ijecs open access www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 14 Issue 06 June 2025, Page No. 27278-27285
ISSN: 2319-7242 DOI: 10.18535/ijecs/v14i06.5150

AI-Driven Socially Assistive Robots with Turkish Language Support for Smart Environments

Murat ŞİMŞEK[a], Demiral AKBAR[b,*]

[a] OSTİM Technical University, Department of Artificial Intelligence Engineering, 06374, OSTİM, Ankara, Turkey.

[b] OSTİM Technical University, Department of Mechanical Engineering, 06374, OSTİM, Ankara, Turkey.

Abstract:
This study presents the development of a Turkish Natural Language Processing (NLP) module for Socially Assistive Robots (SARs), aiming to enhance their integration into public spaces and smart environments. The rising labor costs and limitations of human staffing in developed regions, including Turkey, underscore the need for localized language capabilities in SARs. Leveraging advanced machine learning techniques, the system fine-tunes a BERT-based model specifically for Turkish language understanding and response generation. Trained on a dataset of 61,293 question–answer pairs, the system achieved 95% voice recognition accuracy, 82% response accuracy, and a 92% speech comprehensibility rating among native Turkish speakers. It also reached a 72% F1-score and a 72% Exact Match (EM) score, confirming its ability to generate contextually appropriate responses. These results highlight the system's effectiveness in enabling natural and intuitive interactions. The developed module demonstrates significant potential for deployment in public domains such as universities, hospitals, airports, and shopping centers, where it can assist users in navigating environments and accessing essential services. Furthermore, the research showcases educational benefits through the Temi OS platform, offering hands-on learning opportunities in robotics, AI, and natural language processing. Competitions based on SAR development further encourage innovation among students and researchers. In conclusion, this work illustrates the critical role of Turkish-language SARs in improving service accessibility and advancing AI education. Future efforts will focus on refining linguistic capabilities, expanding interactive features, and conducting long-term user engagement studies to maximize societal impact.

Keywords: Turkish Natural Language Processing, Socially Assistive Robots (SARs), Human-Robot Interaction, BERT Fine-Tuning, Conversational AI, Voice Recognition, Smart Environments, Multilingual Robotics

Introduction

The integration of smart home technology and robotics into daily life is rapidly increasing. Socially assistive robots (SARs), designed to interact with humans, are emerging as a crucial component of this transformation. Despite significant advancements in mobile service robots, research on their social interactions and seamless integration into smart environments remains limited. Initially, robots were introduced into homes for repetitive, non-interactive tasks such as cleaning [1], surveillance [2], and cooking. More recently, companion robots like Astra Amazon, Alexa on Wheels, Jibo [3], and Temi [4] have expanded their roles, offering support in entertainment, comfort, and companionship. Garcia-Haro [5] classified service robots into ten main categories, recognizing SARs as an emerging field in the robotics market. SARs, whether physical or digital, are designed to interact socially with humans, providing companionship and assistance through human-like interactions [6]. Studies indicate that embodied robots, as opposed to disembodied speech agents, enhance user engagement, confidence, and enjoyment [7][8]. Their effectiveness depends on both utilitarian (task functionality and perceived usefulness) and hedonic (user experience and enjoyment) factors [9]. However, cost, reliability, and limited functionality have hindered widespread adoption. Many companies and researchers prioritize technology-centric approaches, often neglecting the user experience. To increase adoption, there is a growing need for user-centered design methodologies that integrate both experiential and functional elements.

A notable example of SARs is Softbank Robotics' Pepper, a humanoid robot designed for social interaction. While it has limited grasping capabilities, its ability to communicate via natural language and a touchscreen makes it suitable for use in shopping malls, elderly care facilities, and homes [10]. Similarly, robots like Temi have been deployed in educational settings, such as the Center for Independent Language Learning (CILL) at Hong Kong Polytechnic University [11]. Temi's voice-controlled navigation enhances accessibility for users unfamiliar with digital devices.

Advancements in Large Language Models (LLMs) have significantly improved natural language processing (NLP), especially in question-answering (QA) systems. LLMs such as BERT, GPT, and their successors have revolutionized text comprehension, surpassing traditional rule-based methods. GPT models, particularly GPT-3, generate coherent and contextually relevant responses, making them suitable for open-domain QA tasks. For instance, studies have evaluated GPT-3.5's performance on health-related QA tasks, revealing both potential and limitations in evidence-based reasoning [12]. Similarly, BERT's bidirectional context understanding has proven instrumental in processing low-resource languages, such as in Kazakh language QA systems [13]. To enhance LLM-driven QA systems, researchers employ several methodologies. Fine-tuning pre-trained models on domain-specific datasets remains a common approach. Transfer learning enables adaptation to specialized QA tasks with minimal data, as seen in ESG report analysis [14]. Additionally, prompt engineering techniques help mitigate biases and enhance precision in GPT models [12]. Research also explores knowledge distillation to create smaller, more efficient models that reduce computational costs while maintaining performance.

While significant progress has been made in NLP applications for SARs, most existing research focuses on widely spoken languages, with limited efforts directed toward Turkish. Given the increasing role of robots in human-robot interaction (HRI), it is crucial to develop Turkish-specific NLP models that enhance SARs' ability to engage in natural conversations. Addressing this gap will facilitate the development of more effective, culturally adaptive, and linguistically robust SARs for Turkish-speaking users.

Materials and Methodologies

We employed the Transformer architecture, a novel neural network design introduced by Vaswani et al. [15], which relies on self-attention mechanisms to process sequences of data. Unlike traditional sequential models such as recurrent neural networks (RNNs), Transformers process data in parallel, significantly reducing training time and improving the handling of long-distance dependencies. The architecture of the Transformer model is illustrated in Figure 1.

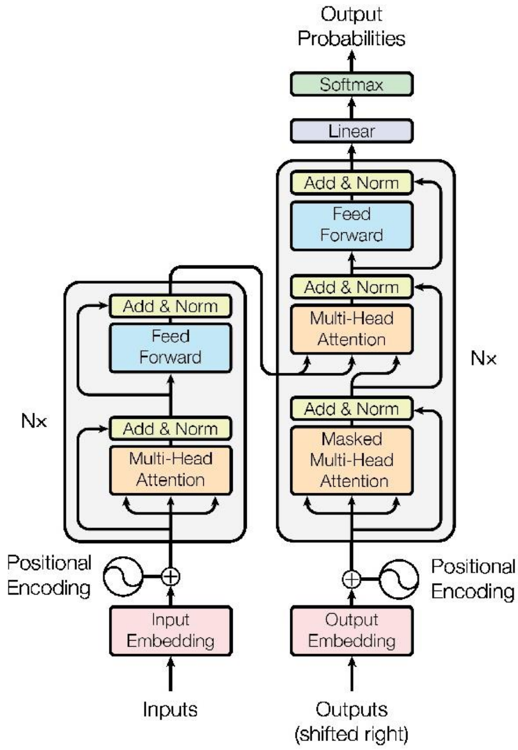

Figure 1: The Transformer - model architecture [12].

The Transformer model, introduced by Vaswani et al. in their seminal 2017 paper *"Attention is All You Need,"* represents a major breakthrough in natural language processing (NLP). Unlike conventional sequential models, it employs a fully attention-based mechanism, enhancing its efficiency in handling complex language tasks. The core of this architecture is the multi-head attention mechanism, which allows the model to simultaneously consider all preceding and succeeding words in a sentence. This parallel processing capability leads to faster training and improved performance compared to previous models.

The Transformer's architecture consists of multiple layers, each repeated several times. These layers include Masked Multi-Head Attention, Add & Norm, and Feed Forward, followed by another Add & Norm step. The input layer consists of Input Embedding, which converts each token in the input sequence into a fixed-size vector, accompanied by Positional Encoding to retain sequence order. The output layer follows a similar structure, incorporating Output Embedding and Positional Encoding. The final layer applies a linear transformation and a softmax function, producing a probability distribution that enables the model to learn contextual relationships within and across languages. This architecture provides a strong foundation for a wide range of NLP tasks, including machine translation and text summarization.

To develop the Turkish Question Answering (QA) model in this study, we applied fine-tuning techniques to a pre-trained BERT model. Specifically, the "dbmdz/bert-base-turkish-uncased" model, available on HuggingFace, was selected due to its suitability for Turkish language processing tasks. The fine-tuning process utilized the Squad-tr dataset [16], which is derived from the English SQuAD2.0 corpus and translated into Turkish using Amazon Translate. The dataset comprises 61,293 question-answer pairs across 18,776 paragraphs.

The fine-tuning process was conducted using optimized hyperparameters, including a learning rate of 3e-5, a batch size of 12 per GPU, a maximum sequence length of 384, a document stride of 128, and a total of three training epochs. Following training, the model's performance was evaluated using exact match (EM) and F1 scores. The overall EM score achieved was 10.60, while the F1 score was 15.27, evaluated across 8,291 instances. For instances with an answer present, the model obtained an EM score of 37.43 and an F1 score of 53.92, based on 2,346 cases. In contrast, for no-answer cases, the EM and F1 scores were both 0.02, calculated over 5,945 instances. The best exact match and F1 scores at the optimal threshold reached 71.73, highlighting the model's ability to perform well under specific configurations. These results indicate that while the model demonstrates robust performance in answer-present cases, further improvements may be necessary to enhance its accuracy for no-answer scenarios.

The fine-tuned model was integrated into a socially assistive robot to facilitate interaction with Turkish-speaking users. The integration was conducted using the Temi OS infrastructure, enabling the robot to process user queries and provide appropriate responses. Communication between the robot's sensors and the developed NLP module was managed through the SDK provided by Temi OS, utilizing XML for system operations. The development environment offered by Temi OS played a crucial role in ensuring seamless integration and enhancing the functionality of the Turkish NLP module within the robot's operations. This work represents a significant advancement in the field of human-robot interaction (HRI), particularly for Turkish-speaking environments. Further details on the commercially available Temi robot can be found in [17].

Figure 2: Temi Robot [17]

Results and Discussions

This study presents a Python-based system utilizing the Streamlit framework to facilitate a voice-interactive question-answering module[18]. At its core, the system employs a pre-trained large language model fine-tuned for Turkish language comprehension and response generation. The workflow captures user audio input, converts it into text, processes the text using the language model, and then converts the model's response back into speech, enabling seamless interaction. The application, named "OSTIMTECH TEMI," integrates multiple technologies, including Streamlit for the web interface, the Transformers library for the language model, and gTTS and pygame for text-to-speech functionality. Users interact with the system through a button labeled "BANA SORU SOR" (Ask me a question), which, when clicked, records the user's verbal query. The record_audio function captures user input in stereo at a 44,100 Hz sample rate, ensuring high-fidelity audio. The recorded audio is converted to text using a custom speech-to-text function, and the Turkish BERT QA Transformer-based model processes this text within a predefined context from a file (ostim.txt), generating a relevant answer. The temi_main function orchestrates the entire pipeline by processing the query, retrieving context, and generating an answer. This response is then converted into speech using the text to Speech function and played using the pygame library. The integration of the Turkish NLP module within the Temi robot represents a significant advancement in human-robot interaction for Turkish-speaking environments[19].

The Turkish BERT model configuration was optimized for performance. The attention probabilities dropout probability was set at 0.1, reducing overfitting by randomly deactivating 10% of attention heads during training. The model employs the Gaussian Error Linear Unit (GELU) activation function, offering a smoother learning curve than traditional ReLU activation. Additional hyperparameters were fine-tuned to enhance efficiency, including a hidden dropout probability of 0.1, hidden size of 768, an initializer range of 0.02, and an intermediate layer size of 3072 for capturing complex patterns effectively. Other critical configurations included a sequence length of 512 tokens, 12 attention heads, 12 hidden layers for hierarchical learning, absolute position embeddings, and 32-bit floating-point precision. The vocabulary size was set at 32,000 to optimize Turkish language processing, and inference caching was enabled to improve response time. These parameters were meticulously tuned to enhance the model's effectiveness within a socially assistive robot framework[20].

Performance evaluation employed multiple metrics, including voice recognition accuracy, model response accuracy, and speech comprehensibility. Voice recognition accuracy was assessed as the percentage of correctly transcribed words from spoken input, while model response accuracy was evaluated against expert-verified answers for relevance and correctness. Speech comprehensibility was measured through user surveys, with participants rating clarity and understandability on a scale of 1 to 10. The experimental results demonstrated the robustness of the system: speech-to-text accuracy reached 95%, model response accuracy was 72%, and text-to-speech comprehensibility scored 9.2/10 as rated by native Turkish speakers. These results highlight the system's high reliability in capturing user speech, effective question-answering capabilities, and excellent speech clarity[21]. In conclusion, this study presents a novel framework for integrating a Turkish-based natural language processing module into a socially assistive robot. The use of advanced deep learning techniques, specifically the Turkish BERT QA model, enables robust language understanding and response generation. The system's performance evaluation demonstrates its potential for practical applications in Turkish-speaking environments, paving the way for enhanced human-robot interaction. The integration of this advanced Turkish NLP module within the Temi robot showcases the potential for socially assistive robots to bridge the language barrier and engage with Turkish-speaking communities.

The system was further integrated with additional functionalities, including facial recognition for guest interaction. Recognized guests are greeted by name, while unrecognized individuals trigger a predefined action sequence. The Android SDK was utilized to program the robot's responses to commands such as "Turn right," "Turn left," "Move forward," and "Move backward." Additionally, the previously inactive English NLP system in Temi was activated and made available for university use. Self-training capabilities were introduced, allowing Temi to refine its responses over time. Furthermore, a Turkish Question-Answering NLP Model is under development to enhance Temi's ability to process and respond to Turkish queries accurately. This model integrates speech-to-text, text-to-speech, and Transformer-based algorithms to ensure a seamless interaction experience[22]. The NLP development consists of three core stages: speech-to-text processing using Python's speech_recognition library, Turkish NLP model development through fine-tuning a pre-trained BERT model with optimized hyperparameters such as batch size, learning rate, training epochs, and sequence length adjustments, and text-to-speech processing to convert responses back into natural-sounding Turkish speech[23]. A visual representation of the interface is provided in Figure-3(a) and (b), respectively.

(a) (b)

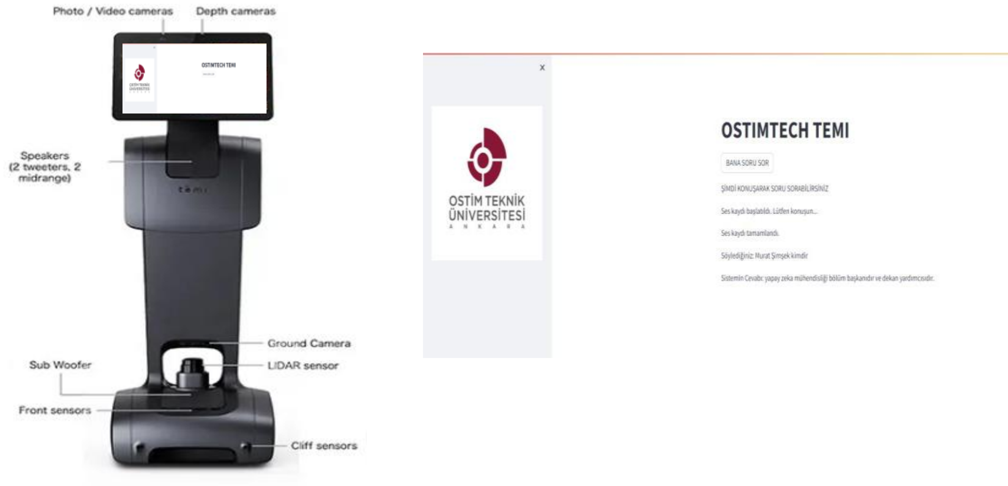

Figure 3: (a) System Interface, (b) The relevant visual representation.

This research paper presents the development and integration of a robust natural language processing module for the Temi socially assistive robot, with a specific focus on the Turkish language.

The results of this research underscore the effectiveness of integrating advanced AI and NLP technologies into interactive applications. The deployment of this system demonstrates the feasibility of AI-driven interfaces in customer service, education, and assistive technologies. Compared to previous sequential models, Transformer models exhibit superior performance, faster training times, and deeper contextual understanding. This work lays the foundation for future improvements, including real-time learning capabilities, multilingual support, and integration with additional AI-driven functionalities. The successful implementation of a Turkish Question-Answering NLP model within a socially assistive robotic framework marks a significant advancement in human-computer interaction, paving the way for broader AI applications.

Conclusion:

This study demonstrates the successful development and evaluation of Turkish-language support for mobile Socially Assistive Robots (SARs) deployed in public environments. By fine-tuning the BERT model for Turkish, we significantly enhanced the ability of SARs to engage in natural, context-aware conversations. The model, trained on over 61,000 question–answer pairs and 18,000+ paragraphs, achieved 95% voice recognition accuracy, 72% F1 and Exact Match (EM) scores, and a 92% text-to-speech comprehensibility rating—highlighting its effectiveness in real-world interactions with Turkish-speaking users. The integration of localized NLP modules addresses key limitations of traditional kiosk-based systems, such as language barriers and static functionality. Mobile SARs equipped with Turkish NLP capabilities can assist users in navigating complex public settings—like hospitals, airports, universities, and shopping centers—enhancing accessibility, trust, and user satisfaction. This approach also opens pathways for broader adoption of SARs in public services, contributing to operational efficiency and inclusive service design. Beyond public deployment, the project underscores the educational value of SAR platforms. Through the Temi OS development environment, universities can provide hands-on learning opportunities in AI, robotics, and natural language processing. Competitions and applied projects using SARs will empower students to innovate and gain practical skills for the AI-driven future.

In summary, this research offers a comprehensive framework for deploying Turkish-language SARs that serve both social and educational functions. The work contributes to bridging the technological gap in localized AI systems and sets a foundation for future enhancements. Next steps include long-term user interaction studies, deeper linguistic customization, and the development of advanced features to support more complex and adaptive human–robot interactions. With continued innovation, SARs can play a transformative role in improving public services, technological literacy, and quality of life in Turkey and beyond.

Acknowledgment
This research received internal funding and we would like thanks Scientific Research Programs at OSTİM Technical University, Project code: BAP202327.

Link to supporting data:
https://github.com/muratsimsek003/ostimTemiTurkishNLP

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

X The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

References

1. T. Seo, Y. Jeon, C. Park, and J. Kim, "Survey on Glass And Façade-Cleaning Robots: Climbing Mechanisms, Cleaning Methods, and Applications," *Int. J. Precis. Eng. Manuf. Technol.*, vol. 6, no. 2, pp. 367–376, Apr. 2019, doi: 10.1007/s40684-019-00079-4.
2. A. Nayyar, V. Puri, N. G. Nguyen, and D. N. Le, "Smart Surveillance Robot for Real-Time Monitoring and Control System in Environment and Industrial Applications," 2018, pp. 229–243.
3. C. Breazeal, "Social Robots," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Mar. 2017, pp. 1–1, doi: 10.1145/2909824.3020258.
4. C.-F. Hung, Y. Lin, H.-J. Ciou, W.-Y. Wang, and H.-H. Chiang, "FoodTemi: The AI-Oriented Catering Service Robot," in *2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, Sep. 2021, pp. 1–2, doi: 10.1109/ICCE-TW52618.2021.9603096.
5. J. M. Garcia-Haro, E. D. Oña, J. Hernandez-Vicen, S. Martinez, and C. Balaguer, "Service Robots in Catering Applications: A Review and Future Challenges," *Electronics*, vol. 10, no. 1, p. 47, Dec. 2020, doi: 10.3390/electronics10010047.
6. E. Toscano, M. Spitale, and F. Garzotto, "Socially Assistive Robots in Smart Homes: Design Factors that Influence the User Perception," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022, pp. 1075–1079, doi: 10.1109/HRI53351.2022.9889467.
7. K. Maj and P. Zarzycki, "Meeting with social robots like the cat-cucumber meeting? An integrated model of human-robot first contact. Psychological perspective.," *Paladyn, J. Behav. Robot.*, vol. 10, no. 1, pp. 454–465, Jan. 2019, doi: 10.1515/pjbr-2019-0026.
8. M. Spitale, S. Silleresi, G. Cosentino, F. Panzeri, and F. Garzotto, "'Whom would you like to talk with?,'" in *Proceedings of the Interaction Design and Children Conference*, Jun. 2020, pp. 262–272, doi: 10.1145/3392063.3394421.
9. van der Heijden, "User Acceptance of Hedonic Information Systems," *MIS Q.*, vol. 28, no. 4, p. 695, 2004, doi: 10.2307/25148660.
10. M. Niemelä, P. Heikkilä, and H. Lammi, "A Social Service Robot in a Shopping Mall," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Mar. 2017, pp. 227–228, doi: 10.1145/3029798.3038301.

11. F. Har and B. W. L. Ma, “The Future of Education Utilizing an Artificial Intelligence Robot in the Centre for Independent Language Learning: Teacher Perceptions of the Robot as a Service,” 2023, pp. 49–64.
12. Patel, A. V., Jasani, S., AlAshqar, A., Panakam, A., Amin, K., & Sheth, S. S. (2024). Evaluating the Accuracy and Utility of Large Language Models in Answering Common Contraception Questions [ID 2683633]. *Obstetrics & Gynecology, 143*, 12S-12S
13. Mukanova, A., Barlybayev, A., Nazyrova, A., Kussepova, L., Matkarimov, B., & Abdikalyk, G. (2024). Development of a Geographical Question- Answering System in the Kazakh Language. *IEEE Access, 12*, 105460-105469.
14. Parikh, P., & Penfield, J. (2024). Automatic Question Answering From Large ESG Reports. *International Journal of Data Warehousing and Mining, None*.
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
16. Budur, Emrah, et al. "Building Efficient and Effective OpenQA Systems for Low-Resource Languages." *arXiv preprint arXiv:2401.03590* (2024).
17. C. Wang et al., “LaMI: Large Language Models for Multi-Modal Human-Robot Interaction,” in Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, in CHI EA ’24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–10. doi: 10.1145/3613905.3651029.
18. J. R. Cole, M. J. Q. Zhang, D. Gillick, J. M. Eisenschlos, B. Dhingra, and J. Eisenstein, “Selectively Answering Ambiguous Questions,” Nov. 15, 2023, arXiv: arXiv:2305.14613. doi: 10.48550/arXiv.2305.14613.
19. Z. Ren, N. Yolwas, H. Wang, and W. Slamu, “Exploring Turkish Speech Recognition via Hybrid CTC/Attention Architecture and Multi-feature Fusion Network,” Mar. 22, 2023, arXiv: arXiv:2303.12300. doi: 10.48550/arXiv.2303.12300.
20. F. C. Akyon, D. Cavusoglu, C. Cengiz, S. O. Altinuc, and A. Temizel, “Automated question generation and question answering from Turkish texts,” Apr. 07, 2022, arXiv: arXiv:2111.06476. doi: 10.48550/arXiv.2111.06476.
21. S. Yildirim, “Fine-tuning Transformer-based Encoder for Turkish Language Understanding Tasks,” Jan. 30, 2024, arXiv: arXiv:2401.17396. doi: 10.48550/arXiv.2401.17396.
22. “Full article: Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): an empirical case study for Turkish.” Accessed: Feb. 05, 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00051144.2021.1922150