HELLENIC REPUBLIC
MINISTRY OF DEVELOPMENT AND INVESTMENT
GENERAL SECRETARIAT FOR RESEARCH AND INNOVATION
HELLENIC FOUNDATION FOR RESEARCH AND INNOVATION

The research project is implemented in the framework of H.F.R.I call "3rd Call for H.F.R.I.'s Research Projects to Support Faculty Members & Researchers" (H.F.R.I. Project Number: 26465)

3rd Call for H.F.R.I.
Research Projects to support
Faculty Members and Researchers
Sub-action II

Project Title

Compassionate, Adaptive, Robust and Explainable Robots

Project Duration

03 November 2025 – 02 November 2029
48 Months

Project Acronym

RoboCARE

Project No

26465

Deliverable No

D2.1

Deliverable Title

Tech Specs

Deliverable Completion Date

02 August 2026

Table of Contents

Table of Figures

Table of Tables

No table of figures entries found.

Document Revision History

Version	Date	Notes
0.1	15/05/2026	Initial draft
0.9	30/06/2026	Internal review
1.0	16/07/2026	Final version

List of Acronyms

Acronym	Meaning
AI	artificial intelligence
FoV	Field of View
GAN	generative adversarial networks
GDPR	General Data Protection Regulation
HRI	human–robot interaction
ISO	International Organization for Standardization
NLP	Natural Language Processing
ORCA	Optimal Reciprocal Collision Avoidance

PACE	PERCEIVE–ADAPT–CONVEY–EXHIBIT
ROI	Return on Investment
ROS	Robot Operating System
SFM	Social Force Model
Sim2Real	simulation to real-world environments
SotA	state-of-the-art
STT	Speech-to-Text
TO	Technical Objective
UC	Use Case
VPR	Visual Place Recognition
WP	Work Package
XAI	Explainable AI

Executive Summary

This deliverable presents the technical specifications established during the initial phase of the RoboCARE (Compassionate, Adaptive, Robust and Explainable Robots) project and provides the scientific and technological foundations that will guide the design, development, integration, and validation of the project's robotic framework. Developed within Work Package 2 (WP2), Deliverable D2.1 consolidates the outcomes of Tasks T2.1 and T2.2 by reviewing the state of the art in socially-aware robotics, defining representative use cases, and specifying the functional and non-functional requirements of the RoboCARE system.

The deliverable begins by examining the current landscape of socially-aware robotic navigation and human–robot interaction, identifying the major challenges, opportunities, and best practices reported in recent literature. Particular emphasis is placed on socially-aware navigation, benchmark design, evaluation methodologies, ethical, social and economic considerations, and the principles of Trustworthy Artificial Intelligence. These analyses establish the conceptual framework within which RoboCARE will operate and define the design principles that will govern the development of socially compliant robotic behaviour.

Building upon this foundation, the deliverable introduces a comprehensive set of use cases covering both domestic and crowded environments. These representative scenarios describe how the robotic system should perceive contextual information, interpret human behaviour and emotions, adapt its navigation and interaction strategies, communicate its decisions through explainable dialogue, and continuously improve through long-term learning. Each use case is accompanied by detailed operational specifications, including actors, preconditions, postconditions, execution flows, interfaces, and success criteria, providing a clear blueprint for the implementation and validation of the RoboCARE framework.

The final part of the deliverable defines the core functional modules that comprise the RoboCARE system. These include perception modules for visual place recognition, speech understanding, emotion recognition, and action recognition; adaptation modules supporting socially-aware behaviour, long-term behaviour understanding, and lifelong learning; communication modules for explainable artificial intelligence and interactive natural language processing; together with data governance protocols, AI auditing mechanisms, and non-functional system requirements. Collectively, these components are organized according to the PACE (PERCEIVE–ADAPT–CONVEY–EXHIBIT) paradigm, which forms the architectural foundation of the RoboCARE project and enables the development of compassionate, adaptive, robust, and explainable robotic systems.

Overall, Deliverable D2.1 establishes the scientific rationale, technical specifications, and system-level requirements that will guide the subsequent implementation activities of RoboCARE. By integrating advances in socially-aware navigation, human-centred artificial intelligence, explainability, lifelong learning, and trustworthy robotics into a unified technical framework, the deliverable provides a solid basis for the development of robotic systems capable of operating safely, transparently, and effectively in real-world human-centred environments.

Introduction

The RoboCARE project aims to develop compassionate, adaptive, robust, and explainable robotic systems capable of operating safely and effectively in human-centered environments. The project focuses on socially-aware robotic navigation and interaction, enabling robotic agents to perceive environmental and social context, adapt their behavior accordingly, communicate their intentions and decisions transparently, and continuously improve through long-term interaction and learning.

Within this context, Work Package 2 (WP2) is responsible for establishing the scientific and technical foundations of the RoboCARE framework. In particular, Task T2.1 investigates the challenges, opportunities, and state-of-the-art approaches in socially-aware robotics and human-robot interaction, while Task T2.2 defines the technical specifications, data handling principles, and trustworthiness requirements that will guide the design and implementation of the project technologies.

This deliverable, D2.1 – Tech Specs, presents the outcomes of these activities. First, it reviews the current state of the art in socially-aware navigation, highlighting key challenges, opportunities, evaluation methodologies, ethical considerations, and Trustworthy AI principles relevant to the RoboCARE vision. Based on this analysis, a set of representative use cases is defined, covering both domestic and crowded environments, which will serve as reference scenarios throughout the project.

Furthermore, the deliverable specifies the main functional modules that constitute the RoboCARE framework. These modules address perception capabilities, including visual place recognition, action recognition, emotion recognition, and natural language understanding; adaptation capabilities, including socially-aware navigation, long-term behavior understanding, and lifelong learning; and communication capabilities, including explainable AI, emotion elicitation, and interactive natural language processing.

The overall framework follows the RoboCARE PACE paradigm, consisting of the PERCEIVE, ADAPT, CONVEY, and EXHIBIT pillars. Through this architecture, the robotic agent is expected to perceive its environment and users, adapt its behavior according to contextual and social information, communicate its intentions and reasoning transparently, and demonstrate socially appropriate behavior during real-world operation.

Finally, this deliverable establishes the initial technical specifications, data governance principles, auditing considerations, and non-functional requirements that will guide the subsequent development, integration, and validation activities of the RoboCARE project.

1. Challenges and Opportunities of Socially-Aware Robotic Solutions

1.1. Socially Aware Navigation: Overview

Socially-aware navigation refers to the ability of mobile robots to move safely and efficiently in environments shared with humans while respecting social norms, personal space, and contextual cues. Unlike traditional robotic navigation, which primarily optimizes for collision avoidance and path efficiency, socially-aware navigation incorporates human-centered considerations such as proxemics, comfort, intent prediction, group behavior, and social etiquette. Over the past decade, extensive research has explored this interdisciplinary field, drawing from robotics, psychology, anthropomorphism and human–robot interaction (HRI) to enable robots to operate more naturally in human coexistance spaces such as domestic and crowded environments. Foundational and recent surveys [3, 8, 9, 10, 11, 12] define and highlight the evolution of this domain from early rule-based models of human space to modern learning-based and data-driven approaches that account for dynamic, multi-agent interactions and long-term social compliance.

"A socially navigating robot acts and interacts with humans or other robots, achieving its navigation goals while modifying its behavior so the experience of agents around the robot is not degraded or is even enhanced."~[3]

Building on these comprehensive studies of human-aware navigation, proxemics-based modeling, core system challenges, and conflict avoidance strategies, we derive a structured understanding of the field that highlights the best practices, guidelines, its maturity and its open problems. In particular, these surveys collectively motivate a synthesis of (i) the key challenges and opportunities in developing robust, scalable, and socially compliant robotic navigation systems, and (ii) the best practices and state-of-the-art solutions that have emerged for socially-aware navigation in complex, real-world environments.

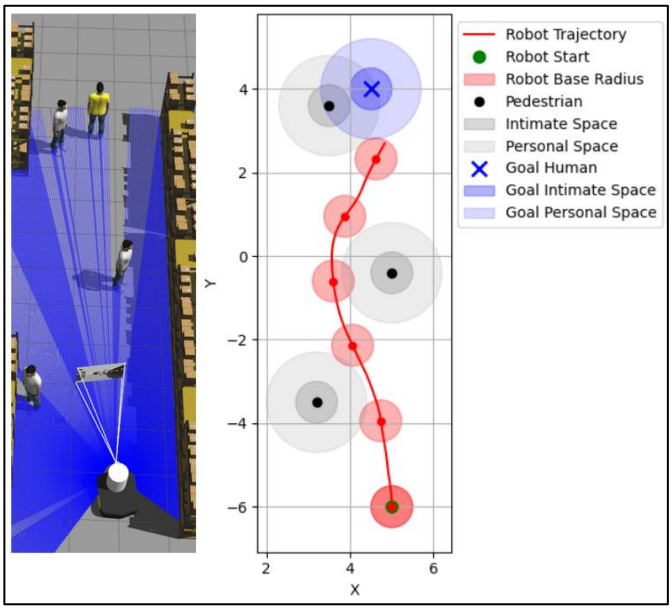

Figure 1. Social Navigation proxemic compliance paradigm: Avoidance of people outside their personal proxemic zone.

1.2. Best practices and State-of-the-Art solutions in Socially-Aware Navigation

Socially aware robotic systems operate at the intersection of robotics, artificial intelligence (AI), and human-centered design, introducing both significant challenges and promising opportunities. A key challenge lies in developing navigation systems that can effectively operate in dynamic, multi-agent environments while respecting human expectations and social norms. At the same time, these systems offer opportunities to improve safety, efficiency, and accessibility in domains such as healthcare, logistics, and public services.

From a technical perspective, state-of-the-art (SotA) approaches to social-aware navigation can be broadly categorized into model-based, learning-based, and generative methods, typically integrated within a perception–prediction–planning pipeline. Classical model-based approaches, such as the Social Force Model (SFM) and geometric methods including velocity obstacles, Optimal Reciprocal Collision Avoidance (ORCA), and proxemics-based planning, explicitly encode human interaction rules but often struggle in highly dynamic or unstructured environments. In contrast, imitation learning and inverse reinforcement learning enable robots to learn socially compliant behavior from human demonstrations, although these methods are often limited by data requirements and generalization challenges.

More recent deep learning-based approaches, including recurrent neural networks, graph neural networks, and transformer architectures, focus on modeling complex multi-agent interactions and predicting pedestrian trajectories. Generative models, such as generative adversarial networks (GAN) and diffusion-based approaches, further improve the ability to capture uncertainty and produce diverse socially plausible trajectories. Reinforcement learning, particularly in multi-agent settings and deep reinforcement learning setup enables end-to-end policy optimization; however, it requires carefully designed reward functions to ensure socially acceptable behavior and can suffer from instability during training, and great amounts of resources and time to efficiently trained. Emerging foundation and multimodal models, including vision-language architectures, represent a new direction toward more generalizable and semantically aware navigation systems.

Despite these advances, several challenges remain. A central difficulty is balancing **task efficiency with social compliance**, as optimal navigation paths may conflict with human comfort or social norms. Additionally, transferring policies from simulation to real-world environments (Sim2Real) remains a significant barrier due to differences in dynamics, perception noise, and human behavior variability. Ensuring robustness to distribution shifts is therefore a critical research problem.

Best practices identified in recent literature, emphasize several key directions. These include integrating prediction and planning within unified frameworks, explicitly modeling multi-agent interactions, and incorporating uncertainty-aware decision-making. Furthermore, balancing efficiency with socially defined cost functions is essential for real-world deployment. Robust evaluation in both simulation and real-world environments, using standardized benchmarks and clearly defined social metrics (such as proxemic violations, collision rates, and comfort measures), is considered essential for reliable assessment.

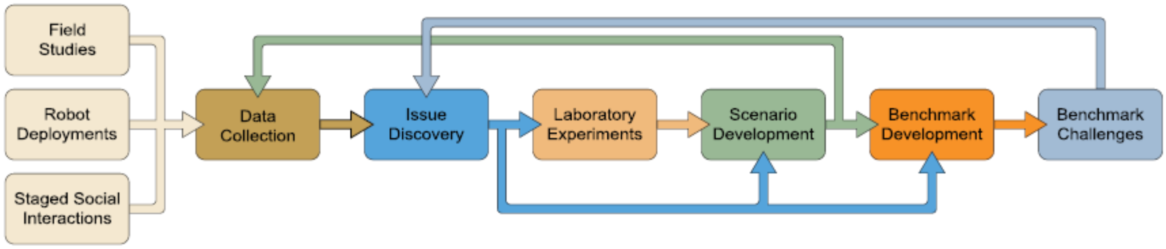

Figure 2. Lifecycle of development of a socially aware navigation system [3].

Additionally, the use of diverse and dynamic simulation environments that capture realistic physical interactions and social scenarios, as well as incorporating human-in-the-loop evaluation throughout the development lifecycle. Real-world trials and subjective human feedback are critical for validating system performance beyond quantitative metrics. Explainable AI (XAI) methods are increasingly important for improving transparency and trust, while robustness to distribution shifts remains essential for deployment. Overall, socially aware robotics requires not only technical innovation but also careful consideration of human factors, ethics, and system-level validation.

1.3. Social Navigation Scenario Guidelines

A **social navigation scenario** provides a structured and reusable way to define and study HRI in real or simulated environments. Rather than treating each encounter as an isolated event, scenarios formalize *what is happening, who is involved, and how success is defined*, enabling consistent data collection, benchmarking, and comparison across studies. As proposed by Francis et al., a scenario is defined through a clear three-step process: first, the **scenario is explicitly specified** in terms of its research context, intended robot task, expected human behavior, and success metrics, ensuring it is both identifiable in data and reproducible in experiments. Second, the definition is **evaluated for commonality, flexibility, and fitness to purpose**, making sure it aligns with established scenario families, remains broad enough to capture real-world behavioral variation, and effectively elicits the target interactions without over-constraining behavior. Finally, the scenario is **clearly communicated using a structured "scenario card" format**, including metadata, detailed definitions of environment and behaviors, and a usage guide that specifies labeling rules, success and failure conditions, and contextual considerations. This structured approach ensures that socially-aware navigation research remains reproducible, comparable, and grounded in realistic human–robot interaction settings while supporting scalable dataset generation and robust evaluation [3].

As summarized in [3], these scenarios are typically grouped according to the *scientific objective* they serve, which helps researchers select appropriate benchmarks and design meaningful experiments.

Pedestrian navigation scenarios focus on low-density environments where robots interact with a small number of humans at a time, often in structured spaces such as corridors, doorways, or intersections. These include common interaction patterns such as frontal approaches, overtaking, right-angle crossings, blind corner encounters, and narrow doorway passages. Such scenarios primarily evaluate local interaction safety, comfort, and social compliance in relatively simple but fundamental settings.

Crowd navigation scenarios, in contrast, address high-density and highly dynamic environments where emergent group behaviors and flow interactions become critical. Here, robots must handle situations such as parallel or perpendicular pedestrian traffic, random or circular crossings in open spaces, and even crowding conditions where the robot becomes surrounded. These scenarios emphasize global coordination, flow adaptation, and robustness in complex multi-agent systems.

Finally, **interaction scenarios** extend beyond pure navigation by introducing task-driven constraints that require higher-level social engagement. Examples include joining or leaving human groups, object handover tasks, question answering, and continuous observation activities. These scenarios evaluate not only motion efficiency but also social intent understanding, task coordination, and human-aware decision-making.

Together, these categorized scenarios provide a shared vocabulary for social navigation research, enabling consistent benchmarking, reproducible experimentation, and systematic comparison of socially-aware robotic behaviors across different levels of complexity.

Social Navigation Scenario Card	
Scenario Metadata	
Scenario name	Frontal Approach
Scenario description	A robot and a human approach head-on in a passable space.
Scientific purpose	Low-density pedestrian scenario applicable indoors and outdoors.
Scenario Definition	
Geometric layout	A space wide enough for the robot and human to pass each other.
Intended robot task	The robot navigates from one side of the space to the other
Intended human behavior	The human navigates in the opposite direction of the robot.
Scenario Usage Guide	
Success metrics	SR, (No) Collisions
Quality metrics	P2: Comfort, P3: Legibility, P4: Politeness
Ideal outcome	Robot goes around humans in a socially acceptable manner.
Failure modes	1. Robot collides with human 2. Robot fails to exit in time limit
Labeling criteria	1. Robot and human face each other 2. Robot and human move toward each other at the start of episode 3. Sufficient clearance exists for robots and humans to pass each other

Figure 3. Socially Aware Navigation scenario card example [3].

1.4. Social Navigation Benchmark Guidelines

A **good social navigation benchmark** must go beyond simply testing whether a robot can avoid obstacles and instead provide a structured, reproducible, and socially meaningful evaluation of human–robot interaction. As outlined by [3], an effective benchmark should first **evaluate social behavior explicitly**, ensuring that metrics capture human comfort, social compliance, and interaction quality rather than only traditional navigation objectives such as path efficiency or success rate. It should also **include a rich set of quantitative metrics**, combining objective measures (e.g., distance, collision rate, trajectory efficiency) with validated subjective evaluations such as human-rated comfort or perceived safety.

In addition, a benchmark should **provide clear baselines for comparison**, including simple heuristic policies and, when possible, state-of-the-art methods, so that performance improvements can be meaningfully interpreted. It must also be **efficient, repeatable, and scalable**, enabling consistent results across many trials and making real-world or simulated evaluation practical for the research community. Furthermore, it should **ground human evaluations in real human data**, either through user studies or validated learned models that approximate human perception, ensuring that "social performance” reflects actual human responses. Finally, it should **use well-validated evaluation instruments**, such as standardized questionnaires and empirically verified metrics, alongside the appropriate parametric or non-parametric statistical tests, so that results are reliable, comparable, and scientifically sound across studies.

Together, these principles ensure that social navigation benchmarks not only measure robot performance but also reflect human experience and acceptance, enabling fair comparison of methods and accelerating progress toward socially intelligent robotic systems.

1.5. Open Challenges in Social Navigation

Open challenges in socially-aware robot navigation span from technological and human-centered, to societal dimensions, reflecting the interdisciplinary nature and complexity of the field. A central challenge is the **fundamental understanding of human behavior**, including cultural variability, individual preferences, and evolving social norms. Robots must not only navigate safely but also interpret intentions, predict actions, and adapt to cooperative or ambiguous human behaviors in real time. A social robot that runs on some form of AI should not only achieve its goals through pure prediction but also try to understand and predict its actions consequences. This includes handling abnormal or adversarial interactions such as obstruction or vandalism, where robust and adaptive decision-making becomes essential for maintaining safety and functionality.

Another major challenge lies in **urban integration and regulation**. Current cities are not inherently designed for autonomous systems, and differences in infrastructure—such as narrow historical streets versus modern wide urban layouts—directly influence robot deployment strategies. Establishing dedicated regulatory frameworks, spatial segregation zones, and coexistence rules with other agents (e.g., bicycles, vehicles, pedestrians) is necessary for scalable and safe deployment of social robots in public spaces.

From an AI and decision-making perspective, key challenges include **situational awareness, conflict detection, and predictive reasoning**. Robots must continuously interpret dynamic environments, detect potential human-robot conflicts early, and anticipate future trajectories to enable proactive rather than reactive behavior. This requires deeper integration of perception, prediction, and planning modules that can reason over both short-term interactions and long-term cooperative tasks.

Finally, **evaluation remains one of the most critical unresolved challenges**, as traditional navigation metrics fail to capture social dimensions such as trust, comfort, and acceptability. Developing robust evaluation frameworks that incorporate subjective human feedback alongside objective performance measures is essential for aligning robotic behavior with human expectations. Overall, these challenges highlight that future progress in socially-aware navigation will depend not only on improved algorithms, but also on advances in human understanding, regulation, and evaluation methodologies.

1.6. Ethical-Social-Economic Considerations

Real-world studies in socially-aware navigation must be conducted under strict methodological and ethical standards to ensure both scientific validity and participant well-being. First, researchers must follow established **human subject research guidelines**, including informed consent, privacy protection, and institutional ethical approval, ensuring that participants are fully aware of the study’s purpose and potential risks. Second, **safety must be preserved at all times**, meaning the robot’s behavior should be rigorously tested to avoid physical harm, discomfort, or psychological stress to humans in shared environments. Third, studies should **report methods and experimental setups in a reproducible manner**, including environment conditions, robot configurations, evaluation metrics, and participant demographics, so that results can be independently validated. Finally, data collection must be driven by **clear scientific objectives**, ensuring that every logged interaction, questionnaire, or observation directly supports well-defined research questions rather than opportunistic or excessive data gathering.

As humans will be involved in any real world experimental study an approval of the Ethics Committee and informed consent of the humans must be obtained. The ethical-social-economic considerations highlighted below are derived from the works of [1,2,3] and grounded by the three laws of Asimov [4]. A robot may not injure a human being, or, through inaction, allow a human being to come to harm. A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law. A robot must protect its own existence, except where such protection would conflict with the First or Second Law.

Social robots raise significant ethical concerns, particularly regarding the potential to influence or manipulate human behavior. Such systems should not shape users' decisions in ways that are unfair, coercive, or not fully understood by the user. For instance, a care robot deployed in a hospital setting should assist patients in making informed choices, rather than subtly pressuring them toward specific outcomes. This concern aligns with broader discussions in [1,2], where the autonomy and influence of intelligent systems are critically examined.

Another major issue concerns privacy and data protection. Socially aware robots often rely on sensitive data such as facial recognition, voice input, emotional cues, and location tracking. Ethical questions arise regarding what data is collected, who has access to it, and whether it is stored securely with informed consent. For example, a home assistant robot should not record private conversations or transmit them to third parties without explicit permission. These concerns are consistent with established ethical frameworks and are reinforced by regulatory requirements such as the General Data Protection Regulation, which governs the collection and processing of personal data within the European context.

Closely related is the issue of dual-use and potential misuse. Technologies designed for beneficial purposes, such as healthcare or assistive robotics, may also be repurposed for surveillance or military applications. Ethical design must therefore anticipate not only intended uses but also possible harmful applications. This forward-looking perspective is also reflected in current research, including [3], which emphasizes responsible evaluation and deployment of robotic systems in social environments.

Legal and regulatory considerations are equally important. Compliance with existing frameworks, such as data protection laws and safety standards, is essential to ensure responsible deployment. For instance, adherence to ISO 13482 helps ensure safe human–robot interaction. Additionally, emerging AI regulations, particularly risk-based classification approaches in the European Union, aim to ensure that robotic systems are deployed in a lawful and ethical manner. Clear liability frameworks must also be established to determine responsibility in cases of system failure or harm, addressing whether accountability lies with manufacturers, developers, users, or deploying organizations.

Beyond legal compliance, broader societal impacts must be considered. The increasing use of robotics may lead to job displacement and economic restructuring, as well as changes in human relationships due to automation. In sensitive domains such as elder care, there is a risk that robots may reduce meaningful human interaction or compromise human dignity. Ethical deployment requires ensuring that robots complement rather than replace human care, and that users are treated with respect rather than as passive recipients of automated services.

Transparency is another key ethical requirement. Individuals should be aware when they are interacting with a robot and understand, at least at a basic level, how it makes decisions. It is ethically problematic for robots to conceal their artificial nature or imitate human behavior in deceptive ways. This concern connects to the broader "black-box" problem in artificial intelligence, where decision-making processes are not easily interpretable.

Finally, safety and fairness must be central to the design of social robots. Systems must be engineered to prevent both physical harm, such as accidents during interaction, and psychological harm, including manipulation or emotional distress. Furthermore, AI-driven robots may inherit biases from their training data, raising concerns about unequal treatment based on age, gender, ethnicity, or disability. Ensuring fairness and inclusivity is therefore essential for ethical human–robot interaction.

The evaluation of social robot navigation requires a multidimensional approach that balances technical performance with human-centered considerations. A primary requirement is **safety**, ensuring that the robot avoids causing harm to itself, nearby humans, or other agents in the environment. Beyond physical safety, **comfort** is also essential, as robots should minimize behaviors that may cause stress, annoyance, or psychological discomfort to people in their vicinity.

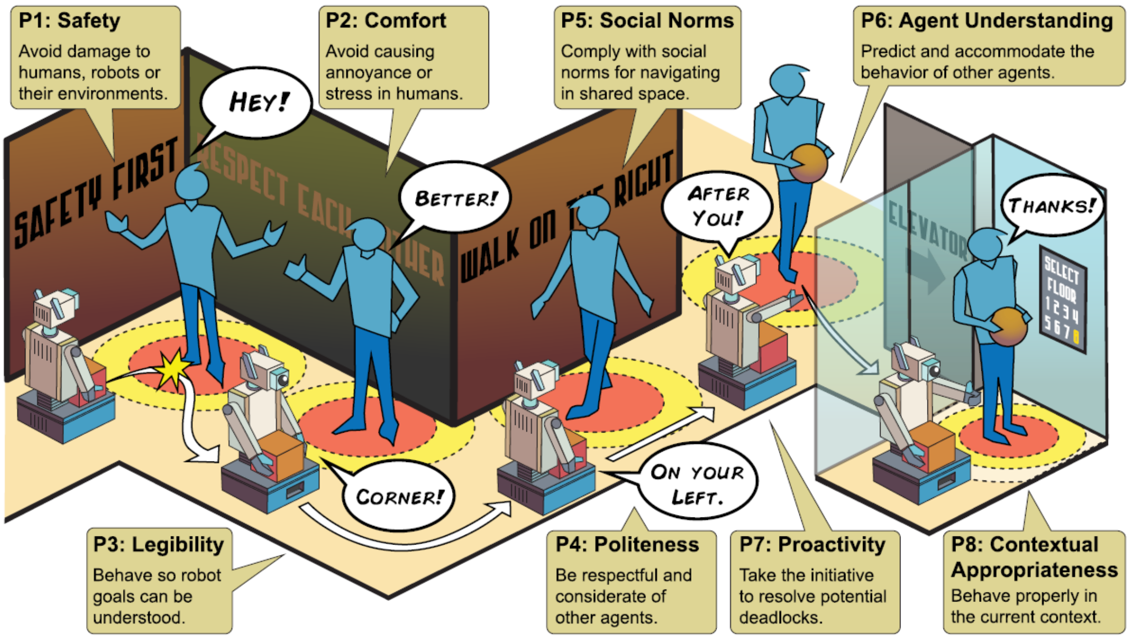

Figure 4. Eight principles for socially-aware robot navigation: safety, comfort, legibility, politeness, social competency, agent understanding, proactivity, and contextual appropriateness. [3]

Equally important is **legibility**, which refers to the robot's ability to behave in ways that make its intentions and goals understandable to surrounding agents, including both humans and other robots. This transparency in motion and decision-making helps foster trust and predictability in shared environments. In addition, robots must adhere to **social norms**, such as respecting personal space (proxemics) [5], adapting to cultural expectations, and using appropriate nonverbal cues like gaze behavior during interactions. These aspects are critical for seamless integration into human-centered spaces.

Closely related to social norms is the principle of **politeness**, which requires robots to act in a respectful and considerate manner toward others. This includes yielding space when appropriate and avoiding intrusive or disruptive actions. Furthermore, effective navigation depends on **agent understanding**, meaning that the robot should be capable of anticipating and adapting to the behavior of nearby agents, enabling smoother and more natural interactions.

Another key dimension is **proactivity and contextual appropriateness**. Social robots should not merely react passively but should take initiative when necessary, for example, by resolving navigation deadlocks or adjusting behavior based on environmental and social context. This requires awareness of dynamic conditions, including changes in crowd behavior, spatial constraints, and situational norms.

Finally, there is an inherent **trade-off between efficiency and social compliance**. While robots are expected to complete tasks effectively and within reasonable time constraints, purely optimal navigation strategies may conflict with socially acceptable behavior. Therefore, socially aware navigation requires balancing task performance with adherence to human expectations and comfort, ensuring that efficiency does not come at the expense of social acceptability.

The economic implications of socially aware robot navigation span multiple dimensions, reflecting both the opportunities and costs associated with deploying such systems across different sectors. Socially aware robots are increasingly relevant in domains such as healthcare, logistics, and service industries, where interaction with humans is frequent and unavoidable. In these contexts, organizations must evaluate the **cost–benefit trade-off** between deploying robotic systems and relying on human labor, particularly in environments like warehouses or hospitals where efficiency and safety are critical.

A significant economic factor lies in the **development and deployment costs** of socially aware robots. Compared to basic autonomous systems, these robots require more advanced sensing, perception, and machine learning capabilities to interpret and respond to human behavior. This increases both hardware and software complexity, leading to higher initial investment costs. Additionally, **operational efficiency** may be affected, as socially compliant navigation often requires slower movements or less direct paths in order to respect human comfort and social norms. This creates a trade-off between productivity and social acceptability, which organizations must carefully balance.

Long-term considerations include **maintenance and scalability**, as socially aware systems often rely on continuous data collection, updates, and model retraining to remain effective in dynamic environments. These requirements can significantly increase operational costs, particularly when scaling up to large fleets of robots. At the same time, the **labor market impact** must be taken into account. While such systems may partially replace human roles in certain sectors, they can also create new opportunities in areas such as system supervision, maintenance, and AI development, contributing to a broader restructuring of the workforce.

Another key aspect is the **return on investment (ROI)**. Organizations must assess whether the added value of socially aware capabilities—such as improved safety, enhanced user experience, and increased customer satisfaction—justifies the additional costs. This is closely linked to **adoption and user acceptance**, as the economic success of these systems depends heavily on public trust and willingness to interact with robots. Socially compliant behavior can enhance acceptance, thereby increasing the likelihood of successful deployment and market viability.

Moreover, **infrastructure and integration costs** represent an additional economic consideration. The deployment of socially aware robots may require modifications to existing environments, such as redesigning spaces in hospitals, warehouses, or public areas to support safe and effective human–robot interaction. These adjustments can further influence the overall cost and feasibility of implementation.

Finally, significant effort should be devoted to encouraging public acceptance and effective cooperation with robots in shared environments. As robots become more human-like in both appearance and behavior, people may respond to them more positively; however, this also introduces challenges such as the *uncanny valley* effect [6], where overly human-like robots can cause discomfort. Therefore, further research is needed to better understand how to increase human acceptance, foster trust, and support smooth human–robot collaboration. In addition, users should be appropriately informed and trained to interact with robotic systems, ensuring safe, efficient, and socially acceptable cooperation.

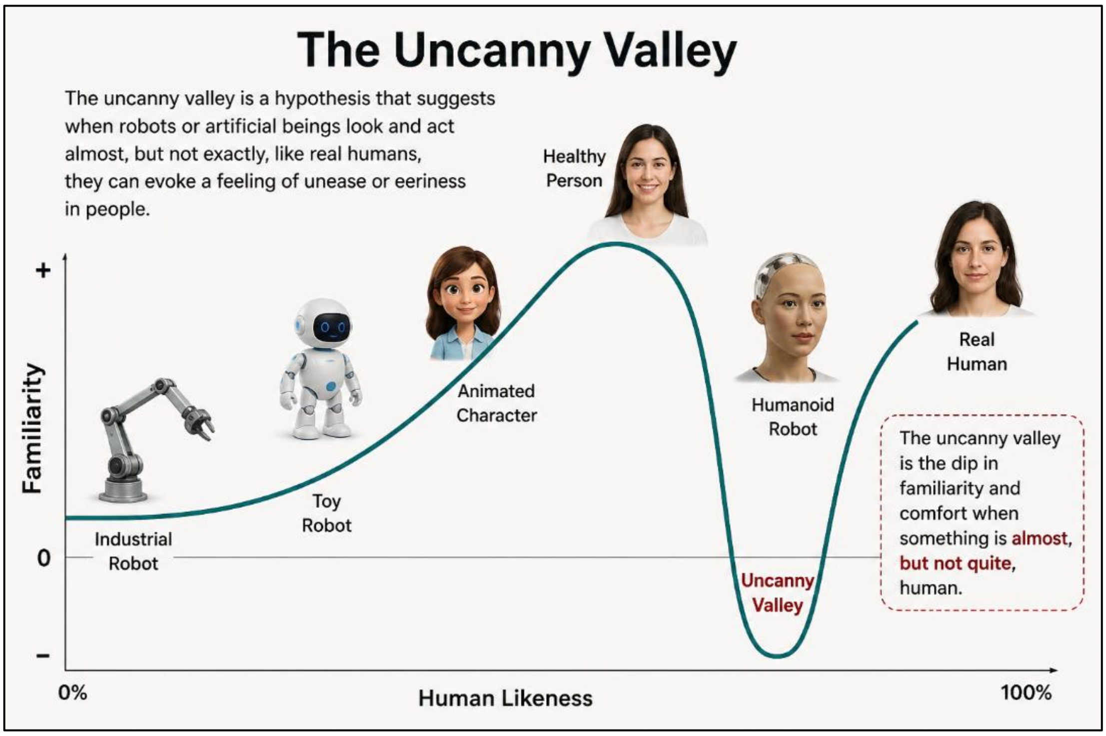

Figure 5. Anthropomorphism Uncanny-Valey effect in social robots.

1.7. Trustworthy AI Principles

Beyond ethical, social, and economic considerations, the RoboCARE framework adopts the principles of Trustworthy Artificial Intelligence to ensure that robotic decision-making remains transparent, robust, accountable, and aligned with human values. Since RoboCARE operates in environments involving continuous interaction with humans, particularly in domestic and crowded settings, trustworthiness is considered a fundamental design requirement rather than an optional system feature.

A primary requirement is transparency and explainability. Users should be able to understand the rationale behind the robot's actions and navigation decisions. To support this objective, RoboCARE incorporates Explainable AI (XAI) mechanisms capable of providing human-understandable explanations regarding environmental perception, behavioral adaptations, navigation decisions, and interaction strategies. Such explanations are expected to improve user trust, acceptance, and cooperation.

Another important requirement is robustness and reliability. Socially-aware robotic systems must continue operating safely under uncertain and dynamic conditions, including sensor failures, environmental changes, unexpected human behaviors, or incomplete information. Therefore, RoboCARE will employ confidence estimation, uncertainty-aware decision-making, and conservative fallback behaviors whenever system confidence falls below acceptable thresholds.

The framework also promotes human oversight and autonomy preservation. The robot is intended to assist rather than replace human decision-making. Human users should always retain the ability to intervene, override, redirect, or terminate robot actions whenever necessary. In situations involving uncertainty or conflicting objectives, the system should prioritize human safety, comfort, and explicit user preferences.

Fairness and inclusiveness constitute another important dimension of trustworthy robotics. The system should avoid systematic performance differences across diverse user groups and environmental conditions. Although RoboCARE is not intended to make high-stakes decisions about individuals, efforts should be made to evaluate model performance across different demographic and contextual settings whenever representative datasets are available.

Finally, accountability and auditability are essential requirements for responsible deployment. RoboCARE will maintain logs of relevant perception outputs, adaptation decisions, user interactions, and explanation requests to support system evaluation, debugging, and future auditing activities. Such records contribute to transparency and facilitate continuous improvement of socially-aware robotic behaviors.

Collectively, these principles ensure that RoboCARE aligns with contemporary Trustworthy AI guidelines and supports the development of robotic systems that are socially acceptable, human-centric, safe, and accountable.

2. Use Cases

2.1. Use Case 1: Domestic Environment

2.1.1. UC 1.1: Place-Aware Navigation Adaptation (Room2Room Intelligence)

This scenario demonstrates the robotic agent's ability to **adapt its navigation behavior based on the recognized environment**, enabling context-aware operation across different indoor spaces.

Figure 6. The robot transmits from Room A to Room B

As the robot moves within the environment, visual sensory input is continuously processed to identify the current place (e.g., office, corridor, living room). The recognized place is treated as a **high-level contextual feature**, which is propagated to downstream modules responsible for socially-aware navigation. Based on this contextual information, the robot dynamically adjusts its navigation behavior. For instance, in open spaces, the robot may increase its velocity and follow more direct paths, while in constrained or shared environments, it may reduce speed and adopt more cautious navigation strategies. These adaptations ensure **safe, efficient, and context-aware navigation** during room-to-room transitions.

This scenario highlights the integration of **Perception (TO1)** with visual place recognition, and **Adaptation (TO2)** with social-aware adaptation of navigation behavior, demonstrating how environmental understanding directly influences robotic decision-making.

Actors:

Primary Actor: Robotic Agent

Secondary Actors: Visual Place Recognition Module, Natural Language Processing Module, Social-Aware Adaptation Module.

Preconditions:

1. Camera sensor is active and calibrated.
2. The visual place recognition model is trained on at least 6 distinct environment classes.
3. The operational environment contains known (trained) place categories.
4. Communication between perception and navigation modules is established.

Postconditions:

1. A semantic label (place class) is produced for the current scene.
2. The label is encoded as a feature (e.g., integer ID) and published to other system modules.
3. Navigation behavior is adapted based on the recognized place (e.g., reduced speed in corridors).

Success Guarantee:

- The system correctly classifies the current environment and provides the corresponding feature.
- The system provides a confidence score associated with the prediction.

Minimal Guarantee: If confidence is below a predefined threshold (<0.5), the system outputs an ``unknown'' label ID: 0, that will run default safe control navigation.

Trigger: Execution of a navigation task or activation of autonomous movement.

Main Success Scenario:

1. The robot receives a navigation command from point A to point B.
2. The navigation system computes a global path using the occupancy map.
3. The perception module processes incoming visual sensory data.
4. The system classifies the current scene into a semantic place category (place ID).
5. The place ID is transmitted to the navigation system.
6. The navigation system adapts its behavior based on the recognized place.
7. The robot continues navigation with context-aware adjustments either in global or local planner/controller.
8. The system continues monitoring and updating the feature provided by Visual Place Recognition model.

Alternative Flow: Recognition Failure

- The perception module cannot determine a specific place.
- label "unknown" ID: 0, is assigned.
- The system continues operation using default conservative behavioral parameters.

Interfaces:

1. Subscribes to image streams from the camera.
2. Publishes place recognition output (label, ID, confidence) to the navigation module.

2.1.2. UC 1.2: Action-Aware Behavioral Adaptation (Action2Action Response)

This scenario demonstrates the robotic agent's ability to **adapt its behaviour based on changes in human activity**, as perceived through the action recognition module.

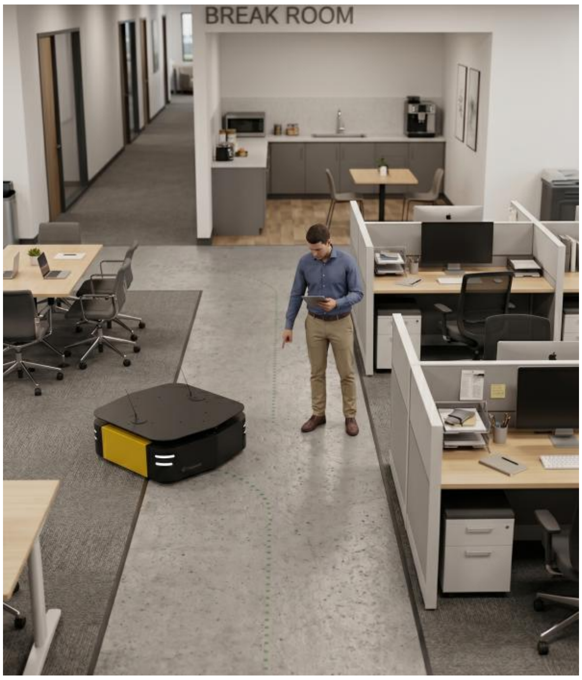

Figure 7. The robot adapts according to human's actions transitions

Within the robot's frontal field of view (FoV), human actions are monitored and classified (e.g., sitting, standing, walking). The emphasis of this use case is on how **the initial social setting along with the detected transitions between actions** are exploited to trigger social-aware behavioural adaptation. When a **change in human activity is successfully perceived**, the robot dynamically adjusts its navigation and interaction strategy. For instance, a transition from a seated to a standing or walking state indicates increased human mobility and potential movement within the environment. In response, the robot may: reduce its speed, re-plan its navigation path, increase interpersonal distance, or temporarily yield priority to the human.

These adaptations ensure **safe, socially compliant, and context-aware navigation**, particularly in shared domestic environments. The recognized action is treated as a **contextual feature**, which is propagated to the social-aware adaptation module (TO2), enabling modulation of: locomotion parameters (e.g., velocity, trajectory), proxemic behaviour (e.g., distance regulation), interaction readiness (e.g., approaching vs. avoiding).

This scenario highlights the integration of **Perception (TO1)** and **Adaptation (TO2)**, demonstrating how real-time understanding of human activity directly influences robotic decision-making and behaviour.

Actors:

Primary Actor: Robotic Agent

Secondary Actors: Action Recognition Module, Natural Language Processing Module, Social-Aware Adaptation Module.

Preconditions:

1. Camera sensor is active, calibrated, and streaming reliably.
2. The action recognition model is trained on at least 6 distinct human actions.
3. The environment contains a human participant performing observable actions.

4. Communication between perception, action recognition, and downstream modules is established.

Postconditions:

1. A semantic label representing the human action.
2. The label is encoded as a feature (e.g., integer ID) and published to downstream modules.
3. Context-aware behaviours are updated based on the recognized action (e.g., speed reduction, proxemic adjustments, task prioritization).

Success Guarantee:

- The system correctly classifies the observed human actions.
- The recognized actions are encoded and transmitted successfully to downstream modules for behaviour adaptation.

Minimal Guarantee: If recognition confidence is low, the system outputs an ``unknown'' label ID: 0, that will run default safe control navigation.

Trigger: Detection of a human in the robot's FoV or initiation of a task requiring action awareness.

Main Success Scenario:

1. A human is detected in the robot's FoV.
2. The human is recognized by the perception model
3. The perception and pose estimation module capture image data and extracts joint positions.
4. Temporal sequences of joint positions are fed into the action recognition model.
5. The model classifies the human action into one of the trained categories, e.g, talking in group of people, sitting on a chair, making a gesture of stop or come, etc.
6. The recognized action is encoded as a feature (e.g, action ID) and transmitted to downstream modules.
7. Downstream modules adapt robot behaviour according to the recognized action.
8. The system continues monitoring and updating action recognition continuously.

Alternative Flow A:

- Sensor failure – frame stream unavailable, system outputs ``unknown'' action label.
- Unseen action – Human performs an action that is not present in the training set, or with low confidence level, system outputs ``unknown'' label and continues safe operation.

Interfaces:

- Subscribes to frame streams.
- Processed, feed and receive joint estimation from pose estimation module.
- Recognize the human action/s inside the robot's FoV.
- Publishes recognized action features (action ID, label, confidence) to downstream modules.

2.1.3. UC 1.3: Emotion-Driven Social Adaptation (Affective Feedback Learning)

This scenario demonstrates the robotic agent's ability to **adapt its behavior based on human emotional responses to its own actions**, enabling a human-in-the-loop learning process for socially-aware interaction.

Figure 8. The robot adapts based on human's non-verbal feedback

In this use case, the robot initially executes a navigation or interaction behavior (e.g., approaching a user, entering a shared space, or initiating engagement). The human's **emotional reaction and behavioral response** to this action are then perceived through multi-modal sensing, including facial expressions and body language (e.g., stop gesture, withdrawal, approach).

The system combines **emotion recognition outputs**, i.e., continuous affective state, and **observed human actions**, e.g., gestures, posture, movement, to interpret the human's feedback regarding the robot's behavior.

Based on this interpretation, the robot dynamically adjusts its behavior policy. For example:

- if the robot approaches and perceives **discomfort or aversion**, it increases interpersonal distance or stops,
- if the user responds with **positive affect or welcoming gestures**, the robot maintains or reinforces its behavior,
- if a **negative emotional response is combined with a prohibitive action** (e.g., stop gesture), the robot immediately halts and re-evaluates its strategy.

Crucially, these interactions are not treated as isolated events. The system calibrates this feedback based on the **long-term behavior understanding module**, which aggregates interaction history and emotional responses to construct a behavioral profile of the user.

Furthermore, a **lifelong learning mechanism with human-in-the-loop feedback** updates the robot's decision-making policy over time. For instance, if the robot repeatedly learns that approaching a specific user during rest periods leads to negative reactions, it adapts its future behavior to avoid similar actions. The emotional and behavioral responses are thus treated as **implicit feedback signals**, guiding the continuous refinement of the robot's social navigation and interaction strategies.

This scenario highlights the integration of **Perception (TO1)**: multi-modal emotion and action recognition, and **Adaptation (TO2)**: behavior policy update through feedback and lifelong learning, demonstrating how human responses directly shape the robot's future behavior in a personalized and socially compliant manner.

Actors:

Primary Actors: Robotic Agent, Human

Secondary Actors: Multi-Modal Emotion Recognition Module, Long-Term Behavior Understanding Module, Lifelong Learning Module.

Preconditions:

1. Camera sensor is active, calibrated, and streaming reliably.
2. Action and Emotion Recognition models are activated and produced results
3. Long-term behavior understanding module is running providing extra information
4. Lifelong learning module provides extra information
5. Produced features are fed to the social-aware adaptation module
6. Robot executes task
7. Lifelong learning module is refined based on the interaction and feedback acquired

Postconditions:

1. Human emotional states are estimated and encoded as features.
2. Social-aware adaptation module adjusts robot behavior accordingly.
3. Lifelong learning module updates the system based on explicit or implicit human feedback.

Success Guarantee:

- The system accurately estimates human emotions and integrates them with action-based signals.
- Robot behavior is adapted in a context-aware, socially appropriate manner.

Minimal Guarantee: If emotion or action recognition fails, the system defaults to safe, neutral behavioral parameters.

Trigger: Detection of human presence in robot's operational area or activation of a human-interaction task.

Main Success Scenario:

1. Robot detects a human in its FoV.
2. Perception modules captures data.
3. Emotion recognition module estimates current human emotional state.
4. Action recognition module estimates current human activity.
5. Features from emotion and action modules are integrated and sent to the social-aware adaptation module.
6. Robot adjusts proxemic and task behavior based on combined context-aware inputs and long-term behavior understanding module.
7. Lifelong learning module updates system parameters based on human feedback.
8. System continuously monitors, updates, and adapts interaction behavior.

Alternative Flow A:

- If emotion estimation is ambiguous, robot behavior relies on long-term behavior understanding and default safety rules.
- If action recognition is ambiguous, robot defaults to safe movement and monitoring until clearer signals are available.
- Sensor failure – frame stream unavailable, system outputs ``unknown'' action label.
- Can not recognize emotion state, system outputs "unknown"

Interfaces:

- Subscribes to sensors data streams.
- Communicates with emotion recognition, action recognition, long-term behavior understanding, and social-aware adaptation modules.
- Publishes feature vectors representing human emotional and action states.
- Receives feedback signals for lifelong learning updates.

2.1.4. UC 1.4: Explainable Human-Robot Dialogue (NLP & XAI Interaction)

This scenario demonstrates the interaction between a human user and the robotic agent through natural spoken language, enabling intuitive communication and transparent access to the robot's internal state.

Figure 9. The robot replies to user queries

The system is able to capture audio input via a microphone and convert it into text using a speech-to-text (STT) module. Moreover, a natural language processing (NLP) module extracts user intent and relevant contextual information. Based on the interpreted command or query, the robotic agent can execute actions, e.g., navigation or task-related commands, provide informative responses, e.g., current position, perceived environment, or system status, or explain its decisions and behaviour.

A key aspect of this use case is the integration of **Explainable AI (XAI)** capabilities. The robot is able to generate **human-understandable explanations** regarding information, such as why a specific action was taken, which contextual factors (e.g., perceived environment, human activity), influenced its decision, and how its internal state evolves over time. Additionally, the system can **externalize aspects of its internal state**, e.g., confidence levels, feature contribution, or internal processing steps, enabling the user to better understand and trust the robot's behaviour. This bidirectional interaction allows users not only to **query the robot**, but also to **inspect its decisions**, supporting transparency, trust, and effective collaboration.

This scenario highlights the integration of **Perception (TO1)** with speech-to-text and language understanding, **Adaptation (TO2)** with interaction-driven updates through user input and **Convey (TO3)** with explainability, emotion elicitation and communication of internal reasoning, demonstrating a seamless and transparent human-robot interaction framework in domestic environments.

Actors:

Primary Actors: Human User, Robotic Agent

Secondary Actors: Speech-to-Text Module, Natural Language Processing, Social-Aware Adaptation Module, Lifelong Learning Module, Interactive Natural Language Processing, Explainable AI Module, Emotion Elicitation Module

Preconditions:

1. Microphone is active and functioning correctly.
2. Speech-to-text module is initialized and capable transcription.
3. NLP module is trained for query handling.
4. Perception Modules initialized.
5. Communication between NLP and downstream modules is established.

Postconditions:

1. The user's spoken input is correctly interpreted as a command or query.
2. The corresponding system action is executed or a response is generated.

Success Guarantee:

1. The system correctly interprets the user's intent and executes the appropriate action or provides an accurate response.

Minimal Guarantee: If the input cannot be confidently interpreted, the system requests clarification or outputs a default response (e.g., ``I did not understand the command'').

Trigger: User provides a spoken command or query.

Main Success Scenario:

1. The user provides a spoken command or question (e.g., ``Go to the kitchen'', ``Where are you?'', ``What do you see?'', "What emotional states do you detect?").
2. The microphone captures the audio input.
3. The speech-to-text module converts the audio into text.
4. The NLP module processes the text and extracts the user query and relevant parameters.
5. The system determines whether the input corresponds to a command or a query.

6. If it is a command, the appropriate module is activated (e.g., navigation system for movement).
7. If it is a query, the relevant information is retrieved (e.g., robot position, perceived objects, the current FoV information (frame)).
8. The response is delivered to the user via text.

Alternative Flows:

1. Speech recognition failure – audio cannot be transcribed reliably.
2. NLP ambiguity – query cannot be clearly determined.
3. Module unavailability – requested subsystem (e.g., navigation) is not available.

Interfaces:

1. Subscribes to audio input from microphone.
2. Communicates with NLP module for intent extraction.
3. Extracts the useful information like current frame or position from odom
4. Communicates with navigation and perception modules, if necessary, to extract results.
5. Presents response to user in PC.

2.2. Use Case 2: Crowded Environment

2.2.1. UC-2.1: Crowd-Aware Behavioral Adaptation (Empty2Crowded Room Response)

This use case demonstrates the robotic agent's ability to adapt its navigation and interaction behavior when transitioning between environments with different social densities, following user instructions, e.g., provided through NLP modules.

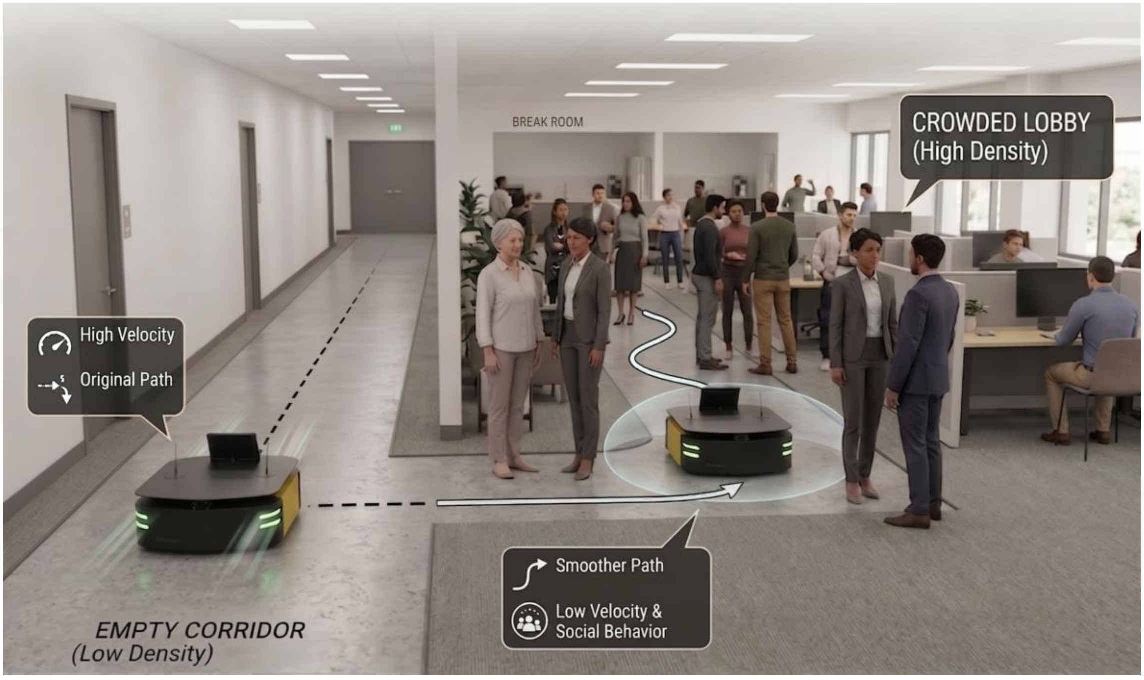

Figure 10. The robot transmits from empty Room A to crowded Room B

A user may instruct the robot to navigate from location A to location B through natural language. During navigation, the robot moves across environments with varying contextual and social characteristics, such as transitioning from an empty corridor or room into a crowded shared space populated by multiple humans.

The robotic agent continuously analyzes visual sensory input and environmental context to recognize the current place, estimate crowd density and dynamically adapt its navigation behavior. Information extracted from the perception modules is propagated to downstream socially-aware adaptation systems, enabling context-aware modifications including velocity adjustment, trajectory re-planning, interpersonal distance regulation, obstacle avoidance, and socially compliant navigation in crowded environments.

For example:

- in empty or low-density spaces, the robot may adopt faster and more direct navigation behaviors,
- while in crowded environments, it reduces speed, increases proxemic awareness, and performs smoother path adaptation to safely navigate among humans.

This scenario highlights the integration of **Perception (TO1):** visual place recognition, and natural language processing, **Adaptation (TO2):** socially-aware navigation.

Actors:

Primary Actors: Robotic Agent, Human Users

Secondary Actors: Visual Place Recognition Module, Action Recognition Module, Social-Aware Adaptation Module

Preconditions:

1. Camera sensors are active, calibrated, and streaming reliably.
2. NLP modules are initialized and capable of real-time processing.
3. Visual Place Recognition, Emotion Recognition, and Action Recognition models are trained and operational.
4. Communication between perception, adaptation, and navigation modules is established.
5. An occupancy map of the environment is provided to the system.
6. The environment contains recognizable landmarks and human entities.

Postconditions:

1. The robot successfully interprets the navigation command and reaches the target destination.
2. Navigation behavior is dynamically adapted according to environmental and social context.
3. Context-aware navigation and interaction behaviors are executed and recorded.
4. Perception, emotion, and action features are continuously updated and shared with downstream modules.

Success Guarantee:

- The system correctly interprets user commands and executes them as intended.
- The robot safely navigates across environments with varying crowd density.

- Context-aware adaptations (e.g., speed, path planning, proxemics) are appropriately applied based on perception environmental cues.

Minimal Guarantee: If perception or command interpretation fails, the robot defaults to safe navigation behavior, reduces locomotion speed, and requests clarification from the user.

Trigger: A user provides a command via NLP interface requesting navigation and/or guidance actions.

Main Success Scenario:

1. The user issues a command through the NLP interface.
2. NLP module extracts the intent and relevant parameters.
3. The robot analyzes the surrounding environment using visual and contextual perception modules.
4. The Visual Place Recognition module identifies the current environmental context.
5. Human activity and crowd density are continuously estimated through perception modules.
6. The socially-aware navigation module dynamically adjusts locomotion behavior according to the perceived context.
7. The navigation system computes a path and adjusts locomotion parameters based on social and environmental context.
8. The robot safely navigates from low-density to crowded environments while adapting speed, trajectory, and proxemic behavior.
9. Information is continuously updated and propagated to downstream modules for future adaptation and learning.

Alternative Flows:

- If the robot detects excessive uncertainty in perception or localization, it switches to conservative navigation behavior and requests user clarification.
- If crowd density exceeds predefined safety thresholds, the robot reduces speed or temporarily pauses navigation until safe movement is possible.

Interfaces:

- Communicates with the NLP module for intent extraction.
- Receives perception data from camera and other sensors.
- Publishes recognized place, emotion, and action features to downstream modules.
- Controls navigation and social-aware adaptation modules to execute guided behavior.

2.2.2. UC-2.2: Place-and-Crowd-Aware Behavioral Adaptation (CrowdActions2Place Re-Action)

In this use case, an autonomous agent adapts to its spatial behavior based on both the proxemic and environmental context of its surroundings. Using Visual Place Recognition, the agent identifies the type of space it is in—such as a meeting room, hallway, or workspace.

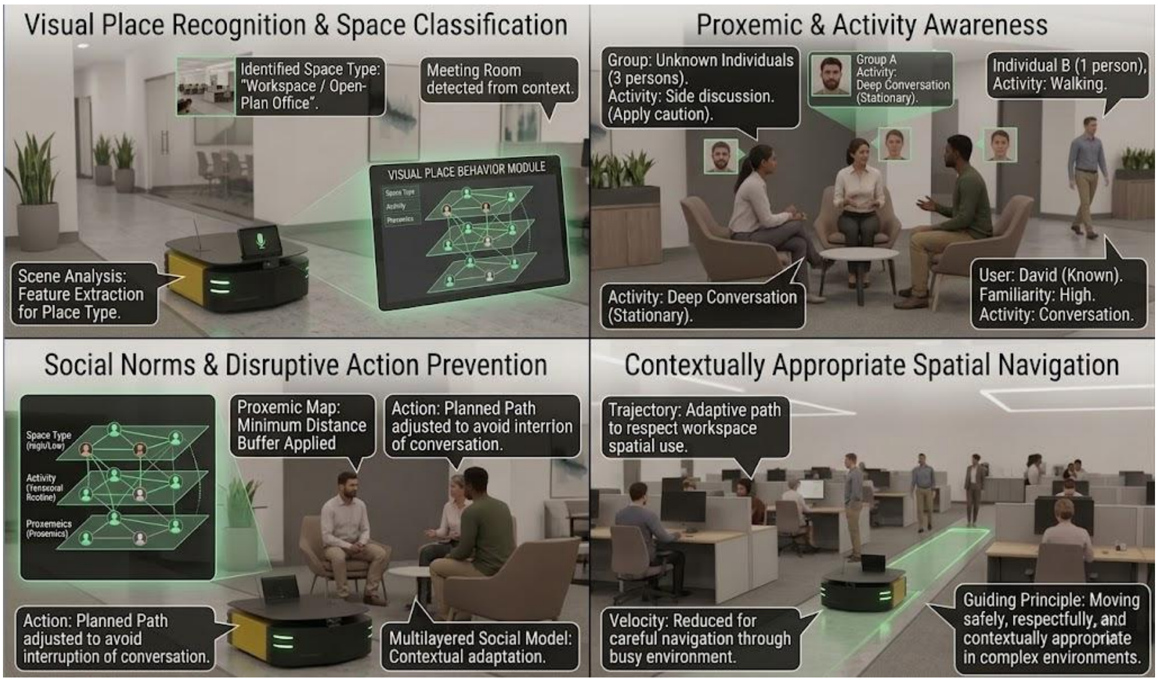

Figure 11. The robot adapts navigation in crowded places shifts

By understanding where group of people are located and what they are doing—speaking, moving, or performing tasks—the agent adjusts its position and movement to respect social norms and avoid disruption. For example, it may maintain distance from a group engaged in conversation in a lounge area, or navigate carefully through a busy workspace without interfering with ongoing activities. This integration of place context, proxemic awareness, and action-based social adaptation enables the agent to move safely, respectfully, and contextually appropriately in complex human environments.

This scenario highlights the integration of **Perception (TO1)** with place and action recognition as well as crowd detection, and **Adaptation (TO2)** with social-aware navigation, demonstrating a seamless and transparent human-robot interaction framework in domestic environments.

Actors:

Primary Actors: Robotic Agent, Human

Secondary Actors: Visual Place Recognition Module, Action Recognition Module, Social-Aware Adaptation Module.

Preconditions:

1. Camera and LiDAR sensors are active, calibrated, and streaming reliably.
2. Visual Place Recognition (VPR) and crowd detection modules are initialized and operational.
3. Action recognition models are trained and ready to process real-time human activity.
4. Robot has map data or prior experience of the environment for context-aware navigation.

Postconditions:

1. Robot successfully identifies the type of space it is operating in.
2. Robot adjusts its movement trajectory based on crowd distribution and proxemic rules.
3. Social-aware adaptation module logs behavioral adjustments for lifelong learning updates.

Success Guarantee:

1. Robot navigates safely through the environment while maintaining socially appropriate distances from humans.
2. Crowd actions and spatial context are correctly integrated into robot decision-making.

Minimal Guarantee: Robot defaults to safe navigation without causing disruption even if place recognition or crowd action estimation fails.

Trigger: Detection of a crowd or entry into a new spatial environment requiring proxemic adaptation.

Main Success Scenario:

1. Robot detects a crowd and perceives the environment using camera and LiDAR sensors.
2. VPR module identifies the type of space (e.g., hallway, meeting room, workspace).
3. Action recognition module classifies crowd activities such as speaking, walking, or stationary tasks.
4. Robot integrates place type, crowd actions, and proxemic rules to plan a socially aware trajectory.
5. Robot navigates while continuously monitoring crowd movements and space constraints.
6. Lifelong learning module updates adaptation parameters based on observed outcomes and human feedback.

Alternative Flows:

1. If VPR fails, robot relies on prior map knowledge or environmental features to infer place type.
2. If some humans are occluded or unrecognized, robot maintains maximum safe distance and monitors until clear signals are available.
3. Sudden crowd movements or obstructions cause robot to halt or re-plan trajectory dynamically.
4. Sensor malfunction triggers fail-safe behavior: stop movement and signal for human assistance if necessary.

Interfaces:

1. Subscribes to camera, LiDAR, and audio sensor streams.
2. Publishes socially aware trajectory plans to the navigation module.
3. Communicates crowd action and spatial context features to social-aware adaptation module.
4. Receives feedback and updates from lifelong learning and long-term behavior modules.

2.2.3. UC-2.3: Behavioural-Aware Social Adaptation (Behavioural Familiarity Response)

This use case demonstrates the robotic agent's ability to adapt its navigation and interaction behavior according to the familiarity and long-term interaction history associated with people present in the environment.

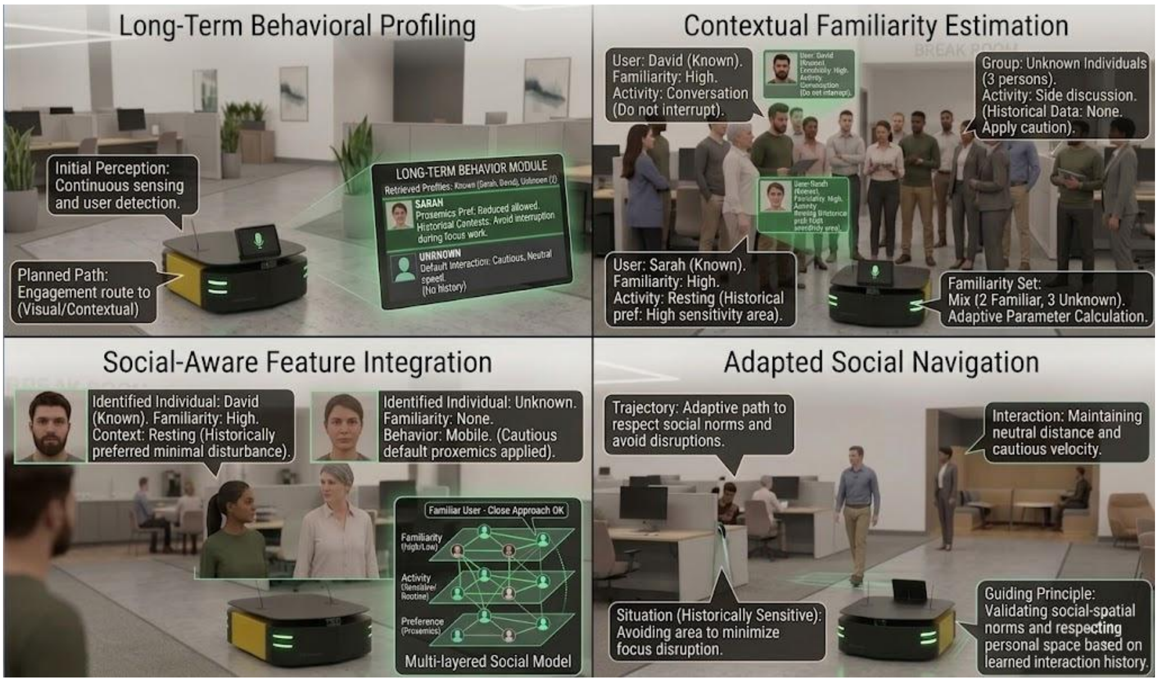

Figure 12. The robot navigates through familiar and unknown humans

As the robot operates in socially dynamic environments, it continuously perceives nearby individuals and groups using visual and contextual sensory information. Beyond detecting human presence, actions, and emotional states, the system incorporates information generated by the Long-Term Behavior Understanding Module, which maintains behavioral profiles derived from historical interactions with known users.

Using this information, the robotic agent estimates familiarity levels with detected individuals, historical interaction patterns, preferred proxemic behaviors, and context-specific behavioral tendencies. These high-level contextual features are propagated to the Social-Aware Adaptation Module, enabling the robot to dynamically adapt its behavior according to the social context and familiarity of the surrounding humans.

For example, the robot may approach a familiar user more naturally and allow reduced interpersonal distance, avoid interrupting a known user during historically sensitive situations (e.g., resting or working), maintain neutral and cautious navigation behavior around unknown individuals, or adapt its trajectory to avoid disrupting groups engaged in conversation.

This capability combines **Perception (TO1)** through emotional and action understanding and **Adaptation (TO2)** with long-term behavioral modeling, and adaptive decision-making, allowing the robot to navigate complex social environments in a personalized, socially appropriate, and goal-directed manner.

Actors:

Primary Actors: Robotic Agent, Human Users

Secondary Actors: Emotion Recognition Module, Action Recognition Module, Long-Term Behaviour Understanding Module, Lifelong Learning Module, Social-Aware Adaptation Module.

Preconditions:

1. Camera sensor is active, calibrated, and streaming reliably.

2. Emotion recognition, action recognition, and crowd detection models are operational.
3. Communication between perception, adaptation, and social-aware modules is established.
4. Environment contains a group of humans for social interaction.

Postconditions:

1. Long-term behaviour understanding and actions are recognized and encoded as features.
2. Robot adapts its behavior according to perceived social context.
3. Interaction decisions are logged and shared with downstream modules.

Success Guarantee:

- System accurately estimates crowd actions and long-term behaviour patterns.
- Robot behaviour is socially appropriate, safe, and goal-directed.

Minimal Guarantee: If perception fails, robot defaults to safe navigation and waits for further input.

Trigger: Detection of humans or groups within the operational environment, or initiation of a socially-aware interaction task.

Main Success Scenario:

1. Robot detects a crowd in its operational area.
2. Perception module estimates humans' positions and actions.
3. The Long-Term Behavior Understanding Module retrieves familiarity and historical interaction information associated with known individuals.
4. Contextual features are propagated to the Social-Aware Adaptation Module.
5. The robot dynamically adjusts navigation behavior, proxemics, and interaction policies according to familiarity and social context.
6. The robot safely navigates and interacts without disrupting nearby individuals or groups.
7. Interaction outcomes are continuously monitored and logged for future adaptation and lifelong learning.

Alternative Flows:

- If individuals are unknown, the robot adopts neutral and cautious navigation behaviour.
- If familiarity confidence is low, the robot adopts defaults navigation behaviour.
- If perception uncertainty increases, the robot reduces speed and maintains safe distances.
- Sensor failure – unavailable sensory input results in fallback safe-navigation policies.

Interfaces:

- Subscribes to image streams.
- Receives familiarity and historical interaction information from the Long-Term Behavior Understanding Module.
- Publishes action, and familiarity features to downstream adaptation modules.
- Communicates with social-aware adaptation and lifelong learning modules for future behaviours.

2.2.4. UC-2.4: Explainable Human-Robot Communication (Robot2User Communication)

This use case describes the interaction between a human user and the robotic agent through natural language, via a personal computer.

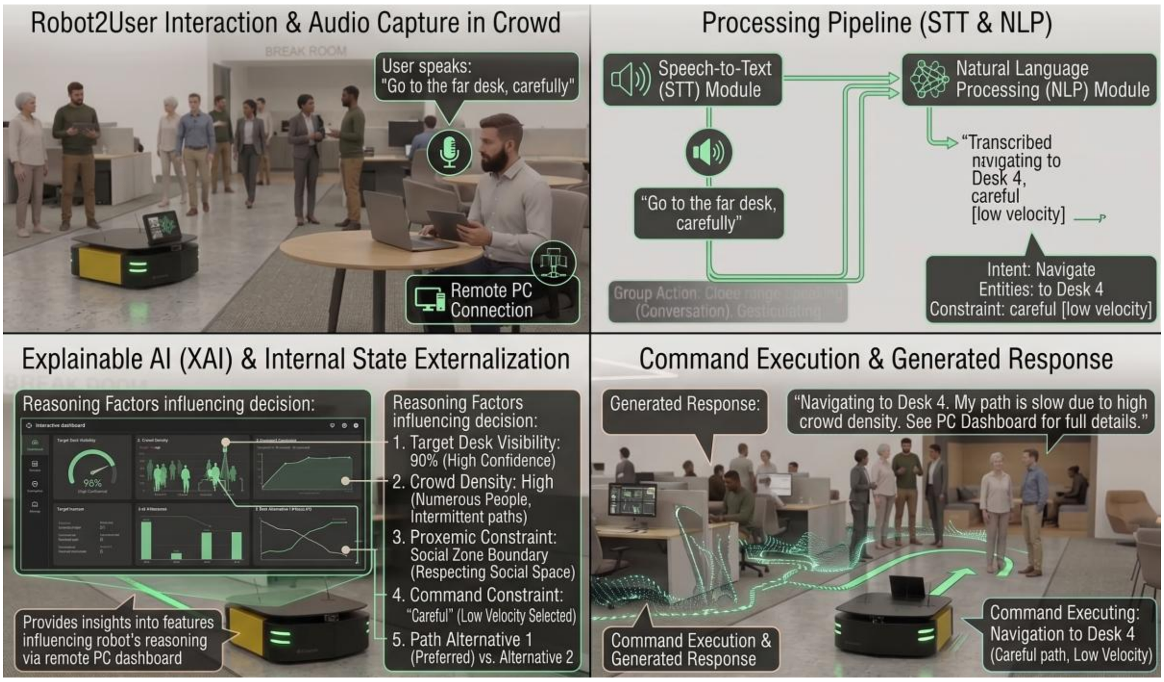

Figure 13. The robot replies to user queries in crowded scenes

The scenarios included are crowded. The system captures audio input via a microphone and converts it into text using a speech-to-text (STT) module. The transcribed input is processed by a natural language processing (NLP) module to extract user intent and relevant entities. Based on the interpreted command or query, the system either executes an action (e.g., navigation command), generates an appropriate response (e.g., reporting position or perceived environment), or provides insights into the features and variables that influence robot's reasoning while having the ability to model and externalize the internal state of the robotic agent via the remote pc connection. This enables intuitive and natural human-robot interaction.

This scenario highlights the integration of **Perception (TO1)** with speech-to-text and language understanding, **Adaptation (TO2)** with interaction-driven updates through user input and **Convey (TO3)** with explainability and communication of internal reasoning, demonstrating a seamless and transparent human-robot interaction framework in domestic environments.

Actors:

Primary Actors: Robotic Agent, Human

Secondary Actors: Action Recognition Module, Social-Aware Adaptation Module, Interactive Natural Language Processing Module, Explainable AI Module, Emotion Elicitation Module

Preconditions:

1. Microphone is active and functioning correctly connected to a PC.
2. Speech-to-text module is initialized and capable of real-time transcription.
3. NLP module is trained for query handling.
4. Perception Modules initialized.

5. Communication between NLP and downstream modules (navigation, perception) is established.

Postconditions:

1. The user's spoken input is correctly interpreted as a command or query.
2. The corresponding system action is executed or a response is generated.

Success Guarantee:

1. The system correctly interprets the user's intent and executes the appropriate action or provides an accurate response.
2. A confidence score is associated with the recognized intent.

Minimal Guarantee: If the input cannot be confidently interpreted, the system requests clarification or outputs a default response (e.g., ``I did not understand the command'').

Trigger: User provides a spoken command or query.

Main Success Scenario:

1. *The user provides a spoken command or question (e.g., ``Go to the kitchen'', ``Where are you?'', ``What do you see?'', "What emotional states do you detect?").*
2. *The microphone captures the audio input.*
3. *The speech-to-text module converts the audio into text.*
4. *The NLP module processes the text and extracts the user intent and relevant parameters.*
5. *The system determines whether the input corresponds to a command or a query.*
6. *If it is a command, the appropriate module is activated (e.g., navigation system for movement).*
7. *If it is a query, the relevant information is retrieved (e.g., robot position, perceived objects, the current FoV information (frame)).*
8. *The response is delivered to the user via text.*

Alternative Flows:

1. The user issues a command (e.g., ``Go to the office''), and the system directly triggers the navigation module providing goal coordinates.
2. The user asks a question (e.g., ``What do you see?''), and the system retrieves the current FoV frame passes and communicates perception results, e.g, i see the inside of an office (VPR result) with two people speaking two each other (action recognition result) being happy (emotion recognition result).
3. Speech recognition failure – audio cannot be transcribed reliably.
4. NLP ambiguity – intent cannot be clearly determined.
5. Module unavailability – requested subsystem (e.g., navigation) is not available.

Interfaces:

1. Subscribes to audio input from microphone.
2. Communicates with NLP module for intent extraction.
3. Extracts the useful information like current frame or position from odometry
4. Communicates with navigation and perception modules, if necessary, to extract results.
5. Presents response to user in PC.

3. Technical Specifications

3.1. Visual Place Recognition Module

Purpose and Role

The Visual Place Recognition (VPR) Module is responsible for identifying and distinguishing indoor environments and contextual locations within the operational space of the robotic agent.

Main Functionalities

- Indoor place identification across more than five semantic categories (e.g., corridor, office, conference room, lounge)
- Confidence estimation associated with each place prediction
- Support for navigation through integration with the ROS2 navigation stack

Input Data

- Visual sensory data from onboard cameras
- Robot localization information (e.g. odometry/ pose estimation)

Output Data

- Place labels and corresponding integer IDs
- Confidence scores associated with each prediction

Interactions with Other Modules

Receives place label and confidence score to modulate navigation behaviour (Social-Aware Adaptation Module)

3.2. Speech-to-text Module

Purpose and Role

The Speech-to-Text Module is responsible for converting spoken language input into structured textual representations. It enables verbal communication between users and the RoboCARE system, supporting downstream language understanding, dialogue management, and interaction capabilities.

Main Functionalities

- Speech recognition and transcription
- Noise Filtering

Input Data

- Raw audio signals from microphones

Output Data

- Transcribed text output
- Timestamp-aligned word or sentence segments

Interactions with Other Modules

The module provides transcribed text to the Interactive Natural Language Processing (INLP) Module.

3.3. Natural Language Processing Module

Purpose and Role

The Natural Language Processing (NLP) Module is responsible for understanding user intentions expressed through natural language and transforming them into structured robot commands. The module enables users to verbally instruct the RoboCARE system to perform navigation and task-oriented actions within the operational environment.

Main Functionalities

- Natural language understanding
- Extraction of task-relevant entities and parameters (e.g., destination locations, action types)
- Mapping user requests to executable robot actions and navigation goals
- Command validation and ambiguity handling

Input Data

- Transcribed text from the Speech-to-Text Module
- User query from sensory input
- Available place labels

Output Data

- Structured action commands
- Navigation goals and destination identifiers
- User queries to the Interactive Natural Language Processing Module

Interactions with Other Modules

Receives transcribed user utterances either from user input or the Speech-to-Text Module and converts them into executable commands for the robot. Provides navigation goals to the Navigation System, user queries to the Interactive Natural Language Processing Module, as well as may utilize place information from the Visual Place Recognition Module to resolve destination references (e.g., "Go to Room A").

3.4. Multi-modal Emotion Recognition Module

Purpose and Role

The multi-modal emotion recognition module analyzes multiple communication signals, such as facial landmarks, RGB-camera input, voice, and gestures, to identify a user's emotional state. It enhances the system's ability to understand emotions accurately, support adaptive interactions, and monitor behavioral patterns over time.

Main Functionalities

- Data collection from multiple sources (e.g. face, voice, sensors)
- Continuous emotional state estimation

Input Data

- Sensory data, e.g., RGB, depth, audio
- Post-processed data, e.g., facial landmarks, human pose estimation

Output Data

- Valence and arousal estimations
- Current emotional state

Interactions with Other Modules

The module depends on action recognition module and its outputs are consumed by the Long-term Behavior Understanding Module.

3.5. Action Recognition Module

Purpose and Role

The Action Recognition Module is responsible for identifying and interpreting human actions and activities from visual sensory data. It enables the system to understand user behavior, supporting higher-level reasoning, context-aware, and adaptive robot responses within RoboCARE.

Main Functionalities

- Human action and activity recognition
- Classification of short-term and/or long-term activities
- Detection of activity transitions
- Higher-level reasoning, e.g., crowd detection.

Input Data

- Video streams from onboard or external cameras
- Depth data for spatial and motion understanding
- Skeleton pose – joint data
- Temporal sequences of frames

Output Data

- Action/Activity labels (e.g., walking, sitting, cooking)
- Confidence scores for recognized actions
- Detected activity transitions

Interactions with Other Modules

The module provides semantic activity understanding to the Social-Aware Adaptation Module and the Lifelong Learning Module.

3.6. Social-Aware Adaptation Module (adapt both to place and to action transition)

Purpose and Role

The Social-Aware Adaptation Module is responsible for translating perceived environmental and social context into adaptive navigation guidelines. It receives high-level semantic information from perception modules (place, action, emotion) and generates modulations to locomotion parameters, proxemic boundaries, and interaction policies to ensure safe, socially compliant, and goal-directed robot behavior in shared human environments.

Main Functionalities

- Modulation of navigation parameters, e.g., speed, trajectory, and interpersonal distance
- Social compliance adaptation (e.g., yielding space, respecting personal proxemics)
- Context-aware policy selection based on environmental and social cues

Input Data

- Place label and confidence score (from Visual Place Recognition Module)
- Human action labels and transitions (from Action Recognition Module)
- Emotional state estimates (from Emotion Recognition Module)
- Long-term behaviour profiles (from Long-Term Behaviour Understanding Module)
- User feedback and interaction history (from Lifelong Learning Module)

Output Data

- Recommended locomotion velocity (linear and angular)
- Proxemic distance adjustments (personal space buffer)

Interactions with Other Modules

See input data

3.7. Long-term Behavior Understanding Module

Purpose and Role

The Long-term Behavior Understanding Module analyzes user behavior and emotional patterns over extended periods to build dynamic behavioral profiles. It processes historical interaction, emotional, and contextual data to identify trends, habits, and behavioral changes.

Main Functionalities

- Support personalized and context-aware human – robot interaction
- Behavioral pattern and Emotional profile generation of users
- Personalized response support

Input Data

- Emotional states / Valence – arousal estimations
- User interaction patterns

Output Data

- Long-term behavioral and emotional profiles
- Predictive behavioral insights

- Context-aware user state assessments

Interactions with Other Modules

The module depends on Multi-modal emotion recognition module and provides information to Lifelong learning and Social-Aware Adaptation modules.

3.8. Lifelong Learning Module

Purpose and Role

During an HRI event, the robot performs an action and subsequently observes human/s feedback, which may be conveyed through emotional responses, gestures, facial expressions, or other social signals. Based on this feedback, the robot adapts its behavior immediately and refines its learning module to enhance the appropriateness and effectiveness of future interactions with the specific individual involved.

Main Functionalities

- Assessment of internal state, as well as external perceived feedback and context
- Modification of robot's behavior
- Personalized learning based on individual's preferences and interaction patterns
- Continuous refinement of robot's behavior

Input Data

- Robot's executed action following its learned navigation policy
- User interacting feedback through Emotion Recognition Module
- Social and environmental context from Visual Place Recognition and Action Recognition Modules

Output Data

- Adapted robot behavior
- Updated personalized user interaction policy
- Improved future action selection at the specific individual interaction pattern

Interactions with Other Modules

Receives from Emotion Recognition Module, Visual Place Recognition and Action Recognition Modules, as well as Social-Aware Adaptation and Lifelong Learning Modules.

3.9. Interactive Natural Language Processing Module

Purpose and Role

The Interactive Natural Language Processing (INLP) Module is responsible for enabling bidirectional communication between users and the RoboCARE system. The module supports the generation of human-readable explanations regarding robot actions and decisions, while also allowing users to interactively query the robot about its behaviour, intentions, and task execution status.

Main Functionalities

- Natural language generation of robot explanations and responses
- Interactive question answering regarding robot actions, decisions, and navigation behaviour
- Dialogue management
- Context-aware response generation based on robot state and task execution history

Input Data

- User queries interpreted by the Natural Language Processing Module
- Robot status, task execution information, and navigation data
- Contextual information from PERCEPT and ADAPT modules

Output Data

- Human-readable textual explanations and responses
- Dialogue responses addressing user questions and requests
- Clarifications regarding robot decisions, actions, and current operational state

Interactions with Other Modules

Receives interpreted user queries from the Natural Language Processing Module and accesses information from navigation, PERCEIVE: Visual Place Recognition, Action Recognition and Emotion Recognition Modules, and ADAPT components: Social-aware Adaptation and Lifelong Learning Modules to formulate responses. Provides generated explanations and dialogue content to the user.

3.10. Emotion Elicitation Module

Purpose and Role

The Emotion Elicitation Module is responsible for expressing the robot's internal state. It evaluates incoming information according to predefined appraisal rules and determines which emotion (e.g., joy, anger, sadness, or fear) should be activated. In such a way, the robot's behavior and decision-making is externalized, enabling more adaptive and human-like interactions.

Main Functionalities

- Evaluates internal, e.g., battery level, executed action confidence, as well as external information, such as human actions or emotion.
- Generates appropriate internal affective state.
- Dynamic expression the robot's emotional responses.

Input Data

- LifeLong Learning Module, Social Aware Adaptation Module
- Robot's internal state estimation
- Battery Status

Output Data

- Robot internal emotional state

Interactions with Other Modules

The module depends on Lifelong Learning and Social Aware Adaptation Modules and provides information to the user regarding the robot’s internal emotional state.

3.11. Explainable AI Module

Purpose and Role

The Explainable AI (XAI) Module provides transparent and human-understandable explanations of the robot's decisions, behaviours, and internal state. The module aims to increase user trust and acceptance by exposing the key factors that influence the robot's reasoning and adaptation processes.

Main Functionalities

- Generation of explanations for robot actions and navigation decisions.
- Identification and presentation of the most influential factors contributing to a decision.
- Interpretation of outputs produced by perception, adaptation, and learning modules.
- Support for natural-language explanation requests through the Interactive Natural Language Processing Module.

Input Data

- Outputs from the PERCEIVE pillar, e.g., Visual Place Recognition, Action Recognition, and Emotion Recognition Modules.
- Outputs from the ADAPT pillar, e.g., Social-Aware Adaptation and Lifelong Learning Modules.
- User queries received through the NLP/INLP Modules.

Output Data

- Human-readable explanations of robot behaviour and decisions.
- Ranked lists of influential features, variables, or observations.

Interactions with Other Modules

The Explainable AI Module receives information from PERCEIVE, ADAPT, CONVEY components and provides explanation outputs to the Interactive Natural Language Processing Module and the end user.

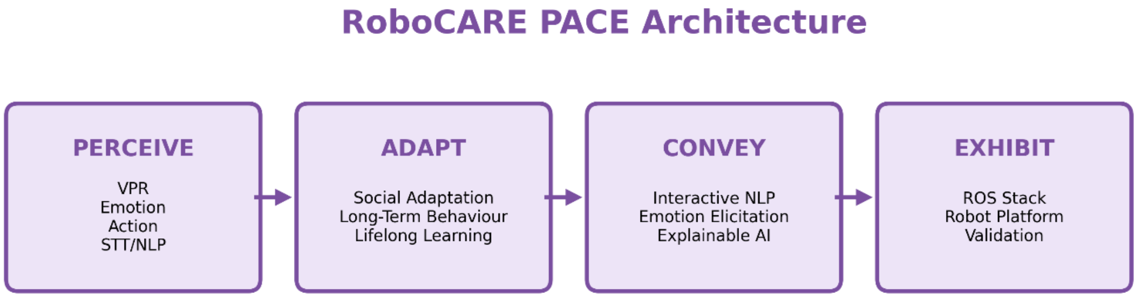

Figure 14. RoboCARE PACE pipeline

3.12. Data Handling and Governance Protocol

Purpose and Role

The RoboCARE framework relies on multimodal sensory information acquired from onboard robotic and off-board computer sensors, such as RGB cameras, depth sensors, microphones, LiDAR devices, robot odometry, and user interaction logs. The purpose of this protocol is to define the procedures governing data acquisition, storage, processing, sharing, auditing, and deletion throughout the lifecycle of the project.

Data Storage and Logging

All system outputs shall be stored through ROS bag files and structured databases. The system shall maintain:

- Sensor logs
- Robot state logs
- Decision logs
- Explainability logs
- User feedback logs

Each interaction episode shall receive a unique identifier enabling traceability and future auditing.

Privacy and GDPR Compliance

Data collection involving human participants shall follow GDPR principles:

- Data minimization
- Purpose limitation
- Transparency
- Right to erasure

Personal identifiers shall not be stored unless explicitly required by approved experimental protocols. Whenever possible, anonymized representations (e.g., skeletal joints, embeddings, heatmaps) shall be preferred over raw recordings.

Data Retention

Experimental data shall be retained only for scientific evaluation purposes and shall be removed following project retention policies and institutional ethical guidelines.

3.13. AI Auditing and Trustworthiness Framework

Purpose and Role

RoboCARE adopts a Trustworthy AI approach aiming to ensure transparency, fairness, robustness, accountability, and human oversight in robotic decision-making. The auditing framework provides continuous monitoring of AI modules throughout operation.

Auditable Components

The following modules shall support auditing: Visual Place Recognition, Emotion Recognition, Action Recognition, Lifelong Learning, Social Adaptation, NLP & Interactive NLP, Explainable AI

Auditing Metrics

The following categories shall be continuously monitored, based on the challenge and the field of each component.

Technical Performance

- Accuracy
- Precision
- Recall
- Latency

Robustness

- Sensor failure tolerance
- Missing data handling
- Distribution shift resilience
- Environmental adaptation

Fairness

When datasets permit, module performance will be evaluated across different user groups and environmental conditions to identify potential biases.

Human Oversight

The system shall maintain, e.g., confidence estimation, decision traceability, human override capability. Whenever confidence falls below predefined thresholds, the robotic agent shall request clarification or switch to conservative behavior.

3.14. Non-Functional Requirements

Purpose and Role

Besides functional capabilities, RoboCARE shall satisfy the following non-functional requirements.

Performance

Performance targets will be evaluated through the project KPIs and validation activities defined in WP5.

Reliability

The framework shall continue operation under dynamic events, e.g., partial sensor failures, communication interruptions, temporary perception degradation, using fallback and safe-mode strategies.

Scalability

The architecture can support, e.g., addition of new perception modules, additional robot platforms, new social navigation scenarios, without modification of the core framework.

Interoperability

All modules shall communicate through ROS interfaces and standard message definitions to facilitate portability across ROS-compatible robotic systems.

Explainability

Explainable AI module will provide traceable and explainable outputs regarding decisions affecting navigation behavior.

4. Conclusions

This deliverable presented the initial technical specifications of the RoboCARE framework and established the scientific and technological foundations that will guide the subsequent development activities of the project. Through the analysis of the state of the art in socially-aware navigation and human-robot interaction, the deliverable identified the main challenges, opportunities, evaluation methodologies, and ethical considerations associated with the deployment of robotic systems in human-centered environments.

Based on this analysis, a set of representative use cases was defined, covering both domestic and crowded environments. These scenarios capture the core functionalities envisioned by RoboCARE, including context-aware navigation, action and emotion understanding, adaptive behavior generation, lifelong learning, explainable decision-making, and natural language interaction. The use cases provide a common reference framework for the design, implementation, integration, and evaluation of the project technologies.

The deliverable further specified the principal functional modules that constitute the RoboCARE architecture, organized according to the PERCEIVE, ADAPT, CONVEY, and EXHIBIT (PACE) paradigm. Technical specifications were provided for perception, adaptation, learning, communication, and explainability components, together with data governance principles, AI auditing considerations, and non-functional requirements related to reliability, scalability, interoperability, and trustworthiness.

Overall, D2.1 establishes a common technical understanding among project partners and provides the baseline specifications required for the implementation of the RoboCARE framework. The outcomes of this deliverable will directly support the development of the conceptual framework, system architecture, algorithmic components, and experimental validation activities planned in the subsequent work packages and deliverables.

Ultimately, the RoboCARE framework aims to advance the development of socially-aware robotic systems capable of operating safely, transparently, and adaptively in real-world human environments, promoting trustworthy and effective human-robot collaboration.

References

[1] Lin, P., Abney, K., & Bekey, G. A. (2011). *Robot ethics: The ethical and social implications of robotics*. MIT Press.

[2] Lin, P., Abney, K., & Jenkins, R. (2017). *Robot ethics 2.0: From autonomous cars to artificial intelligence*. Oxford University Press.

[3] Francis, A., Pérez-D'Arpino, C., Li, C., Xia, F., Alahi, A., Alami, R., Bera, A., Biswas, A., Biswas, J., Chandra, R., Chiang, H.-T. L., Everett, M., Ha, S., Hart, J., How, J. P., Karnan, H., Lee, T.-W. E., Manso, L. J., Mirsky, R., Pirk, S., Singamaneni, P. T., Stone, P., Taylor, A. V., Trautman, P., Tsoi, N., Vázquez, M., Xiao, X., Xu, P., Yokoyama, N., Toshev, A., & Martín-Martín, R. (2025). Principles and guidelines for evaluating social robot navigation algorithms. *Journal of Human-Robot Interaction, 14*(2), Article 34. https://doi.org/10.1145/3700599

[4] Asimov, I. (1950). *I, Robot*. Gnome Press.

[5] Hall, E. T. (1966). *The hidden dimension*. Doubleday.

[6] Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81.

[8] Kruse, T., Pandey, A. K., Alami, R., & Kirsch, A. (2013). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems, 61*(12), 1726–1743.

[9] Rios-Martinez, J., Spalanzani, A., & Laugier, C. (2015). From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics, 7*(2), 137–153.

[10] Mavrogiannis, C., Baldini, F., Wang, A., Zhao, D., Trautman, P., Steinfeld, A., & Oh, J. (2023). Core challenges of social robot navigation: A survey. *ACM Transactions on Human-Robot Interaction, 12*(3), 1–39.

[11] Singamaneni, P. T., Bachiller-Burgos, P., Manso, L. J., Garrell, A., Sanfeliu, A., Spalanzani, A., & Alami, R. (2024). A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research, 43*(10), 1533–1572.

[12] Mirsky, R., Xiao, X., Hart, J., & Stone, P. (2024). Conflict avoidance in social navigation—a survey. *ACM Transactions on Human-Robot Interaction, 13*(1), 1–36.