Implementation of IndoBERT-Based Chatbot for Information and Marketing Services on the Griya Alam Mandiri Website

Wahyu Sukma Jayawardhana[1*], Daniel Rudiaman Sijabat[2], Mukhlis Amien[3]

[1,2,3] Bhinneka Nusantara University, Raya Tidar Street No. 100, Karangbesuki, Sukun District, Malang City, East Java 65146, Indonesia

Keywords

Chatbot; Digital Marketing; IndoBERT; Natural Language Processing; UAT

***Correspondence Email:**
wahyusukmajw@gmail.com*

Abstract

Property information services at Griya Alam Mandiri currently rely on manual WhatsApp interactions during working hours, often leading to delayed responses for prospective buyers. This study aims to implement a Natural Language Processing (NLP) based chatbot using the IndoBERT model on the housing website to automate 24/7 customer service. The methodology involves extracting 3,082 training data samples augmented with typos and informal language, model fine-tuning, and benchmarking against comparative architectures. The evaluation results show that the IndoBERT Base model highly excels in intent classification with an accuracy of 95.85% and an F1-Score of 95.87%, outperforming Indonesian RoBERTa, IndoBERTweet, and XLM-RoBERTa. This system was successfully integrated into a Client-Server architecture connecting a React interface and a FastAPI backend. The final evaluation through User Acceptance Testing (UAT) involving 50 respondents achieved an overall feasibility percentage of 82.50% (Very Satisfied). In conclusion, the IndoBERT chatbot is proven effective in improving the speed of real-time information services and successfully reduces administrative workloads. Future research suggestions include integrating Large Language Models (LLM) with the Retrieval-Augmented Generation (RAG) method to generate more dynamic responses.

1. Introduction

The housing and residential sector in Indonesia serves as an essential foundation for national economic progress, regional development, and societal well-being. Increased urbanization, consistent population growth, and the community's demand for proper housing continue to fuel residential property requirements annually. This continuous demand aligns seamlessly with the regulatory framework governing housing accessibility, financial subsidies, and development safeguards established by national policies [1]. In the contemporary digital era, property marketing relies significantly on online engagement, establishing digital platforms, social networks, and responsive websites as essential mechanisms to interface with prospective property purchasers [2]. Digital marketing strategies are no longer optional but mandatory for developers looking to expand their market reach and maintain a competitive edge in a saturated real estate market. However, practical operations at Perumahan Griya Alam Mandiri demonstrate that customer service workflows are severely constrained by traditional methodologies. Currently, all customer inquiries ranging from basic unit pricing and location details to complex mortgage requirements and legal certifications are handled via manual communication over WhatsApp during limited operational hours. This localized infrastructure results in systemic bottlenecks, significantly prolonged response times, and an overloaded administrative staff. When prospective buyers inquire during evenings, weekends, or national holidays, the delay in response frequently results in a loss of customer interest, directly impacting the effectiveness of the digital marketing funnel and potentially causing revenue loss.

Adopting Artificial Intelligence (AI) frameworks combined with Natural Language Processing (NLP) within corporate customer service has emerged as a crucial approach to elevate speed, precision, and customer satisfaction metrics [3]. Past implementations have successfully deployed NLP virtual assistants to optimize diverse processing applications including public administration portals [4], academic operations [5], and general chatbot implementations [6]. Chatbots have evolved significantly from simple rule-based ELIZA-like structures to advanced conversational AI agents capable of contextual reasoning [7]. In the property sector, previous studies explored the implementation of a real estate chatbot using the Einstein Bot platform to increase information accuracy and booking efficiency [8]. While successful in streamlining operations, the reliance on a proprietary closed-source platform limits scalability and customization. Similarly, other research developed a property marketing information system featuring a chatbot, but the natural language understanding capabilities were restricted to exact keyword matching or basic problem-solving in e-commerce [9], [10]. These historical chatbot approaches relied heavily on rule-based architectures, decision trees, or simple pattern-matching algorithms, which often failed when confronted with human linguistic nuances, typos, or context-heavy sentences [11].

Recent breakthroughs have transformed these workflows towards complex deep learning models to better preserve natural conversational context and hidden intents. The introduction of the Bidirectional Encoder Representations from Transformers (BERT) model by Devlin et al. [12] has revolutionized natural language understanding by eliminating the constraints of unidirectional text processing. BERT's ability to interpret context from both the left and right sides of a token simultaneously via the self-attention mechanism has demonstrated outstanding performance across wide NLP evaluations, including complex intent classification, automated grading systems, and semantic context extraction [13], [14]. The complexity of the Indonesian language, characterized by rich morphology, extensive use of affixes, and a highly dynamic informal vocabulary, presents unique challenges for NLP systems. Wilie et al. [15] introduced IndoNLU, a benchmark for evaluating Indonesian natural language understanding, which included the release of IndoBERT. Their findings proved that language-specific pre-training on massive local datasets vastly outperforms generalized multilingual models. Furthermore, Koto et al. [16] developed IndoBERTweet, specifically adapting the BERT architecture for Indonesian social media text to handle extreme vocabulary mismatches. Intent classification is a fundamental NLP task where user input is mapped to a predefined category or action. Previous research successfully implemented BERT for an automatic short answer grading system in Indonesian [17], achieving an F1-Score of 0.95, and in various implementations such as premarital education services [18] a clinical setting to classify distinct intents for a dental clinic chatbot [19], significantly reducing initial consultation bottlenecks.

Despite these computational advances, there is a distinct lack of comprehensive empirical benchmarks assessing these advanced transformer models within the specific contextual environment of the Indonesian property market. Most current property chatbots still utilize basic NLP parsing or closed platforms. To address this critical research gap, this study develops and implements an interactive IndoBERT-based chatbot on the Griya Alam Mandiri web portal to fully automate client-side engagement twenty-four hours a day, seven days a week. The core novelty of this research centers on the extensive empirical benchmarking of the IndoBERT Base model against alternative transformer architectures, including Indonesian RoBERTa, IndoBERTweet, and XLM-RoBERTa, While recent studies have benchmarked IndoBERT for educational contexts such as question generation [20], its empirical comparative performance specifically within the noisy, informal chat environment of the Indonesian property sector remains a critical gap that this study resolves.. The successful implementation of this system aims to significantly lower human agent workloads, ensure instant response generation, and maximize overall digital marketing throughput for the real estate sector.

2. Research Methods

This research was conducted systematically through an Agile-based lifecycle, emphasizing data collection, rigorous model training, and an integrated system architectural deployment. The research took place at Perumahan Griya Alam Mandiri, located in Pakis, Malang, East Java, over an eight-month period. The methodology encompasses system architecture design, dataset gathering, NLP pipeline configuration, evaluation metrics design, and user acceptance testing, all executed in sequential phases to ensure the validity and reliability of the final model.

2.1 System Architecture Design

The application architecture leverages a modern, highly scalable Client-Server topology. The system was bifurcated into three distinct technological environments to ensure maximum responsiveness, an approach similarly adopted in designing robust land information systems [11]. The integration of these architectural components is depicted in Figure 1.

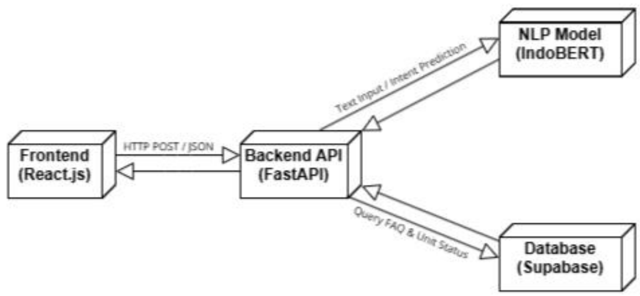

Fig. 1 Client-Server System Architecture integrating Frontend, API, and NLP

The frontend environment was constructed using the React library coupled with Vite as the fast build tool. The frontend utilizes TypeScript to guarantee code safety and reduce runtime errors. The interface provides a floating, interactive chat window component overlaid directly on the official property website. It manages user sessions and asynchronously sends text payloads to the server. The backend environment was engineered utilizing the Python-based FastAPI framework running on an ASGI server. This asynchronous server acts as the primary orchestrator. Upon receiving an HTTP POST request containing the user's message it executes text normalization and tokenization a crucial step for NLP-based IT program chatbots and dialogue systems [14]. The database environment operated on Supabase, a powerful Backend-as-a-Service utilizing a relational PostgreSQL database. This layer dynamically stores the FAQ knowledge base, housing block layouts, and real-time unit availability statuses. By decoupling the data layer from the model, administrative staff can update property prices or mark a unit as sold via the Supabase dashboard without requiring any model retraining.

This architecture is specifically designed to seamlessly support the entire user journey from initial inquiry to customer service handover. The interaction workflow initiates when a prospective buyer visits the Perum Griya Alam Mandiri official website and clicks the floating chatbot icon. This action triggers the display of the interactive chat pop-up interface. From there, the user can start asking various questions regarding housing information, such as unit availability, pricing, or specific block locations. The backend system then receives this input and rigorously evaluates the query against the trained NLP model to find a matching intent. If an appropriate match is confidently identified, the chatbot instantly retrieves and displays the correct predefined answer based on the given questions. Conversely, if the query is too complex or does not match any trained intent, the system executes a graceful fallback strategy. It displays a polite message stating that the question has no matching answer, and immediately provides an option to proceed to a human Customer Service representative. If the user chooses to click the WhatsApp button at the end of the chat, the system automatically forwards the conversation context to the WhatsApp application, ensuring no potential lead is lost due to unanswered queries. The complete operational workflow detailing how users interact with this established system is illustrated in Figure 2.

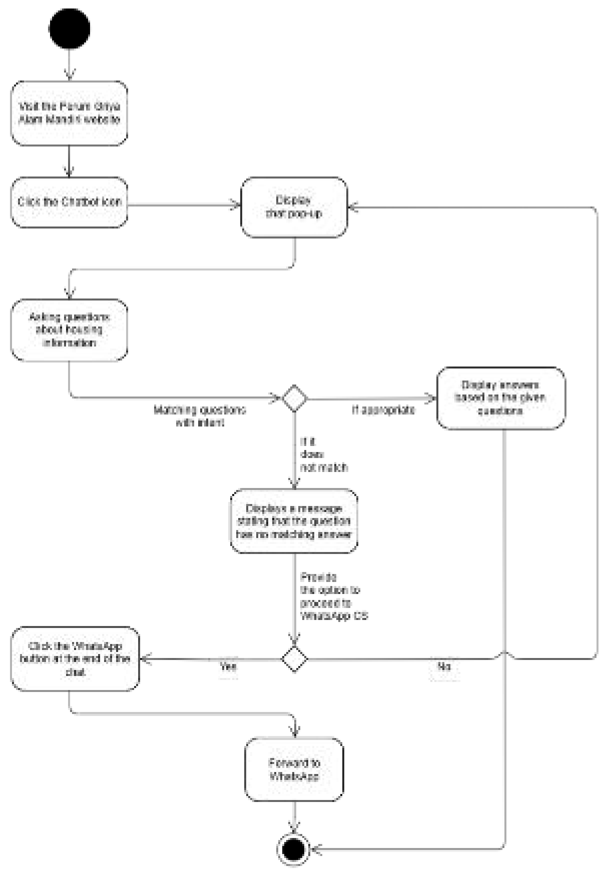

Fig. 2 Workflow of the Chatbot System in Processing User Intent

2.2 Data Collection and Preprocessing

The underlying domain knowledge base was meticulously extracted from official sales documentation, master property catalogs, and historical WhatsApp client conversation logs provided by the Griya Alam Mandiri marketing division. The raw data was manually mapped into 20 distinct operational intent categories. These categories encompassed diverse property queries, including prices, specific units, locations, mortgage requirements, legality, facilities, and basic conversational elements like greetings and closings. Initially, the dataset consisted of 1,927 raw, grammatically correct question samples. Because Indonesian conversational text, particularly in digital customer service, is highly prone to extreme abbreviations, informal slang, and typographical errors, robust data augmentation techniques were introduced to shield the downstream deep learning model from real-world chat noise. The augmentation pipeline programmatically injected spelling mistakes, such as swapping adjacent keyboard characters or dropping vowels, and replaced formal words with informal colloquial modifications. This strategy expanded the dataset by an additional 1,541 augmented samples, generating a final robust evaluation corpus containing 3,082 fully labeled training rows. To ensure unbiased evaluation, the dataset was rigorously split into an 80 percent training set and a 20 percent validation set using stratified sampling to maintain class distribution.

2.3 NLP Pipeline and Model Setup

Recent breakthroughs have transformed intent classification towards complex deep learning models. The introduction of the Bidirectional Encoder Representations from Transformers (BERT) model by Devlin et al. [12] revolutionized natural language understanding by eliminating the constraints of unidirectional text processing. The core NLP classification engine was constructed over the pre-trained IndoBERT Base parameters established by the IndoNLU benchmark. The IndoBERT architecture consists of 12 transformer layers, 12 attention heads, and a hidden size of 768, totaling approximately 124.5 million parameters. It was trained specifically on the massive Indo4B corpus containing over 250 million Indonesian sentences spanning news, Wikipedia, and social media [15]. For the intent classification task, a linear sequence classification head was appended to the base BERT model, taking the pooled output of the special classification token as the aggregate sequence representation. The tokenization process utilized the WordPiece algorithm operating on a vocabulary size of 30,522. The complete intent classification pipeline demonstrating the tokenization and bidirectional attention process is shown in Figure 3.

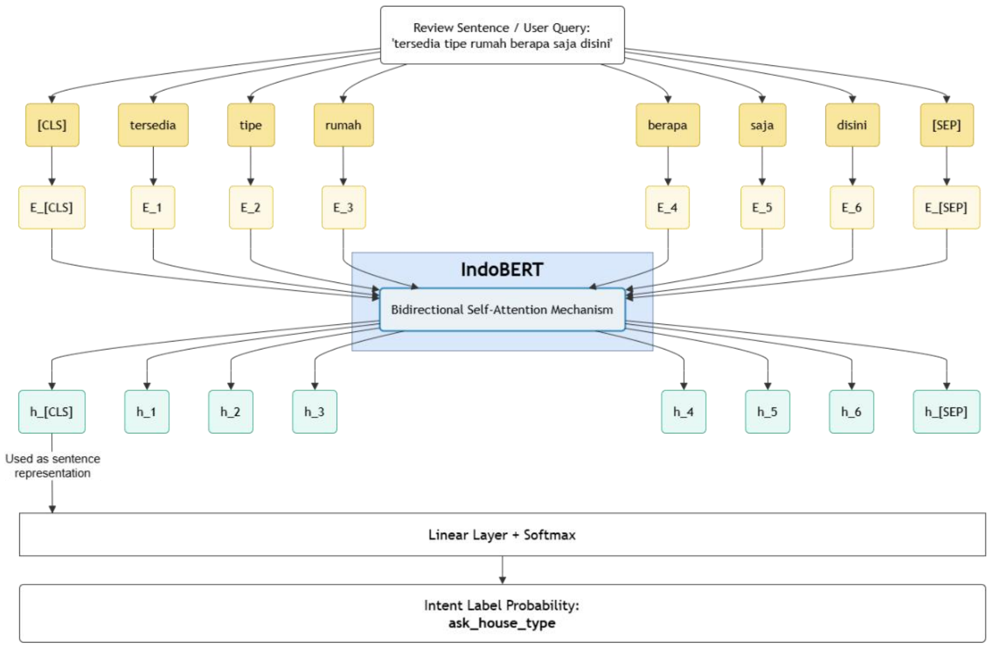

Fig. 3 IndoBERT-based intent classification pipeline demonstrating tokenization, bidirectional attention processing, and sequence representation extraction for a property-related user query

Fine-tuning workflows were deployed using the Hugging Face Transformers package executed on Google Colaboratory leveraging NVIDIA Tesla T4 GPUs to accelerate tensor computations. To guarantee a fair and objective benchmarking process against comparative models such as Indonesian RoBERTa, IndoBERTweet, and XLM-RoBERTa, standard hyperparameter controls were rigidly enforced across all training sessions. The complete hyperparameter configurations used in this study are detailed in Table 1.

Table 1. Model Training Hyperparameter Configuration

Hyperparameter	**IndoBERT Base**	**Indonesian RoBERTa Base**	**IndoBERTweet**	**XLM-RoBERTa**
Learning Rate	4e-5	2e-5	5e-5	2e-5
Batch Size	16	16	16	16
Max Sequence Length	128	128	128	128
Optimizer	AdamW	AdamW	Adam	AdamW
Weight Decay	0,01	0,1	0,01	0,01
Warmup	0,06 (ratio)	0,2 (ratio)	500 steps	500 steps

Total Epochs	5	5	5	5
Loss Function	Cross-Entropy	Cross-Entropy	Cross-Entropy	Cross-Entropy

The hyperparameters included a learning rate of 4e-5 with a linear warmup ratio of 0.06, a batch size of 16 per device, a maximum sequence length of 128 tokens, a weight decay of 0.01, and optimization via the AdamW algorithm across exactly 5 training epochs. The loss was calculated using the standard Cross-Entropy loss function, which is mathematically suited for mutually exclusive multi-class classification tasks.

2.4 Respondent Characteristics and UAT

To evaluate operational feasibility beyond purely technical mathematical indicators, a User Acceptance Testing trial was conducted involving 50 human evaluators. The respondent base was carefully selected to simulate real-world traffic, comprising 5 percent internal company agents, such as administration and marketing staff, to verify domain accuracy, and 95 percent general public users across various age demographics representing prospective property buyers. Evaluators were instructed to interact naturally with the chatbot prototype integrated into the staging website, asking questions regarding unit prices, block locations, and mortgage requirements. Quantitative feedback was subsequently captured via a structured questionnaire utilizing a 5-point Likert scale ranging from very dissatisfied to very satisfied. The assessment focused heavily on four primary indicators, specifically interface usability, information accuracy, system responsiveness, and overall user satisfaction, to ensure that the system not only functions technically but is also acceptable to the target users, consistent with estabilished chatbot acceptance studies [21].

3. Result and Discussion

The results section presents the research findings objectively and systematically, detailing the model convergence, classification performance, robustness against informal language, comparative benchmarking analysis, and the outcomes of the user acceptance testing based on the formulated methodology.

3.1 Training Convergence and Dynamics

The IndoBERT fine-tuning cycle exhibited exceptional model convergence over the 5 training epochs. During the initial epoch, the training loss started at 0.3861 with a validation accuracy of 92.48 percent. By the second epoch, the training loss dropped dramatically to 0.1034, indicating that the model's attention mechanisms were rapidly adapting to the distinct linguistic features and syntactic structures of the property dataset. At the conclusion of the 5th epoch, the model achieved a remarkably low cross-entropy training loss of 0.0194 and a peak validation loss of 0.1958. Crucially, the accuracy and loss curves for both the training and validation splits moved in parallel, converging tightly without manifesting any symptoms of parameter overfitting where training loss approaches zero while validation loss spikes. The complete trajectory of these metrics is illustrated in Figure 4.

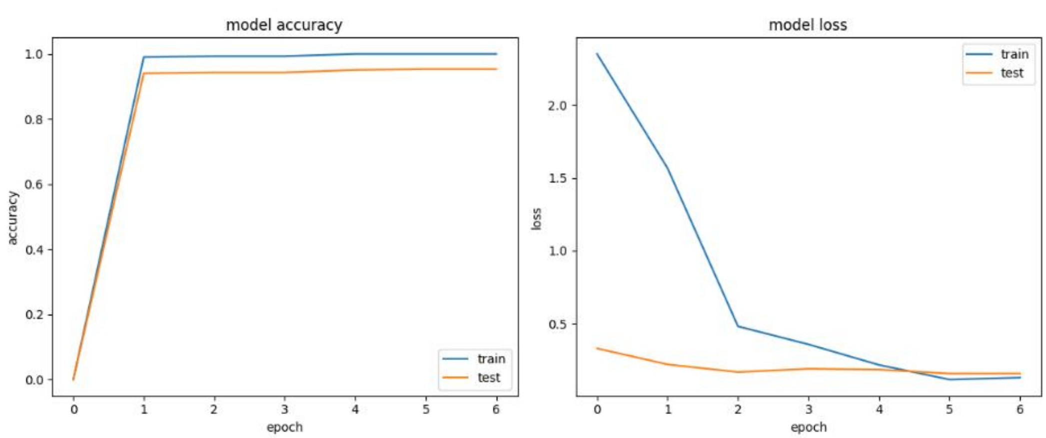

Fig. 4 Training Curves: Accuracy trends rising uniformly to 95.85% and Cross-Entropy loss dropping continuously to 0.0194 over 5 epochs across Train and Test splits

This confirms that the IndoBERT architecture, combined with the augmented dataset, forms excellent generalization capabilities, preparing it to confidently handle entirely novel, unseen textual inputs from real users.

3.2 Intent Classification Performance

Evaluation of the test partition using the multi-class classification report demonstrated top-tier intent matching capabilities. As shown in Figure 4, the training curves indicate stable convergence. Furthermore, Table 2 presents the key performance metrics.

Table 2. Key Performance Metrics for IndoBERT Intent Classification

Intent	Precision	Recall	F1-Score	Support
ask_specific_unit	0.9920	0.9880	0.9900	249
ask_unit_location	0.9592	0.9592	0.9592	49
greeting	0.9167	1.000	0.9565	11
ask_payment	0.7000	0.8750	0.7778	8
Average	0.9635	0.9585	0.9587	386

The IndoBERT model yielded a general classification accuracy of 95.85 percent, a weighted average precision rate of 96.35 percent, a recall performance of 95.85 percent, and a final aggregated F1-Score of 95.87 percent. This performance aligns with and even exceeds other notable BERT implementations, such as the automatic short answer grading system developed by Wijaya [17] which achieved an F1-Score of 0.95, and pre-marital education chatbots [18]. A deeper analysis reveals the model's structural strengths: for critical administrative and financial inquiries, such as mortgage details, legal certificates, and hidden fees, the model achieved perfect precision and recall scores of 1.000. For the most frequently accessed intent regarding specific units, which contained a massive 249 support samples, the model registered a near-flawless F1-Score of 0.9900, outperforming the efficiency of previous medical-domain BERT chatbots that handled 49 distinct intents [19]. Minor misclassifications only occurred in minority classes with extremely limited training data. For example, the fallback intent scored an F1-Score of 0.6667. These drops are mathematically expected when support sizes are in the single digits, causing minor overlaps in semantic vector space. However, these isolated anomalies do not disrupt the aggregate operational integrity of the chatbot system, as the core business logic remains highly accurate.

3.3 Robustness Against Informal Language

To evaluate the model's resilience against real-world chat noise, the system's performance was specifically analyzed on the augmented dataset partition containing injected typographical errors and colloquial phrasing, reflecting the well-documented complexity of informal Indonesian dialects [22]. Traditional rule-based models often fail catastrophically under these conditions. However, the IndoBERT Base model maintained exceptionally high confidence scores when processing misspelled words, such as swapped keyboard characters, and extreme abbreviations. This robustness is directly attributed to the subword tokenization strategy using the WordPiece algorithm. When confronted with informal slang, the tokenizer effectively breaks down unknown or misspelled words into recognizable sub-tokens, largely preventing Out-Of-Vocabulary (OOV) errors. Consequently, the chatbot seamlessly understands and accurately classifies prospective buyers' intents even when their inputs completely lack proper grammatical structure, ensuring a frictionless digital customer service experience.

3.4 Comparative Benchmarking Analysis

To scientifically validate the choice of the primary architecture, an extensive comparative matrix was analyzed against alternative pre-trained transformer models. All models, including Indonesian RoBERTa Base, IndoBERTweet, and XLM-RoBERTa, were subjected to the identical augmented dataset, the same data split, and uniform hyperparameter controls to ensure an equitable testing environment. The benchmarking results, as summarized in Table 3, clearly indicate that IndoBERT Base achieves a superior position across all metrics.

Table 3. Performance Benchmarking Across Pre-trained Architectures

Model	Accuracy	Loss	F1-Score
IndoBERT Base	0.9585	0.1958	0.9587
Indonesian RoBERTa Base	0.8550	0.4899	0.8488
IndoBERTweet	0.8800	0.4956	0.8592
XLM-RoBERTa	0.6750	1.1134	0.6526

IndoBERT Base scored an accuracy of 95.85 percent and the lowest validation loss of 0.1958. This significantly outperformed IndoBERTweet adapted for social media text by Koto et al. [16] with an accuracy of 88.00 percent, Indonesian RoBERTa Base with an accuracy of 85.50 percent, and the multilingual XLM-RoBERTa with an accuracy of 67.50 percent. The superiority of IndoBERT can be attributed to several critical architectural and corpus-level factors. The classic BERT encoder block, which utilizes the Next Sentence Prediction objective during pre-training, displays stronger structural stability for contextual property question-answering formats compared to RoBERTa designs, which intentionally omit this objective. Furthermore, although users employ informal slang, the core entities in property inquiries are strictly formal terms. IndoBERT, trained on a comprehensive formal and informal mix, grasped this better than IndoBERTweet, which is heavily biased towards Twitter syntax and emojis. The dramatic failure of XLM-RoBERTa highlights the issue of capacity dilution, where a broad multilingual model trained on numerous languages struggles to maintain high precision for localized, domain-specific Indonesian text, leading to poor boundary separation between property-specific intents.

3.5 User Acceptance Testing Analysis

While mathematical metrics prove technical capability, the practical field trial evaluates real-world viability. Figure 5 displays the interactive chat interface implemented on the website.

Fig. 5 Chat Interface Mockup: Clean conversational bubble layout embedding suggestion pills and automatic fallback CTA options for customer support handover

The UAT scored a total system feasibility rating of 82.50 percent, matching the designated "Very Satisfied" quality standard. Breaking down the specific indicators, System Usability marked the highest standalone evaluation metric with an average score of 4.3 out of 5. This verifies that the React-based frontend interface, the floating chat bubble layout, and the interactive suggestion pills can be intuitively operated by general public profiles without technical friction. The backend connection framework validated rapid response speeds, scoring 4.2 for Responsiveness, proving that the FastAPI layer and PyTorch inferencing can execute real-time model predictions without noticeable latency, effectively solving the original problem of delayed WhatsApp replies and offering a more scalable solution than earlier Einstein Bot implementations [8]. Information Accuracy scored slightly lower at 3.9, which is a known characteristic of classification-based NLP. When users input highly complex, multi-turn questions that span multiple intents, the model occasionally struggles to isolate a single intent, triggering the fallback mechanism. Nevertheless, the Overall User Satisfaction scored a solid 4.2 out of 5. This provides definitive proof that the chatbot successfully fulfills the commercial and operational expectations of the target demographic, effectively serving as an autonomous digital marketing agent.

4. Conclusions

This research successfully engineered and deployed an interactive, IndoBERT-driven chatbot to automate property client management workflows on the Griya Alam Mandiri website. The implemented Client-Server deployment successfully ensures continuous communication capabilities, significantly reducing the manual overhead and operational fatigue previously burdening the administrative staff. Through rigorous scientific benchmarking against state-of-the-art transformer models, the IndoBERT Base configuration was proven to be the absolute optimal architecture for the Indonesian property domain. Supported by Devlin's bidirectional attention framework and the robust IndoNLU dataset, it achieved a stellar intent detection Accuracy of 95.85 percent and an F1-Score of 95.87 percent, conclusively outperforming baseline alternatives such as IndoBERTweet, Indonesian RoBERTa, and the generalist XLM-RoBERTa. Furthermore, extensive data augmentation strategies involving typographical errors ensured the model remained highly resilient and robust against informal customer vocabulary and noisy inputs. The practical viability of the system was corroborated by a User Acceptance Testing feasibility score of 82.50 percent, indicating high public satisfaction regarding the interface's usability, accuracy, and response speed. The integration of this system effectively modernizes the digital marketing funnel for Griya Alam Mandiri, preventing lead leakage due to slow response times. For future development, researchers should investigate expanding conversational naturalness beyond static classification. Integrating Large Language Models via Retrieval-Augmented Generation frameworks could allow the chatbot to generate highly dynamic, multi-context responses based on vast document databases. Additionally, deploying the core backend across omnichannel messaging ecosystems, such as WhatsApp Business APIs and Instagram Direct, would centralize all customer service interactions into a single, highly intelligent platform.

5. References

[1] Menteri Perumahan dan Kawasan Pemukiman, “Peraturan Menteri Perumahan dan Kawasan Permukiman Republik Indonesia Nomor 5 Tahun 2025 tentang Besaran Penghasilan dan Kriteria Masyarakat Berpenghasilan Rendah serta Persyaratan Kemudahan Pembangunan dan Perolehan Rumah,” 2025.

[2] D. Susanto and Z. Zsa Izazi, “Analisis Strategi Pemasaran Dengan Media Sosial Marketing Pada Penjualan Properti di Brighton Real Estate,” 2024.

[3] M. Iqbal and M. Noor Hasan Siregar, “Implementation of Natural Language Processing for Chatbots in Customer Service,” 2024. [Online]. Available: http://creativecommons.org/licenses/by-sa/4.0/

[4] J. Khatib Sulaiman, M. Abid Nadzif, and U. Stikubank Semarang, “Penggunaan Teknologi Natural Language Processing dalam Sistem Chatbot Untuk Peningkatan Layanan Informasi Administrasi Publik,” *Indonesian Journal of Computer Science Attribution*, vol. 13, no. 1, pp. 2024–1227, 2024.

[5] N. Afifa, R. E. Saputra, and R. A. Nugrahaeni, “Implementasi NLP Pada Chatbot Layanan Akademik Dengan Algoritma Bert Implementation Of NLP On Academic Service Chatbot With Bertalgorithm,” 2023.

[6] G. Caldarini, S. Jaf, and K. McGarry, “A Literature Survey of Recent Advances in Chatbots,” *Information (Switzerland)*, vol. 13, no. 1, Jan. 2022, doi: 10.3390/info13010041.

[7] D. S. Yulianto and C. Taurusta, “Artificial Intelligence Virtual Assistant Using Natural Language Processing Methods,” *JICTE (Journal of Information and Computer Technology Education)*, vol. 8, no. 2, pp. 36–41, Oct. 2024, doi: 10.21070/jicte.v8i2.1664.

[8] F. Febrianto and R. D. Putri, “Implementasi Chatbot Sebagai Agen Perumahan untuk Meningkatkan Efisiensi dan Akurasi Informasi Menggunakan Einstein Bot,” *Jurnal Teknologi Informatika dan Komputer*, vol. 9, no. 1, pp. 320–327, Mar. 2023, doi: 10.37012/jtik.v9i1.1503.

[9] S. Hariyanto, I. Fenriana, S. D. Putra, and D. Lasut, “Perancangan Virtual Assistant Chatbot Berbasis Website sebagai Alat Promosi dan Dukungan Pemasaran,” *Jurnal Multidisiplin*, vol. 2, no. 1, 2023.

[10] J. Gao, A. P. Opute, C. Jawad, and M. Zhan, “The influence of artificial intelligence chatbot problem solving on customers’ continued usage intention in e-commerce platforms: an expectation-confirmation model approach,” *J. Bus. Res.*, vol. 200, Nov. 2025, doi: 10.1016/j.jbusres.2025.115661.

[11] O. Perdana Sihombing *et al.*, “Perancangan Chatbot ‘SiBePeN’ Sebagai Asisten Virtual Layanan Informasi Pertanahan,” 2024. [Online]. Available: https://cloud.google.com/dialogflow/es/docs/intents-

[12] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019. [Online]. Available: https://github.com/tensorflow/tensor2tensor

[13] D. G. N, M. V. G, and V. T A, “Intent Classification using BERT for Chatbot application pertaining to Customer Oriented Services,” European Alliance for Innovation n.o., Jan. 2022. doi: 10.4108/eai.7-12-2021.2314563.

[14] Z. Ran, Y. Chen, and Q. Jiang, “Intent Recognition in Dialogue Systems,” in *Proceedings of the 4th Asia-Pacific Artificial Intelligence and Big Data Forum, AIBDF 2024*, Association for Computing Machinery, Inc, Apr. 2025, pp. 165–173. doi: 10.1145/3718491.3718520.

[15] B. Wilie *et al.*, “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” Oct. 2020, [Online]. Available: http://arxiv.org/abs/2009.05387

[16] F. Koto, J. H. Lau, and T. Baldwin, “IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization,” Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.04607

[17] M. C. Wijaya, “Automatic Short Answer Grading System in Indonesian Language Using BERT Machine Learning,” *Revue d’Intelligence Artificielle*, vol. 35, no. 6, pp. 503–509, Dec. 2021, doi: 10.18280/ria.350609.

[18] F. R. Fatonah, D. S. Maylawati, and E. Nurlatifah, “Chatbot Edukasi Pra-Nikah berbasis Telegram Menggunakan Bidirectional Encoder Representations From Transformers (BERT),” *Jurnal Algoritma*, vol. 21, no. 2, pp. 29–40, Nov. 2024, doi: 10.33364/algoritma/v.21-2.1657.

[19] Irsyad Ibadurrahman, Fahmi Yusuf, and Heru Budianto, "IMPLEMENTASI CHATBOT BERBASIS NATURAL LANGUAGE PROCESSING (NLP) MENGGUNAKAN MODEL BERT UNTUK MENINGKATKAN PELAYANAN KLINIK GIGI," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 325–334, Jul. 2025, doi: 10.36341/rabit.v10i2.6061.

[20] H. Jati, Y. Indrihapsari, P. Setialana, D. Wijaya, S. A. Ardy, and D. D. N. Ardiansyah, "IndoBERT for educational assessment: comparative analysis of transformer models in Indonesian question generation," *IAES International Journal of Artificial Intelligence*, vol. 15, no. 2, pp. 1804–1813, Apr. 2026, doi: 10.11591/ijai.v15.i2.pp1804-1813.

[21] M. Goli, A. K. Sahu, S. Bag, and P. Dhamija, "Users' Acceptance of Artificial Intelligence-Based Chatbots: An Empirical Study," *International Journal of Technology and Human Interaction*, vol. 19, no. 1, 2023, doi: 10.4018/IJTHI.318481.

[22] A. Fikri Aji *et al.*, "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," Long Papers, 2022.