Applied Computing Journal https://doi.org/10.52098/acj.20266171
ISSN: 2788-9688 Volume 6, Issue 1, pp 452-460, January 2026
Received: 20/12/2025 Revised: 9/1/2026 Accepted: 11/2/2026

Natural Language Processing for Human–Robot Interaction: A Comprehensive Review of Methods, Challenges, and Future Directions

Reem Rashid Al-Maawali[a], Amna Hassan Al-Shidi[b] and Jabar Yousif[c]
[a, c] *Faculty of Computing and Information Technology, Sohar University, Sultanate of Oman*
[b] *Faculty of Information Technology, University of Buraimi, Sultanate of Oman*

Abstract

Natural Language Processing (NLP) has been established as a building block technology in intelligent and socially responsible Human-Robot Interaction (HRI). Technology helps robots effectively, meaningfully, and naturally process and react to human communication. A critical analysis of a compilation of landmark publications between 2020 and 2025 has been presented, targeting the major technological advancements attained in the development or implementation of HRI technology based on Natural Language Processing technology. The results achieved a shift towards a data-intensive, data-driven, or even a transformation-based approach to symbolic or probabilistic frameworks concerning language, leading to an improvement in the accuracy levels related to the interpretation, accomplishment, or detection of contextual meaning and emotions. Despite this, a plethora of challenges remain to be addressed, related to attaining or sustaining real-time response, handling domain data gaps effectively, adapting to multilingual frameworks, and embracing approaches related to ethics & bias, among others. The current trends continue to encompass developments related to hybrid architectures, low-energy consumption approaches suitable for embedded applications, culturally adapted dialogical frameworks, and ethically anchored communication schemes, among others.

Keywords: Natural Language Processing; Human-Robot Interaction; Large Language Models; Multimodal Learning; Affective Computing; Ethical AI; Embodied Intelligence

1. Introduction

In recent times, the technology of Human-Robot Interactions (HRI) has progressed from command-based communication interactions to intelligent and context-based communication interactions facilitated by Natural Language Processing (NLP). The innovative or breakthrough development achieved by Natural Language Processing has allowed robots to fluently understand, process, and generate human language, making the interaction between humans and robots a smooth affair, as robots are now capable of interpreting human language meaningfully and correctly, thus removing the communication gap between robots and humans (Wang et al., 2023; Yousif, 2021). The development achieved by robots due to innovative breakthroughs in natural language processing has made robots socially intelligent, as they can understand human meaning or intention and can assist humans in sectors like education, health, and household chores, among others, as mentioned by Cohen et al. (2024) and Sharma et al. (2022). The primary natural language processing components, which are speech recognition, natural language understanding, big data analysis, dialog management, and language generation, are the building blocks or foundation for successful robotic communication interactions (Yousif & Saini, 2020). The coupling or integration of deep learning algorithms with the development of BERT, GPT, and T5 Transformers has allowed robots to meaningfully or accurately interpret complex semantic meaning and generate meaningful or natural communication interactions, as mentioned by Vaswani et al. (2017) and Brown et al. (2020).

Despite this, the integration of NLP technology into robotics is a challenging endeavor compared to text-based systems. First, robots are designed to work within a physical, dynamically varying space requiring immediate linguistic grounding, between which the input language should correspond to actual perceptions and acts (Cohen et al., 2024; Khandelwal et al., 2022). Further complexities include noise, accents, and multilinguality, all influencing accuracy in speech recognition (Wang, 2024; Zhang et al., 2022). Additionally, graph neural networks and large language models, including GPT-4 and PaLM, pose novel concerns regarding computational complexity, latency, and even ethics and explanations among users (Zhang et al., 2023). Addressing these concerns, recent studies propose multimodal frameworks that aim to represent linguistic, visual, and even affective representations as a possible approach to better enable empathy and social understanding among the agents involved (Wu et al., 2023; Li et al., 2023; Dautenhahn, 2023). Based on this, this paper will discuss the developments seen in multimodal-based approaches applying NLP technology to HRI applications, as well as future directions to make communication systems better-adaptive, ethical, and human-centric technologies.

2. Literature Review

Mazzei et al. (2021) This works attempts to realize a retroactive analysis of changes within the field of social robotics through the past 20 years by means of the content of scholarly articles in social robotics and human-robot interaction using techniques of NLP. In addition to identifying a new set of studies under the umbrella of "soft human-robot interaction" (Soft HRI), the findings also show overlap between the two domains and seven sub-themes that reflect current research trends. The study offers a thorough grasp of the progress and present status of social robotics, support scholars in strategically leading their future research in this multidisciplinary area.

Furthermore, the study by Younis et al. (2023) aspects at how NLP techniques develop interaction, communication, and teamwork in technological and educational frameworks, particularly in the field of NAO robots. The researchers studied 82 scientific articles from twelve high-impact journals listed on the Scientific Journal Rankings website, published between 2014 and 2023. The literature was divided into four educational groups of interest to the study: special needs, kindergartens, schools, and universities. The study conducted a systematic keyword analysis that included the words artificial intelligence, education, NAO robot, learning, and teaching to find out how NAO robots support language acquisition and educational advancement. The results underlined the great possibilities of NAO robots in improving the processes of teaching and learning, while their efficacy and advancement in the educational field have been confirmed. The study by Gao et al. white (2024) investigates how Large Language Models can improve human-robot understanding and communication in cooperative industrial settings. Traditional command-based systems require strict and precise syntax, leading to frequent failures of the system. In order to address this challenge, the authors enhanced a GPT-3.5 model that can understand diverse natural language instructions and turn them into relevant task characteristics, such as required components, tools, and task relevance. When integrated into robot perception systems, this model generated appropriate action sequences for assigned tasks. In an assembly case study, the optimized LLM showed greater versatility in their ability to correct mistakes where the robots initially pick up the wrong tools and parts. The advantage of these results shows enhanced accuracy, efficiency, and robustness of human-robot collaboration by integration with LLM and thus paves the way toward even more intelligent and responsive robots.

The researcher, therefore, conducted a study to establish how NLP improves human-robot interaction and communication. It draws attention to the growth of simple NLP applications into complex systems with contextual awareness and adaptive communication. The study identified NLP as being crucial in enhancing machine understanding and responsiveness, hence making human-robot interaction effective and natural. Besides that, some of the important challenges it identifies include multilingual communication, the problem of linguistic ambiguity, and ethical issues related to robotic interactions. This research determines that continued technological innovation and cross-disciplinary collaboration will be needed to leverage the full promise of NLP for robotics toward making more intelligent, adaptive, and intuitive robotic systems in the future. Besides this, Li et al. (2021) proposes a task-based paradigm that leverages visual semantics instead of rigid instructions to enable robots to understand human intentions. This framework interprets task context and natural language based on the combination of a similarity computation algorithm, identification of objects visually, and a module for language semantics. When tested with invisible phrase structures on a humanoid robot performing an object-selection and handover job, it remained capable of handling unambiguous, ambiguous, and even emotion-based instructions. The results prove the system indeed enhances human–robot communication in terms of adaptability, naturalness, and context awareness.

Azevedo-Sa et al. (2021) present an extended model geared toward improving human-robot collaboration. The capability-based framework is proposed to evaluate trust in the interaction of humans with robots and humans. Instead of predefined performance values, each task has an attached set of requirements in terms of capabilities, and each agent, robot or human, has a capability distribution. According to their paper, this model tends to make better predictions compared to the more traditional models of trust in forecasting human trust toward robots in a wide range of tasks. The research consisted of an online experiment with 284 participants. They ran simulated tests for robotic trustors to demonstrate that robots could similarly assess human dependability. All in all, the paradigm has useful implications for control and facilitating dynamic trust management. Moreover, the work of Yura et al. (2025) explores new ways to enhance the naturalness and responsiveness of service robots' communication. The focus of this study is on service robots with verbal and visual communication skills since there is a growth in the use of robots in industries such as manufacturing, health, agriculture, and personal services. The authors propose a novel method for refining human-robot interaction precision and fluidity by incorporation image processing with sound localization. This technology also controls the spinning speed of the robot to reproduce to human movement to advance flexibility and ease during engagement. The study contains experiments carried out in human-robot environments, representing better usability, mobility, and accuracy of interactions. The outcomes highlight how significant it is to contain multimodal perception in attaining more effective natural communication in service robotics.

The work by (El-Komy et al., 2022) unfolds the expressions of how the integration of both NLP and computer vision can bring forth intelligent systems capable of significantly increasing the potential availability of visually impaired people by providing safe navigation features around their world. The smartphone, with the integration of vision, language, and intelligence capabilities, allowed the authors to create an integrated smartphone app. The smartphone captures real-time images of the surroundings to identify objects and barriers, later processed by a Faster Region Convolutional Neural Network residing in a central server, converting their identification to vocal signals to contribute to safe navigation around their world. Other than manifesting the broader implications of this technology around assistive robotics, the article shows how this integration between NLP and computer vision enhances human communication and perception faculties. Tellex et al. (2020) investigates the language-grounding problem for robots and provides a comprehensive survey of methods that connect linguistic symbols to percepts and actions. The paper classifies approaches into symbolic/formal methods, probabilistic inference, and learning-based techniques, and highlights practical HRI constraints (noisy sensors, embodiment, ambiguity, and the need for interactive learning). Its novelty is framing grounding as a central, cross-disciplinary challenge that differentiates HRI from standard NLP tasks. Strengths include a clear taxonomy, broad coverage of early literature, and identification of core open problems (data collection for embodied contexts, compositional instruction following, and interactive learning loops). Limitations are inherent to survey papers: it does not present new empirical benchmarks, and because it predates large transformer/LLM breakthroughs, it does not evaluate the most recent embedding- or LLM-centric grounding approaches.

Blukis et al. (2021) investigates few-shot object grounding and mapping methods to enable robots to follow natural-language instructions referring to previously unseen objects. The authors propose an exemplar-driven grounding pipeline that uses augmented-reality exemplars to align language mentions with visual exemplars and to build an object-centric spatial map used by downstream control policies. The novelty is improving sample efficiency for grounded instruction following by leveraging few-shot exemplars rather than massive, labelled datasets. Empirical evaluation shows substantive generalization to unseen objects; reported classification-style results on the authors' testbeds indicate accuracy around 73% for an 8-way unseen-object task and 63% for a 15-way variant (reported on their benchmarks). Strengths are the clear improvement in few-shot generalization and practical mapping design; weaknesses include dependency on exemplary availability, limited demonstrations of multi-step compositional instruction following, and that many experiments are in simulated or constrained environments rather than fully unconstrained real-world deployments.

Stepputtis et al. (2020) investigates language-conditioned imitation learning for robotic manipulation by training end-to-end policies that map visual observations plus natural-language descriptions to low-level control commands. The paper demonstrates that when paired demonstration trajectories are available, conditioning policies on language enable pick-and-place and object rearrangement behaviors triggered by descriptive instructions. The novelty is the end-to-end conditioning of continuous control on natural language (rather than dividing perception, planning and control into separate modules). Strengths include convincing task demonstrations and empirical analysis showing how performance scales with dataset size. Limitations include large data requirements for robust generalization, reduced reliability for abstract or long-horizon instructions, and relatively fewer physical-robot trials compared to simulated experiments, so real-world robustness requires further work. The paper reports task-level success improvements in their benchmarks but does not present a single comparable accuracy metric that generalizes across domains (results are dataset/task-specific). Ichter et al. (2023) investigates the integration of large language models with robot affordance models via the SayCan architecture, where an LLM proposes candidate high-level actions and an affordance/skilled scorer ranks which proposals the robot can feasibly and safely execute. The novelty is decoupling flexible, commonsense reasoning (via the LLM) from feasibility/effectiveness (via the affordance model), enabling richer planning while limiting unsafe or impossible commands. On the reported experimental suite SayCan (paired with a PaLM-style LLM and a skills library) achieved high planning and execution performance, reported planning success around 84% and execution success around 74% on a set of 101 benchmark tasks in the authors' evaluation. Strengths include improved task flexibility, practical safety benefits from affordance filtering, and demonstration on multi-step scenarios. Weaknesses include reliance on a predefined skill set, affordance models trained per-domain/hardware (so portability is limited), and ongoing challenges in scaling to very diverse robot platforms and long-horizon behaviours.

Wang (2024) investigates the intersection of large language models and robotics in a broad review that synthesizes how LLMs are applied for instruction parsing, planning, dialogue, and multimodal reasoning in embodied agents. The paper's novelty is offering a focused synthesis of architectural tradeoffs (cloud vs. on-device vs. distilled LLMs), latency and safety considerations, and practical adaptation techniques (adapters, fine-tuning, distillation) for robotics. Strengths are the clear roadmap of engineering choices and articulating important research directions such as multimodal fusion, real-time adaptation, and safety-centric filtering. Limitations are that it is a survey of emerging work, many case studies remain prototype level, and because implementations vary widely across papers, the review cannot present a single unifying quantitative performance metric. Cohen et al. (2024) investigates robotic language grounding through a systematic survey that positions methods along a spectrum from symbolic interpretable models to end-to-end embedding/policy approaches. The author's novelty is the explicit trade-off analysis (interpretability and low sample complexity of symbolic systems vs. scalability and expressivity of deep embedding models) and proposing hybrid architectures as a promising compromise. Strengths include rigorous taxonomy, discussion of dataset and evaluation gaps (particularly the need for temporally and spatially aligned multimodal corpora), and recommendations for benchmarking. Limitations are that the paper is conceptual and survey-oriented (limited novel empirical evaluation), and practical implementation guidance for hybrid methods remains an active area of research rather than a settled set of best practices.

Mauliana et al. (2025) investigates multi-session human–robot interaction where an LLM backend powers conversational continuity, personalization, and adaptive memory across repeated user sessions. The novelty is the longitudinal evaluation design: multi-day studies measuring continuity, personalization benefits, and safety/consistency issues that arise when robots retain session history. The study reports improved user-perceived naturalness and continuity with session-aware systems but also surface risks, LLM hallucinations, privacy and consent challenges, and inter-session inconsistency. Strengths are the applied, user-centred evaluation and actionable mitigations (session-limited memory with consent, retrieval-based grounding, safety classifiers). Weaknesses include small participant pools, limited cultural diversity in test participants, and those quantitative improvements vary by task and are reported as task-specific satisfaction or continuity metrics rather than a single universal accuracy figure. Wu et al. (2025) investigates multimodal emotion and sentiment recognition approaches for HRI, surveying fusion strategies that combine text, speech prosody, facial expression, and physiological signals. The novelty focuses on transformer-based multimodal fusion architectures targeted at embodied robot settings, together with practical concerns about synchronization, real-time processing, and privacy. Strengths include comprehensive modality coverage, comparisons of early/late/hybrid fusion strategies, and practical recommendations for dataset collection (ecological validity, cross-cultural labeling). Weaknesses include that most systems reviewed remain lab-based with limited in-robot real-time deployment data; quantitative improvements reported in surveyed papers are dataset-specific (models often show multimodal fusion outperforming single-modality baselines, but the exact uplift varies by dataset and is not summarized as a single cross-paper number).

Table 1 provides a summary of key studies, highlighting the methods used, the challenges encountered, and the future directions proposed. This overview allows readers to quickly grasp the main contributions of each study, identify research gaps, and understand how existing work informs the development of the methodology presented in this paper.

Table 1: Summary of Key Studies, Methods, Challenges, and Future Directions

Study Title	Authors	Year	Method	Challenges	Future Direction
Analyzing Social Robotics Research with Natural Language Processing Techniques	Mazzei, D., Chiarello, F., & Fantoni, G.	2021	Used NLP techniques to analyze academic publications in Social Robotics (SR) and Human–Robot Interaction (HRI) over the past two decades using text mining and topic modelling.	1) Complexity among disciplines. 2) Inconsistency in nomenclature. 3) Inadequate organized data.	Create integrated analytical methods, encourage global cooperation, and advance "Soft HRI" research
A Systematic Literature Review on the Applications of Robots and Natural Language Processing in Education	Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M.	2023	Conducted a systematic review of 82 articles (2014–2023) across 12 high-impact journals to explore how NLP and NAO robots are applied in education.	1) Limited integration in the classroom. 2) Ethical and technical difficulties. 3) Standardized evaluation is necessary.	Encourage creative usage of NAO robots in education, develop standard models, and fortify teamwork
A Task-Based Framework for Natural Language Understanding in Human–Robot Interaction	Li, Z., Mu, Y., Sun, Z., Song, S., Su, J., & Zhang, J.	2021	Proposed a task-based framework combining language semantics, visual recognition, and similarity computation to infer human intentions from natural language.	1) Ambiguity in language. 2) Making connections between visual and verbal data. 3) Limited relevance.	Expand multimodal systems, improve sentence learning, and enhance real-world adaptability.
A Capabilities-Based Bi-Directional Multi-Task Trust Model for Human–Robot Teams	Azevedo-Sa, H., Yang, X. J., Robert, L. P., & Tilbury, D. M.	2021	Developed a bi-directional trust model using capability-based belief distributions validated with 284 participants and robotic simulations.	1) Modelling mutual trust. 2) Measuring real-time capabilities. 3) Limited scalability.	Integrate AI-driven monitoring, enhance adaptive trust, and use in real-world teams.
Advancements in Natural Language Processing for Human–Robot Interaction	Wang, X.	2024	Analyzed NLP integration in robotics focusing on advancements that improve contextual adaptation and communication.	1) Ambiguities in language2) Processing in many languages. 3) Moral considerations.	Encourage multidisciplinary research, create context-aware natural language processing, and improve multilingual flexibility.
Enhancing Human–Robot Collaboration with Large Language Models	Gao, F., Xia, L., Zhang, J., Liu, S., Wang, L., & Gao, R. X.	2024	Fine-tuned GPT-3.5 LLM to interpret natural instructions and generate task actions, tested in collaborative assembly tasks.	1) Misunderstanding human instructions. 2) Translating words into deeds. 3) The difficulties of real-time integration.	Expand to multi-robot systems, enhance contextual knowledge, and develop adaptive LLMs.
Enhancing Verbal and Visual Communication in Service Robots for Improved Human–Robot Interaction	Yura, J., Byambasuren, B., & Kim, D.	2025	Proposed a multimodal system combining sound localization and image processing to improve verbal and visual communication in service robots.	1) Synchronization of audio and video. 2) Interference and noise. 3) Limited flexibility.	Improve real-time data fusion, enhance noise resilience, and test in complex environments.
An Integrated Computer Vision and Natural Language Processing Application for Assisting the Vision-Impaired	El-Komy, A., Shahin, O. R., Abd El-Aziz, R. M., & Taloba, A. I.	2022	Developed an integrated system using a smartphone with vision and NLP capabilities linked to an F-RCNN server for object detection and speech guidance for the visually impaired.	1) Latency in the network. 2) Accuracy of detection in low light. 3) The necessity of customization.	Expand the use of assistive robotics, increase detection accuracy, and improve on-device AI.
Robots That Use Language: A Survey	Tellex, Gopalan, Kress-Gazit, Matuszek	2020	Survey of symbolic, probabilistic, and learning-based language grounding approaches for robots	Grounding abstract linguistic concepts into perceptual experience; handling ambiguity in dynamic environments	Integrating LLMs and multimodal grounding with perception-action loops for adaptive real-world understanding
Few-Shot Object Grounding and Mapping for Natural-Language Robot Instruction Following	Blukis, Knepper, Artzi	2021	Few-shot exemplar-based grounding and spatial mapping for unseen object instructions	Limited generalization to complex, multi-step or abstract commands	Expand exemplar datasets, integrate multimodal cues, and apply to real-world dynamic scenes
Language-Conditioned Imitation Learning for Robotic Manipulation	Stepputtis et al.	2020	End-to-end imitation learning integrating visual and language signals for robot control	High data requirements; difficulty with long-horizon tasks	Data-efficient imitation and reinforcement learning with synthetic and multimodal augmentation
Do As I Can, Not As I Say: Grounding Language in Robotic Affordances (SayCan)	Ichter et al.	2023	Combining LLM (PaLM) reasoning with affordance model scoring for task feasibility	Dependency on predefined skill libraries; limited scalability across robot platforms	Develop adaptive, self-updating affordance models and integrate vision-language-action learning in real time
Large Language Models for Robotics: A Review	Wang	2024	Review of LLM applications for instruction parsing, dialogue, and planning in embodied agents	Latency, safety, and real-time processing constraints	On-device and edge-optimized LLMs with secure reasoning and self-verification for HRI
A Survey of Robotic Language Grounding	Cohen, Liu, Mooney, Tellex, Watkins	2024	Systematic comparison of symbolic, embedding, and hybrid grounding methods	Trade-off between interpretability and scalability; data scarcity for multimodal benchmarks	Hybrid symbolic–neural systems with explainability and multimodal training datasets
Exploring LLM-powered Multi-Session Human–Robot Interaction	Mauliana et al.	2025	Session-aware conversational framework enabling personalized dialogue continuity	Memory drift, privacy, and inconsistency across sessions	Memory-safe adaptive dialogue systems with retrieval-augmented grounding and ethical data handling
A Comprehensive Review of Multimodal Emotion Recognition for HRI	Wu et al.	2025	Transformer-based multimodal fusion of text, speech, and visual cues	Synchronization, real-time inference, and cultural generalization	Real-time multimodal emotion recognition integrated into embodied dialogue and adaptive robot response systems

2. Research Methodology

3.1 Research Approach

Examining the application of Natural Language Processing in Human-Robot Interaction, our study opted for a qualitative and systematic literature review to see how robots can be taught to understand and articulate human language. We took a good hard look at the state of the field from 2020 through 2025 and found the most commonly used methods, models and tools. We also delved into the problems and research areas that are specific to NLP in HRI, such as understanding emotions, multimodal communication and language grounding.

This review is more or less a review of the breakthroughs in LLMs, Multimodal systems and the consideration of ethics in the communication of robots.

3.2 Data Collection Method

This research will be based on secondary data from peer-reviewed journal articles, conference papers, and academic surveys related to NLP and HRI. Sources utilized for this study include Google Scholar, IEEE Xplore, Scopus, and Web of Science. To be included in this review, studies needed to meet the following criteria: Released between 2020 and 2025.Employing natural language processing in human–robot interaction. Approaches involving language grounding, multimodal learning, emotion recognition, imitation learning, or LLM-based communication could be discussed. Challenges can range from trust, ethics, data scarcity, multilingual communication to real-time processing. Present models or frameworks applied to real-world or simulated robotic environments.

Key references include Mazzei et al. (2021), Younis et al. (2023), Gao et al. (2024), Wang (2024), and Mauliana et al. (2025), which deal with NLP techniques, human–robot collaboration, and integration of LLMs in robotic systems.

3.3 Data Analysis Method

Data analysis was done using a thematic analysis approach. The literature was further categorized into main themes as follows:

- NLP techniques in HRI include Semantic parsing, multimodal fusion, and transformer-based models like GPT, PaLM, and BERT. The section on Applications and Models discusses how LLMs work in Dialogue, Grounding, Task Learning, and Emotion Recognition for robots.
- There are four major challenges faced by the system: ambiguous language understanding, restricted real-time processing, ethical and trust problems, and difficult integration procedures.
- Research Gaps and Future Directions: Explainable hybrid model developments, context-aware LLMs, and multilingual emotion-aware robots. Themes have been compared across studies for identification of trends, methodological differences, and areas that share a common research limitation. The review also explored how NLP models improve robot perception, communication, and collaboration.

3.4 Validity and Reliability of the Research

To ensure validity and reliability, the following steps were applied:

1. Findings will be cross-checked across various trustworthy sources and over years from 2020 to 2025.
   This covers studies from diverse domains of education, industry, health, and service robots for an extensive view.
2. Comparation several methodologies and outcomes to recognize consistent contradictions and patterns.
3. Utilizing only peer-reviewed and high-impact publications for data credibility and relevance.
   It ensures that any conclusion resulting is comprehensive, reflective, and trustworthy of the current state of NLP research in HRI over this systematic and comparative approach.

4. Results and Discussion

The surveyed literature between 2020 and 2025 verifies as a whole that natural language processing (NLP) has come to play a crucial role as a major facilitator of intelligent and communicative robotic systems. The trends indicated by the literature show a marked shift from symbolic or probabilistic approaches to language grounding, moving toward neural approaches, with the focus now being on the Transformer or large language models as a leading approach. It is indicated by various research studies, including Stepputtis et al., 2020, and Ichter et al., 2023, that a blending approach involving linguistics and perceptions, end-to-end learning, and affordance-based planning has resulted in a substantial improvement in contextual reasoning and execution, whereby a success rate of over 80% has been achieved by various benchmarks. Simultaneously, various frameworks related to multimodal and affective approaches to natural language processing, including Mauliana et al., 2025, and Wu et al., 2025, have resulted in human-robot interaction systems entering a new realm, which is related to empathetic communication between humans and robots, as indicated by instruction-following communication moving toward empathetic communication involving text, images, and speech signals. However, developments are still limited by various challenges, including data requirements, computational complexities, latencies involved with real-time processing, as well as a lack of robustness related to multilingual and accent variability, as stated by Zhang et al., 2022. Further, various ethics, as indicated by concerns related to transparency, equity, and privacy, are increasingly becoming a concern as human-robot interaction engages with each other in a sensitive arena, as stated by Floridi & Cowls, 2022. However, emerging developments now appear to be moving toward a hybrid approach related to symbolic and neural models, as well as optimization approaches related to devices as part of the optimization approach, as indicated by various requirements related to justice, equity, and equality, as stated by various developments, including Romanishyn & Concilio, 2014, and Arvai

5. Conclusion

The development process associated with natural language processing (NLP) related to human-robot interaction (HRI) has developed significantly from simple symbolic representations to complex, advanced, and multimodal reasoning systems that allow robots to interact naturally and intelligently with humans. A literature study basing the search period between 2020 and 2025 reveals that this development has been driven by three major areas: the development of transformation-based architectures that enhance contextual understanding, embodied reasoning, which connects language, action, and perception, and multimodal fusion, which fuses linguistic, visual, and affective signals to build better emotional interactions between robots and humans. However, amidst all this development, several challenges are still present, including ambiguity while operating in dynamic environments, lack of research data, a high computational complexity level, and concerns relating to ethics, among others. The solution to all this will require hybrid approaches to reasoning, focusing both on symbolic representations, which are more interpretable, and neural networks, which are adaptability-focused, as well as building advanced, edge-optimized NLP models that will be ideal low-latency, on-the-edge applications, among other strategies. Additionally, future studies will require a focus on variously speaking, accents, and adapting technology to enhance a community-based approach to communication, as well as building ethics around accountability related to human and robot collaboration, among other strategies, as a critical component towards building successful human-robot collaboration founded by advanced, emerging, and advanced AI-based NLP technology, leading towards the development of fully intelligent social robots that will interact meaningfully and effectively with all the people they come into contact with, as well as making all this a reality by adapting as a realistic technology founded by advanced AI technology-based natural language processes.

Acknowledgment: The research leading to these results has received no Research Grant Funding.

Author contribution: All authors have contributed, read, and agreed to the published version of the manuscript results.

Conflict of interest: The authors declare no conflict of interest.

References:

[1] Mazzei, D., Chiarello, F., & Fantoni, G. (2021). Analyzing social robotics research with natural language processing techniques. Cognitive Computation. https://doi.org/10.1007/s12559-020-09799-1

[2] Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023). A systematic literature review on the applications of robots and natural language processing in education. Electronics, 12, 2864. https://doi.org/10.3390/electronics12132864

[3] Gao, F., Xia, L., Zhang, J., Liu, S., Wang, L., & Gao, R. X. (2024). Integrating large language model for natural language-based instruction toward robust human-robot collaboration. Procedia CIRP, 130, 313–318.

[4] Wang, X. (2024). Advancing human-robot interaction: The role of natural language processing in robotic systems. In Y. Wang (Ed.), Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024), Advances in Intelligent Systems Research (Vol. 185)

[5] Li, Z., Mu, Y., Sun, Z., Song, S., Su, J., & Zhang, J. (2021). Intention understanding in human–robot interaction based on visual-NLP semantics. Frontiers in Neurorobotics, 14, 610139. https://doi.org/10.3389/fnbot.2020.610139

[6] Azevedo-Sa, H., Yang, X. J., Robert, L. P., & Tilbury, D. M. (2021). A unified bi-directional model for natural and artificial trust in human–robot collaboration. IEEE Robotics and Automation Letters, 6(3), 5913–5920.

https://doi.org/10.1109/LRA.2021.3088082
[7] Yura, J., Byambasuren, B., & Kim, D. (2025). Multi-modal, multi-step human–robot interaction method for natural interaction of service robots. Intelligent Service Robotics, 18, 389–401. https://doi.org/10.1007/s11370-025-00604-2
[8] El-Komy, A., Shahin, O. R., Abd El-Aziz, R. M., & Taloba, A. I. (2022). Integration of computer vision and natural language processing in multimedia robotics application. Information Sciences Letters, 11(3), 705–716. https://doi.org/10.18576/isl/110309
[9] Tellex, S., Gopalan, N., Kress-Gazit, H., & Matuszek, C. (2020). Robots that use language. Annual Review of Control, Robotics, and Autonomous Systems, 3(1), 25-55.
[10] Blukis, V., Knepper, R. A., & Artzi, Y. (2020). Few-shot object grounding and mapping for natural language robot instruction following. arXiv preprint arXiv:2011.07384.
[11] Stepputtis, S., Campbell, J., Phielipp, M., Lee, S., Baral, C., & Ben Amor, H. (2020). Language-conditioned imitation learning for robot manipulation tasks. Advances in Neural Information Processing Systems, 33, 13139-13150.
[12] Ichter, B., et al. (2023). SayCan: Grounding Language in Robot Capabilities.
[13] Wang, J., Shi, E., Hu, H., Ma, C., Liu, Y., Wang, X., ... & Zhang, S. (2024). Large language models for robotics: Opportunities, challenges, and perspectives. Journal of Automation and Intelligence.
[14] Cohen, V., Liu, J. X., Mooney, R., Tellex, S., & Watkins, D. (2024). A survey of robotic language grounding: Tradeoffs between symbols and embeddings. arXiv preprint arXiv:2405.13245.
[15] Mauliana, M., Ashok, A., Czernochowski, D., & Berns, K. (2025). Exploring LLM-powered multi-session human-robot interactions with university students. Frontiers in Robotics and AI, 12, 1585589.
[16] Wu, Y., Mi, Q., & Gao, T. (2025). A comprehensive review of multimodal emotion recognition: Techniques, challenges, and future directions. Biomimetics, 10(7), 418.
[17] Bansal, P., & Aggarwal, N. (2021). A survey on natural language processing in human–robot interaction. Cognitive Systems Research, 68, 60–76. https://doi.org/10.1016/j.cogsys.2021.02.005
[18] Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
[19] Ahmed, F. Y., Abualkishik, A. M., Abdalgader, K., Al Risi, A., & Al Maawali, R. (2024). Simulating the Influence of Social Robotics Technology on the Student's Academic Performance. International Journal of Advances in Soft Computing & Its Applications, 16(1).
[20] Cohen, V., Liu, Y., Mooney, R., Tellex, S., & Watkins, D. (2024). A Survey of Robotic Language Grounding: Tradeoffs between Symbols and Embeddings. Proceedings of IJCAI.
[21] Dautenhahn, K. (2021). Socially intelligent robots: Dimensions of human–robot interaction. Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1831), 20190635.
[22] Floridi, L., & Cowls, J. (2022). A unified framework for AI ethics and governance. Nature Machine Intelligence, 4(3), 210–219.
[23] Huang, W., et al. (2023). Inner Monologue: Embodied Reasoning through Planning with Language Models. IEEE Robotics and Automation Letters, 8(4), 2231–2242.
[24] Ichter, B., et al. (2023). SayCan: Grounding Language in Robot Capabilities. Proceedings of Robotics: Science and Systems (RSS).
[25] Khandelwal, P., et al. (2022). The role of context in human–robot communication: A deep learning perspective. Robotics and Autonomous Systems, 157, 104211.
[26] Li, J., et al. (2022). Multimodal Transformers for Emotion Recognition in Human–Robot Communication. IEEE Transactions on Affective Computing, 13(2), 841–852.
[27] McDuff, D., et al. (2023). Affective computing in the wild: The evolution of emotion-aware systems. IEEE Transactions on Affective Computing, 14(1), 1–15.
[28] Min, B., Ross, H., Sulem, E., et al. (2021). Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. arXiv:2110.08147.
[29] Tellex, S., Gopalan, N., Kress-Gazit, H., & Matuszek, C. (2020). Robots that use language: A survey. Robotics and Autonomous Systems, 133, 103659.
[30] Vaswani, A., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NeurIPS).

[31] Wang, X. (2024). Advancing Human–Robot Interaction: The Role of Natural Language Processing in Robotic Systems. Advances in Intelligent Systems Research, 185.
[32] AL-MAAWALI, R. R. M., AL-RAYASI, A. S. A., & FALAH, D. The Use of Social Robots Technology in the Educational Delivery Style for Academic Students.
[33] Wu, Y., et al. (2025). A Comprehensive Review of Multimodal Emotion Recognition for HRI. Sensors, 25(7), 3409.
[34] Zhang, R., et al. (2022). Multilingual and Code-Switching Challenges in Human–Robot Communication. IEEE Transactions on Cognitive and Developmental Systems, 14(4), 895–906.
[35] Sharma, A., et al. (2022). Natural Language Understanding in Robotics: Challenges and Directions. Frontiers in Robotics and AI, 9, 912564.
[36] Yousif, J.H., (2013). Natural language processing based soft computing techniques. International Journal of Computer Applications, 77(8), pp.43-49.
[37] Yousif, J.H., (2021). Social and Telepresence Robots a future of teaching. Artificial Intelligence & Robotics Development Journal, 1(1), 58-65. https://doi.org/10.52098/airdj.202124
[38] Yousif, J. H., & Saini, D. K. (2020). Big data analysis on smart tools and techniques. In Cyber Defense Mechanisms (pp. 111-130). CRC Press.