Conceptualizing LLM-Based Annotation of Social Attitudes in Long-Term Human-Robot Interaction

Aurora Toma[1,*], Davide Lofrese[1,*], Claudia Chiapparino[1], Vincenzo Solfrizzi[1], Maria Paola Marino[1], Liuzzi Rosalba[1], Fabiana Ricchitelli[1] and Claudio Colonna[1]

[1]University of Bari Aldo Moro, Italy

Abstract

Long-term Human-Robot Interaction generates large amounts of conversational data, whose manual annotation remains a major bottleneck for longitudinal analysis. This paper presents a research vision that explores the use of Large Language Models as automated annotators of social attitude signals in human-robot dialogue.

The proposed framework combines a taxonomy of communicative signals with a prompt-based annotation pipeline capable of assigning both categorical labels and a continuous social attitude score at the utterance level. The approach is grounded in a realistic deployment scenario involving older adults interacting with a social companion robot in domestic environments.

The paper further outlines a validation strategy based on expert annotation and discusses how scalable annotation could enable the analysis of temporal dynamics in human-robot relationships. Finally, key challenges and open questions are identified, including prompt design, contextual dependency, and annotation subjectivity.

Keywords

Human-Robot Interaction, Social Attitudes, Large Language Models, Longitudinal Study, Socially Assistive Robots

1. Introduction

Personal social robots are increasingly being introduced into domestic environments to support older adults in everyday life, particularly in frailty, where continuous assistance and engagement are needed [1]. Beyond functional support, these systems are expected to establish meaningful and sustained interactions with users, making the understanding of relational and social dynamics a central challenge in human–robot interaction (HRI).

Personal social robots are increasingly being introduced into domestic environments to support older adults in everyday life, particularly in frailty contexts, where continuous assistance and engagement are needed [1]. Beyond functional support, these systems are expected to serve as companions that promote psychosocial well-being, reduce loneliness, and encourage cognitive and emotional stimulation through regular interaction [2, 3, 4, 5], making the understanding of relational and social dynamics a central challenge in HRI.

In such scenarios, the ability to analyze interaction dynamics over time is crucial for understanding the effectiveness of robotic companions. The way users relate to a robot, through their openness, the information they share, and their emotional and evaluative responses, can be described in terms of their *social attitude* toward the robot. These attitudes are not static; rather, they evolve across repeated interactions, reflecting increasing familiarity, the emergence of conversational routines, and, in some cases, the formation of a perceived social bond [6]. Understanding these changes over time is therefore essential for both theoretical research and the design of long-term interactive systems. In this work, we adopt the term social attitude to denote the observable manifestation of this relational stance, conceptualized along a continuum from openness (warmth, engagement) to closure (distance, disengagement).

Joint Proceedings of the ACM UMAP Workshops 2026, UMAP 2026, June 8–11, 2026, Gothenburg, Sweden

*Corresponding author.
a.toma21@phd.uniba.it (A. Toma); d.lofrese1@alumni.uniba.it (D. Lofrese); claudia.chiapparino@uniba.it (C. Chiapparino); vincenzo.solfrizzi@uniba.it (V. Solfrizzi); m.marino52@studenti.uniba.it (M. P. Marino); liuzzirosalba@libero.it (L. Rosalba); f.ricchitelli2@phd.uniba.it (F. Ricchitelli); claudiocolonna24@gmail.com (C. Colonna)

To support the analysis of these relational dynamics in long-term Human-Robot Interaction (HRI), it is necessary to adopt methodologies capable of handling large-scale conversational data while preserving fine-grained interpretability.

In longitudinal deployments, interactions between users and robots may occur daily for weeks or months, producing extensive dialogue corpora. Within these datasets, relevant phenomena, such as expressions of affect, personal disclosure, or evaluations of the robot, are conveyed through subtle linguistic and pragmatic cues that require detailed examination at the level of individual conversational turns [7]. However, manual annotation at this scale is time-consuming, costly, and difficult to reproduce, especially when multiple annotators are needed to ensure reliability [8]. As a result, existing studies are often constrained to smaller samples or simplified analyses, limiting the ability to capture the temporal richness of long-term interactions.

Recent progress in Large Language Models (LLMs) offers a potential way to address this limitation. These models have demonstrated strong performance in natural language understanding and classification tasks, achieving levels of annotation reliability comparable to, or exceeding, those of human annotators in several domains [9]. In addition, their ability to follow complex instructions expressed in natural language makes them suitable for tasks that require nuanced interpretation, such as identifying socially meaningful signals in dialogue.

This paper outlines a research vision that explores the potential use of LLMs as automated annotators of social attitude signals in longitudinal human–robot dialogue. Rather than presenting a fully developed system, the contribution aims to conceptualize a possible approach and motivate its relevance within long-term HRI research. The discussion is grounded in a realistic deployment scenario, in which a social companion robot is used by older adults in their homes over an extended period, producing naturalistic interaction data.

The envisioned framework is structured around three main components: (i) the design of an LLM-based annotation pipeline informed by an established taxonomy of social communicative signals; (ii) a validation strategy based on comparison with expert annotations produced by psychologists; and (iii) the longitudinal analysis of how social attitudes may evolve across repeated interactions. These elements are presented at a conceptual level, to outline feasible research directions rather than provide a complete implementation.

2. Related Works

Socially Assistive Robotics (SAR) has been widely explored to support older adults in everyday life by combining functional assistance with social interaction. In domestic and care settings, social robots have been shown to promote engagement, support daily routines, and improve well-being [10, 11, 3, 2]. However, their long-term effectiveness depends on the ability to establish meaningful and adaptive relationships with users, which remains an open challenge.

Long-term deployments provide valuable insights into how users develop relationships with robots, including trust, engagement, and self-disclosure. Recent work has highlighted how repeated interactions can foster the emergence of meaningful human-robot relationships (HRRs), characterized by evolving social behaviors and increasing familiarity over time [12]. Similarly, broader reviews on robot companionship have outlined the role of sustained interaction in shaping users' emotional and social responses to robotic systems [13]. Despite these advances, empirical studies remain limited by the difficulty of analyzing large-scale longitudinal interaction data.

In parallel, a growing body of work has explored the use of social robots in elderly care and domestic environments. In these contexts, robots are designed to support not only daily activities but also psychosocial well-being, by promoting engagement and reducing loneliness [14, 15, 16]. Studies conducted in home settings have shown that older adults can develop meaningful forms of interaction with social robots, often involving conversational exchanges and expressions of personal experiences [1, 17]. However, the analysis of such interactions is often based on relatively small datasets or qualitative methods, which limits the ability to capture fine-grained temporal dynamics.

Recent advances in Natural Language Processing have explored the use of LLMs as automated annotators for text analysis tasks, showing promising results in terms of scalability and cost-efficiency. In particular, LLMs have been found to achieve performance comparable to, and in some cases exceeding, that of human annotators in classification and labeling tasks [9, 18, 19]. At the same time, their reliability in subjective and context-dependent annotation remains debated, as performance can vary depending on prompt design and task complexity [20, 21].

Despite these developments, limited work has investigated the use of LLMs for the annotation of socially meaningful signals in HRI, particularly in longitudinal and real-world settings. Existing HRI studies still rely largely on manual annotation or small-scale analyses, which limits the possibility of systematically investigating how social attitude evolves over time. This paper addresses this gap by proposing a research vision that combines LLM-based annotation with longitudinal analysis of user–robot dialogue in ecologically valid environments.

3. Interaction Setting and Motivating Context

The research vision presented in this paper is situated within a concrete, yet still exploratory, application scenario that serves to illustrate the potential context of use and the type of data that could support the proposed approach.

This work has been carried out in the context of the *SISTER* project (SocIal robotS to support biopsychosocial frailTy) [22]. This project focuses on the use of socially assistive robotics to support frail elderly populations in Italy. In the proposed setting, participants are individuals over 65 years of age who interact with the robot in a natural and unstructured way over an extended period (several weeks or months). Besides reminders and daily questionnaires, no strict interaction schedules or predefined tasks have been imposed to preserve ecological validity and encourage spontaneous use of the system in everyday life. Such a design is expected to generate rich and naturalistic conversational data, where social attitude signals may emerge in a more authentic and meaningful manner.

The robotic platform considered in this scenario is described in *Pixie* [23] and it has been integrated on the UBTECH Alpha Mini humanoid robot.[1] The system is equipped with multimodal interaction capabilities, including microphones, speakers, motorized joints, and LED-based facial expressions. Its architecture is designed to support both task-oriented interactions and open-ended conversations, allowing users to engage in dialogue during daily routines. In the envisioned setup, interactions would be automatically recorded and stored as timestamped textual logs through a back-end infrastructure, thus forming a longitudinal dialogue corpus suitable for subsequent analysis.

To better illustrate the type of interaction that could arise in such a context, a short dialogue example is reported below. It is important to note that this excerpt does not correspond to real collected data, but represents a plausible interaction that could potentially occur in a real-world deployment of this kind:

```
2026-02-19T08:33:43
User: Hey Pixie, good morning.
Robot: Good morning Maria! I'm here, ready to help you and keep you
   company. How are you feeling this morning?

2026-02-19T08:34:16
User: This morning I'm fine, I'm feeling good. I've just had
   breakfast and then I'll get started on the housework.
Robot: I'm glad you're feeling well! Was breakfast good? If you'd
   like, we can chat a bit while you do the housework.
```

Even in an illustrative example of this kind, it is possible to identify several types of social attitude signals that are expected to be relevant in the proposed analysis. These may include the expression of emotional states (e.g., *"I'm fine, I'm feeling good"*), the spontaneous sharing of aspects of daily life (e.g.,

[1] https://www.ubtrobot.com/en/consumer/humanoid-robots/alpha-series/alpha-mini

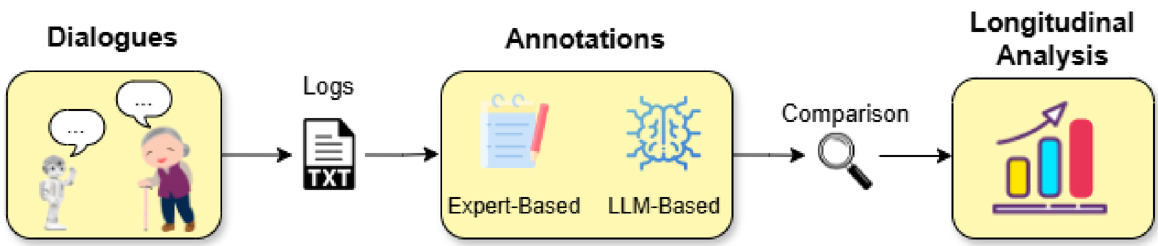

Figure 1: Overview of the proposed LLM-based annotation pipeline for longitudinal human-robot dialogue.

references to breakfast or housework), and the use of the robot's name as a form of personal address. Such signals are not presented here as empirically observed findings, but as representative instances of the phenomena that could reasonably be expected in naturalistic human–robot dialogue.

4. Proposed Annotation Framework and Validation

An overview of the proposed framework is presented in Figure 1. The process begins with raw HRIs, where conversational exchanges are collected and stored as textual logs. These data are then used as input for an annotation stage, where both expert-based and LLM-based approaches are considered in parallel. In the expert-based setting, human annotators label the data according to predefined social communicative categories, while in the LLM-based setting, annotations are automatically generated through prompt-based interaction. The outputs of the two processes are subsequently compared to assess the reliability of automated annotation. Finally, the resulting annotations are used to support longitudinal analysis, enabling the investigation of how social attitude signals evolve over time and giving rise to interpretable interaction trajectories.

4.1. Annotation Scheme

The annotation approach considered in this work is conceptually grounded in the taxonomy of communicative signals introduced by De Carolis et al. [7], which has been used to describe and model social attitude in human–agent interaction. In the context of the present research, this taxonomy is adopted as a reference framework to illustrate how social attitude signals could be systematically identified and categorized at the level of individual utterances.

At a fine-grained level, the scheme includes several pragmatic signal categories taken by De Rosis et al. [24] as described in 1.

While the linguistic signals reported in Table 1 are grounded in prior work, they are not intended to exhaustively represent all aspects of social attitude. In particular, additional conversational phenomena such as *personal address*, *feeling expressions*, *social motivators*, and *social sharing* are also considered, as they capture relevant aspects of how users express relational and social meaning during interaction with a robot.

Rather than being treated in isolation, these low-level signals can be conceptually aggregated into broader performative dimensions, including the degree of self-disclosure, the level of familiarity, the presence of explicit evaluations of the robot, and the overall intensity of conversational engagement. Together, these dimensions provide a structured way to describe the user's relational stance throughout the interaction.

In addition to categorical annotations, the framework includes the assignment of a continuous *social attitude score* to each user utterance, ranging from -1 (indicating disengagement or negative attitude) to $+1$ (indicating positive and socially engaged behavior), with intermediate values representing neutral or task-oriented interaction. The choice of a continuous score, rather than a discrete categorical label, is motivated by the need to capture gradations of relational stance that binary or ordinal classifications would fail to represent. This design is consistent with established frameworks in affective computing

Table 1
Linguistic Signals of Social Attitude.

Signal	Example	Linguistic Features
Friendly self-Introduction	Good morning, nice to meet you. My name is...	Greetings and self-introduction expressions
Colloquial style	”Better safe than sorry”, “Let me think about it”, “Alright, let’s try.”	Paralanguage, spoken language terms, dialectal forms, idiomatic expressions, diminutive or expressive forms
Talks about self	I adore sweets, I like vegetables, When I was young, I was very sporty...	First-person pronouns and verbs expressing knowledge, attitudes, abilities, preferences, or desires
Questions about the agent	Do you have preferences? What do you think about sport?	Second-person pronouns and verbs expressing knowledge, attitudes, abilities, preferences, or desires
Positive/Negative comments	You are very nice! / It’s very kind of you /You are a rude robot...	Agreement/disagreement expressions, evaluation of agent’s politeness, competence, repetitiveness, and understanding ability
Friendly farewell	Thank you and see you soon	Expressions of farewell and thanking

that model emotional and attitudinal states along continuous dimensions, such as the valence-arousal space model introduced by Russell [25] and the PAD model proposed by Mehrabian and Russell [26].

This combination of discrete labels and a scalar score is intended to support both qualitative interpretation of individual conversational turns and quantitative analyses over time. It should be noted, however, that in the current work this scheme is not yet applied to real data, but is proposed as a structured basis for future annotation efforts.

4.2. LLM-Based Annotation Pipeline

Building on this annotation scheme, an annotation pipeline based on LLMs is outlined at a conceptual level. The general idea is to leverage prompt-based interaction with an LLM in order to approximate the annotation process that would otherwise be carried out manually by human experts.

In this pipeline, the model will be provided with a structured prompt containing a description of the notion of social attitude, detailed definitions of each signal category, and a set of illustrative examples demonstrating the expected annotation format. For each target utterance, the LLM will then be asked to identify the relevant communicative signals and assign a corresponding social attitude score, returning the output in a structured and machine-readable form.

A central aspect of this design concerns the role of conversational context. Social signals are often highly dependent on surrounding dialogue, and their interpretation may change significantly when considered in isolation versus within a broader exchange. For this reason, the inclusion of a local dialogue window around the target utterance is expected to play an important role in the quality of the annotations. At the same time, longitudinal interaction scenarios introduce an additional layer of complexity, as users may refer to shared experiences or previously established routines across multiple sessions. The extent to which such long-range context can be effectively incorporated into the annotation process, and whether LLMs can leverage it in a meaningful way, remains an open question that this research aims to address in the future.

4.3. Validation Against Expert Annotation

A further component of the proposed framework concerns the validation of LLM-generated annotations through comparison with human expert judgments. While no validation study is presented in this paper, a possible experimental design is outlined to clarify how the approach’s effectiveness could be assessed in future work.

In such a study, a subset of the dialogue data, collected over an approximately two-month deployment period involving five participants, would be independently annotated by a team of clinical psychologists with expertise in social communication and HRI, following the same annotation guidelines provided to the LLM. The selected subset would be chosen to ensure representativeness across interaction phases (early, middle, and late sessions) and across different participants.

Inter-rater agreement among the experts would be evaluated using established reliability measures, such as Cohen's κ for categorical signal annotations and intraclass correlation coefficients (ICC) for the continuous social attitude score. These metrics would serve both to assess the internal consistency of the expert panel and to provide an upper bound for the expected performance of the automated system.

The comparison between LLM-based and expert annotations would serve multiple purposes. On the one hand, it would provide an indication of whether automated annotation can reach a level of reliability comparable to human agreement, which is a necessary condition for its use in large-scale longitudinal analyses. On the other hand, discrepancies between model outputs and expert judgments could offer valuable insights into the limitations of the approach, as well as into the intrinsic ambiguity of certain communicative signals. To handle annotation uncertainty explicitly, low-confidence model outputs could be flagged for human review, enabling a hybrid annotation workflow in which automation is selectively complemented by expert judgment on ambiguous cases. In this sense, validation is not only conceived as a benchmark for performance, but also as an opportunity to better understand the nature of the annotation task itself.

5. Longitudinal Analysis: Potential Insights

The annotation pipeline outlined in the previous section is intended as a means to enable forms of longitudinal analysis that would otherwise be difficult to conduct at scale. By automating the identification of social attitude signals at the level of individual utterances, it becomes possible to explore how these signals evolve over extended interaction periods.

A first line of investigation concerns the temporal dynamics of social attitude throughout the deployment. It is reasonable to expect that certain communicative behaviors may change as familiarity with the robot increases. For instance, users may gradually engage in more self-disclosure, refer more frequently to shared experiences, or adopt a more informal and personalized interaction style. Similarly, the overall social attitude score may exhibit different trajectories over time, potentially reflecting varying levels of engagement, trust, or relational closeness.

These temporal dynamics are not merely quantitative phenomena, but they reflect deeper relational processes that have been theorized in the literature. The gradual increase in self-disclosure, for instance, is consistent with social penetration theory [27], which describes how interpersonal closeness relationships develops through progressive layers of personal sharing. Similarly, the emergence of trust and perceived social presence in repeated interactions with robots has been documented in longitudinal studies [12, 6], suggesting that users can develop meaningful relational bonds even with non-human agents. By grounding the longitudinal analysis in these theoretical constructs, the proposed framework aims not only to describe interaction patterns but also to interpret them in terms of the underlying relational dynamics they may reflect.

Indeed, analyzing these trajectories at the level of individual participants would make it possible to characterize different patterns of interaction. Some users may show a steady increase in engagement, while others may maintain a stable but neutral attitude, or even display decreasing interest over time. Such variability is particularly relevant, as it may reveal the presence of distinct relational styles in HRI.

These differences may in turn be associated with user-specific factors, such as personality traits, prior familiarity with technology, individual communication preferences, or aspects of the social and psychological context in which the interaction takes place. Although these factors are not directly modeled in the present work, their potential influence highlights the importance of adopting a longitudinal perspective when studying HRR.

In addition to gradual trends, longitudinal data may also reveal more abrupt changes in interaction patterns. Specific moments in which a user's behavior shifts noticeably, for example, a transition from task-oriented dialogue to more socially expressive interaction, or a decrease in engagement following repeated unsatisfactory exchanges, may provide particularly informative insights into the development of the relationship. The possibility of identifying such transitions relies on the availability of fine-grained annotations over large amounts of data, which the proposed pipeline is intended to support.

Beyond analysis, the annotated data could also serve as training material for learning a model of social attitude, enabling the robot to adapt its behavior over time in order to foster a better long-term relationship with the user.

To illustrate the type of analytical output that the proposed pipeline is expected to produce, consider a hypothetical deployment involving a single participant over eight weeks of interaction. The social attitude score assigned to each utterance could be aggregated at the session level (e.g., as a weekly mean) and visualized as a temporal trajectory. Such a representation would allow the identification of gradual trends, such as a progressive increase in score reflecting growing familiarity, as well as more abrupt shifts potentially associated with specific interaction events. Although no empirical results are presented at this stage, this type of visualization constitutes the primary expected output of the longitudinal analysis and will be the focus of future experimental work.

6. Challenges and Open Questions

Despite its potential, the research direction outlined in this paper raises several challenges that remain to be addressed.

First of all, the effectiveness of LLM-based annotation depends strongly on the formulation of the prompt. Designing prompts that consistently produce accurate and well-calibrated annotations across all signal categories requires careful refinements. In addition, there is a risk that prompts become overly tailored to specific examples, which may limit their robustness when applied to new interaction data.

Then, the interpretation of social signals is often highly dependent on the conversational context. While the inclusion of a local dialogue window may improve annotation quality, determining the appropriate amount of context to provide remains an open issue. This challenge is further complicated in longitudinal scenarios, where references to past interactions may span multiple sessions and exceed the effective context capacity of the model.

Finally, the collection and analysis of conversational data in domestic environments raises important ethical concerns. Even if anonymized, such data may include sensitive information related to users' daily lives, emotional states, and personal experiences. Besides anonymization, additional considerations regarding informed consent, data storage, and responsible use of automated analysis tools are equally important.

7. Conclusions

This paper presented a proposal for the use of LLMs as automated annotators of social attitude signals in longitudinal HRI. The proposed framework combines an established taxonomy of communicative behaviors with a prompt-based annotation pipeline, and is situated within a realistic deployment scenario involving social companion robots in domestic environments.

Although no empirical results are provided at this stage of research, the proposed approach could enable the large-scale analysis of conversational data that would otherwise be difficult to annotate manually. In particular, it has been argued that fine-grained, utterance-level annotation may support the investigation of temporal dynamics in HRRs, including gradual trends, individual differences, and moments of change in user engagement.

At the same time, several challenges have been identified, highlighting the exploratory nature of the proposed direction. Issues related to prompt design, contextual interpretation, annotation subjectivity, and ethical considerations remain open and require systematic investigation.

Future work will focus on the implementation and empirical evaluation of the proposed pipeline, including validation against expert annotations and application to real-world interaction data. More broadly, this research aims to contribute to the development of scalable methodologies for studying the evolution of social attitudes in long-term HRI and learning a model to be used by a social robot to adapt its behavior accordingly.

Declaration on the use of Generative AI

During the preparation of this work, the authors used ChatGPT in order to: grammar and spelling check, paraphrase and reword. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] C. L. Kok, C. K. Ho, T. H. Teo, K. Kato, Y. Y. Koh, A novel implementation of a social robot for sustainable human engagement in homecare services for ageing populations, Sensors 24 (2024) 4466.

[2] A. Toma, D. Lofrese, G. Palestra, B. N. De Carolis, Towards personal robotics: An analysis of the state of the art, in: SOCIALIZE 2025, CEUR Workshop Proceedings, 2025.

[3] K. Blindheim, M. Solberg, I. A. Hameed, R. E. Alnes, Promoting activity in long-term care facilities with the social robot pepper: a pilot study, Informatics for Health and Social Care 48 (2023) 181–195.

[4] D. Macis, S. Perilli, C. Gena, Employing socially assistive robots in elderly care, in: Adjunct proceedings of the 30th ACM conference on user modeling, adaptation and personalization, 2022, pp. 130–138.

[5] N. Gasteiger, K. Loveys, M. Law, E. Broadbent, Friends from the future: a scoping review of research into robots and computer agents to combat loneliness in older people, Clinical interventions in aging (2021) 941–971.

[6] M. Skjuve, A. Følstad, P. B. Brandtzæg, A longitudinal study of self-disclosure in human–chatbot relationships, Interacting with Computers 35 (2023) 24–39.

[7] B. De Carolis, N. Novielli, Recognizing signals of social attitude in interacting with ambient conversational systems, Journal on Multimodal User Interfaces 8 (2014) 43–60.

[8] R. Artstein, M. Poesio, Survey article: Inter-coder agreement for computational linguistics, Computational linguistics 34 (2008) 555–596.

[9] F. Gilardi, M. Alizadeh, M. Kubli, Chatgpt outperforms crowd workers for text-annotation tasks, Proceedings of the National Academy of Sciences 120 (2023) e2305016120.

[10] B. De Carolis, G. Palestra, M. Bochicchio, S. Mazzoleni, Alpha mini social robot as a fitness trainer at home, in: 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN), IEEE, 2024, pp. 1638–1643.

[11] A. Poberznik, S. Merilampi, 13 older adults' experiences with pepper humanoid robot, Tutkimusfoorumi (2019) 148.

[12] G. Laban, A. Kappas, V. Morrison, E. S. Cross, Building long-term human–robot relationships: Examining disclosure, perception and well-being across time, International Journal of Social Robotics 16 (2024) 1–27.

[13] E. Ahmed, O. O. Buruk, J. Hamari, Human–robot companionship: Current trends and future agenda, International Journal of Social Robotics 16 (2024) 1809–1860.

[14] L. B. H. Austbø, I. Testad, M. T. Gjestsen, Using a robot to address the well-being, social isolation, and loneliness of care home residents via video calls: Qualitative feasibility study, JMIR Formative Research 9 (2025) e59764.

[15] J. Leoste, K. Lubi, K. Marmor, K. Kangur, Evaluating social assistive robots in clinical nursing

care: Mixed method pilot study on health care workers' perceptions and adoption, JMIR nursing 8 (2025) e70305.
[16] L. Rettinger, A. Fürst, E. Kupka-Klepsch, K. Mühlhauser, E. Haslinger-Baumann, F. Werner, Observing the interaction between a socially-assistive robot and residents in a nursing home, International Journal of Social Robotics 16 (2024) 403–413.
[17] A. Y. Leung, I. Y. Zhao, S. Lin, T. K. Lau, Exploring the presence of humanoid social robots at home and capturing human-robot interactions with older adults: experiences from four case studies, in: healthcare, volume 11, MDPI, 2022, p. 39.
[18] M. Alizadeh, M. Kubli, Z. Samei, S. Dehghani, M. Zahedivafa, J. D. Bermeo, M. Korobeynikova, F. Gilardi, Open-source llms for text annotation: a practical guide for model setting and fine-tuning, Journal of Computational Social Science 8 (2025) 17.
[19] D. Yu, Towards llm-assisted move annotation: Leveraging chatgpt-4 to analyse the genre structure of ceo statements in corporate social responsibility reports, English for Specific Purposes 78 (2025) 33–49.
[20] Q. Li, L. Cui, L. Kong, W. Bi, Exploring the reliability of large language models as customized evaluators for diverse nlp tasks, in: Proceedings of the 31st International Conference on Computational Linguistics, 2025, pp. 10325–10344.
[21] R. D. Kristensen-McLachlan, M. Canavan, M. Kardos, M. Jacobsen, L. Aarøe, Are chatbots reliable text annotators? sometimes, arXiv preprint arXiv:2311.05769 (2023).
[22] V. Solfrizzi, C. Chiapparino, G. Castellano, B. De Carolis, D. Lofrese, G. Palestra, A. Toma, L. Perla, S. Massaro, M. T. Santacroce, et al., Sister: Social robots to support biopsychosocial frailty, in: Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, 2025, pp. 1–6.
[23] D. Lofrese, A. Toma, G. Palestra, B. N. De Carolis, Pixie: a personal robot for elderly assistance, in: Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter, 2025, pp. 1–3.
[24] F. De Rosis, A. Batliner, N. Novielli, S. Steidl, 'you are sooo cool, valentina!'recognizing social attitude in speech-based dialogues with an eca, in: International Conference on Affective Computing and Intelligent Interaction, Springer, 2007, pp. 179–190.
[25] J. A. Russell, A circumplex model of affect., Journal of personality and social psychology 39 (1980) 1161.
[26] A. Mehrabian, J. A. Russell, An approach to environmental psychology., the MIT Press, 1974.
[27] I. Altman, D. A. Taylor, Social penetration: The development of interpersonal relationships., Holt, Rinehart & Winston, 1973.