The technology for social robots has been growing rapidly, and in the near future, we can expect household robots to be affordable, widely available, and used in homes for a variety of tasks like cleaning and tidying up. All the family members, including parents, children, and grandparents, will communicate with household robots in various ways. Therefore, such robots need to be designed so that anyone can interact with them in an intuitive way to tell them what needs to be done. Current household consumer robots lack this functionality due to their limited capability in achieving natural and intuitive interaction. For example, you cannot "tell" a vacuum cleaning robot to not go to a certain area or to clean certain areas more frequently. The goal of this project is to research such interfaces for natural and intuitive human-robot interaction.

Humans often use nonverbal modalities like gestures and gaze to augment verbal communication. In particular, gaze and gestures play a significant role in dyadic interaction between human partners. For instance, we refer to objects in the environment by pointing at them, either by gaze or gestures. Though there is some research on how gestures are used to supplement determining the referents of ambiguous expressions in the environment, there is very little research on incorporating gaze information into gesture and speech to disambiguate uncertainties in verbal and gestural communication. In this project, we plan to conduct research on methods to better understand human-robot interaction through a collaborative robotic interface for humanoid domestic service robots, where the user can interact with the robot intuitively using multimodal cues of language, gestures, and gaze. We aim to deliver a system where a human partner can tell an assistive robot which objects to pick up and where to put them using words, gestures, and gaze.

We plan to employ a user-centric approach by conducting user studies throughout the project, to model how humans use gestures and gaze naturally to collaborate with a robot. These studies will be conducted for a set of specially prepared scenarios. For example, in one set of scenarios, referred to as table-top scenarios, the user asks the robot to arrange items on a table. The second set of scenarios concerns tidying up a room, where the user tells the robot to bring or move various objects located at different places in the room. The scenarios will be developed both in Virtual Reality (VR) and real-world environments.

The VR settings will be simplified replicas of real-world scenarios. They allow for quick prototyping and initial data collection with naive users in the design phase of the project, besides ensuring human safety in kick-off tests of scenarios. Additionally, VR scenarios will be shared among all the project partners to improve their collaboration before starting the integration on the partner's respective robotic platforms. Therefore, VR scenarios will facilitate synchronization and compatibility across the partner labs. We will then incorporate those models into a multimodal interface for human-robot interaction. For this, we will develop a graph-based representation of visual scenes that include gaze information, and from which gestural information can be extracted through automated analysis of sensor recordings. The novel language module will detect the presence of ambiguous speech parts, such as deictic expressions, to augment them with gaze, gesture, or visual information from graph-based representations to resolve referential ambiguities. This methodology and the user-centered approach will allow researching the affordances of the interface for intuitive interaction with the robot.

For this, the interface will be evaluated by conducting user studies in a real-world environment, for both sets of scenarios, to study the facilitatory and inhibitory effects of employing multiple modalities in communication. The previous research has divergent results about the perceptual and cognitive load introduced by communication modalities of different types and their use in multimodal communication. We will test the hypotheses that incorporating gesture- and gaze-based information in human-robot interaction makes the human user more comfortable and more trusting of the robot. The deliverables of the project will facilitate designing intuitive interfaces for the user to interact with robots in private spaces such as homes, semi-private areas such as educational institutes, and public spaces like malls, airports, museums, and hospitals.