

## Article

# Beyond Binary Dialogues: Research and Development of a Linguistically Nuanced Conversation Design for Social Robots in Group–Robot Interactions

Christoph Bensch <sup>†</sup> , Ana Müller <sup>†</sup> , Oliver Chojnowski  and Anja Richert <sup>\*</sup> 

Cologne Cobots Lab, TH Köln—University of Applied Sciences, 50679 Cologne, Germany; christoph.bensch@th-koeln.de (C.B.); ana.mueller@th-koeln.de (A.M.); oliver.chojnowski@th-koeln.de (O.C.)

<sup>\*</sup> Correspondence: anja.richert@th-koeln.de

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** In this paper, we detail the technical development of a conversation design that is sensitive to group dynamics and adaptable, taking into account the subtleties of linguistic variations between dyadic (i.e., one human and one agent) and group interactions in human–robot interaction (HRI) using the German language as a case study. The paper details the implementation of robust person and group detection with YOLOv5m and the expansion of knowledge databases using large language models (LLMs) to create adaptive multi-party interactions (MPIs) (i.e., group–robot interactions (GRIs)). We describe the use of LLMs to generate training data for socially interactive agents including social robots, as well as a self-developed synthesis tool, knowledge expander, to accurately map the diverse needs of different users in public spaces. We also outline the integration of a LLM as a fallback for open-ended questions not covered by our knowledge database, ensuring it can effectively respond to both individuals and groups within the MPI framework.

**Keywords:** human–robot interaction; large language model; social robot; multi-party interaction; group–robot interaction



**Citation:** Bensch, C.; Müller, A.; Chojnowski, O.; Richert, A. Beyond Binary Dialogues: Research and Development of a Linguistically Nuanced Conversation Design for Social Robots in Group–Robot Interactions. *Appl. Sci.* **2024**, *14*, 10316. <https://doi.org/10.3390/app142210316>

Academic Editor: Antonio Fernández-Caballero

Received: 17 September 2024

Revised: 21 October 2024

Accepted: 7 November 2024

Published: 9 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Motivation and Related Work

Social robots have the ability to engage in dialogues with humans using natural language and find applications in various settings such as work, healthcare, travel, and the service industries [1–4]. However, the prospect of widespread acceptance in these scenarios remains dependent on the ability of robots to understand the nuances of human social interactions, which are indispensable for the establishment of genuine communication patterns [5–7]. Therefore, the development of systems that can adapt to individual preferences while navigating unexpected contextual challenges is essential [8–10]. One of the challenges in developing such systems is that research in human–robot interaction (HRI) has primarily focused on dyadic connections, such as the interaction between a single human and a single robot, which resembles chatbots used in customer support scenarios (e.g., [11]) or personal voice assistants such as Siri (Apple, Cupertino, CA, USA) or Alexa (Amazon, Seattle, WA, USA) [12,13].

### 1.1. Natural Language Understanding in Human–Robot Interaction

Socially interactive agents, including chatbots, voice assistants, and social robots, share a dependency on natural language understanding models (NLU) that are usually trained using carefully curated datasets [14], as having a broad and varied set of training phrases (i.e., potential user input) for each intent is crucial for precisely identifying diverse human intentions. Training data are of the utmost importance to reflect the diversity of potential users, in terms of factors such as gender, occupation, education, and age. However, the acquisition of annotated datasets that exhibit 30 language variations within

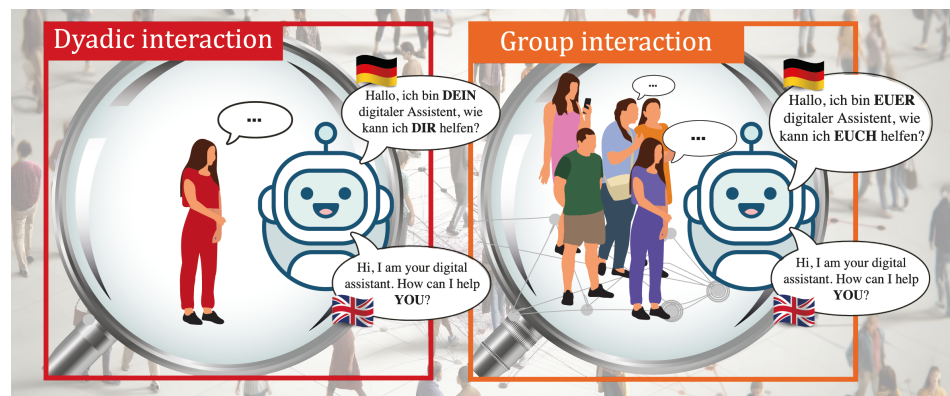
an intent-based system represents a costly and time-consuming process that is not easily scalable due to the need for human input [14]. In previous instances, crowd workers have been employed to categorize and provide annotations for datasets intended for the training of NLU models [15]. This presents a challenge, as the expected quality level was not consistently met [16], and is often associated with precarious working conditions, insufficient compensation, and a lack of social protection for workers. Along with this, attempts have been made to automate the process of generating NLU training data, using intent-based techniques that involve the construction of sentences by selecting synonym words from a predefined list such as '(hello | hi) ( | mighty) world', resulting in variations like 'hello mighty world', 'hello world', 'hi mighty world', and 'hi world' [17]. However, this approach requires a significant amount of manual effort and could lead to a lack of diversity in the training data [16].

With the release of large language models (LLMs) and the emergence of models such as GPT-4 (OpenAI, San Francisco, CA, USA) [18] and LLaMA 3 (Meta, Menlo Park, CA, USA) [19], LLMs are increasingly being used in the field of human-machine interaction (HMI), among many other aspects, with the aim of tackling challenges related to knowledge representation [20] and to generate training data [21] for chatbots, voice assistants, and social robots. In the latter case, the natural language processing (NLP) and generation capabilities of LLMs are employed to generate training data that are more semantically accurate, which accelerates the process and potentially necessitates less manual correction in comparison to intent-based approaches. This is especially beneficial since large language models (LLMs) are often impractical for real-time applications because of significant computational expenses, as well as quality control and compliance requirements; however, they can also serve in generating training data. One potential avenue for utilizing LLMs in this context is to take advantage of their few shot learning capabilities, as demonstrated by Rosenbaum et al. [21]. In this case, the LLM is provided with a single example of a translation or paraphrasing task and is then able to sample a multitude of new and diverse examples from this source. However, it is increasingly apparent that existing LLMs are constrained by a number of challenges, including the generation of hallucinated data, the introduction of prejudices, and the execution of intricate computations.

These challenges encompass elements such as a large number of parameters and substantial memory needs, requiring adaptable approaches for robotics that do not depend on hardware setups such as graphics processing unit (GPU) clusters for inference. Consequently, the technique known as Post-Training Quantization for Generative Pre-trained Transformers (GPTQ) has surfaced as a method to compress models and address these challenges [22]. This development enables LLMs to function on individual consumer GPUs, broadening their practicality in the realm of social robotics, for example in the improvement of communication by the establishment of multi-turn and multi-party interactions (i.e., group-robot interactions) [12,23].

However, most NLU research focuses on the adequate implementation of dyadic interactions (i.e., one agent with one human), as is common in many application scenarios (e.g., chatbots), while research on improving the interactions, behavior, or communication of social robots within groups has been less extensive. The latter is particularly evident in languages other than English, which can be one reason why it has not been adequately addressed. For example, certain languages, such as German (also Polish, French, Italian, etc.), differentiate pronominal usage, i.e., 'you' (singular) and 'you' (plural), resulting in different conversation flows in dyadic versus group interactions, since addressing a group differs grammatically from addressing an individual, unlike in English [24]. As illustrated in Figure 1, the greeting "Hi, I am your digital assistant. How can I help you?" is identical for both group and individual interactions. In contrast, in German, there is a clear differentiation like "Hallo, ich bin dein digitaler Assistent, wie kann ich dir helfen?" versus "Hallo, ich bin euer digitaler Assistent, wie kann ich euch helfen?". In these contexts, group members are linguistically excluded through conversation design if this addresses the persons exclusively in the singular and does not make any adaptations with regard to

the interaction context. An illustration of inclusivity in the manner of addressing groups in English is a salutation such as the greeting “*Hey guys* [...]”, which can be avoided with greater ease than incorporating it into all the responses in the system. These preliminary research and development activities indicate the potential for a paradigm shift, given the limited attention previously devoted to the intricate dynamics of multi-party interactions. Although earlier studies indicate that HRI arises jointly between several users and a robot [8,25,26], the HRI research landscape was dominated mainly by the image of dyadic relationships, that is, the assumption of one-to-one interaction scenarios [10,12,27].



**Figure 1.** Intended design for dyadic versus group interactions, illustrating linguistic nuances in conversation using the German language as a case study. Own illustration with AI-generated background.

### 1.2. The Importance of Groups in HRI

It is crucial to recognize that group interactions and dynamics are complex aspects of social interactions that should not be underestimated. The behavior of individuals in the presence of robots differs when they are in a group setting. These scenarios have a significant impact on the quality and dynamics of the interaction [28,29]. Although interactions with individual users are generally robust, group dynamics can significantly affect the robot’s ability to detect and engage with groups, as their members tend to talk to each other or obscure each other [24]. These findings highlight the ongoing need for research and development (R&D) on group interactions in HRI, particularly in fields such as healthcare, hybrid team production, and public spaces, where collaborative HRI takes place. Ultimately, well-functioning and resilient MPIs can offer benefits in improving HRI, as group members are more inclined to engage with robots [28] and experience a greater sense of ease [30].

### 1.3. Computer Vision in Group–Robot Interactions

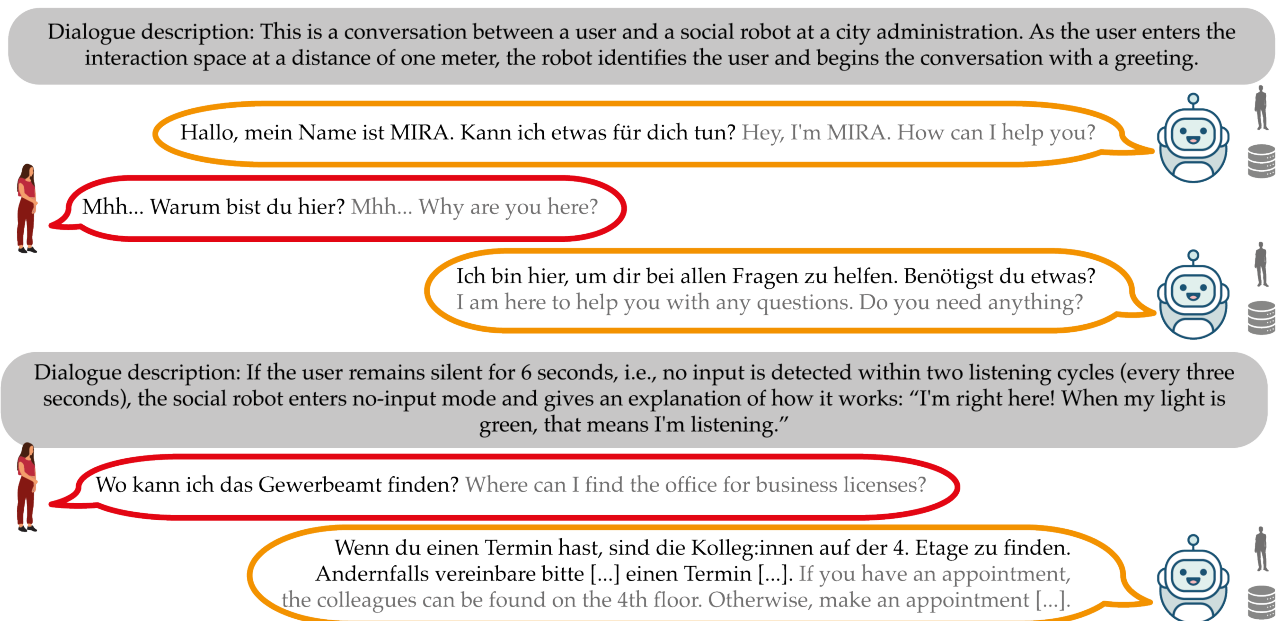
One area where group dynamics are present in HRI is within navigation robots. These robots play a crucial role in guiding people through museum exhibits and identifying group relationships to prevent collisions [28,31]. Advanced systems have been developed to identify group dynamics in real time in various environments using computer vision techniques to analyze images or videos [32,33]. This capability has been further facilitated by progress in the field of real-time object detection, which includes both convolutional neural networks (CNNs) [34] and transformer architectures [35] that are capable of rapidly identifying and/or tracking objects in different scenarios and use cases. For example, *You Only Look Once* (YOLO) models [36] have also been used in the field of navigation robots to test, through a simulation process, to what extent a mobile robot can automatically recognize and navigate around people in unknown environments [37]. However, while YOLO models achieve a good balance between speed and accuracy, these applications still tend to emphasize the avoidance of interactions rather than improving the interaction capabilities of social robots.

In this work, we investigate the development and technical integration of a conversation design that is sensitive to group dynamics and adaptable, taking into account the nuances of linguistic differences between dyadic and group interactions using the German language as a case study. We elaborate on how robust person and group identification is achieved through the utilization of YOLO, as well as the enhancement of knowledge repositories through LLMs to develop adaptive MPIs guided by person identification. In the following, we will present the development of the linguistically nuanced conversation design along with results of the technical strategies presented through an exemplary dynamic dialogue, before discussing hurdles and future directions.

## 2. Materials

### 2.1. The Use Case

Our goal is to enable empathetic interactions between users and socially interactive agents, including social robots in public spaces. This work focuses on integrating a social robot designed to function as a multilingual interactive assistant to provide real-time support and information, helping users navigate administrative processes, offering directions within public buildings, and responding to a variety of inquiries. In this research, we designed and evaluated our system within a German city administration, aiming to aid citizens with diverse requirements in both individual and group interactions with the social robot, Furhat (Furhat Robotics, Stockholm, Sweden). By utilizing computer vision (CV), advanced NLU, and large language models (LLMs), our social robot can engage in context-aware dialogues, adjusting in real time to dyadic and group settings. At the beginning of the R&D efforts for this project, a workshop was conducted involving city administration staff from multiple departments. The objective of this workshop was to identify visitors' needs by leveraging the expertise of employees from various departments within city administration, such as citizen services, the immigration office, the youth welfare office, and the registry office. Additionally, the workshop included an overview of knowledge generation and data creation for socially interactive agents to tailor the system to meet the unique demands of the use case and for the co-creation of a knowledge database with the experts in the areas on site. Subsequently, the employees themselves compiled the knowledge in a knowledge database, focusing particularly on the insights of employees who regularly interact with citizens (i.e., users), mirroring the emphasis of the workshop. The information on navigation (e.g. the location of restrooms) and specific knowledge on departments was collected from the employees, containing a few training examples and a single response each. After the knowledge was scrutinized for formal inaccuracies, it was linked to an existing foundation knowledge database, which contained small talk, information on robotics, AI, jokes, and additional connections, such as a weather application programming interface (API). Following the merging of the datasets, the knowledge database consisted of 222 potential user intents. In the autumn of 2023, we conducted usability and system tests within the town hall, engaging actual users—citizens with genuine issues. For the tests, the robot was placed in the entrance area of the city hall. The following dialogue provides an example of an interaction between the robot, MIRA, and a user in the context of a city administration. Each block represents a dialogue turn, with the speaker and system specifications noted at the top. German statements are provided in black, while their English translations are in gray below. This exemplary dialogue will be further elaborated upon in the paper as part of the discussion on designing nuanced conversational designs for multi-party interactions (Figure 2).



**Figure 2.** This is a conversation between a user (red) and a social robot (orange) in a city administration setting. Each speech bubble is bilingual (German/English). Icons next to the robot indicate a knowledge database response and single user mode. The dialogue continues in Section 3.1.

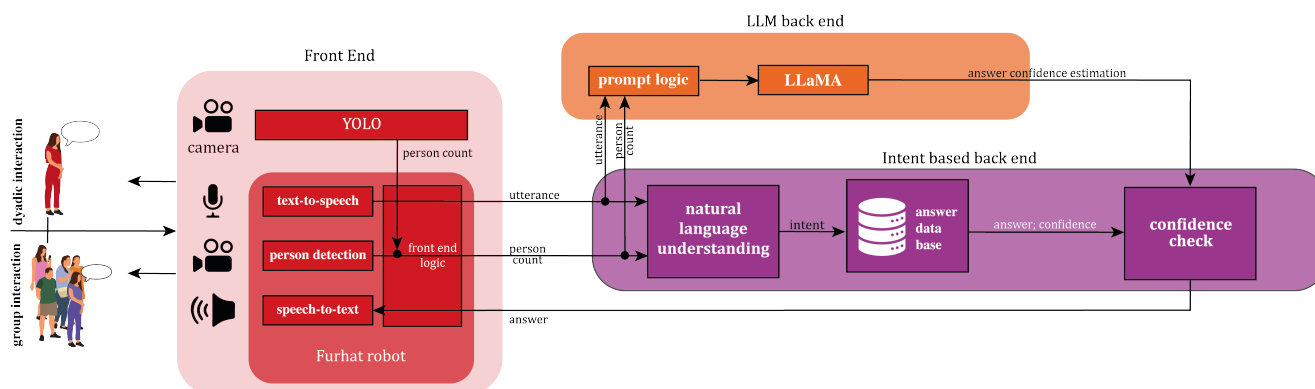
## 2.2. The Social Robot

As detailed in Section 2.1, the social robot is required to fulfill several criteria for successful deployment in public areas. It should operate in various languages and recognize people within its interaction range, allowing it to distinguish between single and group interactions. Furthermore, the robot must provide scripted responses to specified intents and also produce impromptu reactions.

We incorporated the social robot Furhat as the front end, interfacing with three largely autonomous, self-developed subsystems functioning as the back end (for a detailed view of the architecture, refer to Figure 3):

1. Robot front end: This subsystem serves as the user interface, managing most necessary interaction functions, including automatic speech recognition (ASR), person detection, and text-to-speech (TTS) output. It also manages some dialogue management tasks, such as initiating and concluding conversations, and managing speaker turns. The front end is connected to a tablet with flag buttons that allow users to select the system language. In this work, we used a Furhat robot, which provides essential components such as ASR, person detection, and TTS. However, the Furhat robot's built-in person detection had limitations, particularly when individuals were not facing the robot directly or were looking in different directions, which often disrupted dialogue. To address this issue and improve the robot's ability to detect multiple users simultaneously, we integrated a YOLO model for more robust multi-party detection.
2. Intent-based back end: This core subsystem includes a NLU model and a related knowledge database. The database stores predefined subject areas (intents) set by example user queries, along with fixed, preformulated responses. These intents address specific user inquiries, such as identifying the right contact for an issue, providing directions to the nearest restrooms, engaging in small talk, and integrating APIs like a weather forecast.
3. LLM back end: In the event that queries are not aligned with the predefined parameters, this subsystem is engaged. The intent-based back end forwards the user's statement, dialogue history, user count, and language to the LLM server, which then generates an appropriate prompt to produce a free response to the current user query. For this, LLaMA 2 (Meta, USA) [38] in its 13 billion-parameter quantized version,

designed to run on consumer-grade graphics cards [22], was integrated and used as a fallback. The selection of LLaMA 2 was made according to the Open LLM Leaderboard [39], where the LLaMA 2 model was ranked the top in summer 2023, just prior to the execution of the field system assessments.



**Figure 3.** Architectural sketch of our social robot, with the corresponding subsystems. Own illustration.

The speech-to-text (STT) module of the front end transcribes spoken language into written text. This textual representation is then relayed to the intent-based back end. The intent-based back end coordinates the conversation and includes modules for NLU, dialogue management, and natural language generation (NLG). For NLG, the system has access to the knowledge database. The overall latency of the social robot is crucial for conducting real-time dialogues with users. The intent-based back end is equipped with a Google Dialogflow NLU model that deciphers the user's intent by evaluating the input against a response database. In parallel, the LLM back end (LLaMA 2) serves as a secondary back end, stepping in to provide alternative responses when the intent-based AI's confidence level does not meet a predefined threshold, ensuring that the social robot remains effective even when faced with out-of-knowledge queries. To determine whether the response from the intent-based back end or the LLM back end is chosen, the confidence level of the NLU model is decisive. Knowledge database responses are prioritized as they are site-specific, co-creatively developed with employees, and have been human-reviewed for relevance and accuracy. The selected response is then converted back to speech (TTS) and transmitted to the user by the robot, aiming for a natural and intelligible voice output. The way responses are generated in the social robot's architecture varies greatly between the intent-based and the LLM back end. While the intent-based back end adopts a data-driven approach, using predefined responses that match the identified user intents, the LLM operates on a prompt-driven methodology. This approach necessitates the integration of all essential system information, such as the robot's name or behavioral norms, directly into the LLM's prompts. In contrast, in the intent-based back end, this system information must be embedded within the predefined responses that correlate with the recognized intents. To ensure that the social robot delivers coherent responses across both platforms, it is imperative that these two back end systems are meticulously aligned and synchronized.

This modular design allows each component to be optimized independently, as described in the following chapters, while ensuring that they function cohesively to meet the diverse and dynamic needs of users in public spaces.

### 3. Development and Results of the Multi-Party Dialogue Management

#### 3.1. Integration of Multi-Party Person Detection

Although interactions with individual users are generally robust, group dynamics can significantly affect the robot's ability to detect and engage with humans [24]. The Furhat robot features person detection that is based on facial recognition. However, in scenarios where users face or obstruct each other without making eye contact with the robot, there can be challenges, as this often results in disruptions during multi-party interactions, where

individuals may obscure each other or engage in side conversations, leading to errors in person detection.

Therefore, we used the YOLOv5m model [40]. In contrast to facial recognition systems, the YOLOv5m excels in detecting individuals, whether fully or partially visible, from various camera angles (i.e., egocentric and exocentric from the robot [31]). Using CNN architecture, the YOLO models are capable of making bounding-box and class probability predictions in a single network pass. This design feature renders YOLO models particularly well suited to edge computing applications, where the minimization of latency and the processing of data locally are of paramount importance. Among the various YOLO models, the YOLOv5m model emerged as the most suitable option due to its ability to deliver optimal performance at a processing speed of 60 frames per second (FPS) [40]. This decision was also influenced by hardware constraints, specifically the limitation to edge hardware with only 4GB of GPU VRAM. The YOLOv5m model could effectively operate within these parameters while prioritizing human detection.

Expanding upon the YOLOv5m architecture, we developed a real-time person detection system tailored to accurately discern the number of people within the robot's interaction space and determine their participation in the conversation. To further enhance the effectiveness of the social robot, the person detection simultaneously evaluates three camera angles to accurately determine the presence of individuals. In initial testing, person detection errors were primarily due to people in the background, who were not involved in the conversation, being mistakenly counted as users. In compliance with the General Data Protection Regulation (GDPR), we carefully selected a location within the city administration to effectively minimize any background presence. We rigorously adhered to the principle of data minimization, aligning both with legal requirements and ethical guidelines. Although it was not always possible to completely remove background elements, we made every effort to centralize them to the lowest extent possible, incorporating this approach into our framework. Therefore, adjustments were required depending on the deployment environment, mainly involving fine-tuning the interaction radius and adapting the person detection to ambient lighting conditions. To address lighting issues, we adjusted the brightness, gain, and exposure settings of each camera to ensure consistent image quality, minimizing over- and underexposure across different lighting conditions. This iterative calibration process improved the detection accuracy. These modifications ensured optimal performance, maintained high accuracy in person detection in different settings, and contributed to the overall effectiveness of multi-party interactions. This capability strengthens the robustness of the social robot, especially in scenarios involving groups where individuals may be partly obstructed by others. Our system analyzes the area of the bounding boxes of detected individuals to assess whether they are actively involved in the interaction. Adjustable threshold parameters define the minimum area size required for a person to be considered part of the conversation. Figure 4 illustrates the multi-camera person detection system. In the left camera view, three individuals are detected; two are marked in blue, indicating that their bounding-box areas exceed the threshold and are thus considered users of the system. The third individual is marked in red, signifying a smaller area and classification as a non-user. The middle camera accurately detects both participants, and the right camera view further validates the detections, with one participant marked in red due to a smaller bounding box area, indicating background status. This feature plays a crucial role in adapting interactions to fit the social context, allowing a distinction between one-on-one and group interactions, thereby influencing the design of the conversation. The robot's person detection relies primarily on its egocentric camera for this task, which is further supported and confirmed by exocentric cameras, enabling it to adapt its conversational framework in response to the dynamics of the group.



**Figure 4.** Detection of individuals at the city hall through multiple cameras evaluates engagement by examining the area of bounding boxes. Blue boxes represent engaged individuals, while red boxes signify either non-users or background objects. Exocentric camera outputs are shown in the left and right images, and the center video is from Furhat’s egocentric camera. Participant faces are anonymized to safeguard their privacy.

The system continuously transmits the number of detected users to our front end. This information is relayed to both the NLU and the LLM whenever a complete user utterance is detected. This concurrent data acquisition is crucial as it allows the back ends to adapt responses to the evolving dynamics of the interaction. For example, the social robot can seamlessly alternate between responses tailored for individuals and those designed for groups if another person enters the interaction space during a session. This fluidity in response is essential, as it mirrors the natural ability of humans to adapt their communication based on social context, ensuring that the robot can similarly maintain a natural flow of conversation and provide a personalized and engaging user experience.

In the sample conversation provided below, the robot switches to MPI mode after recognizing multiple users. The dialogue example highlights that the concept of group-sensitive conversation design is applied not only in German but also in English. In contrast to the frequent use of personal pronouns (such as ‘ihr’/‘euch’) in the German dialogue, the English conversation design refers to the group with, e.g., ‘guys’. While in German, the group is addressed directly using personal pronouns, in English, there is a noticeable distancing effect when terms like ‘children’ are used. Only the formulation ‘guys’ again refers to the address of the group (Figure 5).

Dialogue description: Another user [second user] joins the interaction space, the social robot transitions from single mode to MPI mode and adjusts the conversation design.



Hallo, weißt du ob Kinder im Jugendzentrum essen können?  
Hello, do you know if children can eat at the youth centre?

Das Jugendzentrum hat verschiedene [...] Workshops für Kinder [...]. Ihr könnt daran teilnehmen [...]. Meldet euch einfach an [...]. Gibt es noch etwas, was ich für euch tun kann? The youth center has various cooking [...] workshops for children [...]. You can participate in them [...]. Just sign up [...]. Is there anything else I can do for you guys?



**Figure 5.** This is the continuation of the dialogue from Section 2.1, where a second user joins the interaction space, prompting the social robot to switch to MPI mode. The second users intents are shown in purple. Icons beside the robot indicate MPI mode and a knowledge database response. The conversation will be continued in Section 3.3.



When our system was evaluated within the city administration using actual users, who naturally exhibited demographic and physical diversity, it was found that the (group) person detection system functions reliably across multiple scenarios. To optimize performance, we adjusted the interaction radius and fine-tuned the person detection thresholds based on the bounding box areas. These parameters are adaptable for each deployment environment. Using the egocentric camera (i.e., the middle one) and using the side cameras for confirmation, we improved detection accuracy. The system reliably distinguished users from non-users by analyzing the size of the detected bounding boxes and validating detections across multiple camera views. This multi-perspective analysis significantly reduced errors caused by background individuals being mistakenly counted as users.

### 3.2. Development of the Knowledge Expander

Unlike a pre-trained LLM, the NLU model within the intent-based back end requires an initial training phase to function effectively. This variability is crucial because it relies on word embeddings to determine intents, which helps achieve high confidence in intent recognition. We developed a LLM-based synthesis tool, the knowledge expander, to accurately map the various needs of different users in public spaces. This ensures sufficient variation in NLU training while reducing the required manual effort. Generally, at least 20 training examples per user intent and language are needed to train the NLU model effectively. Ideally, this would mean that, with knowledge expander, only one human-written example sentence per user intent would be required to generate at least 20 training examples for it [41]. In a multilingual setting, a single English example sentence could also be used to generate 20 training examples in different languages.

However, training our NLU model with generated examples based on one human-created example per intent was not successful. This is attributed to the fact that, in training a NLU model, not only is the number of training examples critical, but also the quality, the variance between the examples, and the conceptual distance between different user intents. In our case, the training examples generated with the knowledge expander were not sufficiently diverse, leading to numerous errors during NLU training that could not be manually corrected. By analyzing these errors, we observed that the examples generated by the LLM based on a single human-written example sometimes differed by only one word, such as “Can I use the toilet?” versus “Can I use the toilets?”. This issue was addressed by providing three manually created examples per intent in knowledge expander, which varied more significantly in phrasing. In this manner, the LLM was able to more accurately comprehend the intent’s meaning, consequently producing a greater degree of diversity in the training examples.

Taking into account the hardware conditions, the GPT-3.5 [18] API was used, given the substantial quantitative need for the generated training data. The knowledge database utilized in the context of the city administration included 222 different intents. After expansion, we received 19,621 ( $M = 88.38$ ) training examples per language in German, English, French, Italian, Spanish, Polish, Russian, and Turkish. This expansion resulted in a total of 156,864 training examples across all eight languages. Given the GPU memory constraints, the locally running model LLaMA 2 could not be parallelized, which would result in the generation of these examples lasting several days. Although the generation speed of a single training example via the GPT-3.5 API does not significantly exceed that of LLaMA 2, using GPT-3.5 allows for parallelization. Since user intents could be expanded independently, we utilized an arbitrary number of threads to generate training examples. As a consequence of the limitations of the API in terms of the number of requests that can be made per unit of time, this generation method still required several hours to complete.

When creating the 156,864 training examples in eight languages, some conflicts still required manual resolution during the NLU model training. The aforementioned conflicts were attributable to the existence of intents that were inherently analogous, such as “Reisepass\_abholen” (pick up passport) versus “Reisepass\_beantragen” (apply for passport). This resulted in the generation of analogous training examples that could be applied to both

intents, thus precipitating errors during NLU training. It was also observed that the training instances generated sometimes lacked sufficient variability or were even identical. To mitigate this problem, we increased the number of generated training instances and employed post-processing to filter out examples that were too similar, using the Levenshtein distance with a threshold of  $lev(a, b) \geq 0.9$ , thus enhancing the diversity and relevance of the training dataset.

A manual check of the training examples showed that some phrases were unusual. Although this generated sentences similar to those used by children or non-native speakers, it led to errors when translated into other languages. For example, the English sentence “Are you able to understand me?” was incorrectly translated into German as “Hörst du meine Ansprache?” (“Do you hear my speech?”), which was then erroneously expanded as “Kannst du meinen Vortrag hören?” (“Can you hear my lecture?”). To address this, we found that better results were achieved by translating the manually created seed sentences first and then expanding them in the target language, rather than expanding the original sentences and translating the entire set. This approach ensured that the nuances of each language were better captured during the expansion process. It is evident that, when using the knowledge expander for translations and responses, manual corrections are necessary to ensure quality. Addressing these challenges is crucial to ensure accurate and consistent interaction with the social robot.

### 3.3. Development of Linguistically Nuanced Intent-Based Responses

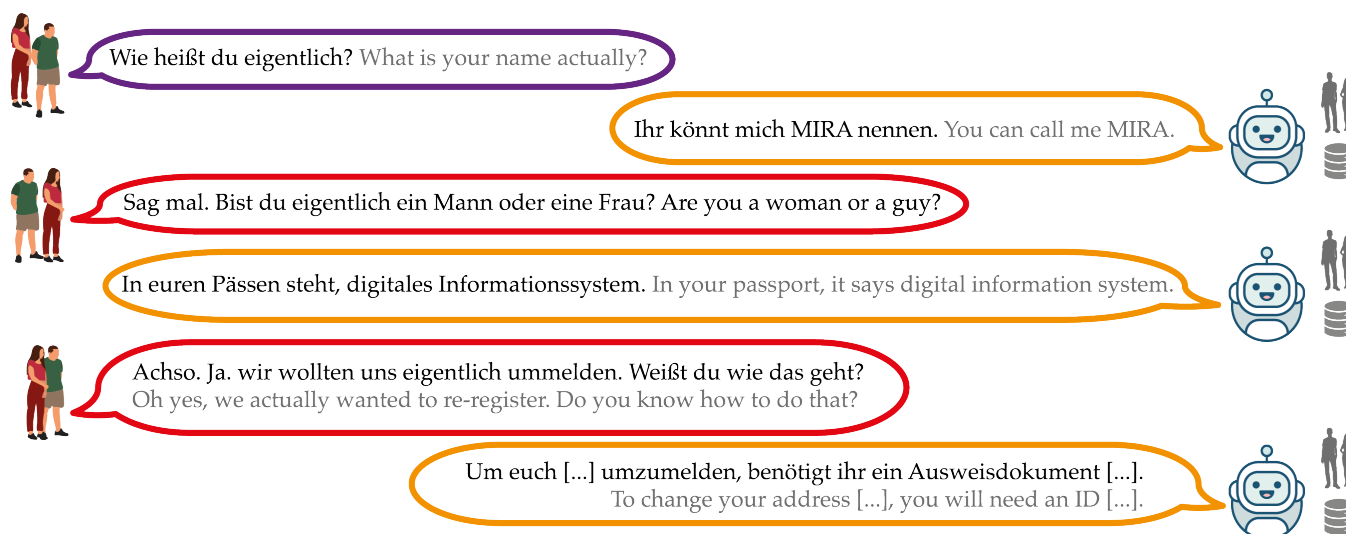
To provide linguistically nuanced responses in dyadic and group interactions, the knowledge expander was also used to generate variations of system responses (i.e., to achieve variation in conversation design) and responses for groups, since the hand-crafted foundation responses within the knowledge database were oriented toward individual users. We synthesized plural responses by instructing the GPT-3.5 to address a group of people based on singular responses (Figure 6):

GPT 3.5 Prompt: Formulate **num-synonyms** different synonyms in simple english<sup>a</sup> that say the same thing as the following sentences in the plural. The synonyms should address a group of people from the perspective of one person: **example-sentences**

<sup>a</sup>The language was adapted to one of the eight languages used as required.

**Figure 6.** The GPT 3.5 prompt to generate multiple English synonyms for plural sentences, addressing a group.

For example, the German sentence “Du kannst mich MIRA nennen” (“You can call me MIRA”) was expanded into different paraphrases such as “Mein Name ist MIRA” (“My name is MIRA”) as well as different languages and into the plural for MPis: “Ihr könnt mich MIRA nennen.” (“You can call me MIRA”) (see example dialogue below). In total, the expanded knowledge database responses consisted of 1764 responses for the 222 intents with  $M = 7.25$  different responses per intent, language, and context of dialogue (dyadic versus group). However, the challenge lies in crafting high-quality responses that accurately reflect the desired system behavior and transmitted information. The responses generated were carefully examined to ensure quality. However, we will address the challenges encountered in the following dialogue (Figure 7).



**Figure 7.** This is the continuation of the dialogue from Section 3.1. Icons beside the robot indicate MPI mode and knowledge database responses. The conversation will be continued in Section 3.4.

As evidenced by the exemplary dialogue, certain translations and extensions were erroneous, resulting in difficulties in generating responses aligned with the intended intentions. For example, the response to the question about the gender of the robot was sensibly translated from German into English but was not transferred to MPI mode in German. Here, “In my passport, it says digital information system” in German MPI mode became “In euren Pässen steht, digitales Informationssystem.” (“In your passports it says digital information system.”). We discovered the same issue with the German no-input phrase (see Section 2): “Ich werde zuhören, wenn ich grün leuchte” (“When my light is green, that means I’m listening.”) was incorrectly transferred to MPI mode as “Wenn ihr grün leuchtet, höre ich zu” (“If you light up green, I’m listening”).

### 3.4. Development of Linguistically Nuanced LLM Responses

The prompts used for the integration of LLaMA 2 were customized to the handcrafted knowledge database to ensure that the responses from the LLM matched those from the intention-based back end. The prompt describes an interaction between a (group of) user(s) and a social robot called MIRA. MIRA is designed as a helpful assistant and responds briefly and precisely, in a maximum of two sentences. Based on the person detection results, we used variant prompts for dyadic and group interactions. This prompt would guide the LLM to generate responses that are helpful, precise, and still prevent the effect of ‘too long did not listen’.

Based on an average response length of two sentences, which corresponds to about 24 to 40 words, the reference implementation of LLaMA 2 13B with  $\sim 8$  words per second takes 3–5 s to generate a response. To shorten the response time, the model was quantized using the exl2 algorithm [22], improving performance to  $\sim 50$  tokens per second. EXL2 allows for mixed-precision quantization, meaning that different quantization levels (2 to 8 bits per weight) can be applied to various parts of the model, optimizing for both memory efficiency and performance. For our system, we used the 13B version of LLaMA 2 with 6 bits per weight, which reduced memory usage while maintaining response quality. EXL2 automatically selects the optimal quantization settings by minimizing the quantization error, ensuring that the model performs well even with reduced bit precision. This approach allowed us to reduce the original size of the 13B model from approximately 26 GB to 10.5 GB, allowing us to run the model on a single GPU while still achieving stable outputs within the constraints of our hardware. Furthermore, the generation process was modified with a beam search, further increasing the performance to  $\sim 100$  tokens per second. After testing the system in a real-world environment, the average response time for the LLM-

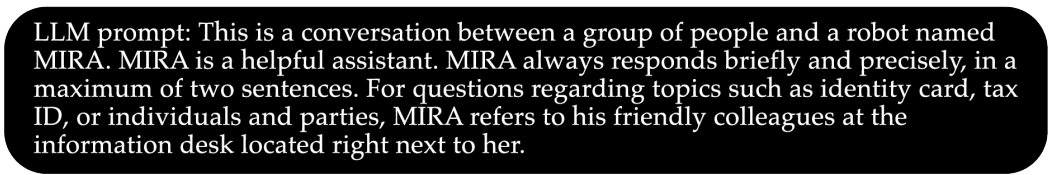
based system was found to be 1.04 s (SD = 0.41), with a minimum response time of 0.34 s and a maximum of 1.89 s. For the whole system, including the LLM, the mean response time was 1.57 s (SD = 1.02), with a minimum response time of 0.46 s and a maximum of 4.98 s. In order to facilitate a cohesive dialogue, the LLM was provided with the preceding conversation. As the responses were generated in real time, it was not possible to assess their quality in advance, in contrast to the approach used with the knowledge database. However, the generation of responses was guided in the desired direction by our LLM prompt to avoid the production of excessively lengthy responses. Additionally, the beam search was supplemented with a series of stopwords, which constrained the model and prevented the hallucination of user input. As can be seen in the following dialogue (Figure 8), it was also necessary to use the LLM prompt to limit hallucinations when errors occurred in the NLU model.



**Figure 8.** This is the continuation of the dialogue from Section 3.4. Icons beside the robot indicate MPI mode vs. single user mode as well as LLM vs. knowledge database responses.

As detailed in Section 3.2, some of the intents were very close to each other (such as collecting/applying for an ID card). Combined with unusual phrasing or faulty speech recognition, this could result in questions that the intent-based system was supposed to answer being passed on to the LLM, leading to unhelpful responses. Although maintaining a brief LLM prompt was essential to ensure quick response times, we determined that, in this specific scenario, it was necessary to make the prompt more restrictive to prevent such

errors. Consequently, we added an additional sentence to the prompt to direct users to staff members, which led to the following final prompt (Figure 9).



LLM prompt: This is a conversation between a group of people and a robot named MIRA. MIRA is a helpful assistant. MIRA always responds briefly and precisely, in a maximum of two sentences. For questions regarding topics such as identity card, tax ID, or individuals and parties, MIRA refers to his friendly colleagues at the information desk located right next to her.

**Figure 9.** Final LLM prompt.

#### 4. Conclusions, Limitations, and Future Directions

This work outlines the development of an adaptive conversation design system aimed at improving the capabilities of social robots in managing multi-party interactions (i.e., group–robot interactions) and addressing their linguistic nuances, particularly the distinction between singular and plural forms of address, which is necessary in many languages, such as German.

A significant capability of our system is its ability to discern between individual and group interactions through the utilization of a robust person detection system, based on the YOLOv5m model. This feature enables the social robot to adapt its conversational strategy in accordance with the number of users, which is a crucial capability in public spaces where robots frequently interact with multiple individuals simultaneously. The social robot’s approach to handle the social dynamics of group settings represents a step forward compared to systems that focus on one-to-one interactions.

The knowledge expander tool, developed as part of this work, leverages LLMs, like GPT-3.5, to automate the generation of diverse training examples for NLU models. This approach significantly reduces the manual effort involved in the creation of the dataset while ensuring that the social robot can handle a wide range of user intents in different languages and interaction contexts. By dynamically expanding responses, also in both singular and plural forms, the knowledge expander plays a crucial role in enhancing the social robot’s ability to handle interactions with both individuals and groups. This expansion ensures that the social robot can generate contextually appropriate responses tailored to the specific social dynamics of the interaction.

The modular architecture of the system, comprising a robot front end, an intent-based back end, and a LLM back end, enables independent optimization of each component while ensuring that they work effectively. The front end manages key interaction functions, the intent-based back end handles predefined responses and NLU tasks, and the LLM back end generates real-time responses for out-of-knowledge queries. The interaction among these components is critical to the social robot’s overall performance. For example, person detection provides real-time data to both the NLU and LLM back ends, allowing the social robot to adjust its response strategy based on the detected number of users. The intent-based back end addresses predefined knowledge with high accuracy, while the LLM back end supplements this by providing contextually appropriate responses for out-of-knowledge queries, thereby maintaining responsiveness in dynamic interaction scenarios. It is particularly important to ensure that prompts fed to the LLM back end are carefully aligned with the knowledge of the intent-based back end. Since the LLM serves as a fallback, it should avoid answering queries already covered by the intent-based system to minimize the risk of generating hallucinated responses. However, when testing the system in the wild, we still discovered the limitation of LLM hallucinations: if the answers were not within the scope of the knowledge database and could not be clearly assigned by the LLM, the LLM would sometimes give supposedly inexplicable answers. This was mostly due to incomplete STT phrases, when the system did not capture the whole user intent, likely to happen when there was a lot of ambient noise in the city hall. Although we tried to counteract this factor with fallbacks, where the system admits not to know the answer, in some cases there were still hallucinations generated by the LLM, a

topic that needs to be explored more deeply at a broad interdisciplinary level of science. A constraint of our research so far is our focus on English and specifically German as a case study. This choice is important due to our project's involvement with robotics and socially interactive agents in public service environments in Germany and to advance research in conversational design beyond English. German is particularly suitable because it has more complex grammar and syntax than languages like English. Nevertheless, the diverse nature of languages and the current limitations in the ability of generative artificial intelligence (AI) to handle many other languages, particularly low-resource ones, restrict the broader applicability of our findings.

Several areas offer the potential for further development. Enhancing the social robot's ability to distinguish between different speakers (speaker diarization) in group settings could improve its management of multi-party interactions. The integration of advanced voice recognition and turn-taking algorithms may facilitate the generation of more accurate and contextually appropriate responses. It is of the utmost importance to guarantee that these improvements are implemented in a manner that adheres to the strictest standards of user privacy, particularly when handling sensitive voice data in public settings.

Furthermore, the integration of LLM-based turn-taking mechanisms has the potential to markedly enhance the social robot's responsiveness by more accurately determining when a user has concluded their utterance. By leveraging the capabilities of large language models, the social robot can better predict the natural end of a user's utterance, allowing for quicker and more fluid transitions in conversation. This not only enhances the robot's ability to engage in real-time interactions but also reduces the likelihood of interrupting the user or causing awkward pauses.

Further research could also explore more sophisticated prompting techniques for LLMs to improve the quality and coherence of responses, particularly in multi-turn dialogues where context may change rapidly. These improvements could help maintain the continuity of conversations, especially in complex scenarios.

**Author Contributions:** Conceptualization, C.B. and A.M.; methodology, A.M., C.B., O.C. and A.R.; software, C.B. and O.C.; validation, C.B., A.M. and O.C.; formal analysis, C.B. and A.M.; investigation, C.B., A.M., O.C. and A.R.; resources, C.B., A.M., O.C. and A.R.; data curation, A.M. and C.B.; writing—original draft, C.B., A.M. and O.C.; writing—review and editing, A.M. and A.R.; visualization, A.M. and C.B.; supervision A.R.; project administration A.M. and A.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Federal Ministry of Education and Research of Germany in the framework FH-Kooperativ 2-2019 (project number 13FH504KX9).

**Institutional Review Board Statement:** The R&D activities were reviewed and approved by the Ethics Research Committee of TH Köln—University of Applied Sciences (application no. THK-2023-0004) on 29th of June 2023.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this work are available on request from the corresponding author.

**Acknowledgments:** We thank our collaboration partners DB Systel GmbH and Kreisstadt Bergheim.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
API	Application programming interface
ASR	Automatic speech recognition
CNN	Convolutional neural network

CV	Computer vision
GDPR	General Data Protection Regulation
GPU	Graphics processing unit
GRI	Group–robot interaction
HRI	Human–robot interaction
LLM	Large language model
MPI	Multi-party interaction
NLU	Natural language understanding
NLP	Natural language processing
R&D	Research and development
STT	Speech-to-text
TTS	Text-to-speech
YOLO	You Only Look Once

## References

- Breazeal, C. Social robots for health applications. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 5368–5371. [CrossRef]
- Cerrato, L.; Campbell, N. Engagement in dialogue with social robots. In *Dialogues with Social Robots: Enablements, Analyses, and Evaluation*; Springer: Singapore, 2017; pp. 313–319.
- Jayaraman, S.; Phillips, E.K.; Church, D.; Riek, L.D. Social Robots in Healthcare: Characterizing Privacy Considerations. In Proceedings of the Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, USA, 11–15 March 2024; HRI '24, pp. 568–572. [CrossRef]
- Webster, C.; Ivanov, S. *Robots in Travel, Tourism and Hospitality: Key Findings from a Global Study*; Zangador: Varna, Bulgaria, 2020.
- Hameed, I.A.; Tan, Z.; Thomsen, N.B.; Duan, X. User Acceptance of Social Robots: A Case Study. In Proceedings of the International Conference on Advances in Computer-Human Interaction, Venice, Italy, 24–28 April 2016.
- Williams, M.A. Robot Social Intelligence. In Proceedings of the 4th International Conference on Social Robotics (ICSR 2012), Chengdu, China, 29–31 October 2012; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7621. [CrossRef]
- Correia, F.; Melo, F.S.; Paiva, A. Group Intelligence on Social Robots. In Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Republic of Korea, 11–14 March 2019; pp. 703–705.
- Sabanovic, S.; Michalowski, M.P.; Simmons, R. Robots in the wild: Observing human-robot social interaction outside the lab. In Proceedings of the 9th IEEE International Workshop on Advanced Motion Control, Istanbul, Turkey, 27–29 March 2006; pp. 596–601.
- Šabanović, S.; Reeder, S.M.; Kechavarzi, B. Designing robots in the wild: In situ prototype evaluation for a break management robot. *J. Hum. Robot. Interact.* **2014**, *3*, 70–88. [CrossRef]
- Oliveira, R.; Arriaga, P.; Paiva, A. Human-robot interaction in groups: Methodological and research practices. *Multimodal Technol. Interact.* **2021**, *5*, 59. [CrossRef]
- Adam, M.; Wessel, M.; Benlian, A. AI-based chatbots in customer service and their effects on user compliance. *Electron. Mark.* **2021**, *31*, 427–445. [CrossRef]
- Addlesee, A.; Cherakara, N.; Nelson, N.; Hernández García, D.; Gunson, N.; Sieńska, W.; Romeo, M.; Dondrup, C.; Lemon, O. A Multi-party Conversational Social Robot Using LLMs. In Proceedings of the Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, USA, 11–14 March 2024; HRI '24, pp. 1273–1275. [CrossRef]
- Porcheron, M.; Fischer, J.E.; Reeves, S.; Sharples, S. Voice interfaces in everyday life. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–12.
- Lin, C.W.; Auvray, V.; Elkind, D.; Biswas, A.; Fazel-Zarandi, M.; Belgamwar, N.; Chandra, S.; Zhao, M.; Metallinou, A.; Chung, T.; et al. Dialog Simulation with Realistic Variations for Training Goal-Oriented Conversational Systems. *arXiv* **2020**, arXiv:2011.08243.
- Bapat, R.; Kucherbaev, P.; Bozzon, A. Effective crowdsourced generation of training data for chatbots natural language understanding. In *Proceedings of the Web Engineering: 18th International Conference, ICWE 2018, Proceedings 18, Cáceres, Spain, 5–8 June 2018*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 114–128.
- Parrish, A.; Huang, W.; Agha, O.; Lee, S.H.; Nangia, N.; Warstadt, A.; Aggarwal, K.; Allaway, E.; Linzen, T.; Bowman, S.R. Does Putting a Linguist in the Loop Improve NLU Data Collection? *arXiv* **2021**, arXiv:2104.07179.
- Monta, M.; Androulakis, S. Intent-Utterance-Expander. 2017. Available online: <https://github.com/miguelmota/intent-utterance-expander> (accessed on 22 August 2024).
- Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. GPT-4 Technical Report. *arXiv* **2024**, arXiv:2303.08774.
- Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Yang, A.; Fan, A.; et al. The Llama 3 Herd of Models. *arXiv* **2024**, arXiv:2407.21783.

20. Kwon, M.; Hu, H.; Myers, V.; Karamcheti, S.; Dragan, A.; Sadigh, D. Toward grounded social reasoning. *arXiv* **2023**, arXiv:2306.08651.
21. Rosenbaum, A.; Soltan, S.; Hamza, W. Using Large Language Models (Llms) to Synthesize Training Data. 2024. Available online: <https://www.amazon.science/blog/using-large-language-models-llms-to-synthesize-training-data> (accessed on 17 August 2024).
22. Frantar, E.; Ashkboos, S.; Hoefler, T.; Alistarh, D. GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers. *arXiv* **2023**, arXiv:2210.17323.
23. Paetzel-Prüsmann, M.; Kennedy, J. Improving a Robot's Turn-Taking Behavior in Dynamic Multiparty Interactions. In Proceedings of the Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, Stockholm, Sweden, 13–16 March 2023; HRI '23, pp. 411–415. [[CrossRef](#)]
24. Müller, A.; Richert, A. No One is an Island—Investigating the Need for Social Robots (and Researchers) to Handle Multi-Party Interactions in Public Spaces. In Proceedings of the 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Busan, Republic of Korea, 28–31 August 2023; pp. 1772–1777.
25. Fraune, M.R.; Nishiwaki, Y.; Šabanović, S.; Smith, E.R.; Okada, M. Threatening Flocks and Mindful Snowflakes: How Group Entitativity Affects Perceptions of Robots. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; HRI '17, pp. 205–213. [[CrossRef](#)]
26. Šabanović, S. We're in This Together: Social Robots in Group, Organizational, and Community Interactions. In Proceedings of the 8th International Conference on Human-Agent Interaction, Virtual Event, 10–13 November 2020; pp. 3–4.
27. Abrams, A.M.; der Pütten, A.M.R.v. I–C–E Framework: Concepts for Group Dynamics Research in Human-Robot Interaction: Revisiting Theory from Social Psychology on Ingroup Identification (I), Cohesion (C) and Entitativity (E). *Int. J. Soc. Robot.* **2020**, *12*, 1213–1229. [[CrossRef](#)]
28. Reig, S.; Luria, M.; Wang, J.Z.; Oltman, D.; Carter, E.J.; Steinfeld, A. Not Some Random Agent: Multi-person interaction with a personalizing service robot. In Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, 23–26 March 2020; pp. 289–297.
29. Faria, M.; Melo, F.S.; Paiva, A. Understanding robots: Making robots more legible in multi-party interactions. In Proceedings of the 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Vancouver, BC, Canada, 8–12 August 2021; pp. 1031–1036.
30. Fraune, M.R.; Šabanović, S.; Kanda, T. Human group presence, group characteristics, and group norms affect human-robot interaction in naturalistic settings. *Front. Robot. AI* **2019**, *6*, 48. [[CrossRef](#)] [[PubMed](#)]
31. Taylor, A.; Chan, D.M.; Riek, L.D. Robot-Centric Perception of Human Groups. *J. Hum.-Robot Interact.* **2020**, *9*, 15. [[CrossRef](#)]
32. Pathi, S.K.; Kiselev, A.; Loutfi, A. Detecting Groups and Estimating F-Formations for Social Human–Robot Interactions. *Multimodal Technol. Interact.* **2022**, *6*, 18. [[CrossRef](#)]
33. Luber, M.; Spinello, L.; Silva, J.; Arras, K.O. Socially-aware robot navigation: A learning approach. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Algarve, Portugal, 7–12 October 2012; pp. 902–907. [[CrossRef](#)]
34. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012*; Pereira, F., Burges, C., Bottou, L., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; Volume 25.
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
36. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
37. Dinh, M.C.; Tien, N.T.; Tuyen, T.M.; Xuan, N.V.; Anh, P.T.Q.; Bay, H.V.; Truong, X.T. Socially Aware Robot Navigation Framework: Automatic Detecting and Autonomously Approaching People in Unknown Dynamic Social Environments. In Proceedings of the 2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS), Hanoi, Vietnam, 27–29 November 2023; pp. 751–756.
38. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. Llama 2: Open Foundation and Fine-Tuned Chat Models. *arXiv* **2023**, arXiv:2307.09288.
39. Sutawika, L.; Gao, L.; Schoelkopf, H.; Biderman, S.; Tow, J.; Abbasi, B.; Fattori, B.; Lovering, C.; Phang, J.; Thite, A.; et al. EleutherAI/lm-Evaluation-Harness: Major Refactor. 2023. Available online: <https://zenodo.org/records/10256836> (accessed on 18 August 2024).
40. Jocher, G.; Stoken, A.; Borovec, J.; Liu, C.; Hogan, A.; Diaconu, L.; Ingham, F.; Fang, J.; Wang, M.; Gupta, N.; et al. Ultralytics/yolov5: V3.1—Bug Fixes and Performance Improvements. 2020. Available online: <https://zenodo.org/records/4154370> (accessed on 7 August 2024).
41. Sabharwal, N.; Agrawal, A. Introduction to Google Dialogflow. In *Cognitive Virtual Assistants Using Google Dialogflow: Develop Complex Cognitive Bots Using the Google Dialogflow Platform*; Apress: Berkeley, CA, USA, 2020; pp. 13–54. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.