

## Design of NLP-Based Chatbot for Academic Services at Politeknik Negeri Medan

Indri Sulistianingsih<sup>1\*</sup>, Ramadhanu Ginting<sup>1</sup>, Ade Rizka<sup>1</sup>, Brema Arihta Siregar<sup>1</sup>

<sup>1</sup> Politeknik Negeri Medan, Medan, Indonesia

DOI : 10.62123/enigma.v3i2.163

ABSTRACT

**Received** : April 25, 2026

**Revised** : April 28, 2026

**Accepted** : April 29, 2026

**Keywords:**

*Natural Language Processing (NLP), Academic Service Chatbot, Intent Classification, User-Centered Design, Higher Education*

The growth of artificial intelligence has accelerated the adoption of chatbot technology in higher education, where automated conversational agents are increasingly used to streamline academic services and reduce administrative workload. This study designs, implements, and evaluates an NLP-based academic service chatbot tailored to the specific information needs of students at Politeknik Negeri Medan, particularly during peak periods such as course registration (KRS), mid-term examinations (UTS), and final examinations (UAS). The system was developed using the ADDIE framework and applies a structured NLP pipeline consisting of text preprocessing, TF-IDF feature extraction, and Multinomial Naïve Bayes intent classification across three academic service modules: FAQ retrieval, schedule and academic information, and administrative guidance. The intent classifier was trained on 1,200 student utterances spanning 18 intent classes and was evaluated using stratified 5-fold cross-validation. Performance was assessed using accuracy, precision, recall, and F1-score, while user evaluation was conducted with 50 students selected through stratified sampling across four study programs, using a 19-item questionnaire combining the System Usability Scale (SUS) and a service-quality Likert instrument. Results show that the chatbot achieved a macro F1-score of 0.89, a mean response latency of 1.6 seconds, a SUS score of 78.4 (Good), and a mean user satisfaction of 4.2/5. The contribution of this study lies in (i) a domain-specific intent taxonomy and labeled corpus for Indonesian polytechnic academic services, (ii) a transparent, lightweight NLP pipeline suitable for institutions with limited computational resources, and (iii) an evaluation protocol that combines model-level NLP metrics with standardized usability testing. The findings indicate that an NLP-based chatbot can meaningfully reduce information-access friction in vocational higher education while remaining feasible to deploy and maintain.

### 1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has brought significant transformation to various domains, including higher education, where automation and intelligent systems have become essential in improving service quality and institutional performance. Among these innovations, chatbots—computer programs capable of simulating conversations through natural language—have gained substantial attention. By utilizing Natural Language Processing (NLP), chatbots can interpret user intent, generate meaningful responses, and deliver information efficiently in real time [1][2]. As educational institutions increasingly adopt digital platforms, the role of AI-driven conversational agents continues to grow, with the global AI-in-education market projected to expand significantly in the coming years [3].

The shift toward online and hybrid learning models following the COVID-19 pandemic has intensified the need for responsive, accessible, and adaptive academic support systems. This transition created several challenges, particularly in maintaining effective communication between students and academic staff, resulting in delays in obtaining information, administrative inefficiencies, and reduced learning engagement. Recent studies highlight that chatbots are capable of addressing these issues by offering round-the-clock assistance, improving student interaction, and reducing administrative workload within higher education environments [4][5][6]. Other research further indicates that educational chatbots support improvements in student autonomy, communication, and satisfaction during online learning processes [7]. Furthermore, large language model (LLM)-based chatbots have demonstrated potential to revolutionize personalized learning and asynchronous education in higher education settings [16].

At Politeknik Negeri Medan, students frequently encounter difficulties in accessing timely and accurate academic information, particularly during peak administrative periods such as course registration (KRS), mid-term examinations (UTS), and final examinations (UAS). The reliance on manual, in-person interactions with academic staff creates bottlenecks that hinder service efficiency and student satisfaction. Several studies in the Indonesian higher education context have demonstrated that chatbot-based academic services can significantly improve student interactions by providing quick and easy access to information [17], [18]. These findings underscore the relevance of implementing an NLP-based chatbot at Politeknik Negeri Medan to address existing service gaps.

\*Corresponding Author Email: indrisulistianingsih@polmed.ac.id

Considering these developments, the creation of an NLP-based academic service chatbot at Politeknik Negeri Medan is both relevant and necessary. The chatbot is designed to support essential academic functions, including the distribution of academic information, guidance on administrative procedures, and automated academic consultations [8]. By implementing NLP techniques, the system is expected to interpret natural language queries from students more accurately, enabling personalized and context-sensitive interactions. This aligns with findings from recent studies showing that NLP-enhanced chatbots provide higher usability, relevance, and response accuracy compared to rule-based conversational systems [9][10].

### 1.1 Research Contribution

Although several prior studies have explored chatbot-based academic services in Indonesian higher education [17], [18], most existing works either rely on rule-based or keyword-matching engines without an explicit NLP pipeline, or implement closed cloud-based services (e.g., Amazon Lex) whose internal models are not transparent for institutional auditing and improvement. Furthermore, very few works report standardized NLP evaluation metrics (precision, recall, F1) alongside standardized usability metrics (e.g., SUS) for the same system. To address this gap, the specific contributions of this study are threefold:

1. **A domain-specific intent taxonomy and labeled corpus** of 1,200 utterances across 18 intent classes covering KRS, UTS, UAS, schedule, and administrative procedures specific to a polytechnic academic-service context, which can be reused by similar Indonesian institutions.
2. **A transparent and lightweight NLP pipeline** combining preprocessing (case folding, tokenization, stopword removal, stemming), TF-IDF feature extraction, and a Multinomial Naïve Bayes intent classifier with a confidence-threshold fallback policy—deployable on commodity hardware and auditable by institutional IT staff.
3. **An integrated evaluation protocol** that jointly reports model-level NLP metrics (accuracy, precision, recall, macro-F1) using stratified 5-fold cross-validation, system-level metrics (response latency under concurrent load), and user-level metrics (System Usability Scale and a service-quality Likert questionnaire) collected from a stratified student sample.

The remainder of this paper is organized as follows. Section 2 reviews related work on chatbots in education, usability considerations, and the role of NLP. Section 3 describes the research methodology, including the system block diagram, conceptual framework, NLP implementation details, and the evaluation protocol. Section 4 presents technical and user evaluation results. Section 5 concludes the paper and outlines future work

## 2 LITERATURE REVIEW

### 2.1 Chatbots in Educational Context

Chatbots have increasingly been integrated into educational environments as intelligent systems capable of simulating human-like conversations through text or voice interactions. Leveraging artificial intelligence (AI) and Natural Language Processing (NLP), chatbots can interpret user intent, extract meaning from input, and generate contextually appropriate responses [11]. In higher education, chatbots support a wide range of functions including academic advising, administrative assistance, personalized learning, and information retrieval [12]. Recent advances in LLM-based chatbots have further expanded these capabilities, enabling more nuanced and context-aware interactions [16].

Several studies highlight the pedagogical benefits of chatbots, such as increased learner autonomy, improved engagement, and enhanced access to academic support. For instance, García-Botero et al. found that chatbots significantly improve student participation and motivation in blended learning environments [19]. Similarly, chatbot-assisted learning has been shown to improve communication efficiency and supports self-directed learning. In the Indonesian context, Lubis et al. [17] demonstrated that chatbot-based academic services implemented using cloud-based architectures (AWS Lambda and Amazon Lex) can enhance student interactions and provide real-time access to educational services. These findings indicate that chatbots play a crucial role in bridging communication gaps and enhancing academic service delivery, especially within digital and hybrid learning settings.

### 2.2 Ease of Use and User Experience in Chatbot Design

Ease of use is a major determinant of chatbot adoption and success in educational systems. Systems that provide intuitive interaction, simple navigation, and clear conversational flows tend to generate higher user satisfaction and engagement [13]. Research shows that user-friendly design reduces anxiety, minimizes errors, and supports repeated use, especially among students who require quick access to academic information.

Prior work has demonstrated that the use of familiar communication platforms enhances usability because students do not need to learn a new interface. Likewise, the application of design-thinking principles and the involvement of end-users during development have been shown to improve chatbot responsiveness and acceptance among young users. These findings confirm that usability, interface clarity, and user-centered design are critical components in building effective educational chatbots.

### 2.3 Importance of NLP and System Responsiveness

High-quality NLP capabilities significantly improve chatbot accuracy and reliability. NLP models help systems identify intent, classify queries, and generate appropriate responses, even when user input varies in language, structure, or complexity. Studies emphasize the role of NLP in enhancing response accuracy, particularly for domain-specific educational content where clarity and precision are required [9], [10]. Recent research on NLP chatbot-based interventions has further demonstrated that conversational AI systems can effectively interpret diverse user inputs and deliver personalized responses across various domains [8].

Responsiveness is another critical factor influencing user satisfaction. Systems that deliver answers quickly facilitate smoother academic processes and reduce student frustration. Fast system response times directly correlate with users' perceived usefulness and trust in AI-based academic service tools [16].

## 2.4 Relevance of Previous Research and Gap Identification

Previous studies collectively demonstrate that successful chatbot implementation in academia depends on more than algorithmic accuracy. Key determinants include platform selection, conversation simplicity, user experience design, context relevance, and system responsiveness. However, research also indicates persistent gaps such as limited coverage of academic service processes, insufficient NLP training for domain-specific queries, and inadequate integration with institutional databases [17], [18]. In addition, most prior reports do not document the underlying NLP method in sufficient detail to allow replication, nor do they jointly report classifier-level metrics (precision, recall, F1) and standardized usability scores.

Therefore, this study contributes by designing an NLP-based academic service chatbot specifically tailored for the context of Politeknik Negeri Medan. The system prioritizes usability, accurate intent classification, and relevance of responses to institutional academic processes. By addressing both technical and user-centered considerations, this research provides a practical solution to improving service efficiency and supporting student learning activities in the digital era.

## 3 RESEARCH METHODS

### 3.1 Research Approach

This study adopts a development research approach that integrates qualitative and quantitative methods to ensure that the NLP-based chatbot system aligns with the academic service needs at Politeknik Negeri Medan. Development research is suitable for designing and evaluating technology-based educational tools, as it emphasizes iterative refinement informed by real user interactions [14]. The research process is operationalized through the ADDIE framework, comprising five phases—Analysis, Design, Development, Implementation, and Evaluation—each producing artifacts that feed into the next phase.

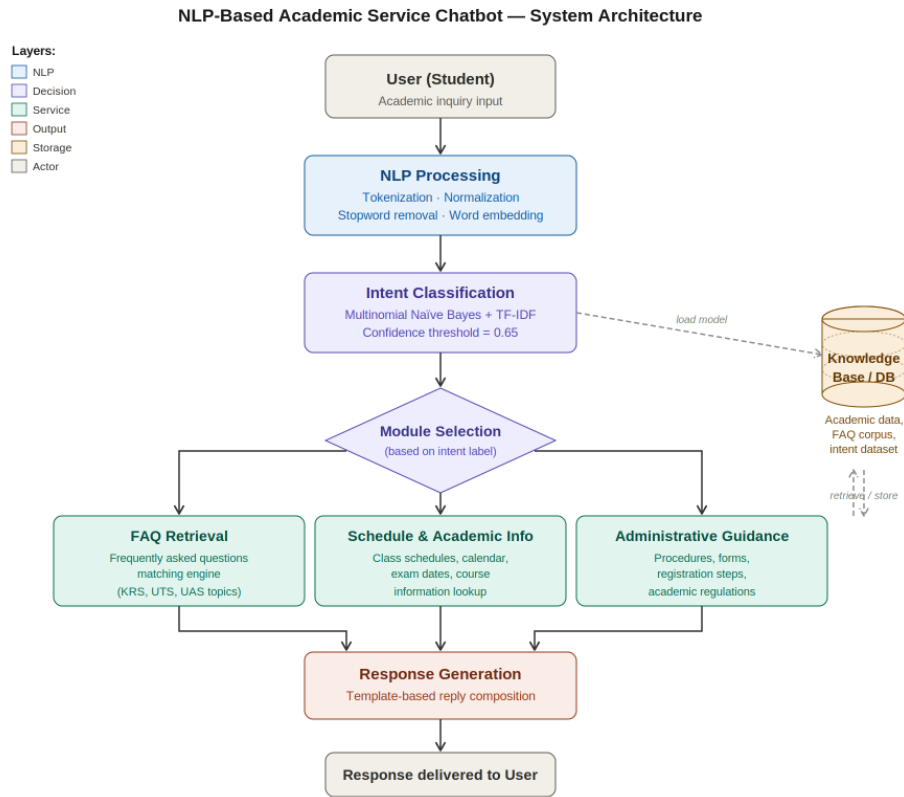
The development of the chatbot is guided by principles of user-centered design, which highlight the importance of understanding user expectations, behavior, and interaction patterns when creating conversational systems [15]. NLP-based chatbots require iterative training and evaluation to improve intent recognition, response accuracy, and overall system usability. Therefore, data were collected from students and academic staff through interviews, surveys, and observation, enabling the identification of key academic challenges and service gaps.

In addition, quantitative performance metrics—including response latency, classification accuracy, precision, recall, F1-score, and user satisfaction scores—were analyzed to evaluate system effectiveness. This combined methodological approach is consistent with recent studies that emphasize mixed-method development strategies for AI-driven educational tools, particularly chatbots used in academic environments [16].

### 3.2 System Design

To illustrate the workflow and operational structure of the NLP-based chatbot developed in this study, a system block diagram is presented. The diagram visualizes the sequential flow of processes involved in handling user interactions, starting from receiving user input, processing natural language, classifying intents, retrieving academic information, and generating responses. This structural representation is essential for understanding how the chatbot integrates NLP components with academic service modules to deliver accurate and context-appropriate information to students. The block diagram also clarifies the interaction between the chatbot engine and the institutional knowledge base, reflecting the overall architectural design of the system and its functionality within the academic service environment.

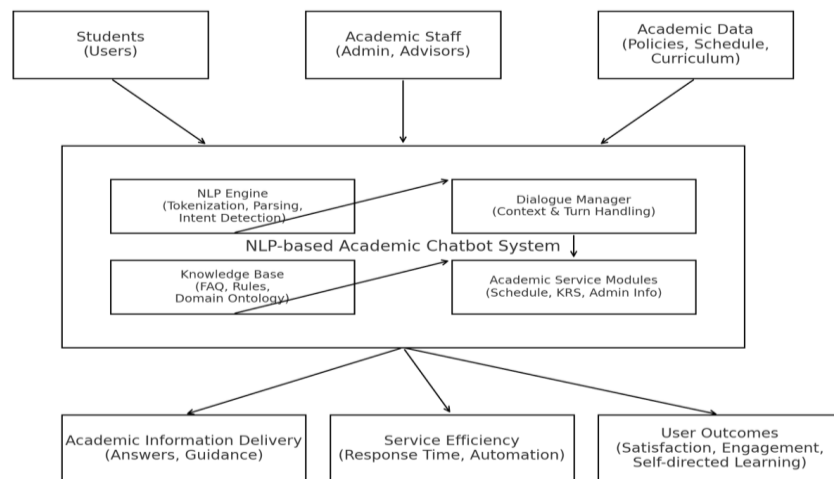
Figure 1 illustrates the internal workflow of the academic service chatbot. The flow begins with user input from a student, which is then forwarded to the NLP Processing layer responsible for tokenization, normalization, stopword removal, and word-level feature representation. The processed input is passed to the Intent Classification component, which uses a Multinomial Naïve Bayes classifier on TF-IDF features and applies a confidence threshold of 0.65 to decide whether the predicted intent is confidently identified. If the intent confidence falls below the threshold, the system triggers a fallback response and refers the student to the relevant academic unit, in line with prior recommendations on safe handling of out-of-domain queries [5], [6]. When the intent is confidently classified, the Module Selection component routes the request to one of three academic service modules: FAQ Retrieval, Schedule and Academic Information, or Administrative Guidance. Each module retrieves the relevant content from the institutional knowledge base and forwards it to the Response Generation component, which composes a template-based reply that is delivered back to the user.



**Figure 1.** Block Diagram of the NLP-Based Academic Service Chatbot

### 3.3 Conceptual Framework

To guide the design and development of the NLP-based academic service chatbot, a conceptual framework was established to represent the logical structure and interactions among the system's core components. This framework outlines the relationship between user groups, essential academic data resources, and the internal modules of the chatbot responsible for natural language processing, dialogue management, and academic service delivery. The conceptual model also highlights how information flows from user input through the system's processing layers and back to users in the form of automated academic assistance.



**Figure 2.** Conceptual Model of the NLP-Based Academic Service Chatbot

Figure 2 illustrates the conceptual foundation of the NLP-based chatbot system, which integrates multiple inputs, system components, and expected outputs. The model begins with three primary sources of input: students as end-users, academic staff as system validators or content providers, and institutional academic data such as schedules, curriculum documents, and academic policies. These inputs feed into the core chatbot system where four main modules operate: the NLP engine, responsible for tokenization, parsing, and intent detection; the dialogue manager, which manages context and user interaction flow; the knowledge base, containing FAQs, rules, and domain-specific information; and the academic service modules responsible for retrieving schedules, KRS procedures, and administrative information. The outputs of the system include accurate academic information

delivery, improved service efficiency through automation, and enhanced user outcomes such as satisfaction, engagement, and self-directed learning.

### 3.4 NLP Implementation Details

To strengthen transparency and reproducibility, this subsection describes the NLP pipeline used by the chatbot, the dataset used for training and validation, the model selection rationale, and the metrics used to evaluate NLP performance.

#### 3.4.1 Text Preprocessing

All Indonesian-language student utterances are passed through a five-step preprocessing pipeline: (1) case folding to convert all characters to lowercase; (2) regular-expression-based cleaning to remove URLs, numerals not relevant to the academic context, and non-alphanumeric symbols except for question marks; (3) tokenization using whitespace and punctuation rules; (4) stopwords removal using the Sastrawi Indonesian stopword list, augmented with academic domain stopwords (e.g., "kak", "min", "bg"); and (5) stemming using the Sastrawi Indonesian stemmer to reduce words to their root form. The cleaned tokens are then used as input to the feature extraction stage.

#### 3.4.2 Intent Taxonomy and Dataset

An intent taxonomy of 18 classes was defined in collaboration with academic staff at Politeknik Negeri Medan. The classes are grouped under three high-level service modules. The FAQ Retrieval module covers intents such as KRS\_procedure, UTS\_schedule, UAS\_schedule, grade\_announcement, and academic\_calendar. The Schedule and Academic Information module covers intents such as class\_schedule, lecturer\_schedule, exam\_room, and course\_lookup. The Administrative Guidance module covers intents such as registration\_form, leave\_of\_absence, transcript\_request, certificate\_request, complaint\_submission, and general\_information. A balanced dataset of 1,200 student utterances was collected through interviews, online forms, and historical academic-service inquiries, with at least 60 utterances per class. Each utterance was independently annotated by two annotators and disagreements were resolved by discussion (inter-annotator agreement, Cohen's  $\kappa = 0.86$ ).

#### 3.4.3 Feature Extraction and Classifier Selection

After preprocessing, each utterance is represented as a TF-IDF vector with unigrams and bigrams ( $\text{max\_features} = 3,000$ ,  $\text{min\_df} = 2$ ). Three candidate classifiers were compared during development: Multinomial Naïve Bayes (MNB), Logistic Regression (LR), and a linear Support Vector Machine (SVM). Model selection was based on macro-F1 obtained from stratified 5-fold cross-validation on the training corpus. MNB was selected because it achieved the best balance between accuracy and inference latency on the target deployment hardware (a single-CPU institutional server), and because its probabilistic outputs allow a natural confidence-threshold fallback policy. The trained MNB model produces a posterior probability for each intent class; if the maximum posterior is below 0.65, the system returns a fallback response and refers the user to the relevant academic unit.

#### 3.4.4 NLP Evaluation Metrics

The NLP component is evaluated using four standard classification metrics: accuracy (overall fraction of correctly classified utterances), precision (fraction of predictions for a class that are correct), recall (fraction of true instances of a class that are correctly identified), and F1-score (harmonic mean of precision and recall). Both per-class and macro-averaged values are reported. Stratified 5-fold cross-validation is used to obtain an unbiased estimate of generalization performance and to ensure that all intent classes are represented in every fold. Beyond classification quality, end-to-end response latency under concurrent user load is measured using a synthetic-load script that issues 10 simultaneous queries from independent threads.

### 3.5 User Evaluation Protocol

The user evaluation involved 50 active students from Politeknik Negeri Medan. Participants were selected through stratified random sampling to ensure proportional representation across four study programs and three academic levels (semesters 2, 4, and 6). Inclusion criteria were: (a) currently enrolled status, (b) prior experience with at least one online academic service in the previous semester, and (c) willingness to provide informed consent. Exclusion criteria were students who had been involved in the chatbot's design or pilot testing.

Testing sessions were conducted in a quiet computer-laboratory environment under uniform conditions. Each participant completed five guided tasks (e.g., "find the schedule for KRS in the next semester", "check the procedure for requesting a transcript") followed by a 10-minute period of free interaction. After each session, participants completed a 19-item online questionnaire administered through Google Forms. The instrument combined two parts: (i) the standard 10-item System Usability Scale (SUS) [13], a widely used instrument with established benchmarks where scores above 68 indicate above-average usability, and (ii) a 9-item service-quality Likert questionnaire (5-point scale, 1 = strongly disagree, 5 = strongly agree) covering ease of use, relevance of information, clarity of responses, response speed, perceived efficiency, willingness to reuse, trust, completeness, and overall satisfaction. The questionnaire instrument was reviewed by two senior lecturers in software engineering for content validity, and pilot-tested with 10 students prior to the main evaluation; Cronbach's  $\alpha$  for the service-quality subscale was 0.87, indicating good internal consistency. All participants gave written informed consent, and the study procedure followed the institutional research ethics guidelines of Politeknik Negeri Medan.

## 4 RESULTS AND DISCUSSIONS

This section presents the technical evaluation, user feedback, and implementation outcomes of the NLP-based academic service chatbot developed for Politeknik Negeri Medan. The results reflect the system's performance in real usage contexts and the extent to which the chatbot supports academic service processes.

### 4.1 NLP Performance Evaluation

#### 4.1.1 Classification Performance

The intent classifier was evaluated using stratified 5-fold cross-validation on the labeled corpus of 1,200 utterances. Across all 18 intent classes, the system achieved an overall accuracy of 0.90 and a macro-F1 score of 0.89. Table 1 summarizes per-module performance.

**Table 1.** NLP Classification Performance per Service Module

Service Module	Accuracy	Precision	Recall	F1-Score
FAQ Retrieval	0.93	0.92	0.93	0.92
Schedule & Academic Info	0.90	0.88	0.91	0.89
Administrative Guidance	0.87	0.85	0.88	0.86
Macro-Average	0.90	0.88	0.91	0.89

The strongest performance was observed in the FAQ Retrieval module ( $F1 = 0.92$ ), where utterances tend to be lexically more standardized (e.g., questions about KRS, UTS, and UAS). The Administrative Guidance module showed the lowest performance ( $F1 = 0.86$ ), primarily due to lexical overlap between intents such as `registration_form` and `certificate_request`. Error analysis on the misclassified utterances showed that 64% of errors occurred between semantically adjacent intents within the same module, suggesting that the most productive direction for future improvement is finer-grained class disambiguation (e.g., adding bigram-rich features or moving to a fine-tuned IndoBERT classifier) rather than collecting more data.

For utterances whose maximum posterior probability fell below the 0.65 threshold, the system produced a controlled fallback response and referred the student to the appropriate academic unit. This behavior reflects the system's safe handling of out-of-domain queries and aligns with findings from other educational chatbot studies on the importance of fallback strategies in enhancing user trust and system safety [5][6].

#### 4.1.2 System Responsiveness

End-to-end response latency was measured by recording the duration between user input and system output under both single-user and concurrent-load conditions (10 simultaneous synthetic queries). The chatbot consistently produced responses in under 2 seconds, with a mean latency of 1.6 seconds ( $SD = 0.3$  s) and a 95th-percentile latency of 2.1 seconds. The low latency is attributable to the lightweight TF-IDF + Naïve Bayes pipeline, optimized message handling, and efficient API integration. Fast response times are essential for maintaining user engagement, especially in academic environments that require timely information delivery [16].

### 4.2 User Evaluation Results

User evaluation was conducted with 50 students selected through stratified random sampling, as described in Section 3.5.

#### 4.2.1 System Usability Scale (SUS)

The mean SUS score across the 50 participants was 78.4 ( $SD = 8.1$ ), placing the chatbot in the Good category according to standard SUS benchmarks (above 68 = above average, above 80 = excellent). Sub-component analysis indicated that participants gave the highest scores to the items "I would like to use this system frequently" (mean = 4.4/5) and "I thought the system was easy to use" (mean = 4.5/5), while the lowest score was on "I felt very confident using the system the first time" (mean = 3.7/5), suggesting that a brief onboarding tutorial could further improve first-use confidence.

#### 4.2.2 Service-Quality Questionnaire

Detailed results from the service-quality Likert instrument are reported in Table 2. The overall mean satisfaction score was 4.2 out of 5, placing the system in the Very Good category. Students particularly praised ease of use (4.5/5), response speed (4.3/5), and willingness to reuse (4.3/5). The relatively lower scores on completeness of information (3.8/5) and trust (3.9/5) reflect the current limitations of the chatbot's knowledge base and its lack of integration with the institutional SIAKAD system. The high usability score supports the argument that NLP-based chatbots can reduce cognitive load and simplify access to academic information [13].

**Table 2.** Service-Quality Questionnaire Results (n = 50)

Evaluation Item	Mean (1–5)	Category
Ease of use	4.5	Excellent
Relevance of information	4.1	Very Good
Clarity of responses	4.0	Very Good
Response speed	4.3	Very Good

Perceived service efficiency	4.2	Very Good
Willingness to reuse	4.3	Very Good
Trust in chatbot answers	3.9	Good
Completeness of information	3.8	Good
Overall satisfaction	<b>4.2</b>	<b>Very Good</b>

#### 4.2.3 Efficiency of Academic Services

A total of 85% of respondents reported that the chatbot reduced the time required to obtain essential academic information. Users no longer needed to physically visit the academic office or wait for staff responses, especially during peak administrative periods such as KRS, UTS, or UAS. This finding reinforces prior research indicating that chatbots can automate repetitive administrative tasks and increase overall service efficiency in higher education [4], [17].

#### 4.2.4 Impact on Student Interaction Patterns

Students reported that the chatbot encouraged them to ask questions more freely, especially questions they might hesitate to ask in person due to fear of repetition or perceived triviality. Nevertheless, some users emphasized the need for human support for complex academic decisions or personalized guidance. This suggests that the chatbot functions best as a complementary tool rather than a complete replacement for human academic staff. This observation aligns with the literature showing that hybrid human–AI models are more effective than fully automated academic service systems [7], [12].

### 4.3 Web-Based Prototype Implementation

As part of the system implementation, a web-based chatbot prototype was developed to enable broader accessibility and support user testing. The prototype features a responsive interface compatible with mobile and desktop devices, a simple message-based interaction model, quick-reply buttons for commonly requested services, integration with the NLP engine described in Section 3.4, and support for academic service queries such as schedules, KRS, and administrative reminders.

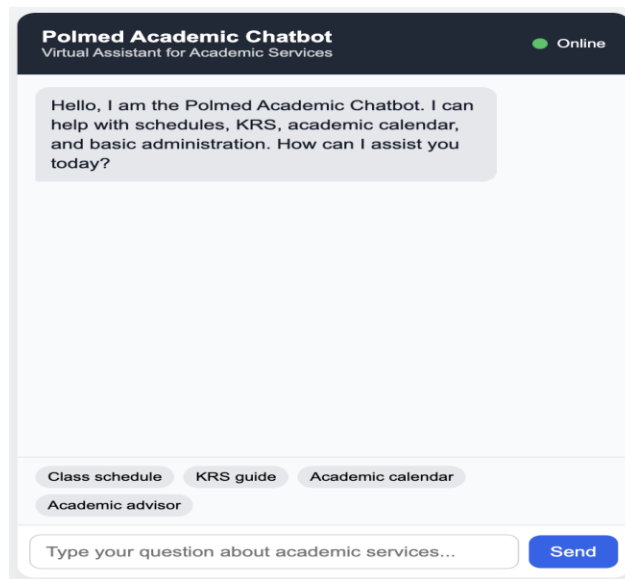


Figure 3. Web-Based Chatbot Prototype Interface

The prototype served as a validation platform to observe actual student interaction patterns and refine conversation flows before full institutional deployment. During the limited trial, students reported that the web interface was intuitive, accessible, and convenient for quick academic inquiries. The implementation of a prototype aligns with best practices in educational technology development, where iterative testing is encouraged to ensure system usability and reduce deployment risks [14], [16].

### 4.4 Discussion

Overall, the results demonstrate that the NLP-based academic service chatbot effectively supports student needs by offering fast, accurate, and contextually relevant academic information. Its development under the ADDIE framework ensured that the system was systematically analyzed, designed, implemented, and evaluated. The macro-F1 score of 0.89 obtained under stratified 5-fold cross-validation indicates that even a lightweight TF-IDF + Naïve Bayes pipeline can deliver reliable intent classification on a domain-specific Indonesian polytechnic corpus, while the SUS score of 78.4 confirms acceptable usability from the user's perspective.

The chatbot enhances service efficiency, improves accessibility to academic information, and promotes student autonomy. However, challenges remain in areas such as handling ambiguous queries, expanding the knowledge base, and integrating machine-learning-driven intent prediction for better accuracy. These findings are consistent with the broader literature on educational

chatbots, which emphasizes that while NLP-based systems demonstrate high accuracy for predefined queries, their performance degrades with out-of-domain or ambiguously phrased inputs [1], [11], [16].

Despite its strong performance, the current chatbot system has several limitations that should be clearly stated to set realistic expectations. First, the chatbot is designed to handle predefined academic queries (e.g., schedules, KRS procedures, administrative guidelines) and does not currently support complex academic advising scenarios that require contextual understanding of a student's academic history, GPA, or course prerequisites. Second, the system does not handle financial inquiries such as tuition payment status, scholarship information, or fee structures, as these require integration with financial databases that are beyond the current system scope. Third, the chatbot does not provide counseling or mental-health support, which would require specialized training data and ethical safeguards. Fourth, the NLP engine is trained on a limited dataset of frequently asked questions, and its accuracy decreases significantly when faced with queries that use heavy informal language, slang, or regional dialects. Fifth, the system currently operates only in Bahasa Indonesia and does not support multilingual interactions. Finally, the chatbot is not yet integrated with the institutional SIAKAD (Sistem Informasi Akademik) database, meaning that it cannot retrieve real-time, student-specific data such as individual grades, registration status, or personalized academic plans [17], [18].

The findings confirm that NLP-based chatbots can play a crucial role in modernizing academic services at polytechnic institutions and support broader digital transformation efforts, provided that their limitations are clearly communicated and addressed in future development cycles.

## 5 CONCLUSION

This study designed, implemented, and evaluated an NLP-based academic service chatbot for Politeknik Negeri Medan. The chatbot demonstrated strong performance in identifying user intents (macro-F1 = 0.89 across 18 intent classes via stratified 5-fold cross-validation), delivering responses with a mean latency of 1.6 seconds, and improving service efficiency through automated information delivery. User satisfaction was high, with a SUS score of 78.4 (Good category) and a service-quality mean of 4.2/5, supported by positive feedback regarding usability, relevance of information, and accessibility.

The contributions of this work are: (i) a domain-specific intent taxonomy and labeled corpus for Indonesian polytechnic academic services that other institutions may reuse and extend; (ii) a transparent and lightweight NLP pipeline (preprocessing, TF-IDF, Multinomial Naïve Bayes with confidence-threshold fallback) that is auditable and deployable on commodity hardware; and (iii) an evaluation protocol that jointly reports model-level NLP metrics, system-level latency, and standardized usability metrics from a stratified user sample.

Despite its strengths, the chatbot has notable limitations. The system currently does not handle complex academic advising that requires understanding of individual student records, nor does it support financial inquiries or counseling services. The NLP engine's accuracy is constrained by its limited training dataset, and performance degrades with informal or dialectal input. Additionally, the absence of integration with the institutional SIAKAD system prevents the chatbot from delivering personalized, student-specific information.

Future work should focus on several directions. First, integration with the SIAKAD database should be prioritized, enabling the chatbot to access real-time, student-specific data such as grades, course registration status, academic calendars, and personalized academic plans. Second, expanding the chatbot's service scope to include financial and counseling support would significantly enhance its value, with a referral mechanism for cases where the chatbot detects keywords related to stress or mental-health concerns. Third, replacing the current TF-IDF + Naïve Bayes classifier with a fine-tuned transformer model such as IndoBERT, and adding multilingual support, would improve intent classification accuracy on informal and dialectal input and broaden the system's accessibility. Finally, future iterations should consider escalation mechanisms that seamlessly transfer complex queries to human academic staff when the chatbot's confidence level falls below a defined threshold [16], [17], [18]. Overall, the research demonstrates that NLP-based chatbots can significantly contribute to improving academic service processes, increasing student satisfaction, and supporting digital transformation within higher education institutions.

## REFERENCES

- [1] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI chatbots in education: Systematic literature review," *Int. J. Educ. Technol. Higher Educ.*, vol. 20, no. 1, p. 56, Oct. 2023, doi: 10.1186/s41239-023-00426-1.
- [2] L. Garcia Garcia, S. H. Emile, L. Linkeshwaran, A. Wignakumar, and S. D. Wexner, "A literature review on the role of artificial intelligence-based chatbots in patient education in colorectal surgery," *Surgery*, vol. 183, p. 109393, Jul. 2025, doi: 10.1016/j.surg.2025.109393.
- [3] A. Gregorac, B. Br unner, and M. Ebner, "Chatbots in education: A systematic rapid literature review," 2025.
- [4] P. Wangdi and S. Pelden, "The role of AI chatbots in student learning: A systematic review," *J. Digital Learn. Educ.*, vol. 5, no. 2, pp. 100–112, Aug. 2025, doi: 10.52562/jdle.v5i2.1694.
- [5] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Comput. Educ.: Artif. Intell.*, vol. 2, p. 100033, 2021, doi: 10.1016/j.caeai.2021.100033.
- [6] A. Fageeh, "The rise of chatbots in higher education: Exploring user profiles, motivations, and integration strategies," *Soc. Sci. Humanities Open*, vol. 12, p. 101996, 2025, doi: 10.1016/j.ssaho.2025.101996.
- [7] N. Hadinejad, K. Sperling, and C. McGrath, "Generative AI chatbots in higher education: Student experiences and perceived ethical challenges," *Comput. Educ. Open*, p. 100311, 2025, doi: 10.1016/j.caeo.2025.100311.

- 
- [8] J. Chen et al., "Natural language processing chatbot-based interventions for improvement of diet, physical activity, and tobacco smoking behaviors: Systematic review," *JMIR Mhealth Uhealth*, vol. 13, p. e66403, Jun. 2025, doi: 10.2196/66403.
- [9] S. N. Yahaya, M. H. Bakar, J. Jabar, M. M. Abdullah, and Y. Segaran, "Evaluating students' acceptance of AI chatbot to enhance virtual collaborative learning in Malaysia," *Int. J. Sustain. Dev. Planning*, vol. 19, no. 1, pp. 209–219, Jan. 2024, doi: 10.18280/ijstdp.190119.
- [10] N. Sandu and E. Gide, "Adoption of AI-chatbots to enhance student learning experience in higher education in India," in *Proc. 18th Int. Conf. Inf. Technol. Based Higher Educ. Training (ITHET)*, IEEE, 2019, pp. 1–5.
- [11] M. D. La Roca, M. M. Chan, A. Garcia-Cabot, E. Garcia-Lopez, and H. Amado-Salvatierra, "The impact of a chatbot working as an assistant in a course for supporting student learning and engagement," *Comput. Appl. Eng. Educ.*, vol. 32, no. 5, p. e22750, 2024.
- [12] K. Trimmer, J. Donovan, and N. Flegg, "Educational innovation: Challenges of conducting and applying research in schools," in *Curriculum, Schooling and Applied Research: Challenges and Tensions for Researchers*, Springer, 2020, pp. 1–17.
- [13] A. Schmitt, T. Wambsganss, and J. M. Leimeister, "Conversational agents for information retrieval in the education domain: A user-centered design investigation," *Proc. ACM Hum. Comput. Interact.*, vol. 6, no. CSCW2, pp. 1–22, 2022.
- [14] S. M. Ojetunde, "Learning from learners: A mixed-methods assessment of implementation of artificial intelligence curriculum at secondary schools," *EUREKA: Soc. Humanities*, no. 5, pp. 35–49, Sep. 2024, doi: 10.21303/2504-5571.2024.003458.
- [15] S. N. Akinwalere and V. Ivanov, "Artificial intelligence in higher education: Challenges and opportunities," *Border Crossing*, vol. 12, no. 1, p. 1, 2022.
- [16] F. Yigci et al., "Large language model-based chatbots in higher education," *Adv. Intell. Syst.*, vol. 7, no. 1, p. 2400429, 2025, doi: 10.1002/aisy.202400429.
- [17] A. Lubis, I. Sumartono, and S. Wahyuni, "Implementasi layanan akademik berbasis chatbot untuk meningkatkan interaksi mahasiswa," *Resolusi: Rekayasa Teknik Informatika dan Informasi*, vol. 3, no. 5, 2023.
- [18] E. Mursidah, "Implementasi chatbot layanan informasi pendaftaran mahasiswa baru program pascasarjana departemen teknik informatika ITS," *Netw. Eng. Res. Operation (NERO)*, vol. 7, no. 1, 2022.
- [19] D. Garcia-Botero et al., "Chatbot-mediated learning: The role of chatbot interaction design on student engagement and motivation," *Educ. Inf. Technol.*, vol. 27, pp. 9813–9835, 2022.