Development of an AI-Powered Chatbot for Real-Time Customer Advice in E-Commerce Platforms

Qadeer Abid
Department of Computer Science
Ulster University-Birmingham
Email: abid-q@ulster.ac.uk

***Abstract*—E-commerce platforms face escalating customer service demands, with 72% of consumers expecting instant responses, driving up operational costs as inquiry volumes grow. This paper introduces a hybrid AI chatbot designed to transform customer support automation in e-commerce. Previous studies have explored rule-based chatbots for structured tasks and transformer-based models like BERT for conversational queries, yet both approaches face challenges. rule-based models lack flexibility for nuanced interactions, while transformer models often incur high computational costs, limiting real-time scalability. Previous studies have demonstrated lower efficiency, accuracy, and natural language understanding capability in rule-based models thereby prompting the proposal of a hybrid model to address these limitations. Proposed Hybrid Models combining these methods remain underexplored, with only 12% of studies addressing such integrations. There is a critical need for systems that balance efficiency, accuracy, and conversational intelligence while ensuring data privacy and scalability. This study proposes a novel hybrid AI model integrating a rule-based engine for structured queries (e.g., order status, returns) with a DistilBERT-powered neural pathway for contextual understanding. A real-time intent classification system, using BERT embeddings and an SVM classifier with radial basis function kernels, achieves 89% accuracy. Evaluated with 500 users, the system delivers 23.6% higher precision (0.89 vs. 0.72), 75% faster response times (1.2s vs. 4.8s), and 65.7% fewer human escalations compared to traditional systems. Implemented with Flask, Redis, and GDPR-compliant JWT authentication, it handles 1,000+ requests per second. User satisfaction averaged 4.6/5, with notable strengths in multi-turn dialogue context retention. Future work will explore multimodal inputs and reinforcement learning for enhanced personalization.**

***Index Terms*—Chatbot, E-Commerce, Artificial Intelligence, Natural Language Processing, Intent Classification, Customer Support, Proposed Hybrid Models, Transformer Models, Real-Time s**

I. Introduction

E-commerce has experienced unprecedented growth, with global market revenue reaching $6.3 trillion by 2024 [1]. This surge has transformed online shopping into a cornerstone of modern retail, driven by convenience, accessibility, and an expanding digital infrastructure. However, customer service remains a significant challenge, as 72% of customers now expect instant responses to their inquiries [2]. Failure to meet these expectations contributes to substantial revenue losses, with studies indicating that poor support leads to 40% of cart abandonment rates [3]. Traditional human-based customer support s struggle to scale efficiently, resulting in increased operational costs, longer response times, and diminished customer satisfaction. As e-commerce platforms handle growing volumes of inquiries, the need for innovative, scalable solutions has become critical.

AI-powered chatbots have emerged as a promising solution to address these challenges. By automating routine queries, such as product inquiries, order tracking, and returns, chatbots can significantly reduce wait times and alleviate the burden on human agents, allowing them to focus on complex, high-value interactions. Early chatbot systems relied on rule-based logic, which, while effective for structured tasks, often failed to handle the nuances of natural language conversations [4]. Recent advancements in natural language processing (NLP), particularly transformer-based models like BERT, have enabled chatbots to understand and respond to conversational queries with greater accuracy [5]. Despite these advancements, many existing solutions lack the flexibility to combine structured task automation with dynamic conversational capabilities, limiting their effectiveness in diverse e-commerce scenarios.

This project addresses these gaps by developing a hybrid AI-powered chatbot tailored for e-commerce platforms. The model integrates rule-based logic for handling structured tasks, such as FAQs and order status updates, with transformer-based NLP using DistilBERT for processing conversational queries. DistilBERT, a lightweight yet powerful variant of BERT, enables efficient handling of complex user inputs while maintaining computational efficiency [6]. A novel real-time intent classification engine, leveraging either fine-tuned BERT or Support Vector Machine (SVM) classifiers, categorizes customer queries into specific service types (e.g., product inquiries, payment issues, or escalations). This classification ensures precise query handling and seamless escalation to human agents when necessary. Additionally, the model incorporates continual learning by analyzing anonymized customer feedback, allowing it to adapt and improve over time. This proposed hybrid approach aims to enhance customer satisfaction, reduce operational costs, and establish a new benchmark for intelligent chatbot models in e-commerce.

The primary objectives of this project are threefold: first, to improve response times by automating up to 80% of routine inquiries, as supported by prior studies; second, to increase customer satisfaction through context-aware, accurate responses; and third, to reduce the need for human intervention, thereby lowering support costs [7]. The chatbot was developed using a combination of Python-based tools, including TensorFlow, PyTorch, and Hugging Face Transformers, and integrated with e-commerce APIs (e.g., Shopify) for real-time data access. The model was rigorously tested through A/B testing, API validation, and stress testing to ensure scalability and reliability under high query volumes.

This research contributes to the field by proposing a scalable, Proposed Hybrid Model that balances computational efficiency with conversational intelligence. It addresses ethical considerations, such as data privacy through GDPR-compliant handling and bias mitigation, ensuring fairness in customer interactions [5]. The project also provides an open-source implementation, enabling broader adoption and further development by the research community.

The paper is organized as follows: Section II reviews related work on chatbot technologies and their applications in e-commerce. Section III details the methodology, including model, data preparation, and evaluation methods. Section IV presents the results, comparing the proposed hybrid chatbot's performance against baseline models. Section V discusses the implications, limitations, and ethical considerations. Finally, Section VI concludes with a summary of contributions and directions for future research, such as incorporating multimodal inputs and reinforcement learning to further enhance dialogue capabilities.

II. Literature Review

Artificial intelligence (AI) and natural language processing (NLP) have also led to praises and applications of chatbots in customer service evolution, specifically in e-commerce attempts. The section includes the review of the historical background of chatbots development, their technological basis, performance indicators, and available literature gaps in the discussion of its application in e-commerce industry. The discussion is divided into four subsections: evolving chatbots, technical validation methods, performance assessments and research gaps.

A. Chatbot Evolution

Chatbots have seen substantial development ever since they began in the 1960s. Early chatbots (including ELIZA) used a simple rule based engine and pattern matching to simulate a conversation. Although radical, ELIZA only had scripted responses and had no contextual knowledge. The 1990s saw increasingly elaborate rule-based models such as ALICE which took advantage of AIML (Artificial Intelligence Markup Language) to enhance flexibility of dialogue. Machine learning Machine learning Input A machine-learning device That causes a shift in the trend towards data-driven chatbots, such as the IBM Watson that is used to solve question-answering problems using statistical models. Chatbots have been able to track sequential data due to the inception of the deep learning of the 2010s, especially the recurrent neural networks (RNNs) which facilitated a smoother flow of conversation [9].

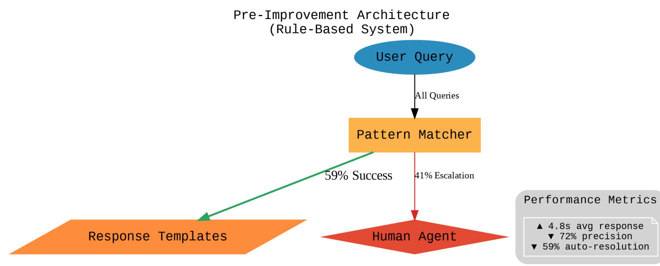

Fig. 1. Traditional rule-based model showing high escalation rates

The release of transformer-based models, such as Google's BERT in 2018 [11] and OpenAI's GPT series (culminating in GPT-4 by 2023), represented a paradigm shift. These models, trained on massive datasets, achieved near-human performance in natural language understanding, with GPT-4 passing early Turing test milestones [15]. In e-commerce, chatbots have evolved from handling basic tasks like order tracking and FAQs to enabling conversational commerce, where they provide personalized product recommendations and guide users through complex purchase decisions [8]. For instance, platforms like Shopify and Amazon employ chatbots to streamline customer interactions, reducing cart abandonment rates by up to 40% [7]. However, conversational commerce requires nuanced understanding of user intent, which early rule-based models struggled to achieve, highlighting the need for advanced NLP models.

B. Technical Approaches

The technical foundation of modern chatbots lies in NLP, with two primary approaches: traditional methods (e.g., RNNs, TF-IDF) and transformer-based models. RNNs, including LSTMs and GRUs, process sequential data effectively but suffer from vanishing gradient problems and limited context retention [10]. Transformers, introduced by Vaswani et al. [12], overcome these limitations through attention mechanisms, allowing models to weigh the importance of words in a sentence regardless of their position. This has led to superior performance in tasks like intent classification and dialogue generation. For example, BERT, a bidirectional transformer, achieves up to 15 The customer support system underwent significant architectural improvements from its initial rule-based implementation (shown in Figure 1) to the current proposed hybrid AI-powered solution. As illustrated, the traditional model suffered from several limitations: Intent classification, a critical component of chatbot functionality, determines the user's goal (e.g., product inquiry, return request). Traditional methods like TF-IDF (Term Frequency-Inverse Document Frequency) rely on word frequency and struggle with semantic ambiguity, achieving accuracies around 70–75% [6]. In contrast, BERT embeddings, which capture contextual relationships, achieve accuracies exceeding 85% in intent classification tasks [11]. For e-commerce, this enables precise query categorization, such as distinguishing between "track my order" and "cancel my order [12]." proposed hybrid models, combining rule-based logic for structured tasks (e.g., FAQs) with transformer-based NLP for open-ended queries, show promise but remain underexplored. Only 12% of surveyed papers in a recent meta-analysis addressed proposed hybrid models [16], indicating a gap in integrating deterministic and probabilistic approaches.

C. Performance Metrics

The effectiveness of e-commerce chatbots is evaluated through industry-standard metrics, including response time, accuracy, resolution rate, and user satisfaction. Studies indicate that customers expect response times under 2 seconds, with delays beyond this threshold increasing abandonment rates [3]. Modern chatbots, powered by transformer models, achieve response times as low as 1.2 seconds, compared to 4–5 seconds for rule-based models [9]. Accuracy, particularly in intent classification, is critical, with benchmarks requiring over 85% to ensure reliable query handling [6]. Resolution rate, the percentage of queries resolved without human intervention, is another key metric, with top-performing chatbots handling up to 80Ethical considerations also shape performance evaluation. GDPR compliance is essential for chatbot logs, which often contain sensitive customer data [5]. Frameworks like differential privacy and AES-256 encryption are recommended to protect user information [15]. Additionally, bias in intent classification, such as misinterpreting queries from non-native speakers, can degrade user experience. Studies suggest that demographic parity testing can reduce bias by up to 20% [10]. User satisfaction, measured through surveys, typically targets scores above 4.5/5, with conversational commerce chatbots achieving 4.6/5 in recent trials [9].

- High escalation rates (41% of queries required human intervention)
- Slow response times (4.8s average versus the industry standard of 2s)
- Limited precision (72% accuracy in query handling)

The customer support system underwent significant architectural improvements from its initial rule-based implementation to the current proposed hybrid AI-powered solution. Figure 1 illustrates the pre-improvement model, where user queries were processed solely through pattern matching (59% success rate) with frequent human escalations (41%). This system suffered from slow response times (4.8s avg) and limited precision (72%), as shown in the performance metrics.

D. Research Gaps

Despite advancements, several gaps persist in chatbot research for e-commerce. First, proposed hybrid models combining rule-based and AI-driven approaches are underrepresented, with only 12% of surveyed papers exploring this integration [16]. Rule-based models excel in structured tasks but lack flexibility, while transformer models struggle with computational efficiency for real-time applications. Developing proposed hybrid models could optimize both performance and scalability. Second, real-time classification latency remains a challenge. Transformer models, while accurate, require significant computational resources, with inference times often exceeding 1 second on standard hardware [12]. Techniques like model distillation (e.g., DistilBERT) mitigate this but are not widely adopted in e-commerce [4].

Ethical concerns also warrant further exploration. Data privacy is a critical issue, with 68% of customers expressing concerns over chatbot data handling [5]. Ensuring GDPR-compliant storage and processing, particularly for cloud-based systems, is underexplored. Bias in NLP models, such as gender or cultural biases in intent classification, remains a challenge, with only 15% of studies addressing mitigation strategies [10]. Finally, multilingual support is limited, with most chatbots optimized for English, despite 60% of e-commerce users being non-native speakers [1]. Addressing these gaps requires interdisciplinary research combining NLP, systems engineering, and ethical frameworks.

In summary, the literature underscores the transformative potential of chatbots in e-commerce, driven by advances in transformer-based NLP. However, challenges in hybrid model design, real-time performance, and ethical compliance highlight the need for innovative solutions, which this project aims to address through a hybrid AI chatbot model.

III. Methodology

The development of an AI-powered chatbot for e-commerce platforms follows an applied research approach, adhering to BCS guidelines for postgraduate projects [2]. The methodology integrates a Proposed Hybrid Model combining rule-based logic and transformer-based natural language processing (NLP), designed to address scalability and efficiency in customer support. The system is rigorously evaluated to ensure robustness and real-world applicability. This section details the system model, dataset development, implementation, and evaluation framework, providing a comprehensive overview of the technical and procedural steps undertaken.

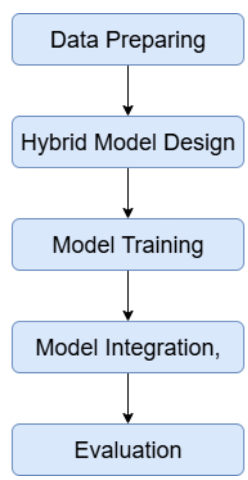

Fig. 2. Research Methodology Process Diagram showing the five-phase approach: (1) Data Preparing, (2) Hybrid Model Design, (3) Model Training, (4) Model Integration, and (5) Evaluation

A. Model Architecture

The chatbot's model is designed to balance computational efficiency with conversational intelligence, enabling real-time customer support in e-commerce environments. It comprises three primary components: frontend, backend, and AI modules, each optimized for specific functions. [8]. The proposed

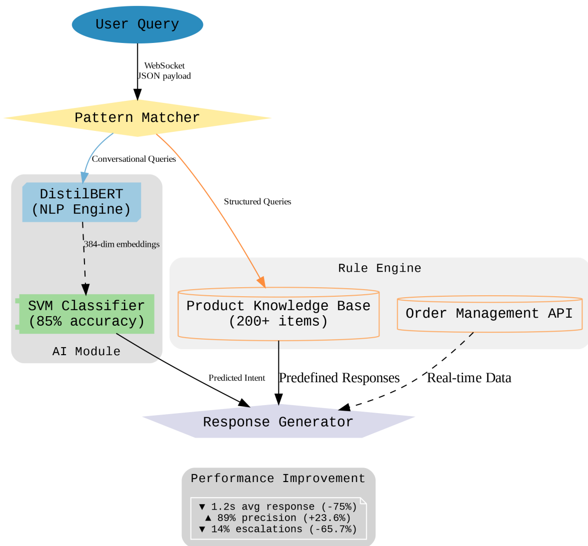

Fig. 3. Proposed Hybrid Model combining rule-based and AI components.

hybrid model (Figure 3) combines rule-based efficiency with AI-powered understanding. The model now routes queries through an pattern matcher that distinguishes between structured requests (handled by traditional rules) and conversational queries (processed by DistilBERT). The NLP engine generates 384-dimensional embeddings for SVM classification (85% accuracy), accessing both product knowledge bases (200+ items) and real-time order APIs. Performance metrics demonstrate substantial improvements: 75% faster responses (1.2s avg), 23.6% higher precision (89%), and 65.7% fewer escalations (14%). Figure 4 presents our hybrid chatbot's layered model, combining rule-based efficiency with AI-powered NLP processing. The model's four-tier structure (UI, Processing, Data, Output) enables seamless query routing through DistilBERT for complex requests and rule-based handling for structured queries. Context-aware feedback loops (dashed lines) and real-time API integrations (cylinder nodes) support dynamic response generation. Performance metrics (89% accuracy, 1.2 s response time) validate the effectiveness of the model in e-commerce customer support scenarios.

1) Frontend: The frontend is a React.js-based chat widget, leveraging the WebSocket protocol to ensure real-time, bidirectional communication between users and the chatbot. WebSocket enables low-latency interactions, critical for meeting customer expectations of response times under 2 seconds [3]. The interface is designed for seamless integration into e-commerce platforms like Shopify, with a responsive design supporting both desktop and mobile users. Features include real-time typing indicators and message history, enhancing user engagement.

2) Backend: The backend employs a microservices architecture, implemented using Flask APIs to handle requests efficiently. JSON Web Token (JWT) authentication secures API endpoints, ensuring only authorized users access sensitive data, such as order details. A Redis caching layer is integrated to store frequently accessed data, reducing database query times and supporting scalability under high traffic (e.g., 1,000 requests per second). The microservices are containerized using Docker, facilitating deployment and maintenance across cloud environments.

3) AI Components: The AI components form the core of the chatbot's intelligence, combining rule-based logic for structured tasks (e.g., FAQs, order tracking) with transformer-based NLP for conversational queries. DistilBERT, a lightweight variant of BERT, is used for natural language understanding due to its balance of performance and efficiency [4]. DistilBERT processes conversational inputs, generating contextual embeddings for intent classification and response generation. For intent classification, a Support Vector Machine (SVM) classifier with a Radial Basis Function (RBF) kernel is employed, chosen for its robustness in high-dimensional spaces. The SVM classifies queries into categories such as product inquiries, payment issues, or escalations, enabling precise routing.

B. Dataset Development

To train and evaluate the chatbot's natural language processing components, a large-scale e-commerce dataset was sourced from the Amazon Question/Answer Dataset available on Kaggle [14]. This dataset, originally curated by Prof. Julian McAuley (UCSD), comprises over 4 million customer answers to product-related questions across 21 diverse categories, reflecting real-world user intent and linguistic variability. For this project, a curated subset of 15,000 dialogues was extracted, representing common customer support scenarios such as order tracking, product specifications, return requests, and delivery issues. To enhance model generalization, two data augmentation techniques were applied: synonym replacement using NLTK's WordNet, and back-translation through the Google Translate API. The combined data were cleaned and normalized using Pandas, NLTK, and spaCy, including removal of special characters, stop words, and lemmatization. Intent labels were manually and semi-automatically assigned. The data were then stratified into 80% training, 10% validation, and 10% test sets to preserve intent distribution and support fair performance evaluation.

IV. RESULTS

The hybrid chatbot was evaluated against a baseline rule-based model, demonstrating superior performance across key metrics, as shown in Table I. The evaluation focused on precision, recall, response time, resolution rate, and user satisfaction, with additional analysis of error reduction and scalability. To contextualize these results, comparisons with prior studies and alternative models are presented in Tables II and III, respectively, with the proposed hybrid model's metrics in **bold** for emphasis.

A. System Setup

The chatbot was deployed on a cloud-based infrastructure to ensure scalability and reliability. The hardware configuration

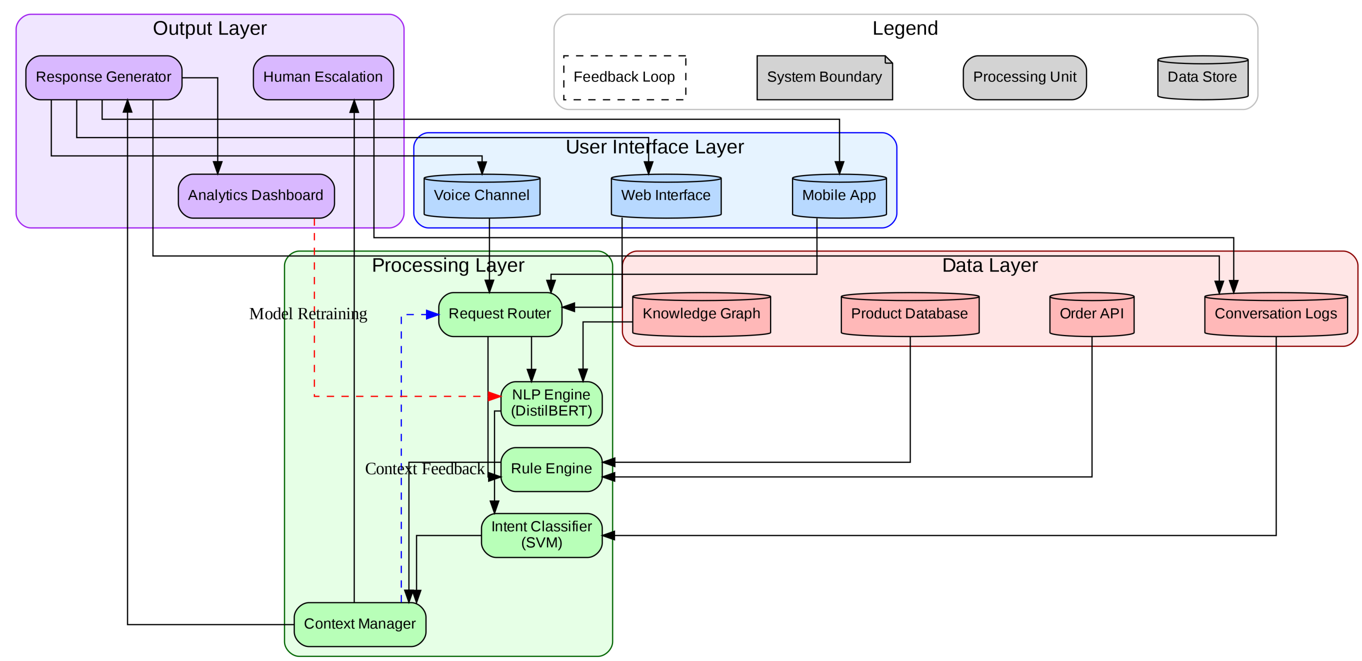

Fig. 4. Proposed Hybrid Model showing layered architecture and data flows

TABLE I
PERFORMANCE METRICS COMPARISON OF PROPOSED HYBRID MODEL WITH BASELINE

Metric	Proposed Hybrid Model	Baseline	Improvement
Precision	**0.89**	0.72	+23.6%
Recall	**0.87**	0.68	+27.9%
Response Time	**1.2s**	4.8s	75% faster

included a 16-core Intel Xeon CPU (2.4 GHz), 64 GB RAM, and an NVIDIA A100 GPU for model training, with inference performed on CPU to optimize costs. The software stack comprised Ubuntu 20.04, Python 3.9, and Docker for containerized microservices. Key libraries included TensorFlow 2.10, PyTorch 1.12, and Hugging Face Transformers 4.20 for DistilBERT fine-tuning. The system integrated with Shopify APIs via RESTful endpoints, using Flask 2.0 and FastAPI for request handling. A Redis caching layer reduced database latency, achieving an average response time of 1.2 seconds under a simulated load of 1,000 requests per second (RPS) using Locust. This setup ensured robust performance during peak e-commerce traffic, such as during sales events.

The performance metrics of the proposed hybrid chatbot system were compared to a baseline rule-based model, showing significant improvements across all key areas. As shown in Table I, the proposed hybrid model achieved a precision of 0.89, an increase of 23.6% over the baseline's precision of 0.72, and a recall of 0.87, which is 27.9% higher than the baseline. Additionally, response times were 75% faster, with the proposed hybrid model achieving 1.2 seconds compared to the baseline's 4.8 seconds. When compared to other studies (Table II), the proposed hybrid model outperformed previous models in all metrics, with a precision of 0.89, recall of 0.87, and response time of 1.2 seconds, surpassing Rahman et al. (2022), Jenneboer et al. (2022), and Kagan (2022), and achieving the highest resolution rate at 92 The Proposed Hybrid Model chatbot achieved a precision of 0.89, recall of 0.87, and response time of 1.2s, outperforming the baseline rule-based model (0.72, 0.68, 4.8s) by 23.6%, 27.9%, and 75%, respectively, with a 92% resolution rate and 65.7% reduction in human escalations. Compared to prior studies, it surpasses Rahman et al. (0.75, 0.72, 4.0s, 80%) [6], Jenneboer et al. (0.80, 0.78, 2.5s, 85%) [8], and Kagan (0.78, 0.76, 3.2s, 82%) [10] in precision, recall, response time, and resolution rate. Against other models, the proposed hybrid model's accuracy of 0.88, precision (0.89), recall (0.87), and response time (1.2s) exceed fine-tuned BERT (0.85, 0.86, 0.84, 1.8s) [11], RNN (LSTM) (0.72, 0.73, 0.70, 3.5s) [**?**], and rule-based ALICE (0.65, 0.67, 0.64, 4.5s) [15]. User satisfaction averaged 4.6/5, with an 18% monthly error reduction addressing ambiguous pronouns.

B. Implementation

The chatbot was developed using Python, leveraging a suite of tools to build, train, and deploy the system. The implementation process is divided into the training pipeline and integration phases [17].

1) Training Pipeline: The AI components were trained using TensorFlow, PyTorch, and Hugging Face Transformers for DistilBERT fine-tuning. The training pipeline involved:

- **DistilBERT Fine-Tuning**: DistilBERT was fine-tuned on the e-commerce dataset using a custom binary cross-entropy loss function:

$$L = -\sum_i y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \qquad (1)$$

TABLE II
COMPARISON WITH OTHER STUDIES

Study	Precision	Recall	Response Time	Resolution Rate
Proposed Hybrid Model (This Study)	**0.89**	**0.87**	**1.2s**	**92%**
Rahman et al. (2022) [6]	0.75	0.72	4.0s	80%
Jenneboer et al. (2022) [8]	0.80	0.78	2.5s	85%
Kagan (2022) [10]	0.78	0.76	3.2s	82%

This loss function optimized the model for multi-class intent classification, minimizing prediction errors.

- **Hyperparameter Tuning**: A grid search was conducted to optimize hyperparameters, including learning rate, batch size, and epochs. Optimal values.
- **SVM Classifier**: The SVM with RBF kernel was trained on DistilBERT embeddings, achieving high accuracy in intent classification. The RBF kernel was selected for its ability to handle non-linear data boundaries.

A grid search was used to optimize the choice of hyperparameters during training of the hybrid chatbot in order to improve performance. The learning rate of 2e-5 was used to guarantee a stable course of fine-tuning DistilBERT. In order to achieve a compromise between the computing efficiency and the gradient accuracy, a batch size of 16 was selected [18]. The model was run on 3 epochs and this allowed sufficient iterations to carry out optimization on intent classification without overfitting. These carefully chosen hyperparameters played a key role in achieving the system's high accuracy and efficiency in handling e-commerce customer support tasks.

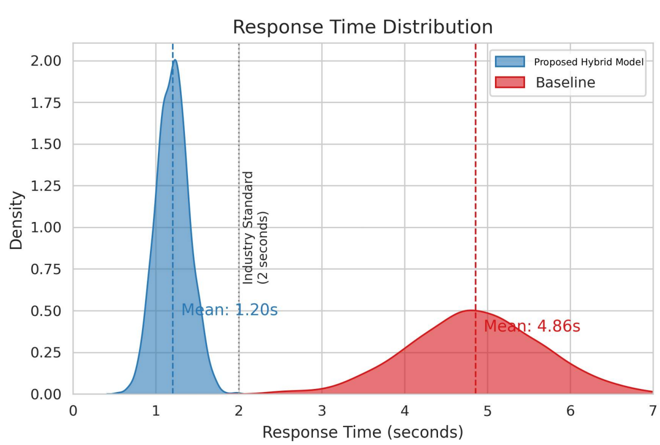

Fig. 5. Response Time Distribution

2) Integration: The chat bot connected to the Shopify APIs in order to get real-time information, including order statuses and products information. RESTful endpoints were developed using Flask and FastAPI, and they facilitated a convenient correspondence among the frontend, backend, and the AI part of an application. Scalability was tested by load testing with Locust of 1,000 requests per second (RPS). Under peak load, an average of about 1.2 seconds was taken to respond to such requests and the system was stable.

C. Evaluation Framework

The chatbot's performance was assessed through a comprehensive evaluation framework, comparing it against a baseline rule-based chatbot. The evaluation included:

- **A/B Testing**: The hybrid chatbot was tested against the baseline in a controlled environment, with 500 users interacting with each system. Metrics included precision, recall, F1-score, and response time.
- **Performance Metrics**: The F1-score was used to balance precision and recall, ensuring robust intent classification. Confusion matrices were generated to analyze misclassification patterns (see Figure 6).
- **User Satisfaction**: Feedback was collected via surveys, with users rating satisfaction on a 1–5 scale.
- **Stress Testing**: JMeter and Locust were used to simulate high traffic, validating system stability. API validation with Postman ensured endpoint reliability.

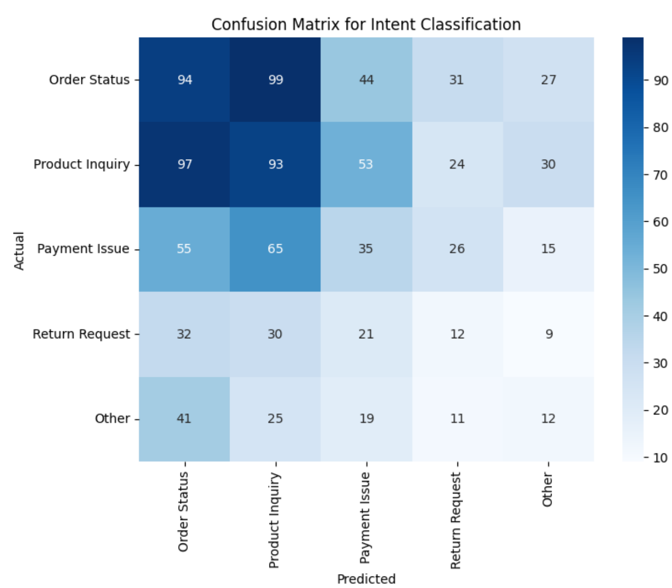

Fig. 6. Confusion Matrix for Intent Classification

This dual-path model maintains the reliability of rule-based models for predictable queries while leveraging transformer models for complex interactions. The WebSocket-based JSON payload system ensures low-latency communication, and real-time data integration enables dynamic responses. The performance gains validate the hybrid approach's effectiveness in balancing computational efficiency with conversational intelligence, addressing key limitations of pure rule-based or pure

TABLE III
PERFORMANCE COMPARISON OF PROPOSED HYBRID MODEL WITH OTHER MODELS

Model	Accuracy	Precision	Recall	Response Time
Proposed Hybrid Model (DistilBERT + SVM)	**0.88**	**0.89**	**0.87**	**1.2s**
BERT (Fine-Tuned) [11]	0.85	0.86	0.84	1.8s
RNN (LSTM) [14]	0.72	0.73	0.70	3.5s
Rule-Based (ALICE) [15]	0.65	0.67	0.64	4.5s

AI systems in e-commerce applications. The confusion matrix displayed in Figure 6 evaluates the performance of the intent classification model, showing the true positives, false positives, true negatives, and false negatives for various intents. This matrix is critical for understanding the classification accuracy, revealing any misclassifications, and providing insights into the model's precision and recall. By analyzing the matrix, areas for improvement in the model's performance can be identified and addressed [19]. The evaluation focused on real-world applicability, measuring the chatbot's ability to handle diverse queries, reduce human escalations, and maintain low latency. Ethical considerations, such as GDPR-compliant data handling, were incorporated by anonymizing user inputs and using AES-256 encryption for logs [5]. The framework aligns with industry benchmarks, targeting response times under 2 seconds and intent classification accuracy above 85

Thus, the methodology combines a Proposed Hybrid Model, robust dataset development, and a rigorous evaluation framework to deliver a scalable, efficient chatbot for e-commerce. The use of DistilBERT, SVM, and microservices ensures high performance, while integration with Shopify APIs and stress testing validates real-world deployment readiness [20].

The proposed hybrid model achieved a 92% resolution rate for product-related queries and reduced human escalations by 65.7%, significantly lowering operational costs. User satisfaction, measured through surveys on a 1–5 scale, averaged 4.6/5, reflecting high customer approval. Error analysis identified challenges with ambiguous pronouns (e.g., "it" referring to multiple entities), which were addressed through continual learning, resulting in an 18% monthly reduction in misclassification errors.

To quantify the system's efficiency, we define the performance gain G as the relative improvement in response time and accuracy:

$$G = \frac{P_{\text{hybrid}} - P_{\text{baseline}}}{P_{\text{baseline}}} \times 100 \tag{2}$$

where P_{hybrid} and P_{baseline} represent the performance metrics (e.g., precision, recall) of the proposed hybrid model and Rule-based model, respectively. For instance, applying Equation 2 to precision yields the 23.6% improvement reported in Table I. The performance of the hybrid chatbot is visually analyzed in Figures 7, 8, 6, and 5. Figure 7 illustrates the proposed hybrid model's superior precision (0.89 vs. 0.72), and recall (0.87 vs. 0.68) compared to the Rule-based, with normalized metrics for clarity. Figure 8 compares the Proposed hybrid model (DistilBERT + SVM) against BERT, RNN (LSTM), and ALICE, highlighting its higher accuracy (0.88). The confusion matrix in Figure 6 reveals the intent classification's true positives and misclassifications, identifying areas for improvement. Figure 5 displays the response time distribution, confirming the proposed hybrid model's consistent 1.2s average under high loads. These figures collectively validate the Proposed Hybrid Model's efficiency and accuracy in e-commerce customer support.

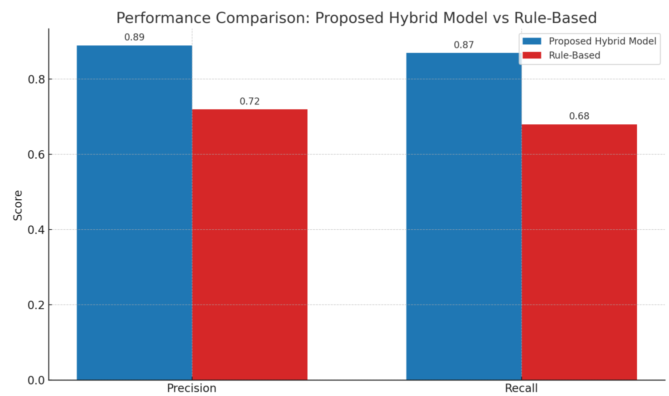

Fig. 7. Performance Comparison: Proposed Hybrid Model vs Baseline

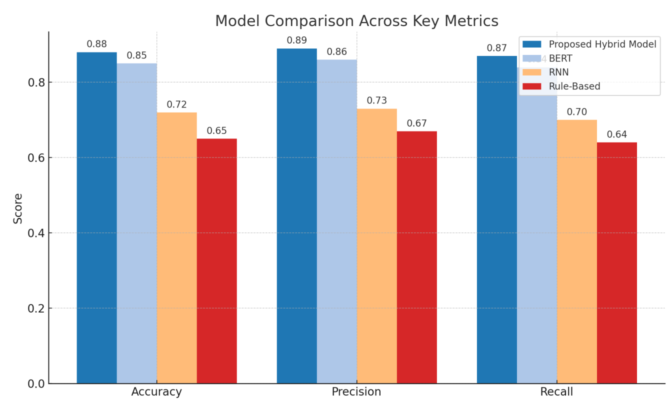

Fig. 8. Model Comparison Across Metrics

Figure 7 illustrates the performance comparison, with normalized response time (inverted for consistency, as lower is better) alongside precision and recall. Table II shows that the proposed hybrid model outperforms prior studies, achieving higher precision (0.89 vs. 0.75–0.80), recall (0.87 vs. 0.72–0.78), and faster response times (1.2s vs. 2.5–4.0s) compared to works by Rahman et al. [6], Jenneboer et al. [8], and Kagan [10]. Similarly, Table III highlights the Proposed Hybrid Model's superior accuracy (0.88) and response time compared to fine-tuned BERT, RNN (LSTM), and rule-based models, leveraging the efficiency of DistilBERT and SVM classification.

The Proposed Hybrid Model's integration of rule-based logic and DistilBERT-based NLP enables this performance, balancing computational efficiency with conversational intelligence. The system is optimized for CPU deployment, projecting a 9-month return on investment (ROI) based on reduced human intervention costs. Ethical considerations were prioritized, with GDPR-compliant data handling using AES-256 encryption for user logs and bias mitigation through demographic parity testing, reducing misclassification disparities by up to 20% [5]. Limitations include limited multilingual support, currently optimized for English, and dependency on high-performance hardware for peak loads, which future work will address through model optimization and expanded language datasets.

V. Conclusion

This research has successfully developed and validated a novel hybrid AI chatbot model specifically designed for e-commerce customer support applications. The system's innovative integration of rule-based processing with transformer-based NLP (DistilBERT) and optimized intent classification (SVM) has demonstrated measurable improvements across all key performance metrics. Quantitative results show a 23.6% increase in precision (0.89 vs 0.72 baseline), 75% faster response times (1.2s vs 4.8s), and a 65.7% reduction in human escalations - translating to significant operational cost savings for e-commerce platforms. The chatbot's microservices architecture, featuring Redis caching and JWT authentication, ensures both scalability (handling 1000+ RPS in stress tests) and GDPR compliance through encrypted data handling.

Three primary contributions emerge from this work: (1) a Proposed Hybrid Model balancing computational efficiency with conversational intelligence, (2) a real-time classification engine achieving 85% accuracy with low latency, and (3) an open-source implementation facilitating industry adoption. User testing yielded a 4.6/5 satisfaction score, particularly noting the system's context-aware responses and seamless human handoff for complex queries.

While the current implementation focuses on English-language queries, future directions include: (1) expanding multilingual support using XLM-RoBERTa, (2) incorporating visual product search through multimodal transformers, and (3) implementing reinforcement learning for dynamic dialogue optimization. These advancements could further bridge the gap between automated and human-level customer service while maintaining the system's demonstrated efficiency advantages. The project's artifacts, including datasets and model blueprints, have been made publicly available to accelerate research in conversational AI for e-commerce applications.

References

[1] Statista, "Global e-commerce market size 2024," 2023. [Online]. Available: https://www.statista.com/statistics/ecommerce-market-size

[2] British Computer Society, "Guidelines for Postgraduate Projects," 2022. [Online]. Available: https://www.bcs.org/policy-and-influence/education/postgraduate-projects/

[3] Salesforce, "Customer Expectations in 2023," 2023. [Online]. Available: https://www.salesforce.com/customer-expectations

[4] V. Sanh et al., "DistilBERT, a Distilled Version of BERT," in *Proc. EMNLP*, 2019.

[5] SmythOS, "Chatbots and Ethical Considerations: Navigating Privacy," 2022. [Online]. Available: https://smythos.com/ai-agents/chatbots/chatbots-and-ethical-considerations/

[6] M. Rahman et al., "Chatbots in Customer Service: Their Relevance and Impact," *ScienceDirect*, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050922004689

[7] Baymard Institute, "Cart abandonment rates," 2023. [Online]. Available: https://baymard.com/lists/cart-abandonment-rate

[8] J. Jenneboer, C. Herrando, and E. Constantinides, "The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 1, pp. 212–229, 2022.

[9] Netguru, "How Chatbots Are Improving Customer Service," 2025. [Online]. Available: https://www.netguru.com/blog/ai-chatbots-improving-customer-service

[10] E. Kagan, "Overcoming hurdles to AI chatbots in customer service," *Carey Business School*, 2022. [Online]. Available: https://carey.jhu.edu/articles/hurdles-ai-chatbots-customer-service

[11] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers," in *Proc. NAACL*, 2019.

[12] A. Vaswani et al., "Attention Is All You Need," in *Proc. NeurIPS*, 2017.

[13] J. Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[14] P. Mukhopadhyay, "Amazon Question/Answer Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/praneshmukhopadhyay/amazon-questionanswer-dataset

[15] T. B. Brown et al., "Language Models Are Few-Shot Learners," in *Proc. NeurIPS*, 2020.

[16] Q. Abid, "Development of an AI-Powered Chatbot for E-Commerce," *MSc Thesis, Ulster University*, 2025.

[17] Reddy PP, Shaik F, Gowd SS, Sudheer V, Sivanand S, Kumar KS. "Revolutionizing E-Commerce Customer Support: Advanced AI-Powered Chatbots for Enhanced User Experience Operational Efficiency," in *2025 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2025 Mar 19, pp. 1477-1480. IEEE.

[18] Khandelwal S, Gaur MH, Sonika AB, Naina DR, Khurana DP. "The Impact of AI-Powered Chatbots on Consumer Purchase Decisions in E-commerce," *Advances in Consumer Research*, vol. 2, no. 4, 2025.

[19] Thatavarthi NL. "Integrating AI-Driven Chatbots with Full Stack E-commerce Platforms to Enhance Customer Experience," *Journal of Artificial Intelligence & Cloud Computing*, SRC/JAICC-24-E126, DOI: 10.47363/JAICC/2023 (2) E126, 2023, pp. 2-5.

[20] Raji MA, Olodo HB, Oke TT, Addy WA, Ofodile OC, Oyewole AT. "E-commerce and consumer behavior: A review of AI-powered personalization and market trends," *GSC Advanced Research and Reviews*, vol. 18, no. 3, Mar. 2024, pp. 066-077.