time, even when a single sensor cannot detect the hand. This sensor fusion methodology might help in museums, specifically when occlusion issues are recursive in this type of environment.

6.2 Gesture Generation

Gesture recognition may reveal an essential feature for social robots that indicate user intentions in certain situations. Similarly, gesture generation adds more cues to convey emotion and makes the robot more human-like when interacting with people. As mentioned earlier, the majority of papers adapt the robot’s gestures while they establish its speech in order to coordinate both elements. External tools adapt the robot’s gestures autonomously, such as the Behavior Expression Animation Toolkit or BEAT [52], which generates a synchronized set of gestures according to an input text. It uses linguistic and contextual information to control body and facial gestures, besides the voice’s intonation. BEAT is composed of different XML-based modules. The language tagging module receives an XML tagged text and converts it into a parse tree with different discourse annotations. The behavior generation module uses the language module’s output tags and suggests all possible gestures; then, the behavior filtering module selects the most appropriate gestures (using the gestures conflict and priority threshold filters). Finally, the behavior scheduling module converts the input XML tree into a set of synchronized speech and gestures, which is ultimately converted into some executive instructions by the script compilation, usable to animate a 3D agent or a humanoid robot. Despite the lack of adaptation of this framework in the museum context, we find this technology relevant and adaptable for tour-guide robots in museums.

Gestures are not limited to hands or arms but can also concern the upper or lower body, such as faces. For example, social robot Doris in [17] possesses a head with eyes and eyebrows for facial expressions. The authors implemented a fuzzy controller to output the different levels of facial expression on Doris to control the emotional aspect of its responses. Similarly, in [38], the humanoid robot Berenson conveyed its emotions (negative or positive) toward artworks through facial expressions. This strategy enabled the robot to indicate to visitors what artworks it appreciated in the museum.

By drawing on examples of robotic services in healthcare, robots can imitate human motions to generate gestures. As explained in [8], the authors introduced an incremental learning framework that combines an incremental learning approach with a deep convolutional neural network to learn and imitate human motion in surgery tasks. In the experiments, the model was used to manage the redundancy control of an anthropomorphic KUKA robot and to perform the collected human motion autonomously. The results showed that their model significantly reduced learning human motion data processing time. There may be potential to employ these methods to allow social robots to imitate human teachers for learning. For instance, guide robots might learn how to present exhibits from actual employees.

7 Behavior Generation

The individual modules presented previously demonstrate how to set up an effective social museum robot. To obtain a complete robot that will guide visitors, interact with them and show them the different exhibits in a natural manner, it is essential to correctly combine these modules (navigation, perception, speech, gestures) to create complete behaviors for social robots and optimize them for museum settings to achieve multi-services (e.g., greet, educate or guide visitors [15]). This combination should enable the development of a robot that is able to integrate all the features depicted above (navigation, perception, speech, and gestures), in order to interact naturally with users.

7.1 Architecture Design to Generate Behaviors

In general, each robot adopts a specific architecture created explicitly to fit their systems and work with their sensors.
This architecture often has a stack-up structure with different layers [17, 19, 22, 24, 27] wherein communications are made between these layers to transmit information or data. For example, Stricker et al. [22] proposed a stackup architecture containing four layers for their guiding robots: (1) a Hardware layer to access values from sensors and actuators of the robot; (2) a Skill Layer to process values from the previous layer to cover whole navigation skills, including map building and localization, as well as person detection and tracking; (3) a Behavior layer to organize and combines each skill of the robot to achieve specific behavior (e.g., guide a user); and (4) a Control layer to orchestrate the behaviors with a finite state machine to achieve the different tasks and interaction functions. Moreover, the robots have access to an Administration server for the different information about the rooms, institute, labs and people working in the building. On the other hand, this server allows to control and observe the state of the robots remotely. This architecture is a simple example that researchers can draw on, for their systems.

With some sensors and systems available, frameworks with their software need to be installed and used to program and set up the robot. That is the case of the camera Kinect [19, 26] or the robot NAO [25], in which modules such as speech recognizer, face detection, face tracking, microphones and speakers are already arranged and useable with a graphical programming tool called Choregraphe or a Python Software Development Kit (SDK).

7.2 Personalized Behaviors

For effective museum robots, some reviewed papers record information from interactions with visitors [24, 27, 48] and personalize the robot’s behavior to user preferences, needs and personalities [3, 13]. This strategy permits the robot to establish a relationship with visitors, especially regular ones. For example, in [24], the robot can access the visitors’ movement history and propose to visitors who seem interested in a specific exhibit to test it. In [27], the authors implemented three stages for each visitor: new, acquaintance and friend. Acquaintance was meant for visitors who returned twice to visit the robot, while friend referred to three visits. Verbal and non-verbal gestures (social distance) were adapted according to these stages. Moreover, the robot changed its explanations whether it had a first contact with the visitor. This enabled the robot to avoid repeating an explanation or provide more detailed information. This approach needs to use a database to record and access these elements on a more technical side. In doing so, management systems such as MySQL or a structured server [21, 22, 24, 27] are necessary to safely record these elements. For example, in [21], a robot service network protocol (RSNP) in each robot is installed for communication. A software component transmits the state of each robot to the RSNP, previously registered in a database. This data was transmitted in an XML format and contained the robot’s state or the action it was executing. Within this server, the authors could organize the task of five robots in a museum: (1) a receptionist robot that greeted visitors; (2) a photography service robot that takes photos for visitors; (3) a questionnaire robot that gave questionnaires to visitors on a tablet screen; (4) an exhibition guide robot; (5) and a patrol robot that observed the environment freely.

The robot’s behavior is also a way to improve its interaction with visitors. For example, in [44], the authors modelled a "Friendliness" behavior for a social robot in a museum, after collecting data from human–human interaction experiments to analyze different friendly behaviors. During the experiments, the museum expert played by the participants greeted visitors and could approach them to explain an exhibit when they seemed interested in it. When they did not explain the exhibits, the participants moved to a waiting position. By analyzing the different data, the authors extracted four variables that define how friendly behaviors can be played at different degrees: (1) "Response time" that defines the time when the confederate takes an action and the participants response to this action; (2) "Approach speed" calculated as the speed for the confederate to approach the visitor; (3) "Individual Distance" defined as the distance between the visitors and the expert that explain an exhibit; and (4) "Attentiveness" defined as the location where the expert waits before continuing with their explanation. Subsequently, the authors implemented the same scenario with the humanoid robot ASIMO in modelling the robot’s behavior according to the four variables. Overall, the participants perceived friendliness and likeability when they adjusted the parameters of the four variables accordingly, even though people had mixed thoughts regarding some variables. This was demonstrated in the "Approach speed" variable, whereby some people had a good impression with a fast approach, while other people felt pressured.

8 Additional Devices to Improve Experiments

Social robots interacting in public spaces such as museums must deal with several issues, including noises and back and forth from users surrounding the robots. In addition to their effective laser and camera sensors enhancing their abilities (perception, navigation, or speech), social robots might use external technologies to improve their interactions with visitors.

8.1 Pole System

Kanda et al. [35] used a vertical bar called "pole system", in which two laser sensors are attached on the top and the bottom. This vertical bar allows the robot to detect and track
humans while also extracting their position and orientations to assume two people belong to the same group. The same process is used in [36], wherein a sensor pole also calculates the positions and orientations of visitors and the robot. Subsequently, the robot analyzes this information to identify what exhibits the users are interested in and engages them by initiating conversation. They are especially effective for estimating visitors’ position and orientation and dealing with many detection issues, such as occlusion. Interestingly, authors have also used several sensor poles to extract features from visitors at different locations, and to extract the level of interest from visitors about the robot or an exhibit [36, 53], which might become an asset when guiding in crowded environments.

8.2 Touch Screen

Many engineers have also integrated a touch screen into their robots in order to facilitate communication with users in an environment where speech recognizers could fail. The Konrad and Suse robots [22] exemplify this, as they used a touch screen-based graphical dialog system to permit visitors to communicate. Visitors could choose a destination and the robots could guide them toward their destinations. The robots could also inform visitors about news concerning the museums or services they could offer. Santos et al. [37] proposed interactive communication based on a screen whereby a GUI displayed information for users and a menu to learn new routes or replay ones already learned. A Kinect camera that recognized human gestures was employed to navigate the menu.

9 Discussion

In the above sections, we have presented a range of papers concerning the employment of social robots in museum settings for different roles and tasks (e.g., tour guide or receptionist). By analyzing the technical methods used, we hope to help researchers design appropriate robots for this context. In the review, we have also determined what techniques and algorithms may be appropriate to use, in order to enable social robots to interact in museum settings and with visitors.

We categorized our work into five sections, defining elements that should influence the robot’s behaviors or decision-making. Some of these elements were especially evident across the reviewed papers (e.g., social navigation, speech) and can be deemed as essential to building an effective social robot. They are also currently subject to further analysis and changes in the field of robotics, such as autonomous speech with ML to build relevant dialogue models for Natural Language Processing (NLP) [54].

It was evident that navigation is utterly essential when designing robots for museums. It enables robots to guide visitors, avoid obstacles and safely avoid people in an often crowded place. As for perception, a tour-guide robot needs to observe its surroundings, detect visitors and track and possibly identify them to personalize the interaction. With these two features, the robot can achieve main guiding tasks for visitors, though it is not able to communicate with them. In that case, an additional speech module is necessary to provide verbal information or explain exhibits to visitors. Moreover, with appropriate gestures, the robot can emphasize its speech or convey emotions. While we have presented the features (navigation, perception, speech, and gestures) distinctly, the overall behaviors may work best when combined. Hence, in the last section we presented some well-known architecture and software employed to enable the communication between the different modules. Further information on the included papers is available in the appendices.

9.1 Approach to Designing Effective Guide Robots

By drawing on the technical methods presented in the review, we can propose a potential social guide robot for museums. Here we present a list of technical methods we found relevant and appropriate in this context by selecting main social features that could influence the interactions between the robot and the visitors according to the sections above. It mainly includes the robot’s mobility, visual perception, speech and gestures:

Social navigation: A guiding robot needs to integrate effective social navigation to move autonomously through the museum and guide visitors. The employment of the ROS navigation stack seems appropriate to define the robot’s navigation and localization [16, 20, 23, 26]. It could be combined with a model that personalizes the robot’s motion as described in [32], to approach visitors interested in the robot or avoid them. We propose using the ROS navigation stack since it integrates all of the necessary functionalities to have an efficient navigation system to avoid, move, and approach visitors in a museum environment. Moreover, we can integrate it into different robot systems, and is publicly available.

Perception: We need to identify and track people for the robot’s perception, which is essential when the robot guides them or explains the exhibits. In this case, we require external sensors such as range cameras or laser scanners [37]. According to the devices employed, the software used might be different, e.g., if we decide to use a Kinect camera, there are already preinstalled software. Additionally, emotion detection could also be added to the system to personalize the robot’s behavior. In this case, probabilistic methods, such as DBN [40] seems the best solution to detect emotions dynamically since it enables the robot to detect emotions in a crowded environment.
Speech: The speech would demand more upstream work to prepare an effective script dialogue that can explain different exhibits. It needs to also be integrated with museum experts/staff and relate to the exhibition being presented. We can control the robot’s speech [25] remotely or employ a RBS [19]. If we have already prepared a script that allows the robot to interact with visitors and answer their requests dynamically, the RBS process seems sufficient. If we would like to add safety and control, we can add a human agent who verifies the robot’s speech or a tablet in case there is an issue. These methods seem safest and the most accurate solutions compared to the results proposed by probabilistic models [47].

Gestures: Like speech, a robot’s gestures can follow a script according to the dialogue. We can also implement a model that generates personalized gestures regarding the situation. We can include the BEAT [52] software to generate personalized gestures regarding the visitors’ utterances if we desire to coordinate gestures with users’ personalities. However, it will require formatting the text from the recognized speech properly into XML tagged text.

Finally, we need to combine these features into our system by specifying the architecture and how each module will be bound and communicate their information. This will depend mainly on the robot used and the different types of sensors available. For example, Choregraphe Software will be employed if the robotic system is a product from Softbank Robotics Europe, such as NAO [25].

We have provided the suggestions above to show justification for specific methods over others. Final choice will usually depend on the application and the system’s tasks.

9.2 Limitations and Strengths

There are several limitations that must be acknowledged. Primarily, this review took a general and non-systematic approach and only included literature published in English. Additionally, the list of models and methods presented in the review are non-exhaustive.

We did also not include methods used in other contexts. Other useful work includes research by Johanson et al. [55] who studied the impact of a humanoid robot as a receptionist in a hospital. The authors developed social skills considered fundamental in human communication to build close relationships, convey emotions, or to increase human attention. This includes (1) self-disclosure, in which the robot stated sentences that were considered as personal statements about the robot; (2) leaning forward, in which the robot leaned (approximately 20°) forward toward a participant when they spoke, while maintaining a neutral standing position during the rest of the interaction; (3) voice pitch, wherein the robot both increased and decreased its voice pitch within the single interaction in order to convey emotions. Another relevant paper by Irfan et al. [56], studied the employment of data-driven approaches for developing a barista robot’s speech in generic and personalized long-term human–robot interaction. The purpose was to implement models that can remember and recall the user’s preferences (e.g., preferred beverage) after few interactions with them. These two papers, along with many others could also inform the development of museum robots and highlights the importance of learning from other contexts.

A strength of the review is the inclusion of and focus on recent publications on social museum robots, which best represent the current technical methods used in robotics. Another is the consideration of some advantages and disadvantages of the methods. This may help to support researchers in their choices and encourage them to contribute to the development of this sector.

9.3 Future Work

Work remains on integrating social robots in public spaces such as museums. Social robots in museums may facilitate cultural access and reduce employee workload, especially in doing repetitive and mundane tasks. At the moment, features such as speech or perception need improvement, to be more efficient. Indeed, in the papers reviewed, we have not seen any social robots possessing autonomous speech to discuss or answer visitor requests. This issue concerns not only social robots but also the sector of NLP that requires additional work to integrate autonomous dialogue on robots. It also includes issues about speech recognition that become more difficult when the robot is interacting in a noisy environment [3].

Perception is also a feature in which researchers should focus more on in the following years, although significant enhancements have been provided in the last years with ML models for visual skills, such as human detection. However, the task becomes more complex when used for unique features, such as emotion recognition or identity recognition [57].

Lastly, we noted a lack of systems that propose personalized services to visitors, although literature exists which shows that people prefer to interact with robots that can adapt their behaviors to their preferences, needs, and personalities [3, 12, 13, 56]. For example, it might be interesting to enable museum guide robots to propose different exhibits to visitors according to their preferences. Some previous reviews have proposed interesting directions to improve social robots [1, 2], and discuss the factors for successful HRI in museum settings extensively [15].
10 Conclusion

To our best knowledge, this review is one of the few that lists methods and algorithms employed for social robots in museums settings while studying the different abilities that make a successful social guide robot. We discuss many key functions and technical methods for museum robots to uncover state-of-the-art models. This led to the discussion of adaptive locomotion and navigation techniques, such as using and processing visual sensor information. Further, accurate communication through verbal cues provides essential information to visitors, while non-verbal cues help to convey emotion or emphasize speech. Evidently, some of the techniques rely on traditional approaches, such as RBS or FSM, while some integrate autonomous processes with state-of-the-art methods in AI, such as ML or Deep Learning. Different software architecture that facilitates communication with the robot’s modules and sensors was also evident. Notably, some software is already available for public use and can extract information from sensors. For instance, the ROS Navigation Stack package provides necessary materials for autonomous localization and navigation when appropriate sensors are available for the robots. These frameworks can be used as a starting point, as they are easily integrable into various robotics platforms. Ultimately, through this review, we hope to support researchers in their technical design and encourage them to contribute to the development of appropriate social museum robots.

Author contributions HSA obtained the funding and conceived the initial idea as the principal investigator of this project. MH located and synthesized the literature. MH drafted the manuscript. All authors discussed the results and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This work was supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2020-0-00842, Development of Cloud Robot Intelligence for Continual Adaptation to User Reactions in Real Service Environments).

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Ahmad M, Mubin O, Orlando J (2017) A systematic review of adaptivity in human-robot interaction. Multimodal Technol Interact 1(3):14
2. Tapus A, Mataric M, Scassellati B (2007) Socially assistive robotics [Grand challenges of robotics]. Robot Autom Mag IEEE 14:35–42
3. Irfan B et al (2020) Challenges of a real-world HRI study with non-native English speakers: can personalisation save the day? pp 272–274
4. Keizer S et al (2014) Machine learning for social multiparty human-robot interaction. ACM Trans Interact Intell Syst 4(3):14
5. Pieská S, et al (2012) Social service robots in public and private environments, pp 190–195
6. Foster ME et al (2019) MuMMER: socially intelligent human-robot interaction in public spaces CoRR, abs/1909.06749.
7. Spagnuolo G et al (2015) Passive and active gravity-compensation of LIGHTarm, an exoskeleton for the upper-limb rehabilitation. In: 2015 IEEE international conference on rehabilitation robotics (ICORR)
8. Su H et al (2022) An incremental learning framework for human-like redundancy optimization of anthropomorphic manipulators. IEEE Trans Ind Inf 18(3):1864–1872
9. Mauri A et al (2019) Mechanical and control design of an industrial exoskeleton for advanced human empowering in heavy parts manipulation tasks. Robotics 8(3):65
10. Land N et al (2020) A framework for realizing industrial human-robot collaboration through virtual simulation. Procedia CIRP 93:1194–1199
11. Vicentini F et al (2020) PIROS: cooperative, safe and reconfigurable robotic companion for CNC pallets load/unload stations. In: Caccavale F et al (eds) Bringing innovative robotic technologies from research labs to industrial end-users: the experience of the European robotics challenges. Springer International Publishing, Cham, pp 57–96
12. Hellou M et al (2021) Personalization and localization in human-robot interaction: a review of technical methods. Robotics 10(4):120
13. Gasteiger N, Hellou M, Ahn HS (2021) Factors for personalization and localization to optimize human–robot interaction: a literature review. Int J Soc Robot
14. Sekmen A, Challa P (2013) Assessment of adaptive human-robot interactions. Know Based Syst 42:49–59
15. Gasteiger N, Hellou M, Ahn HS (2021) Deploying social robots in museum settings: a quasi-systematic review exploring purpose and acceptability. Int J Adv Rob Syst 18(6):17298814211066740
16. Duchetto FD, Baxter P, Hanheide M (2019) Lindsey the tour guide robot—usage patterns in a museum long-term deployment. In: 2019 8th IEEE international conference on robot and human interactive communication (RO-MAN)
17. Vásquez BPEA, Matía F (2020) A tour-guide robot: moving towards interaction with humans. Eng Appl Artif Intell 88:103356–103356
18. Zhang B, Nakamura T, Kaneko M (2016) A framework for adaptive motion control of autonomous sociable guide robot. IEEE Trans Electr Electron Eng 11(6):786–795
19. Pang W-C, Wong C-Y, Seet G (2018) Exploring the use of robots for museum settings and for learning heritage languages and cultures at the chinese heritage centre. Presence 26(4):420–435
20. Cheng F-C, Wang Z-Y, Chen J-J (2018) Integration of open source platform duckieTown and gesture recognition as an interactive interface for the museum robotic guide. In: 2018 27th wireless and optical communication conference (WOCC)

21. Okano S et al (2019) Employing robots in a museum environment: design and implementation of collaborative robot network. In: 2019 16th international conference on ubiquitous robots (UR)

22. Stricker R et al (2012) Interactive mobile robots guiding visitors in a university building. In: 2012 IEEE RO-MAN: the 21st IEEE international symposium on robot and human interactive communication, pp 695–700

23. Ng MK et al (2015) A cloud robotics system for telepresence enabling mobility impaired people to enjoy the whole museum experience. In: 2015 10th international conference on design technology of integrated systems in nanoscale era (DTIS)

24. Shiomi M et al (2007) Interactive humanoid robots for a science museum. IEEE Intell Syst 22(2):25–32

25. Mondou D, Prigent A, Revel A. A dynamic scenario by remote supervision: a serious game in the museum with a nao robot. In: Advances in computer entertainment technology. Springer International Publishing.

26. Bellarbi A et al (2016) A social planning and navigation for tour-guide robot in human environment. In: 2016 8th international conference on modelling, identification and control (ICMIC).

27. Iio T et al (2020) Human-like guide robot that proactively explains exhibits. Int J Soc Robot 12:549–566

28. Pandey AK, Alami R (2010) A framework towards a socially aware Mobile Robot motion in Human-Centered dynamic environment, pp 5855–5860

29. Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1(1):269–271

30. Sakamoto T, Sudo A, Takeuchi Y (2021) Investigation of model for initial phase of communication: analysis of humans interaction by robot. J Hum-Robot Interact 10(2)

31. Helbing D, Molnár P (1995) Social force model for pedestrian dynamics. Phys Rev E 51(5):4282–4286

32. Ratsamee P et al. (2013) Social navigation model based on human intention analysis using face orientation. In: 2013 IEEE/RSJ international conference on intelligent robots and systems

33. Patompak P et al (2016) Mobile robot navigation for human-robot social interaction. In: 2016 16th international conference on control, automation and systems (ICCAS)

34. Hall ET (1966) The hidden dimension/Edward T. Hall. 1966: Doubleday Garden City, N.Y. xii, 201 p., [6] leaves of plates :-xii, 201 p., [6] leaves of plates

35. Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1(1):269–271

36. Yousuf MA et al (2013) How to move towards visitors: a model for museum guide robots to initiate conversation. In: 2013 IEEE RO-MAN

37. Alvarez-Santos V et al (2014) Gesture-based interaction with voice feedback for a tour-guide robot. J Vis Commun Image Represent 25(2):499–509

38. Moualla A et al (2017) Readability of the gaze and expressions of a robot museum visitor: Impact of the low level sensory-motor control. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)

39. Goekley R, FortiZZi J, SimmonRs R (2007) Natural person-following behavior for social robots, pp 17–24

40. Baig MW et al (2015) Perception of emotions from crowd dynamics, pp 703–707

41. Damasio AR (1999) The feeling of what happens: body and emotion in the making of consciousness. Harcourt Brace, pp 386–386

42. Kohonen T, Schroeder MR, Huang TS (2001) Self-organizing maps. Springer, Berlin

43. Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge

44. Huang C-M et al. Modeling and controlling friendliness for an interactive museum robot. In: Robotics: science and systems

45. Andrist S, Spannan E, Mutlu B (2013) Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In: 2013 8th ACM/IEEE international conference on human-robot interaction (HRI)

46. Hartelius E (2008) The rhetoric of expertise

47. Varitimidis S et al (2020) Towards implementing an AI chatbot platform for museums. In: International conference on cultural informatics, communication & media studies, vol 1

48. Graterol W et al (2021) Emotion detection for social robots based on NLP transformers and an emotion ontology. Sensors 21(4):1322

49. Lloyd S (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28(2):129–137

50. Qi W, Su H, Alveretti A (2020) A smartphone-based adaptive recognition and real-time monitoring system for human activities. IEEE Trans Human-Mach Syst 50(5):414–423

51. Ovur SE et al (2021) Novel adaptive sensor fusion methodology for hand pose estimation with multileap motion. IEEE Trans Instrum Meas 70:1–8

52. Cassell J, Vilhjálmsson HB, Bickmore T (2001) BEAT: the behavior expression animation toolkit, pp 477–486

53. Oyama T et al (2013) Tracking visitors with sensor poles for robot’s museum guide tour. In: 2013 6th international conference on human system interactions (HSI).

54. Lopez MM, Kalita J (2017) Deep learning applied to NLP. CoRR. abs/1703.03091.

55. Johanson DL et al (2019) The effect of robot attentional behaviors on user perceptions and behaviors in a simulated health care interaction: randomized controlled trial. J Med Internet Res 21(10):e13667–e13667

56. Irfan B, Hellou M, Belpaeme T (2021) Coffee with a hint of data: towards using data-driven approaches in personalised long-term interactions. Front Robot AI 8:676814

57. Caroppo A, Leone A, Siciliano P (2020) Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults. J Comput Sci Technol 35(5):1127–1146

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mehdi Hellou is an Early Stage Researcher in Robotics and Artificial Intelligence at the Cognitive Robotics Laboratory, University of Manchester. His Ph.D. is being completed as part of the PERSEO project, which has been funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955778. He spent five years at the Sorbonne University’s Polytech Sorbonne school of engineering (previously known as the University of Pierre and Marie-Curie), where he completed his thesis in robotics with six months of internships at SoftBank Robotics Europe in Paris on longterm personalization for human-robot interaction. After he graduated from Polytech Sorbonne, Mehdi started a new Master in Artificial Intelligence at Sorbonne University called ANDROID for Robotics, Decision Making, Reinforcement Learning, Multi-Agent, and Highly Interactive Virtual Environments.

Jong Yoon Lim received his B.E. degree in Computer Science from Yonsei University, Republic of Korea and his M.S degree in Software
Engineering from The University of Auckland, New Zealand. He is interested in natural language processing, computer vision, machine learning, and computational photography, with the goal of building intelligent machines capable of recreating our visual world. Currently, he researches Social Human-Robot Interaction to build intelligent robots and software platforms at the Department of Electrical, Computer and Software Engineering, The University of Auckland, New Zealand. Before joining The University of Auckland, he worked as a Senior Software Engineer for Samsung Electronics.

Norina Gasteiger is a digital health enthusiast, with interests in the design, evaluation and implementation of innovative digital health technologies. Her previous research has included human-robot interaction studies, participatory design of technology and work with extended reality, patient portals, social media and mobile apps. Currently, Norina is a Ph.D. fellow at the University of Manchester, United Kingdom. Her Ph.D. in Health Informatics is funded by the National Institute for Health and Care Research Applied Research Collaboration for Greater Manchester and focuses on upskilling care home workers using virtual reality and augmented reality training. Norina holds a Master’s degree in Public Health from the University of Auckland, in addition to a Bachelor of Health Sciences (Honours) and Bachelor of Art. She frequently collaborates with the multidisciplinary Centre for Automation and Robotic Engineering Science (CARES) group at New Zealand’s University of Auckland.

Minsu Jang is interested in transferring self-supervised and collaborative learning methods to service robots to promote their cognitive capabilities in the real-world. He received his Ph.D. in artificial intelligence from Korea Advanced Institute of Science and Technology in 2015. He has been a researcher at Electronics and Telecommunications Research Institute in South Korea for more than 23 years until now, and worked for several projects in the fields of symbolic reasoning, human-robot interaction, human-care robots, cloud robotics, open-world cognition and lifelong learning. He is a member of the directors of Korea Robotics Society and served as an organizer or as a committee member for international workshops and conferences including ROMAN, IROS and ICSR.

Ho Seok Ahn received his B. S. degree in Information and Communication Engineering from Sungkyunkwan University, Republic of Korea, and his Ph. D. degree in Electrical Engineering and Computer science from Seoul National University, Republic of Korea, in 2005 and 2010, respectively. He was a visiting researcher at the Intelligent Systems Research Institute at Advanced Industrial Science and Technology (AIST), Japan, in 2006. He was a senior researcher at Korea Institute of Industrial Technology (KITECH), Republic of Korea, from 2010 to 2012. He was a lecturer at University of Science and Technology (UST), Republic of Korea, from 2011 to 2012. He was a research scientist at Advanced Telecommunications Research Institute International (ATR), Japan, from 2012 to 2013. He was a postdoc researcher at the University of Auckland, New Zealand, from 2013 to 2015. He is currently a Senior Lecturer at the Department of Electrical, Computer and Software Engineering, University of Auckland, New Zealand, since 2015. He is a Deputy Director of CARES. His research interests include social robots, cultural robots, facial robots, Human-Robot Interaction, artificial emotional systems, android robots, humanoid robots, intelligent service robots, and modular robot systems. He got multiple Best Paper Awards and Robotics Competition Awards. He is a member of IEEE.