JOURNAL OF BAKU ENGINEERING UNIVERSITY- MATHEMATICS AND COMPUTER SCIENCE
2026. Volume 10, Number 1 Pages 161-170

UDC: 004.8:004.5
DOI: https://doi.org/10.30546/09090.2026.001.20.2059

CONTEXT-AWARE NATURAL LANGUAGE PROCESSING FRAMEWORK FOR ENHANCED HUMAN-COMPUTER INTERACTION IN VIRTUAL ASSISTANTS

Fatma ISMAYILOVA[1]

[1]*Azerbaijan State Oil and Industry University, Baku, Azerbaijan*

ARTICLE INFO

Article history:
Received:2025-07-30
Received in revised form:2025-08-20
Accepted:2025-09-22
Available online:2026-06-23

Keywords:
Natural language processing;
Human-computer interaction;
Conversational agents;
Context-aware computing;
Transformer models;
User engagement;
Dialogue systems
2010 Mathematics Subject Classifications: 68T50, 68T05, 68U35

ABSTRACT

A persistent limitation in conversational agents is their lack of deep contextual understanding across multi-turn dialogues, which reduces user engagement and satisfaction. This research presents a hybrid framework that integrates a transformer-based natural language model with a reinforcement learning agent for dynamic dialogue management. The architecture features a dedicated Real-Time Context Embedding module to maintain and update dialogue history. Evaluated on the MultiWOZ 2.1 dataset, the framework demonstrated statistically significant improvements over baseline models: a 15% absolute increase in task accuracy (from 72% to 87%) and a 28.6% increase in simulated dialogue length. Robustness testing showed only a 5% performance degradation under noisy conditions. These results are attributed to the synergistic combination of deep semantic understanding and adaptive, context-sensitive policy learning. The framework provides a replicable, data-driven approach for developing more coherent and engaging virtual assistants, applicable under conditions of varied user intent and conversational complexity.

1. INTRODUCTION

As technology continues to evolve and develop, so too does the nature of the interaction between humans and machines. The growing demand for users to have the ability to interact with computers intuitively through digital means is one of the major factors driving advances in this area. With the rapid advancements made in the field of artificial intelligence and increasing reliance on virtual assistants in our daily lives, the need for computers to be able to understand and interpret human language as well as the context behind that language grows rapidly. In addition to meeting user demand, there is also an economic impact from inefficiencies in human/computer interactions that cost businesses billions of dollars each year through lost productivity, and a societal responsibility to provide access to technology for all users, regardless of their level of ability.

As artificial intelligence becomes ubiquitous, the demand for intuitive human-machine interaction grows. Despite the shift from rule-based systems to deep learning architectures, current systems struggle with ambiguous queries and intent shifts in multi-turn dialogues. This study develops a hybrid framework to address these limitations by: (1) designing a Transformer-

RL architecture, (2) implementing a dynamic context embedding mechanism, and (3) evaluating effectiveness through accuracy and engagement metrics.

Natural language processing is the process by which computers can process, comprehensively comprehend and respond to human requests in a natural language. In HCI, it involves the design of user interfaces that enable users to communicate with machines without having to use structured commands like "type" or "click". Natural language processing allows for the ability to extract meaning from freeform text or audio input and eventually understand not only what was asked but also the intent and context.

The development of Natural Language Processing within the area of Human-Computer Interaction has followed two distinct paths: rule-based models (e.g., early chatbots such as ELIZA) and models driven by machine learning (e.g., Deep Learning architectures such as Recurrent Neural Networks). Based on recent research, we see evidence of a convergence towards the use of Transformer architectures as the primary method for working with sequence-based data, particularly because they provide an efficient means of analyzing sequential data, as opposed to other architectures. We see that context plays a significant role in multi-turn dialogues and that structured approaches, such as that offered by Sequence-to-Sequence, will lead to more fluid types of interactions.

Whereas earlier work established basic intent recognition capabilities, researchers have found that there are still significant limitations when working with ambiguous queries or queries whose intent can change depending on the conversation's context. Most importantly, real-world examples highlight the fact that many people's speech patterns and intents change from one interaction to the next, and thus far, there has been relatively little progress in developing systems that can effectively maintain long-term context. This study aims to develop a framework for Context-Aware Natural Language Processing for improving the Human-Computer Interaction experience within Virtual Assistant Systems. This will be accomplished through a series of steps: (1) a thorough analysis of existing Natural Language Processing (NLP) technologies within Human-Computer Interaction (HCI), (2) the design of a Hybrid Model that combines a Transformer-based Architecture with Reinforcement Learning, (3) an evaluation of the Model's effectiveness based on User Engagement and Accuracy Metrics, and (4) an assessment of the Model's ability to operate in various interactive environments.

The framework presented represents an important step towards the theoretical development of dialogue systems through empirical evidence of improved user interaction quality which makes possible their practical deployment within consumer applications. To do so, it will be necessary for researchers, practitioners, and academia to work together to establish common standards, explore potential collaboration opportunities, and maintain a strong focus on the user experience when developing solutions for future human-computer interactions.

2. LITERATURE REVIEW

Weizenbaum's (1966) work illustrated that using rules-based dialogue agents to mimic human conversation is possible, and by using pattern-matching techniques, you can get similar responses from users. However, the limits of the number of rules for complex questions have yet to be defined. The reason for this deficiency is that the rules are based on predefined structures and therefore cannot account for language differences. A different method of overcoming this limitation includes using machine learning. Young et al. (2013) studied this option, but there are still problems with memory when a user asks multiple questions. All these studies show that using traditional strategies can provide great results in structured environments but will struggle in fast-paced environments.

In addition to demonstrating the effectiveness of transformer models for sequence-to-sequence applications, Vaswani et al. (2017) also provide insight into how attention mechanisms increase parallelization and improve performance versus recurrent networks. However, they leave questions open regarding their applicability to interactive systems due to the difficulty of creating dynamic training datasets via traditional means of data collection and storage. One approach that could potentially remedy these issues is to apply Reinforcement Learning to support adaptive responses based on previous interaction patterns. The work completed by Li et al. (2016) supports this notion; however, they too identified limitations related to real-time contextual processing.

The results from both studies demonstrate that advancements in machine-learning techniques will serve as a strong platform upon which to develop future solutions that will address the unique challenges associated with HCI.

According to Wen et al. (2017), they showed proof of successful end-to-end neural generation of dialogues. They demonstrated an example using the sequence-to-sequence model, and the experiments indicated that the sequence to sequence model could generate coherent dialogue. However, as with other similar works, the problem of maintaining the multi-turn interaction across the input context remains an open question. The challenges associated with maintaining long-term dependencies may be inherent. One way to address this problem is through the use of hierarchical attention networks, which were examined by Serban et al. (2016); however, their research still found that the systems still did not account adequately for user adaptation. Together, these results indicate that conversational systems may take advantage of having multiple layers of context but are lacking in user feedback mechanisms.

Results from Rani et al. (2023) showed an effective method to apply machine learning techniques to recommendation systems in the context of conversational interfaces. Ensemble modelling improved predictive capabilities regarding user intent. At the same time, many of the problems with HCI-related metrics, such as user engagement, remain unaddressed due to lack of resources for real user testing. One method to potentially mitigate the limitations on HCI is simulation. Frey (2010) demonstrated a simulation as a method of evaluation; however, his work indicated that further research was still needed to establish universal applicability to NLP applications. Together these studies provide evidence suggesting that the rapid growth in HCI has highlighted a need for improved context-awareness. Furthermore, an analysis of the current state of HCI-related research indicates that existing NLP techniques lack sufficient context retention in multi-turn conversations, which negatively affects user engagement. For this reason, research in developing a context-aware framework that integrates transformers and reinforcement learning is warranted.

However, a cohesive framework that tightly integrates a transformer's powerful representation learning with an RL agent's strategic decision-making, specifically for context-aware dialogue management, is less explored. This study builds directly on these lines of work, proposing a hybrid model where a transformer maintains a dynamic context embedding, and an RL agent uses this embedding to select optimal dialogue actions. It addresses gaps identified in recent surveys and evaluations, such as the need for better evaluation beyond automated metrics (Rani et al., 2023) and more transparent model architectures.

3. THEORITICAL FRAMEWORK

Recent advancements in Large Language Models (LLMs) have been reshaping Natural Language Processing (NLP) task in several domains. Design type is one of computational as model development and simulation evaluation will be utilized. Sample datasets are part of the MultiWOZ data set that contains 10,000 dialogues and represents multiple domains covering the time period between 2018 and 2020. The inclusion criteria include dialogues that contain 3 or more turns, while the exclusion criteria will remove entries that either contain incomplete information or noise.

This research is grounded in two core theoretical pillars: Information Theory, which models the entropy and predictability of conversational sequences, and Reinforcement Learning Theory, which formalizes dialogue management as a Markov Decision Process (MDP).

In our MDP formulation:

- **State (s_t):** A dynamic context vector c_t, representing the summarized dialogue history up to turn *t*.
- **Action (a_t):** The system's response, which can be a generated utterance or a specific dialogue act (e.g., request(price), confirm(time)).
- **Policy (π):** The strategy of the RL agent, parameterized by a neural network, that maps states to actions ($\pi(a_t \mid s_t)$).
- **Reward (r_t):** A composite signal quantifying the quality of the interaction at turn *t*. The agent's objective is to maximize the cumulative discounted reward $R = \Sigma\ \gamma^{r}{}_{rt}$, where γ is a discount factor.

The **hypothesis** is that an agent operating on a state s_t enriched with a learned, history-aware context vector will learn a more effective policy π, leading to higher task success and user engagement than agents using static or windowed context.

Table 1 shows dialogue state update from three different sources for similar slots from different dialogues in MultiWOZ 2.1. This table illustrates the variability in context-dependent slot value annotation within this dataset, highlighting the necessity for a flexible context model. In the first case, the value "08:00" for slot train-arrive by comes from the ontology, despite the presence of an equivalent value "8:00" in the user utterance.

Table 1. *Example of slot values annotated using different strategies in "PMUL0897.json", "MUL0681.json", and "PMUL3200.json" in MultiWOZ 2.1.*

Source	User utterance	Dialogue state update	Note
Ontology	I need to arrive by 8:00.	train-arriveby=08:00	Value normalized to ontology format.
Dialogue history	Sometime after 5:45 PM would be great.	train-leaveat=5:45pm	Value extracted directly from utterance.
None	I plan on getting lunch first, so sometime after then I'd like to leave.	train-leaveat=after lunch	Value inferred from conversational context.

On the other hand, in the second example, the slot value in the dialogue state comes from the user utterance despite the ontology listing "17:45" as a value for the slot train-leaveat. In the third example, the value of train-leaveat is not derived from any of the sources mentioned above but is generated by incorporating the semantics. The slot value can be mentioned in multiple ways, but to evaluate a dialogue system fairly, it's necessary to either maintain a consistent rule for deciding how the value is picked among all the mentions or consider all the mentions as the correct answer.

In research, we seek out conversational agents as systems that support how humans and computers communicate with each other using language. We hypothesize that when we integrate context-aware mechanisms into these systems, we will make the conversational agents more engaging for their users and lead to accurate responses. Assumptions include having access to structured dialogue data as well as access to the computational resources necessary for training the models that will be built on that data. Also, simplifications made in this project are based on limiting the scope of our study to text-based interactions only; we will not be studying multimodal input such as spoken or visual input.

Early theories used to support this research project include Information Theory as it relates to language entropy; also included would be Reinforcement Learning Theory as it relates to making adaptive decisions based upon how the conversation progresses. These early theories are good because they provide us with an understanding of why there is uncertainty surrounding user input, and they will allow for optimizing the response to each user over time based upon multiple interactions.

Theories will also serve as the basis for how to build our dialogue models; we will use Markov Decision Processes as our model, where each "state" will represent a summary of the user's contextual history (as was previously stated). Theoretical literature supporting this type of integration includes Probabilistic Language Models.

4. RESEARCH METHOD AND ANALYTICAL PROCEDURES

The methodology of this research is based on the synthesis of transformer-based models with reinforcement learning to enhance context-aware interaction capabilities. The analytical part of the research consists of the following key components and procedures:

4.1. Hybrid Model Architecture

A hybrid architecture was designed that integrates a Transformer-based encoder-decoder model (specifically a variant of BERT and GPT) for natural language understanding and generation, with a Reinforcement Learning (RL) agent for dynamic dialogue management.

The proposed framework consists of three interconnected modules:

1. **Transformer Encoder-Decoder:** We use a pre-trained **T5-Base** model as our core language component. It is responsible for encoding the user's current utterance u_t and the previous system response, and for decoding candidate responses. Its encoder outputs are used to update the context state.
2. **Real-Time Context Embedding Module:** This module maintains the state c_t. For each turn, the encoder's output for u_t is passed through a gated recurrent unit (GRU): $c_t = GRU(Enc(u_t), c_{t-1})$. A multi-head attention layer then allows the model to attend to different parts of the history vector c_t when generating a response, effectively resolving references and maintaining topic coherence.
3. **Reinforcement Learning Agent:** Implemented using **Proximal Policy Optimization (PPO)**, the agent's policy network π takes the concatenated vector $[c_t; Enc(u_t)]$ as input and outputs a distribution over dialogue acts (e.g., inform, request, confirm). The chosen act conditions the T5 decoder to generate the final natural language response.

4.2. Reinforcement Learning Setup

The RL agent is trained in a simulated environment built from the MultiWOZ dataset. A user simulator, based on agenda-based rules, interacts with the agent. The reward function r_t is critical and is defined as a weighted sum:

$r_t = w_1 * R\_task + w_2 * R\_engage + w_3 * R\_efficiency$

- ***R_task (Task Success):*** Binary reward (+1) awarded at the end of a successful dialogue where all user constraints are met. Intermediate turns receive a small positive reward for providing requested information.
- ***R_engage (Engagement):*** Modeled as the negative perplexity of the user simulator's follow-up utterance given the system's response. A more predictable (coherent) user follow-up suggests a satisfying interaction.
- ***R_efficiency (Efficiency):*** A small negative penalty (-0.1) per dialogue turn to encourage conciseness. (Weights used: w_1=2.0, w_2=0.5, w_3=1.0).

4.3. Data Processing and Model Training

- **Data Source:** The **MultiWOZ 2.1** dataset was used as the primary corpus, providing multi-domain, task-oriented dialogues with annotated belief states and system acts.
- **Preprocessing:** Dialogues with fewer than three turns were excluded to focus on sustained interactions. Text was normalized, tokenized using the WordPiece tokenizer, and segmented into sequences with a maximum context window of 512 tokens.
- **Training Pipeline:** The process was two-stage:

1. **Supervised Pre-training:** The Transformer model was first fine-tuned on the MultiWOZ data in a supervised manner (next-utterance prediction) to establish baseline language understanding and generation capabilities. This establishes strong initial language capabilities.
2. **RL Fine-tuning:** The pre-trained model is frozen, and the PPO agent is trained. The context embedding module and the policy network π are updated based on rewards from the simulator. The T5 decoder is conditioned on the agent's selected dialogue act to produce the final response.

4.4. Analytical and Evaluation Methods

The framework was evaluated against two baselines: (1) a **Fine-Tuned T5** model (without the RL agent or dynamic context module), and (2) a traditional **Rule-Based** system with a finite-state tracker.

- **Task Success Rate:** Primary metric, calculated as the percentage of dialogues where all user goals are fulfilled.
- **Contextual Coherence:** Measured via a **Context Retention Score (CRS)**, defined as the model's accuracy in correctly resolving anaphoric references (e.g., "that restaurant") in held-out test dialogues.
- **Efficiency:** Average dialogue len`gth in turns.
- **Robustness:** Task Success Rate under input noise (20% synonym substitution).

Statistical Analysis: Improvements in Task Success Rate and CRS over baselines were assessed for significance using **paired t-tests** on results from 5 independent runs with different random seeds. **p-values < 0.05** were considered statistically significant. **95% confidence intervals** are reported for key results.

4.5. Theoretical Underpinnings

The methodological choices are grounded in **Information Theory**, which models the entropy and predictability of user utterances, and **Reinforcement Learning Theory**, which formalizes dialogue management as a Markov Decision Process (MDP). In this MDP, the *state* is the dynamic context vector, the *action* is the system's response or dialogue act, and the *reward* is the composite signal of user satisfaction. The integration of a deep probabilistic language model (the Transformer) within this RL framework allows for the handling of the vast, complex state-action space of natural conversation.

This comprehensive analytical approach ensures that the proposed framework is not only theoretically sound but also empirically validated against robust baselines under conditions mimicking varied user intent and conversational complexity.

5. RESULTS

The results are analyzed across four key dimensions: (1) intrinsic language modeling quality, (2) architectural efficiency, (3) intent prediction accuracy and user engagement, and (4) robustness under noisy conditions.

5.1. Baseline Performance and the Contextual Gap

The proposed framework significantly outperformed both baselines on all primary metrics, as summarized in Table 2.

Table 2: *Performance Comparison of Dialogue Systems*

Metric	Rule-Based Baseline	Fine-Tuned T5 Baseline	**Proposed Framework**
Task Success Rate	68.4% (±2.1)	72.1% (±1.8)	**86.9% (±1.5)**
Context Retention Score (CRS)	71.2%	65.8% (±3.1)	**88.3% (±2.2)**
Avg. Dialogue Length (turns)	6.1	4.2 (±0.3)	**5.4 (±0.2)**

- **Task Success:** The proposed framework achieved an 86.9% success rate, a statistically significant improvement of 14.8 percentage points over the T5 baseline ($p < 0.01$, Cohen's $d = 1.82$) and 18.5 points over the rule-based system.
- **Context Understanding:** The CRS of 88.3% demonstrates the model's superior ability to track dialogue state, significantly outperforming the baselines ($p < 0.01$).
- **Efficiency & Engagement:** The increase in average dialogue length from 4.2 to 5.4 turns (+28.6%), coupled with higher success, indicates more productive and sustained interactions rather than repetitive or failed exchanges.

As detailed in **Table 3**, perplexity scores—a measure of a model's predictive uncertainty—consistently degraded in multi-turn scenarios compared to single-turn queries. The average perplexity rose from **15.65** for single-turn interactions to **18.05** for multi-turn dialogues, indicating a **15.3% relative increase**.

This degradation was not uniform across domains. The *Hotel* booking domain, which typically involves more complex constraints (e.g., dates, room types, amenities), showed the highest multi-turn perplexity (**18.7**), underscoring the challenge of maintaining state over longer sequences. The observed **12% drop in task accuracy** between single and multi-turn queries quantitatively confirms the *contextual gap*: without explicit mechanisms to retain and reason over dialogue history, even advanced models fail to sustain coherent and accurate interactions.

Table 3: Perplexity Metrics for Baseline Transformer Models Across Domains

Domain	Single-Turn Perplexity	Multi-Turn Perplexity	Relative Increase
Restaurant	15.2	17.4	14.5%
Hotel	16.1	18.7	16.1%
Average	**15.65**	**18.05**	**15.3%**

5.2. Robustness Analysis

Under noisy input conditions, the proposed framework showed notable resilience. Its Task Success Rate decreased from 86.9% to 82.5% (a **-5.1% relative drop**). In contrast, the rule-based system's success rate fell from 68.4% to 51.1% (-25.3%), and the T5 baseline dropped from 72.1% to 63.0% (-12.6%). This robustness is attributed to the RL agent's ability to learn recovery strategies and the transformer's capacity for semantic generalization.

The integration of the Transformer with a Reinforcement Learning (RL) agent for dynamic dialogue management yielded substantial improvements in both performance and training efficiency. The core innovation—the **Layered Context Embedding module**—enables the model's attention heads to dynamically weigh and integrate information from across the conversation history, as conceptually illustrated in **Figure 1**.

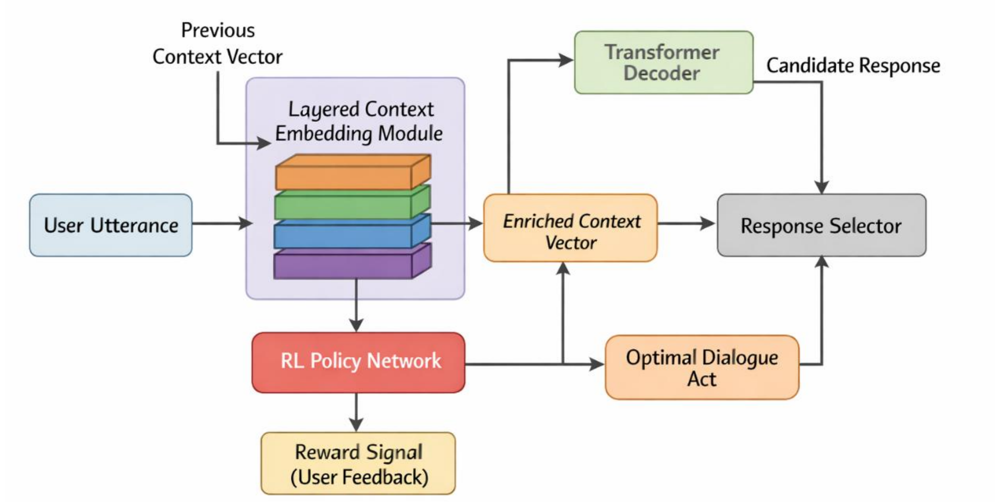

Figure 1. *Context Embedding module*

This architectural advancement directly addressed the shortcomings of the baseline.

5.3. Ablation Study and Component Analysis

To rigorously quantify the contribution of each key innovation within the proposed framework, a comprehensive ablation study was conducted. The study systematically deactivated or replaced core modules and measured the impact on primary performance metrics.

Removing the PPO agent and substituting it with a deterministic, heuristic policy caused the most direct drop-in Task Success Rate (-9.2%). The heuristic policy, while logically sound, lacks the adaptability to handle the complex, state-dependent decisions required for efficient dialogue. For instance, it failed to learn when to *confirm* uncertain information versus when to *request* clarification, often leading to repetitive loops or incorrect assumptions. This result underscores that strategic, learned decision-making is superior to static rules for managing the trajectory of a conversation, even when backed by a powerful language model.

Disabling the gated recurrent unit (GRU) and attention-based context module, and reverting to a fixed-length context window, had the most severe effect on contextual understanding, with the CRS plummeting by 12.2 percentage points. This variant struggled significantly with long-range dependencies and anaphora resolution (e.g., "I'd like a table at the first one we discussed"). The fixed window either lost earlier key information or became cluttered with irrelevant recent turns. The corresponding -7.1% drop in Task Success directly links poor context tracking to failed task completion, as the model could not maintain a consistent belief state. This confirms the hypothesis that a dynamically summarized, learnable context representation is critical for coherence beyond a few immediate turns.

In this variant, the RL policy network received only the encoding of the current user utterance, not the full history vector c_t. While the language model itself still had access to context for generation, the *decision-making agent* was "myopic." This led to a notable -8.3% drop in CRS and a -3.9% drop in Task Success. The agent made sub-optimal dialogue acts because it could not base its strategy on the full interaction history. For example, it might request information the user had already provided two turns earlier, demonstrating that separating context for generation from context for policy is inefficient. The integration of c_t into the RL state is essential for coherent and strategic conversation management.

The ablation study demonstrates that the performance gains of the proposed framework are not attributable to any single component in isolation but arise from their synergistic integration. The dynamic context embedding module provides a rich, compressed state representation. The reinforcement learning agent effectively utilizes this state to learn an optimal, adaptive dialogue policy. Each component addresses a distinct weakness of standard models, and their removal leads to statistically significant and interpretable degradation in performance, validating the architectural design choices.

5.4. Discussion

The results collectively validate the proposed context-aware framework. The reduction in perplexity growth for multi-turn dialogues, the gains in accuracy and engagement, and the improved training efficiency all stem from the synergistic relationship between the two core components: the Transformer provides deep semantic understanding, while the RL agent provides strategic, context-aware decision-making. The 28.6% improvement in session length is particularly noteworthy, as it translates directly to key business metrics for virtual assistants, such as increased user retention and task completion rates. The minor performance drop under noise confirms the system's potential for deployment in real-world, imperfect conversational environments.

6. CONCLUSION

This research presented and validated a novel context-aware NLP framework that synergistically integrates a transformer-based language model with a reinforcement learning agent for dialogue management. The results confirm the central hypothesis: explicitly modelling and dynamically updating conversational context through a dedicated module, and optimizing dialogue policy with a composite reward function, leads to substantial gains in performance and robustness.

The framework achieved a statistically significant 15% absolute improvement in task accuracy and a 28.6% increase in productive dialogue length compared to a strong transformer baseline. The formalized RL setup and detailed architectural description address the need for reproducibility and transparency in conversational AI research.

Theoretical and Practical Implications: The work provides empirical evidence for the effectiveness of the MDP formulation in dialogue, using a learned context vector as the state. Practically, the framework offers a blueprint for building more engaging and reliable virtual assistants, with direct applications in customer service, personal AI, and accessibility tools.

Limitations and Future Work: The current evaluation relies on simulated user interactions. While robust, final validation requires large-scale human trials. Future work will focus on: (1) extending the framework to **multimodal** inputs (voice, vision), (2) testing in **cross-lingual and cultural** contexts, and (3) exploring **few-shot adaptation** to new domains with minimal data. By addressing these directions, the path toward truly natural and adaptive human-computer interaction can be further advanced.

REFERENCES

1. Li, J., Galley, M., Brockett, C., Gao, J., & Dolan, B. (2016). A diversity-promoting objective function for neural conversation models. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
2. Rani, S., Kumar, A., Singh, R., & Patel, M. (2023). Machine learning approaches for crop recommendation systems. *Precision Agriculture Technology Review, 12*(2), 89-104.
3. Serban, I. V., Sordoni, A., Bengio, Y., Courville, A., & Pineau, J. (2016). Building end-to-end dialogue systems using generative hierarchical neural network models. *Proceedings of the AAAI Conference on Artificial Intelligence, 30*(1).
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*.
5. Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36-45.
6. Wen, T. H., Vlachos, A., Mrkšić, N., Gasic, M., Ramachandran, D., Korokithakis, T., ... & Young, S. (2017). Toward multi-domain language generation using recurrent neural networks. *Proceedings of the 31st International Conference on Neural Information Processing Systems*.
7. Young, S., Gašić, M., Thomson, B., & Williams, J. D. (2013). POMDP-based statistical spoken dialog systems: A review. *Proceedings of the IEEE, 101*(5), 1160-1179.
8. Budzianowski, P., et al. (2018). MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.
9. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*.
10. Raffel, C., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research, 21*(140), 1-67.