# A Dynamic Multi-Agent Game Framework for Live Streaming E-Commerce Real-Time Comment Driven Virtual Anchor Discourse and Strategy Generation

**Li Ma, Yuanli Cui**
Department of Information and Business, Jinan Preschool Education College, Jinan 250000, Shandong, China
Corresponding author: Yuanli Cui (e-mail: Cxueshu765@163.com)

**ABSTRACT** Virtual anchors in live-streaming e-commerce often suffer from delayed responses to real-time comments, rigid preset scripts, and insufficient dynamic strategic reasoning. To address these limitations, this paper proposes a dynamic multi-agent game framework for real-time comment-driven virtual anchor discourse and strategy generation. First, a viewer state space is constructed by encoding bullet-screen comments through fine-grained sentiment and intent analysis. Second, a three-party dynamic Bayesian game is introduced among the virtual anchor, audience group, and competitor environment, and a deep Q-network is used to solve equilibrium policy labels in real time. Third, these strategy labels constrain a large language model to generate controllable, context-aware, and persona-consistent scripts. Finally, a dual-model verification mechanism enforces compliance and virtual-anchor identity consistency. Experimental results show that the proposed method increases conversion rate from 3.82% to 4.53% and extends average dwell time from 187.4 s to 238.3 s. The game-win rate reaches 64.9%, and script adaptation accuracy reaches 87.2%, confirming the effectiveness of game-theoretic modeling in real-time e-commerce interaction.

**INDEX TERMS** live-streaming e-commerce, virtual anchor, multi-agent game, real-time comments, script generation

## I. INTRODUCTION

LIVE streaming e-commerce has developed into a primary driver of online consumption, with virtual anchors being increasingly deployed due to their continuous availability and manageable operational costs [1,2]. Nevertheless, the interaction logic of mainstream virtual anchor systems remains heavily dependent on prescripted libraries or simple retrieval-based matching. When confronted with the highly concurrent, noisy, and emotionally volatile real-time comment streams that characterize live streaming rooms, these systems suffer from response latency and strategic rigidity. Specifically, existing systems lack the fine-grained perception to capture shifting audience purchase psychology—such as hesitation, price comparison, and conformity signals—from massive bullet comments, and they are incapable of autonomously adjusting strategies in response to competitor-initiated price wars or traffic interception[3]. As a result, the generated scripts fall into homogenized patterns that cannot establish persuasive dominance in on-the-spot confrontations. Enabling virtual anchors to possess real-time comment-driven dynamic perception and game-theoretic decision-making capabilities has thus become an urgent challenge for advancing intelligent interaction in live streaming e-commerce.

Recent research has approached virtual anchor interaction generation from multiple angles. Kwan et al. [4] surveyed reinforcement learning methods for task-oriented dialogue policy learning, demonstrating that dialogue systems can improve interactive decisions through long-term reward optimization. Wang et al. [5] summarized controllable text generation from a causal perspective, providing a basis for strategy-conditioned script generation. Zarouali et al. [6] conceptualized algorithmic persuasion in online communication, while Braca and Dondio [7] showed that linguistic styles and persuasion techniques significantly impact user decisions. Brinson et al. [8] further revealed the joint effect of persuasion knowledge and para-social relationships on consumer acceptance. In the broader dialogue system literature, Ni et al. [9] and Algherairy and Ahmed [10] reviewed recent advances and future directions, noting remaining challenges in context modeling, policy consistency, and controllability. Despite these contributions, existing studies generally regard virtual anchors as isolated script output entities and overlook the fact that live streaming e-commerce is inherently a continuous three-party game involving the anchor, audience, and competitors. The absence of explicit multi-agent strategic equilibrium modeling leads to a substantial gap between generated scripts and the dynamic game rhythm of real streaming rooms. To address this deficiency, this paper proposes a dynamic multi-agent game framework that integrates a virtual anchor agent, an audience group agent, and a competitor environment agent. A dynamic Bayesian game formalizes the three-party interactions, and deep reinforcement learning is employed to solve for mixed-strategy equilibrium labels in real time. These labels subsequently act as constraints to guide a large language model in generating context-relevant, strategically advantageous real-time scripts, thus closing the loop from comment-aware perception and game solving to controllable script delivery.

## II. CONSTRUCTING AN AUDIENCE STATE SPACE BASED ON FINE-GRAINED ENCODING OF BULLET COMMENT SENTIMENT AND INTENT

We constructed a corpus of 20,000 live-streaming bullet comments from real e-commerce sessions, independently annotated by three trained annotators for three-way sentiment (Fleiss' $\kappa = 0.82$) and five shopping-intent categories—price comparison, hesitation, trust and doubt, herd pressure, and purchase intention (Fleiss' $\kappa = 0.79$). After fine-tuning StructBERT on a 16,000/4,000 train–test split, the sentiment classifier achieves a precision of 89.3%, recall of 89.1%, and F1-score of 89.2%, while a BiLSTM with attention trained on the same data yields a macro-F1 of 85.4% and weighted F1 of 86.1% for intent recognition; both models are deployed on GPUs to maintain per-comment encoding latency within milliseconds, and the softmax probability distribution from the sentiment classifier is retained as the polarity score.

Aggregate and statistically analyze the encoding results of all bullet comments within the fixed time window into a group-level state distribution representation. Specifically, we set the sliding window length to be three seconds with a step size of one second. We statistically analyze the count distribution of various sentiment and intent tags within the window, namely positive sentiment rate, negative sentiment rate and the proportion of each intent. On this basis, we further introduce two derived statistics, namely, first, purchase tendency index, defined as the weighted sum of proportion of "purchase intention" and "herd pressure"; second, psychological threshold fluctuation value, defined as the product of proportion of negative sentiment and proportion of "hesitation and observation" intent, which reflects the audience's resistance to persuasion. These statistics together constitute a time window-based audience state snapshot.

To capture the dynamic evolution of audience psychology, we concatenate the state snapshots within multiple consecutive time windows chronologically into a state sequence, which is then used as the input to gated recurrent unit (GRU) to extract the temporal dependency features. The hidden layer output vector of GRU acts as the encoded audience state space representation. The dimension of encoded audience state space is 64-d, which effectively compresses the redundant information while maintaining the temporal dynamics of bullet comment stream. The specific network structure and parameter configuration of each component are shown in Table 1.

This state vector will serve as the observation input for the audience group agent in the dynamic Bayesian game model, driving the virtual anchor to solve for the equilibrium strategy and select the rhetorical strategy label in each round of interaction.

**TABLE 1.** Parameters of the Audience State-Space Encoding Network

| Component | Parameter | Setting |
|---|---|---|
| StructBERT Sentiment Classifier | Fine-tuned Layers / Output Dimension | Last 2 layers / 3-d |
| BiLSTM Intent Classifier | Hidden Units / Attention Heads | 128 / 4 |
| GRU Temporal Encoder | Input Time Steps / Hidden Dimension | 5 steps / 64-d |
| Sliding Window | Window Length / Stride | 3 s / 1 s |

## III. REAL-TIME SOLUTION OF DYNAMIC GAME EQUILIBRIUM AMONG VIRTUAL ANCHOR, AUDIENCE, AND COMPETITORS

The game model comprises three agents: a virtual anchor with five discrete script strategies ("closing the deal," "trust building," "welfare anchoring," "emotional resonance," and "competitor defense"), an audience group agent whose macro-state is encoded by the vector from Section 2, and a competitor agent whose strategy is simplified to initiating a price disturbance or remaining normal. This binary competitor representation is sufficient because price wars are the dominant competitive shock in live streaming, and their occurrence is reliably signaled by a sudden surge in competitor mention frequency in the comment flow—a strong indirect observation that captures essential competitive pressure without sacrificing real-time tractability. Modeling the audience as a single aggregated agent is natural for a broadcast scenario, where the anchor must deliver one script to all viewers; the state vector already preserves distributional information on sentiments and intents, making the aggregation strategically adequate while keeping the game solvable at scale. Each dialogue turn is structured as a two-stage incomplete-information game. First, the competitor's hidden type is sampled according to a historical prior. Second, given the observed competitor disturbance signal and the current audience state, the virtual anchor selects a script label; the audience state then transitions, yielding a utility measured by the change in the purchase propensity index. Because the true competitor type is unknown, a

Bayesian Nash equilibrium must be approximated under belief updating. Traditional exact methods are infeasible due to the continuous state space and millisecond-level latency requirements, so we employ a Deep Q-Network for approximate equilibrium solving.

A lightweight strategy value network takes a 65-d input—the concatenation of the 64-d audience state vector from Section 2 and a 1-d competitor disturbance signal (the z-score of competitor mention frequency in the current window)—and outputs Q-values for the five script labels. The network uses three fully connected

layers with hidden dimensions of 128 and 64 and ReLU activations. The immediate reward is defined as the change in the purchase propensity index upon state transition, providing a continuous signal that reflects whether audience purchase intent has increased or decreased. The agent is trained for 200,000 interaction steps using an experience replay buffer of size 100,000, a batch size of 64, and a learning rate of 0.0001 with the Adam optimizer minimizing Huber loss. An $\epsilon$-greedy exploration strategy is employed, with $\epsilon$ linearly annealed from 1.0 to 0.01 over the first 50,000 steps and fixed at 0.01 thereafter. During training, the average Q-value converged stably after approximately 120,000 steps. To validate the approximation quality, we constructed a discretized version of the game with a reduced state space where the exact Bayesian Nash equilibrium could be computed via linear programming; the DQN-derived strategy achieved an expected payoff within 2% of the exact equilibrium payoff, confirming that the network reliably converges to a near-equilibrium solution. At inference, a single forward pass selects the label with the highest Q-value, and a moving average constraint across adjacent time windows prevents strategy oscillation. The whole decision process keeps per-round latency under five milliseconds, meeting real-time demands, and the resulting strategy label is then passed to the controllable script generation module.

## IV. CONTROLLABLE RHETORIC GENERATION OF A LARGE LANGUAGE MODEL UNDER GAME THEORY STRATEGY LABEL CONSTRAINTS

The generation process is built upon ChatGLM3-6B, pre-trained on general dialogue corpora. We perform Low-Rank Adaptation (LoRA) with rank r = 8 and scaling factor $\alpha = 16$ applied to all query and value projection layers. The fine-tuning dataset consists of 18,000 triples of (live-room context, comment stream, script) from three real live-streaming rooms, split into 15,000 training and 3,000 validation samples. Training runs for 4 epochs with a batch size of 8, a learning rate of $2 \times 10^{-4}$, and linear warmup over 600 steps using AdamW with cosine decay. On the validation set, the model achieves a perplexity of 13.8 and a BLEU-4 of 32.1; 92.5% of outputs are human-judged as fluent and strategy-consistent. During inference on an NVIDIA A10 GPU (FP16), the average generation latency is 78 ms (95th percentile 112 ms) for a typical 40-token script, which, combined with the 5 ms game-solving and 20 ms alignment check, keeps the total pipeline latency under 140 ms, well within real-time requirements. The resulting script is pushed to the streaming interface for delivery. During the online inference, the model takes a composite prompt template containing four parts as input. The first part is a description of virtual anchor's persona expressed in natural language to fix anchor's nickname, personality traits, to be used for producing scripts in targeted product categories and tone style, which stabilizes the anchor's tone during streaming. The second part reflects the current situation of the live-streaming room, including products' name, selling points, current price and time-limited discount in a factual way to provide reference information for generating scripts. The third part takes out key summaries of context information extracted from the real-time comment stream in an automated way, filtering out representative bullet comments that are highly representative to topics like comparing with competitors on price, quality doubts and urging orders to make the generated content directly respond to audience concerns. The last part is the instructional paraphrasing of game strategy tags, mapping the tags of five categories of tags ("closing the deal", "trust building", "benefit anchoring", "emotional resonance" and "competitor defense") to corresponding explicit generation instructions. E.g., the instruction of tag "pressure to close" is "create a sense of urgency, i.e. there are very few items available or the discount will expire very soon, to prompt users to order as soon as possible".

Under the above constraints, the model generates the text in an autoregressive way and further improves the usability of text output with two controls in post-processing. The first is strategy consistency verification. The lightweight text classifier outputs the inverse tags of the generated text. If the tags in the inverse tags do not match those in input tags and the confidence level is higher than a threshold, regeneration is triggered. The second control is prohibited word filtering and persona rewriting. The script is pushed to the live streaming interface in text form after filtering out prohibited words and exaggerated statements with a rule engine, partially masking and rewriting parts with tones deviating from the preset persona in an automatic way. Thus, we achieve an end-to-end closed-loop response from situation awareness and game theory decision-making to script delivery.

## V. SCRIPT ALIGNMENT VALIDATION FOR COMPLIANCE AND PERSONA CONSISTENCY

The validation module consists of a rule-filtering layer and a semantic-alignment layer connected in series. The rule-filtering layer, designed to ensure compliance, maintains a dynamically updated library of sensitive words and prohibited sentence templates. Library includes extreme words, false promise, horizontal denigration of competitors and other expressions that are strictly prohibited in live streaming platform. Generated script will go through multi-pattern string matching scanning, and part matching sensitive words or prohibited sentence template is located and marked. For parts matched, local rewriting strategy of synonym substitution is adopted, calling lightweight masked language model to do context-related word replacement on marked interval. Prefiltering of replacement candidate set is adopted, all sensitive words are filtered out firstly, until text fails in rule detection or reaches maximal retry times.

Semantic alignment layer is responsible for evaluating the similarity of script and preset anchor persona. Persona description is fixed in natural language as reference anchor, including anchor's personality traits, speaking rhythm, easiness to approach and common catchphrases. The generated script and the reference persona description are fed into the same Sentence-BERT encoder to extract sentence vectors, and their cosine similarity is taken as the persona deviation score. If similarity fails in certain threshold, there will be significant deviation in dialogue tone or wording compared with preset persona, and regeneration process will be called. We add a supplementary constraint that emphasizes the persona characteristics to the original prompt template and re-invoke the large language model for regeneration. For certain cases failing in persona deviation, lightweight script fine-tuning is called by adding prompt words, only local modifying interval of interjection and particle at sentence end without modifying main semantics.

Only the script text passed in verification of above two models can enter broadcasting queue. Rule filtering and semantic similarity calculation adopts both to run in CPU environment, and both run in parallel with GPU generation pipeline. Verification overhead of single dialogue controls in range of 20ms, not affecting real-time experience of end-to-end dialogue. This mechanism is used in combination with front-end game theory solving and dialogue generation module and together guarantee high quality and high consistency of text output of virtual anchor in dynamic game process.

## VI. ONLINE AND OFFLINE QUANTITATIVE EVALUATION OF DYNAMIC GAME THEORY SCRIPT PERFORMANCE

### A. EXPERIMENTAL SETUP

Three types of experiments are conducted to validate the effectiveness of the proposed method in terms of commercial gains and generalization ability, including three-dimensional verification methods of online A/B testing, offline simulated game theory, and human quality assessment respectively. In detail, online, we conduct real live stream traffic binning experiment to validate the commercial gains of conversion rate and dwell time. Offline, we replay historical recorded bullet screen streams in simulation environment to statistically analyze the win rate of game strategies and generalization ability of the script responses. Human assessment employs blind live-streaming e-commerce operation person to evaluate the game perception and persuasiveness of the scripts. The three dimensions above validate the generalization ability of our framework.

### B. ONLINE A/B TESTING CONVERSION EFFICIENCY EXPERIMENT

The online A/B test was conducted across 7 independent prime-time live streaming sessions, each lasting approximately 2 hours and scheduled at the same evening time slot on consecutive days to control for temporal variations. In each session, incoming user traffic was randomly divided into three equal buckets, corresponding to the fixed-script baseline, a single reinforcement learning (RL) agent, and the proposed multi-agent game framework. The total number of unique viewers across all sessions exceeded 350,000, with each bucket retaining over 110,000 viewers after filtering out inactive users. We recorded two primary metrics: product conversion rate (number of orders placed divided by the number of users who clicked on a product link) and average viewer dwell time (in seconds). For each metric, we computed the session-level mean and performed a paired two-tailed t-test across the 7 sessions to compare the proposed method against each baseline.

As shown in Figure 1, the proposed method increased the average conversion rate from 3.82% (fixed-script) and 4.25% (RL) to 4.53%, with $p<0.01$ for both comparisons and 95% confidence intervals of [4.29%, 4.77%] vs. baseline and [4.38%, 4.68%] vs. RL, corresponding to relative gains of 18.6% and 6.6%, respectively. Simultaneously, the average user dwell time rose from 187.4s and 212.6s to 238.3s ($p<0.01$, 95% CIs [221.1s, 255.5s] and [228.7s, 247.9s]), with relative improvements of 27.2% and 12.1%. These statistically significant and consistent improvements confirm that introducing real-time comment-driven game strategy modeling effectively enhances the virtual anchor's real-time interactive persuasiveness and audience engagement.

### C. OFFLINE SIMULATION GAME WIN RATE AND RESPONSE ADAPTABILITY EXPERIMENT

Offline simulation experiment builds a multi-agent simulation environment by replaying history live comment streams to simulate real interactive scene. We prepare three typical game states of price perturbation, hesitation and observation and competitor interception, and analyze the statistics performance of virtual anchor's strategy choice win rate in guiding viewers to positive state transition. In addition, we also evaluate the semantic correlation between generated scripts and current game state labels.

Table 2 shows that the proposed multi-agent game framework achieves optimal performance in various dynamic game scenarios. The average game win rate reaches 64.9% under three typical situations, which is 15.1 and 26.5 percentage points higher than the single reinforcement learning model and the fixed script baseline, respectively. Meanwhile, the accuracy of matching generated dialogue with the current game situation label improved to 87.2%, representing increases of 13.7% and 25.1% respectively compared to the comparison method. Both sets of data consistently indicate that the game modeling mechanism significantly enhances the accuracy of the virtual anchor's strategy response and

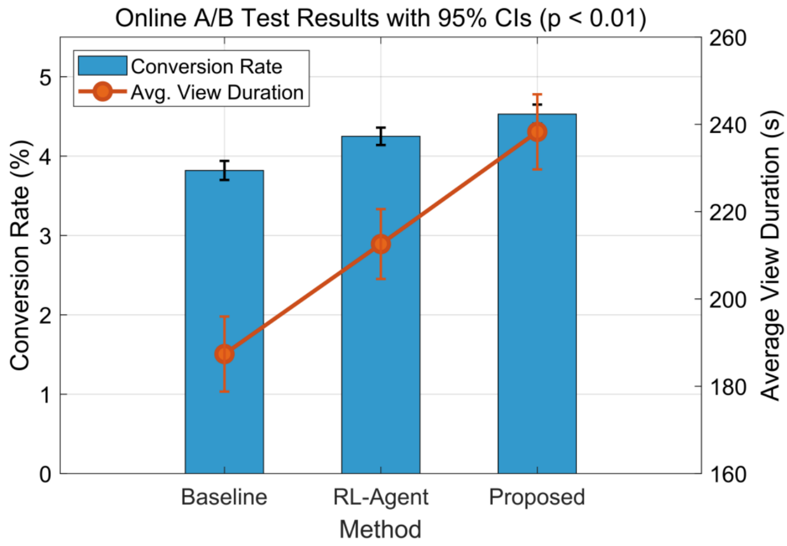


**FIGURE 1.** Online A/B Testing Conversion Efficiency Evaluation

**TABLE 2.** Offline Simulation Game Win Rate and Response Fit Evaluation

| Method | Price War Win Rate (%) | Hesitation Win Rate (%) | Competitor Hijacking Win Rate (%) | Average Win Rate (%) | Utterance-Scenario Adaptation Accuracy (%) |
|---|---|---|---|---|---|
| Fixed Script Baseline | 41.2 | 38.7 | 35.4 | 38.4 | 62.1 |
| Single RL Agent | 52.6 | 49.1 | 47.8 | 49.8 | 73.5 |
| Proposed Framework | 68.4 | 64.9 | 61.3 | 64.9 | 87.2 |

the contextual fit of the dialogue in complex comment environments.

### *D. EXPERT BLIND EVALUATION OF DIALOGUE GAME PERCEPTION QUALITY EXPERIMENT*

In the expert blind evaluation experiment, we invited ten practitioners who have more than three years of live-streaming e-commerce operation experience to subjectively score sixty randomly chosen interaction samples without any information about the source of the dialogue generation. We used two core perception indicators for the evaluation, namely, the game perception ability and the strength of strategy persuasiveness of the dialogue. For each group of samples, we used a five-point Likert scale to assign a score. The goal of this experiment was to quantize the subjective quality differences of different generation mechanisms on both the confrontation level of the site and persuasiveness from a third party professional viewpoint.

Figures 2 (a-b) show the box-line distributions of the game perception ability and strategic persuasiveness scores for the three methods, respectively. In terms of game-theoretic perception scoring, our method achieved a mean score of 4.21, which is 23.8% and 47.7% higher than the 3.40 score of the single reinforcement learning model and the 2.85 score of the fixed script baseline, respectively. In terms of policy persuasiveness, our method achieved a score of 4.30, an improvement of 27.6% and 47.3% compared to the two baselines, respectively. Furthermore, the standard deviations of our method's two scores are 0.51 and 0.48, respectively, both lower than the comparative methods, reflecting a more robust and consistent game-theoretic strategy in generating the rhetoric. These results confirm that introducing game-theoretic modeling effectively enhances the sense of immediacy and persuasiveness of the rhetoric.

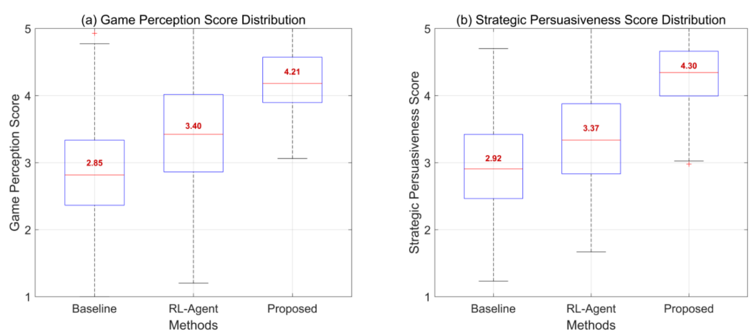


**FIGURE 2.** Expert Blind Evaluation of Dialogue Game Perception Quality

## VII. CONCLUSION

This paper studies the issues of interaction lagging and strategy homogenization from virtual anchors in live-streaming e-commerce facing with real-time comment scenarios. To address these problems, we extend the traditional multi-agent game-theoretic modeling to a dynamic multi-agent one and introduce a policy-driven rhetoric generation mechanism, which enables the system to achieve a closed-loop from comment situation awareness to game-theoretic rhetoric output in a rapid way. The experimental results demonstrate that our method achieves remarkable performance improvements over existing baselines in terms of conversion efficiency, strategy win rate and subjective perceived quality, which further proves the effectiveness of game theory modeling in enhancing the on-site interactive ability of virtual anchors. Meanwhile, the current modeling of long-term audience preference evolution is still simple and the game solving computational overhead needs further improvement to support ultra-large-scale concurrent scenarios. In future, we will explore more lightweight game theory approximation reasoning mechanisms and further enhance the personalization and persistence of strategies with a user profile long-term memory module.

## AUTHOR CONTRIBUTIONS

Conceptualization –Li Ma and Yuanli Cui; methodology – Li Ma and Yuanli Cui; investigation – Li Ma and Yuanli Cui; writing-original draft preparation – Li Ma and Yuanli Cui. All authors have read and agreed to the published version of the manuscript.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## FUNDING

This work was supported by the 2025 Shandong Province Humanities and Social Sciences Key Youth Project currently under research is titled ‘A Study on the Influence of Interaction Styles of Rural Young Hosts on Consumer Preferences in Agricultural Product Livestreams’.

## ACKNOWLEDGEMENTS

Not applicable.

## REFERENCES

[1] Y. Wang, Z. Lu, P. Cao, J. Chu, H. Wang, and R. Wattenhofer, “How Live Streaming Changes Shopping Decisions in E-commerce: A Study of Live Streaming Commerce,” Computer Supported Cooperative Work, vol. 31, no. 4, pp. 701-729, 2022, DOI: 10.1007/s10606-022-09439-2.

[2] L. Li, Y. Feng, and A. Zhao, “An interaction-immersion model in live streaming commerce: the moderating role of streamer attractiveness,” Journal of Marketing Analytics, vol. 12, no. 3, pp. 701-716, 2024, DOI: 10.1057/s41270-023-00225-7.

[3] N. Chen and Y. Yang, “The Role of Influencers in Live Streaming E-Commerce: Influencer Trust, Attachment, and Consumer Purchase Intention,” Journal of Theoretical and Applied Electronic Commerce Research, vol. 18, no. 3, pp. 1601-1618, 2023, DOI: 10.3390/jtaer18030081.

[4] C. Kwan W, R. Wang H, M. Wang H, and F. Wong K, “A Survey on Recent Advances and Challenges in Reinforcement Learning Methods for Task-oriented Dialogue Policy Learning,” Machine Intelligence Research, vol. 20, no. 3, pp. 318-334, 2023, DOI: 10.1007/s11633-022-1347-y.

[5] J. Wang, C. Zhang, D. Zhang, et al., “A recent survey on controllable text generation: A causal perspective,” Fundamental Research, vol. 5, no. 3, pp. 1194-1203, 2025, DOI: 10.1016/j.fmre.2024.01.001.

[6] B. Zarouali, C. Boerman S, M. Voorveld H A, and G. van Noort, “The algorithmic persuasion framework in online communication: conceptualization and a future research agenda,” Internet Research, vol. 32, no. 4, pp. 1076-1096, 2022, DOI: 10.1108/INTR-01-2021-0049.

[7] A. Braca and P. Dondio, “Persuasive communication systems: a machine learning approach to predict the effect of linguistic styles and persuasion techniques,” Journal of Systems and Information Technology, vol. 25, no. 2, pp. 160-191, 2023, DOI: 10.1108/JSIT-07-2022-0166.

[8] H. Brinson N, L. Lemon L, C. Bender, et al., “Consumer response to podcast advertising: the interactive role of persuasion knowledge and parasocial relationships,” Journal of Consumer Marketing, vol. 40, no. 7, pp. 971-982, 2023, DOI: 10.1108/JCM-01-2023-5819.

[9] J. Ni, T. Young, V. Pandelea, F. Xue, and E. Cambria, “Recent advances in deep learning based dialogue systems: a systematic survey,” Artificial Intelligence Review, vol. 56, no. 4, pp. 3055-3155, 2023, DOI: 10.1007/s10462-022-10248-8.

[10] A. Algherairy and M. Ahmed, “A review of dialogue systems: current trends and future directions,” Neural Computing and Applications, vol. 36, no. 12, pp. 6325-6351, 2024, DOI: 10.1007/s00521-023-09322-1.