ACCSCIENCE PUBLISHING

International Journal of Systematic Innovation

ARTICLE

Natural language processing-based development of artificial intelligence-driven autonomous socially assistive robots

Murat Şimşek[1], and **Demiral Akbar**[2*]

[1]Department of Artificial Intelligence Engineering, Faculty of Engineering, OSTİM Technical University, Ankara, Turkey

[2]Department of Mechanical Engineering, Faculty of Engineering, OSTİM Technical University, Ankara, Turkey

***Corresponding author:**
Demiral Akbar
(demiral.akbar@ostimteknik.edu.tr)

Abstract

This study focuses on addressing the growing need for localized language support in socially assistive robots (SARs) due to rising labor costs and the limitations of human labor in developed countries. The research aims to develop a Turkish natural language processing (NLP) module to enhance SARs' social interaction capabilities and integration into smart living spaces. By leveraging advanced machine learning models, specifically XLM-RoBERTa Large (deepset/xlm-roberta-large-squad2), the study evaluated cross-lingual transfer learning for Turkish question answering, addressing specific linguistic challenges, including agglutinative morphology and vowel harmony. The model was evaluated on the Turkish Question-Answering Dataset (TQuAD 2.0) with 2,520 validation examples, achieving 79.37% F1-score and 56.67% exact match score. The research established a methodological framework connecting adaptive NLP design principles with control systems theory, demonstrating how concepts from adaptive fuzzy control and robust neural adaptive control inform the development of more stable and reliable NLP systems for SAR applications. These outcomes highlight the potential of cross-lingual NLP models for SAR applications in Turkish-speaking environments. The research contributes to the field by: (i) evaluating cross-lingual transfer learning for Turkish SAR applications, (ii) demonstrating the effectiveness of XLM-RoBERTa for low-resource language adaptation, (iii) establishing a framework that connects adaptive NLP design with control systems theory for enhanced robustness, and (iv) identifying real-world SAE applications in healthcare, smart homes, and industrial settings. Future work will focus on integrating this NLP module with speech recognition and synthesis components for complete voice-interactive SAR systems.

Keywords: Socially assistive robot; Natural language processing; Large language model; Question-answering; XLM-RoBERTa; Turkish Question-Answering Dataset; Cross-lingual transfer learning; Adaptive control systems

Citation: Şimşek, M. & Akbar, D. (2026). Natural language processing-based development of artificial intelligence-driven autonomous socially assistive robots. *Int J Systematic Innovation*. 10(1):1-10. https://doi.org/10.6977/IJoSI.202602_10(1).0001

Received: December 21, 2024

Revised: December 13, 2025

Accepted: January 5, 2026

Published online: February 13, 2026

1. Introduction

There is a growing interest in smart homes and robotic technology in various aspects of life. However, limited studies have investigated the social interactions of socially assistive robots (SARs) with humans and their integration into smart homes and living spaces. With recent advancements in robotics, mobile service robots are becoming a part of our daily lives, accompanying us in human-interactive environments, such as airports, hospitals, schools, and offices. Initially, robots entered living spaces with the capability to perform relatively limited and repetitive tasks without the need for human interaction, such as cleaning (Seo *et al*., 2019), surveillance (Nayyar *et al*., 2018), and cooking. Recently, companion robots, such as Astra Amazon, Alexa on wheels, Jibo (Breazeal, 2017), and Temi (Hung *et al*., 2021), have begun to spread in living spaces to assist with entertainment and comfort.

García-Haro *et al.* (2021) previously proposed 10 main classes of service robots, identifying SARs as an emerging area in the service robotics market. SARs are physical or digital entities capable of socially interacting and communicating with humans, designed to offer companionship and/or assistance through interactions that resemble human interactions (Toscano *et al*., 2022). Researchers have examined how interacting with an embodied robot, compared to disembodied speech agents, enhances users' confidence and enjoyment (Maj & Zarzycki, 2019; Spitale *et al*., 2020). The effectiveness of SARs depends on their ability to address both utilitarian and hedonic factors. According to Van der Heijden (2004), hedonic systems aim to provide self-fulfilling value to users, in contrast to utilitarian systems that aim to provide instrumental value.

1.1. Previous research in question-answering systems across languages

Question-answering (QA) systems have been developed for various languages using different approaches. For English, the Stanford Question-Answering Dataset (SQuAD) (Rajpurkar *et al*., 2016) established benchmarks that led to significant advances in reading comprehension models. The Bidirectional Encoder Representations from Transformers (BERT) model achieved state-of-the-art results on English QA tasks by leveraging bidirectional context understanding (Devlin *et al*., 2019). For Chinese, researchers developed the Chinese Machine Reading Comprehension (CMRC) dataset and fine-tuned multilingual models, achieving F1 scores above 80% (Cui *et al*., 2019). Arabic QA systems have been developed using similar transfer learning approaches, with AraBERT achieving competitive results despite the language's complex morphology (Alkesaiberi *et al*., 2024). The Kazakh language QA system demonstrated the applicability of BERT-based models for low-resource Turkic languages, achieving high Bilingual Evaluation Understudy (BLEU) scores (Mukanova *et al*., 2024).

Cross-lingual transfer learning has emerged as a powerful approach for low-resource languages. XLM-RoBERTa (Conneau *et al*., 2020), trained on 100 languages including Turkish, has shown remarkable zero-shot and fine-tuned performance across diverse linguistic tasks. Our approach leverages this cross-lingual capability by utilizing the deepset/xlm-roberta-large-squad2 model, which combines the multilingual representations of XLM-RoBERTa with task-specific fine-tuning on SQuAD 2.0. This enables effective transfer to Turkish QA without requiring extensive Turkish-specific pre-training.

Our approach follows similar methodologies to these previous works, specifically: (i) utilizing cross-lingual transfer learning from pre-trained multilingual language models, (ii) evaluating on native Turkish QA datasets (TQuAD 2.0), and (iii) adapting the models for the specific linguistic characteristics of the target language. Our work aims to integrate the natural language processing (NLP) module into a SAR platform for real-time voice-based interaction capabilities.

1.2. Linguistic characteristics of the Turkish language

Turkish presents unique challenges for NLP systems, distinguishing this research from work in other languages. Turkish is an agglutinative language, meaning that words are formed by adding suffixes to root words, creating complex word forms that can express entire sentences in other languages with a single word. For example, the word "*evlerinizden*" (from your houses) contains four morphemes: *ev* (house) + *ler* (plural) + *iniz* (your) + *den* (from). This characteristic significantly increases vocabulary size and data sparsity challenges (Eryiğit, 2014).

Additionally, Turkish exhibits vowel harmony, where suffixes must harmonize with the vowels in the root word according to front/back and rounded/unrounded distinctions. Turkish also has a relatively free word order (subject–object–verb as default but flexible), which affects how context must be processed in transformer models. These linguistic features require specialized tokenization strategies and model adaptations. The XLM-RoBERTa model we employed uses SentencePiece tokenization trained on multilingual data, effectively handling Turkish morphological complexity through subword segmentation (Conneau *et al*., 2020).

1.3. Research contributions

This research makes the following specific contributions to the field:

(i) We evaluate cross-lingual transfer learning using XLM-RoBERTa Large for Turkish QA in the context of SAR applications.
(ii) We demonstrate the effectiveness of the deepset/xlm-roberta-large-squad2 model on the TQuAD 2.0 benchmark, achieving 79.37% F1-score and 56.67% exact match in a zero-shot transfer learning setting without Turkish-specific fine-tuning.
(iii) We establish a methodological framework connecting adaptive NLP design principles with control systems theory, showing how concepts from adaptive fuzzy control (Boulkroune *et al.*, 2017), robust neural adaptive control (Hayakawa *et al.*, 2008), and adaptive backstepping control (Zouari *et al.*, 2013) can inform the development of more stable and reliable NLP systems for robotics applications.
(iv) We provide detailed error analysis and sample predictions to understand model behavior on Turkish linguistic patterns and identify real-world applications across healthcare, smart home, and industrial domains.

2. Materials and methods

We employed the Transformer architecture, a novel neural network design introduced by Vaswani *et al.* (2017), that relies on self-attention mechanisms to process sequences of data. Unlike traditional sequential models such as recurrent neural networks, Transformers process data in parallel, significantly reducing training times and improving the handling of long-distance dependencies.

2.1. XLM-RoBERTa Large model architecture

For our Turkish QA system, we employed the XLM-RoBERTa Large model (deepset/xlm-roberta-large-squad2), a state-of-the-art cross-lingual language model. XLM-RoBERTa (Conneau *et al.*, 2020) is trained on 2.5 TB of filtered CommonCrawl data spanning 100 languages, including Turkish. The model architecture consists of 24 transformer layers with 1,024 hidden dimensions and 16 attention heads, totaling approximately 560 million parameters.

The key advantages of XLM-RoBERTa for our application include: (i) Cross-lingual transfer capability—the model learns language-agnostic representations that transfer effectively to Turkish without language-specific pre-training; (ii) Robust tokenization using SentencePiece with a 250,000 token vocabulary that effectively handles Turkish agglutinative morphology; (iii) Pre-training on diverse multilingual data that captures cross-linguistic patterns beneficial for understanding Turkish syntax and semantics.

The specific model variant (deepset/xlm-roberta-large-squad2) has been additionally fine-tuned on the SQuAD 2.0 English dataset, which includes both answerable and unanswerable questions. This task-specific fine-tuning provides the model with strong QA capabilities that transfer to Turkish through cross-lingual representations.

2.2. TQuAD 2.0

We evaluated our model on TQuAD 2.0, a native Turkish reading comprehension dataset. Unlike machine-translated datasets, TQuAD 2.0 contains originally Turkish content derived from Turkish Wikipedia articles, ensuring natural language patterns and culturally relevant content.

The dataset comprises 14,221 training examples and 2,520 validation examples across diverse topics, including history, geography, science, sports, and culture. Each example consists of a context paragraph, a question, and one or more answer spans extracted from the context. The dataset statistics are presented in Table 1.

Table 1. TQuAD 2.0 statistics

Dataset property	Value
Dataset name	TQuAD 2.0 (erdometo/tquad2)
Training examples	14,221
Validation examples	2,520
Language	Turkish (native)
Source	Turkish Wikipedia

Below are the dataset examples:

(i) Example 1: Context: "*Rollo'nun gelişinden önce popülasyonları Picardy'den veya Franklar olarak nitelendirilen Île-de-France'den farklı değildi. Daha önceki Viking yerleşimcileri 880'lerde gelmeye başlamıştı...*" Question: "*Kim geldiğinde orijinal viking yerleşimcilerine ortak bir kimlik vermiştir*?" Answer: "*Rollo.*"
(ii) Example 2: Context: "*Akdeniz'e ulaşmak için en önde gelen iki Norman ailesi, Hauteville'den Tancred ve Drengot ailesinin soyundan geliyordu...*" Question: "*III. Henry tarafından asillendirilmiş liderin adı nedir*?" Answer: "*Drogo.*"
(iii) Example 3: Context: "*Akdeniz'e ulaşmak için*

en önde gelen iki Norman ailesi, Hauteville'den Tancred ve Drengot ailesinin soyundan geliyordu..." Question: "*Melfi Kontu kimdi*?" Answer: "*William Iron Arm*."

2.3. Model configuration and hyperparameters

The XLM-RoBERTa Large model was configured using the hyperparameters detailed in Table 2 for evaluation.

Table 2. Model configuration parameters

Hyperparameter	Value
Maximum sequence length	384 tokens
Document stride	128 tokens
Batch size	4
N-best size	20
Max answer length	30 tokens
Precision	BF16 (mixed precision)
GPU	NVIDIA A100-SXM4-40GB
Transformers version	4.57.3

The tokenization process used SentencePiece with a vocabulary size of 250,000 tokens. For contexts longer than the maximum sequence length, overlapping chunks were created using a document stride of 128 tokens to ensure answer spans at chunk boundaries could be correctly identified.

2.4. Model evaluation approach

Important Clarification: This research evaluates the pre-trained deepset/xlm-roberta-large-squad2 model in a *zero-shot transfer learning* setting on TQuAD 2.0. No additional fine-tuning was performed on Turkish data. This approach was chosen to:

(i) Assess cross-lingual transfer capability: Evaluate how well multilingual pre-training on 100 languages (including Turkish) enables the model to perform Turkish QA without language-specific fine-tuning.
(ii) Resource efficiency: Demonstrate that effective Turkish NLP capabilities can be achieved without the extensive computational resources required for full model fine-tuning, making this approach more accessible for SAR development.
(iii) Baseline establishment: Provide baseline performance metrics that future research can improve through Turkish-specific fine-tuning or domain adaptation.

The model leverages two levels of pre-training:

(i) Multilingual pre-training: XLM-RoBERTa base training on 2.5 TB of data across 100 languages
(ii) Task-specific pre-training: Fine-tuning on English SQuAD 2.0 for QA capabilities

The evaluation on TQuAD 2.0 demonstrated the model's ability to transfer both multilingual representations and task-specific knowledge to Turkish QA contexts. The TQuAD 2.0 validation set contains only answerable questions (2,520 examples with answers, 0 unanswerable questions). This differs from SQuAD 2.0, which includes unanswerable questions to test model discrimination capabilities.

3. Results and discussion

This section presents the evaluation results of the XLM-RoBERTa Large model on TQuAD 2.0. The evaluation was conducted using the HuggingFace Transformers library with the pre-trained deepset/xlm-roberta-large-squad2 model in a zero-shot transfer learning setting.

3.1. Model evaluation results

The XLM-RoBERTa Large model was evaluated on the TQuAD 2.0 validation set comprising 2,520 examples. The evaluation results are presented in Table 3.

Table 3. TQuAD 2.0 evaluation results

Metric	Score
Total examples evaluated	2,520
Exact match	56.67%
F1 score	79.37%

The F1 score of 79.37% indicates strong performance in extracting relevant answer spans from Turkish context passages. The exact match score of 56.67% reflects the strict nature of this metric, which requires character-perfect matching with ground truth answers. The gap between F1 and EM scores is typical for extractive QA tasks and indicates that the model often identifies partially correct answers that overlap significantly with the ground truth.

3.2. Answer detection performance

Since TQuAD 2.0 contains only answerable questions (all 2,520 validation examples have answers), the results of the confusion matrix for answer detection are presented in Table 4. Table 5 describes answer detection metrics.

The model successfully identified that all questions in the TQuAD 2.0 validation set are answerable, achieving 100% accuracy in answer detection. This result indicates that the model's null-answer threshold is appropriately calibrated for this dataset.

Table 4. Confusion matrix for answer detection

Category	Prediction: No answer	Prediction: Has answer
Actual: No answer	0 (TN)	0 (FP)
Actual: Has answer	0 (FN)	2,520 (TP)

Abbreviations: FN: False negative; FP: False positive; TN: True negative; TP: True positive.

Table 5. Answer detection metrics

Detection metric	Score (%)
Precision	100.00
Recall	100.00
F1 (detection)	100.00
Accuracy	100.00

3.3. Sample prediction analysis

Analysis of model predictions revealed both successful extractions and error patterns. The following examples illustrate the model's behavior:

(a) Correct predictions (exact match=1):

- (i) Q: "*Panthers savunması kaç sayı bırakmıştır*?" Gold: "308." Predicted: "308."
- (ii) Q: "*Jared Allen'ın kaç tane kariyer sack edişi vardır*?" Gold: "136." Predicted: "136."
- (iii) Q: "*Pro Bowl için kaç tane Panthers savunma oyuncusu seçilmiştir*?" Gold: "*dört*." Predicted: "*dört*."
- (iv) Q: "*Hangi oyuncu sezonun en çok top kapmasına sahiptir*?" Gold: "*Kurt Coleman*." Predicted: "*Kurt Coleman*."

(b) Incorrect predictions (exact match=0):

- (i) Q: "*Josh Norman kaç tane top çalmıştır*?" Gold: "*dört*." Predicted: "88." The model extracted a related number from context, but not the correct answer.
- (ii) Q: "*Bu sezon takımdaki en fazla sack etmeyi kim kaydetmiştir*?" Gold: "*Kawann Short*." Predicted: "*Jared Allen*." Entity confusion between players mentioned in the same context.
- (iii) Q: "*Panthers savunması 2015 yılında kaç top çalma ile kayda geçmiştir*?" Gold: "24." Predicted: "(118)." Extracted wrong numerical value from context.

3.4. Error analysis

Analysis of incorrect predictions revealed the following error categories:

- (i) Numerical confusion (approximately 35% of errors): The model sometimes extracted incorrect numbers when multiple numerical values appeared in the context. This was particularly common for statistical questions about sports data, where many numbers were mentioned.
- (ii) Entity confusion (approximately 30% of errors): When multiple named entities of the same type (e.g., player names and location names) appeared in the context, the model occasionally selected the wrong entity. This suggests challenges in resolving coreference and understanding fine-grained semantic distinctions.
- (iii) Partial match errors (approximately 25% of errors): The model extracted answers that overlapped with the ground truth but included extra tokens or missed some tokens. For example, extracting "(118)" instead of "118."
- (iv) Context misalignment (approximately 10% of errors): In some cases, the model extracted text from a different part of the context that seemed superficially related to the question but did not contain the correct answer.

3.5. Theoretical framework: Adaptive control principles for robust natural language processing in socially assistive robots

The design of robust NLP modules for SARs shares fundamental conceptual parallels with adaptive control theory. Both domains address the problem of maintaining stable and reliable system performance under uncertainty, nonlinearities, and time-varying operating conditions. In control systems, these challenges arise from model uncertainties, external disturbances, and unmodeled dynamics; in NLP-driven robotic interaction, they emerge from linguistic ambiguity, morphological variability, contextual uncertainty, and user-dependent interaction dynamics.

3.5.1. Adaptive and robust control analogies in natural language processing systems

Adaptive control strategies, such as adaptive fuzzy control and robust output-feedback control, have been extensively studied for uncertain nonlinear dynamical systems and chaotic synchronization problems (Boulkroune *et al.*, 2017). These methods rely on real-time parameter adaptation and feedback mechanisms to guarantee boundedness and stability in the presence of nonlinear input characteristics and unknown system dynamics.

An analogous requirement exists in SAR-oriented NLP systems. The linguistic input to the NLP module is inherently uncertain and nonstationary, particularly for agglutinative languages such as Turkish. In this context, the self-attention mechanism of transformer-based models can be interpreted as a form of adaptive gain scheduling, where attention weights are dynamically adjusted based on the relevance and uncertainty of linguistic tokens. This adaptive allocation of representational capacity mirrors the role of adaptive control laws in regulating system response under varying operating conditions.

Robust neural adaptive control approaches (Hayakawa *et al*., 2008) further strengthen this analogy. In such systems, neural networks compensate for modeling uncertainties while guaranteeing asymptotic stability. Similarly, the XLM-RoBERTa model leverages learned cross-lingual representations to compensate for linguistic variability and sparse data conditions, thereby maintaining stable QA performance across diverse Turkish sentence structures.

3.5.2. Backstepping control perspective on transformer layer hierarchies

Adaptive backstepping control provides a recursive methodology for stabilizing nonlinear systems by systematically designing control laws from lower-order subsystems to higher-order dynamics (Zouari *et al*., 2013). This layered design philosophy closely aligns with the hierarchical structure of transformer architectures used in NLP.

In transformer-based NLP models, lower layers predominantly capture lexical and syntactic features, while higher layers encode semantic and contextual relationships. This hierarchical abstraction process parallels the recursive stabilization steps in backstepping control, where intermediate virtual control signals are constructed to ensure global system stability. The validation of NLP models on structured benchmarks such as TQuAD 2.0 serves an analogous role to Lyapunov-based stability analysis in control theory, providing empirical guarantees of bounded and reliable system behavior.

The relevance of this analogy is further reinforced by adaptive backstepping control studies in flexible robotic manipulators (Zouari *et al*., 2013), where nonlinearities, elastic dynamics, and actuator uncertainties must be addressed simultaneously. SAR platforms integrating NLP modules face similar multi-domain uncertainties spanning perception, interaction, and actuation layers.

3.5.3. Optimal control and computational resource allocation

Nonlinear optimal control frameworks aim to achieve performance objectives while minimizing resource consumption under system constraints, as demonstrated in industrial applications such as gas compressors driven by induction motors (Rigatos *et al*., 2023). This optimization principle is directly applicable to NLP deployment in real-time robotic systems.

The use of mixed-precision computation (BF16) and constrained sequence lengths in this study represents an implicit optimal control strategy, balancing inference accuracy against computational latency and energy consumption. Such trade-offs are critical for SARs operating under real-time constraints and limited onboard computational resources.

3.5.4. Implications for control-informed socially assistive robot natural language processing design

The control-theoretic interpretation of NLP robustness suggests several design guidelines for future SAR systems:

(i) Closed-loop interaction: Incorporating user feedback mechanisms enables adaptive adjustment of dialogue strategies, analogous to feedback control loops.
(ii) Stability monitoring: Confidence-aware answer selection acts as a stability margin, preventing unreliable or unsafe system responses.
(iii) Adaptive tuning: Domain-specific fine-tuning parallels parameter adaptation in adaptive controllers, allowing performance improvement without loss of stability.
(iv) Fault tolerance: Detection of degraded NLP confidence can trigger fallback interaction modes, similar to fault-tolerant control architectures.

Overall, this framework demonstrates that established principles from adaptive and robust control theory—uncertainty compensation, hierarchical stabilization, and optimal resource allocation—provide a rigorous foundation for designing stable and reliable NLP-driven SARs operating in complex real-world environments.

4. Real-world applications and future directions

4.1. Real-world applications and adaptive socially assistive robot deployment

The integration of localized Turkish NLP capabilities into SARs extends beyond educational settings to numerous real-world applications where human–robot collaboration requires linguistic and cultural adaptation.

4.1.1. Healthcare applications

In healthcare environments, Turkish-speaking SARs can provide crucial support:

(i) Patient monitoring and companionship: SARs equipped with Turkish NLP can conduct routine health assessments through conversational interaction, asking patients about symptoms, medication adherence, and daily activities. The QA capability demonstrated in this research enables the robot to respond to patient queries about medication schedules, dietary restrictions, and appointment information.

(ii) Elderly care: Turkey's aging population (projected to reach 20.8% by 2040) creates significant demand for assistive technologies. SARs with culturally appropriate Turkish language support can provide companionship, remind elderly users to take medications, and alert caregivers to emergencies. The adaptive nature of our NLP system, informed by control theory principles (Section 3.5), enables the robot to adjust interaction complexity based on the user's cognitive status.

(iii) Rehabilitation support: SARs can guide patients through physical therapy exercises using natural Turkish instructions, answer questions about recovery procedures, and provide motivational feedback. The system's ability to handle Turkish's agglutinative morphology ensures understanding of varied phrasing from patients with different educational backgrounds or dialects.

4.1.2. Smart home integration

Turkish-speaking SARs can serve as intelligent interfaces for smart home ecosystems:

(i) Home automation control: Users can issue natural language commands in Turkish to control lighting, temperature, security systems, and appliances. The QA capability allows users to query system status ("What temperature is the living room?") and receive contextualized responses.

(ii) Energy management: SARs can provide real-time information about energy consumption, recommend optimization strategies, and answer questions about utility usage patterns—all in culturally appropriate Turkish communication styles.

(iii) Security and monitoring: Integration with smart home sensors enables SARs to answer questions about home security status, provide alerts about unusual activities, and serve as mobile surveillance platforms with natural language interfaces.

4.1.3. Industrial human–robot collaboration

In manufacturing and service industries, Turkish NLP-enabled SARs facilitate safer and more efficient operations:

(i) Workplace safety: SARs can monitor work environments, provide safety reminders in Turkish, answer worker questions about hazard protocols, and report safety violations. Drawing parallels to adaptive backstepping control for robot manipulators (Zouari *et al*., 2013) and nonlinear optimal control for industrial systems (Rigatos *et al*., 2023), these SARs must balance multiple objectives, including information delivery, physical safety, and operational efficiency.

(ii) Training and onboarding: New employees can interact with SARs to receive training information, ask procedural questions, and practice skills in a low-pressure environment—all conducted in their native Turkish language, which is especially valuable in Turkey's diverse industrial workforce.

(iii) Quality control: SARs equipped with vision systems and Turkish NLP can answer operator questions about quality standards, provide real-time feedback on product specifications, and explain defect classifications in an accessible language.

4.1.4. Adaptive feedback mechanisms

The control systems perspective (Section 3.5) informs how adaptive feedback from users can continuously improve SAR behavior across these applications:

(i) User interaction analytics: Similar to system identification in adaptive control, analyzing user interaction patterns (question types, rephrasing frequency, and interaction duration) provides data for model refinement. For example, if healthcare users frequently rephrase questions about medications, this indicates areas where model fine-tuning is needed.

(ii) Contextual adaptation: Just as adaptive fuzzy control adjusts parameters based on system state (Boulkroune *et al*., 2017), SARs can adjust their language complexity, response verbosity, and interaction pace based on user profiles and real-time feedback. An elderly user might receive simpler, slower responses, while a technical professional receives detailed, technical information.

(iii) Performance monitoring: Implementing confidence-based feedback mechanisms (analogous to stability monitoring in robust neural adaptive control [Hayakawa *et al.*, 2008]) allows SARs to request clarification when understanding is uncertain, improving interaction reliability.
(iv) Multi-domain transfer: Insights from our Turkish NLP development can be transferred to other low-resource Turkic languages (e.g., Azerbaijani, Uzbek, and Kazakh) through similar cross-lingual approaches, expanding SAR accessibility across Central Asia and the Turkic-speaking regions.

4.1.5. Connection to intelligent control and automation

The integration of localized NLP into SARs represents a convergence of intelligent control and human-centered automation:

(i) Hierarchical control architecture: Modern industrial automation employs hierarchical control structures where high-level planning, mid-level coordination, and low-level execution operate at different timescales. Similarly, SAR systems integrate strategic planning (task goals), tactical execution (dialogue management), and operational control (NLP inference), all coordinated through natural language interfaces.
(ii) Human-in-the-loop systems: Unlike fully autonomous systems, SARs exemplify human-in-the-loop intelligent automation where natural language serves as the primary human–machine interface. This aligns with Industry 4.0 principles, emphasizing collaborative rather than replacement automation.
(iii) Safety-critical adaptation: Drawing from safety-critical control systems literature, SARs in healthcare and industrial settings must guarantee bounded performance even under unusual inputs. The theoretical framework connecting adaptive control to NLP robustness (Section 3.5) provides design principles for ensuring fail-safe behavior.

4.1.6. Scalability and deployment challenges

Real-world SAR deployment faces challenges analogous to control systems implementation:

(i) Environmental variability: Like control systems operating under varying conditions, SARs must handle acoustic noise, diverse user accents, and non-standard linguistic inputs. Robust control strategies (Boulkroune *et al.*, 2017; Hayakawa *et al.*, 2008) suggest incorporating conservative design margins and fault detection mechanisms.
(ii) Long-term stability: Deployed systems must maintain performance over extended periods. Continual learning approaches, validated through principles from adaptive control theory, can enable SARs to improve without catastrophic forgetting or instability.
(iii) Integration complexity: SARs integrate multiple subsystems (perception, navigation, manipulation, and dialogue). This multi-component integration parallels multivariable control systems (Hayakawa *et al.*, 2008), requiring careful interface design and coordination mechanisms to ensure system-wide stability and performance.

4.2. Future work

Future research directions include:

(i) Speech integration: Integrating the NLP module with Turkish automatic speech recognition (ASR) and text-to-speech (TTS) systems for complete voice-interactive capabilities.
(ii) Domain-specific fine-tuning: Fine-tuning the model on domain-specific Turkish datasets (medical, elderly care, and industrial) to improve performance in targeted applications.
(iii) Multimodal integration: Combining NLP with computer vision and gesture recognition for richer human–robot interaction.
(iv) Continual learning: Implementing adaptive learning mechanisms that allow the SAR to improve from user interactions while maintaining stability guarantees informed by control theory.
(v) Cross-lingual extension: Extending the approach to other Turkic languages to expand SAR accessibility across Central Asia.

5. Conclusion

This study evaluated cross-lingual transfer learning for Turkish QA in SAR applications. The XLM-RoBERTa Large model (deepset/xlm-roberta-large-squad2) was evaluated in a zero-shot setting on the TQuAD 2.0, achieving an F1 score of 79.37% and an exact match score of 56.67% on 2,520 validation examples without any Turkish-specific fine-tuning.

The results demonstrate the effectiveness of multilingual transformer models for Turkish QA tasks. The model successfully handled Turkish linguistic challenges, including agglutinative morphology and free word order, through its SentencePiece tokenization and cross-lingual representations learned from 100 languages.

Error analysis revealed that the primary challenges are numerical confusion (35%), entity confusion (30%), partial match errors (25%), and context misalignment (10%). These findings suggest directions for future improvement, including domain-specific fine-tuning and enhanced numerical reasoning capabilities.

Significantly, this research established a methodological framework connecting adaptive NLP design with control systems theory, demonstrating how principles from adaptive fuzzy control, robust neural adaptive control, and adaptive backstepping control inform the development of more stable and reliable NLP systems for robotics. This theoretical grounding provides design principles for implementing robust, fault-tolerant SARs that can maintain stable performance under uncertain and dynamic real-world conditions.

The identified applications in healthcare (patient monitoring, elderly care, and rehabilitation), smart homes (automation control, energy management, and security), and industrial settings (workplace safety, training, and quality control) demonstrate the broad potential impact of localized Turkish NLP in SAR systems. The adaptive feedback mechanisms informed by control theory enable continuous improvement while maintaining stability guarantees essential for safety-critical applications.

This work establishes a foundation for developing culturally and linguistically appropriate SARs for Turkish-speaking populations, with implications for other low-resource Turkic languages. Future work will focus on speech integration, domain-specific fine-tuning, and multimodal interaction capabilities to realize complete voice-interactive SAR systems for real-world deployment.

Acknowledgments

None.

Funding

This research was supported by internal funding from OSTİM Technical University Scientific Research Programs (Grant No. BAP202301).

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

Conceptualization: Murat Şimşek, Demiral Akbar
Formal analysis: Murat Şimşek
Investigation: Murat Şimşek
Methodology: Murat Şimşek, Demiral Akbar
Writing–original draft: Murat Şimşek
Writing–review & editing: Demiral Akbar, Murat Şimşek

Availability of data

The data used in this study are publicly available. TQuAD 2.0 (erdometo/tquad2) and the pre-trained model (deepset/xlm-roberta-large-squad2) are available via Hugging Face. Any additional materials (e.g., evaluation configurations/scripts) can be provided by the corresponding author upon reasonable request.

References

Alkesaiberi, A., Alkhathlan, A., & Abdelali, A. (2024). Ara-CANINE: Character-based pre-trained language model for Arabic language understanding. *International Journal on Cybernetics & Informatics,* 13(2), 45-59.

https://doi.org/10.5121/ijci.2024.130204

Boulkroune, A., Hamel, S., Zouari, F., Boukabou, A., & Ibeas, A. (2017). Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities. *Mathematical Problems in Engineering,* 2017, Article 8045803.

https://doi.org/10.1155/2017/8045803

Breazeal, C. (2017). Social robots: From research to commercialization. In: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction; March 6-9, 2017; Vienna, Austria.

https://doi.org/10.1145/2909824.3020258

Conneau, A., Khandelwal, K., Goyal, N., *et al.* (2020). Unsupervised cross-lingual representation learning at scale. In: Jurafsky, D., Chai, J., Schluter, N., & Tetreault, J., (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; July 5-10, 2020; Online. pp. 8440-8451.

https://doi.org/10.18653/v1/2020.acl-main.747

Cui, Y., Liu, T., Che, W., *et al.* (2019). A span-extraction dataset for Chinese machine reading comprehension. In: Inui, K., Jiang, J., Ng, V., & Wan, X. (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); November 3-7, 2019; Hong Kong, China, pp. 5883–5889.

https://doi.org/10.18653/v1/D19-1600

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., & Solorio, T. (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, MN, USA. Volume 1, pp. 4171-4186.

https://doi.org/10.18653/v1/N19-1423

Eryiğit, G. (2014). ITU Turkish NLP web service. In: Bentivogli, L., Koehn, P., & Manning, C. (Eds.), Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics; April 26-30, 2014, Gothenburg, Sweden. pp. 1-4. Association for Computational Linguistics.

https://doi.org/10.3115/v1/E14-2001

García-Haro, J. M., Oña, E. D., Hernandez-Vicen, J., Martinez, S., & Balaguer, C. (2021). Service robots in catering applications: A review and future challenges. *Electronics*, 10(1), Article 47.

https://doi.org/10.3390/electronics10010047

Hayakawa, T., Haddad, W. M., & Hovakimyan, N. (2008). Neural network adaptive control for a class of nonlinear uncertain dynamical systems with asymptotic stability guarantees. *IEEE Transactions on Neural Networks*, 19(1), 80–89.

https://doi.org/10.1109/TNN.2007.902704

Hung, C.-F., Lin, Y., Ciou, H.-J., Wang, W.-Y., & Chiang, H.-H. (2021). Foodtemi: The AI-oriented catering service robot. In: Proceedings of the 2021 IEEE International Conference on Consumer Electronics–Taiwan (ICCE-TW); September 15-17, 2021; Penghu, Taiwan. pp. 1–2.

https://doi.org/10.1109/ICCE-TW52618.2021.9603096

Maj, K., & Zarzycki, P. (2019). Meeting with social robots like the cat-cucumber meeting? An integrated model of human-robot first contact: Psychological perspective. *Paladyn, Journal of Behavioral Robotics*, 10(1), 454–465.

https://doi.org/10.1515/pjbr-2019-0026

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.

https://doi.org/10.1109/ACCESS.2024.3433426

Nayyar, A., Puri, V., Nguyen, N. G., & Le, D.-N. (2018). Smart surveillance robot for real-time monitoring and control system in environment and industrial applications. In: Bhateja, V., Nguyen, B. L., Nguyen, N. G., Satapathy, S. C., & Le, D.-N. (Eds.), *Information systems design and intelligent applications*. Berlin: Springer. pp. 229–243.

https://doi.org/10.1007/978-981-10-7512-4_23

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. In: Su, J., Carreras, X., & Duh, K. (Eds.), Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing; November 1-4, 2016; Austin, Texas, USA. pp. 2383-2392.

https://doi.org/10.18653/v1/D16-1264

Rigatos, G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & Zouari, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. *Results in Control and Optimization*, 11, 100226.

https://doi.org/10.1016/j.rico.2023.100226

Seo, T., Jeon, Y., Park, C., & Kim, J. (2019). Survey on glass and façade-cleaning robots: Climbing mechanisms, cleaning methods, and applications. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 6(2), 367-376.

https://doi.org/10.1007/s40684-019-00079-4

Spitale, M., Silleresi, S., Cosentino, G., Panzeri, F., & Garzotto, F. (2020). "Whom would you like to talk with?": Exploring conversational agents for children's linguistic assessment. In: Santos, O. C., Dillenbourg, P., & Börner, K. (Eds.), Proceedings of the Interaction Design and Children Conference; June 21-24, 2020; London, UK. pp. 262-272.

https://doi.org/10.1145/3392063.3394421

Toscano, E., Spitale, M., & Garzotto, F. (2022). Socially assistive robots in smart homes: Design factors that influence the user perception. In: Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI); March 7-10, 2022; Sapporo, Japan. pp. 1075-1079.

https://doi.org/10.1109/HRI53351.2022.9889467

van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly*, 28(4), 695–704.

https://doi.org/10.2307/25148660

Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017). Attention is all you need. *arXiv*. Preprint.

https://doi.org/10.48550/arXiv.1706.03762

Zouari, F., Ben Saad, K., & Benrejeb, M. (2013). Adaptive backstepping control for a class of uncertain single input single output nonlinear systems. In: Proceedings of the 10th International Multi-Conference on Systems, Signals & Devices (SSD); May 6-10, 2022; Setif, Algeria. pp. 1-6.

https://doi.org/10.1109/SSD.2013.6564134