

Development of an Academic Chatbot Based on Natural Language Processing for Student Information Services at Universitas Bina Darma

Muhammad Arif Rahmad Syahputra^{1*}, Widyanto¹, Muhammad Nasir¹

¹Department of Informatics, Faculty of Science and Technology, Universitas Bina Darma, Palembang, Indonesia

Abstract

Purpose: Repetitive academic inquiries and substantial linguistic variation in informal Indonesian student messages can reduce the consistency and timeliness of university information services. This study develops and evaluates an academic chatbot for student information services at Universitas Bina Darma through WhatsApp.

Methods/Study design/approach: An applied engineering-research approach was employed. An exploratory questionnaire involving 100 active students was used to identify dominant information needs and representative query expressions, while verified institutional documents were used separately to construct controlled service responses. The implemented architecture integrates Baileys as the WhatsApp messaging gateway, Node.js/Express as the orchestration layer, Dialog flow ES for intent and entity recognition, and MySQL for conversation logging and monitoring. The final agent contains 129 custom intents, 22 entities, and 2,572 training-phrase entries. An operational confidence threshold of 0.70 was applied to route uncertain inputs to fallback handling. Evaluation was conducted at four layers: 13 black-box functional scenarios, 24 intent cases representing major service families and out-of-scope input, 15 response-conformity cases, and six end-to-end integration scenarios.

Result/Findings: All black-box scenarios produced the expected outputs. Intent agreement reached 95.83%, response conformity reached 86.67%, and all integration scenarios were successful. The difference between intent agreement and response conformity indicates that correct intent recognition does not necessarily guarantee an equally specific or institutionally appropriate answer when response coverage is incomplete.

Novelty/Originality/Value: The contribution of this study is therefore positioned as applied system engineering rather than a new NLP algorithm. It demonstrates a deployable, confidence-controlled architecture that separates informal-query recognition from verified institutional delivery and evaluates recognition, answer conformity, and technical integration as distinct quality dimensions. Broader intent coverage, human-user evaluation, threshold calibration, load testing, and dynamic integration with institutional information systems remain necessary for wider deployment.

Keywords: Natural Language Processing, Intelligent Conversational System, Academic Information Service, Dialog flow ES, WhatsApp

Received August 2026 / **Revised** September 2026 / **Accepted** September 2026

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



INTRODUCTION

Universities increasingly employ conversational systems to provide immediate responses to repetitive administrative and academic questions. Early higher-education chatbots demonstrated that automated assistants can support both administrative and learning activities [1], while university-admission and frequently asked question systems showed the feasibility of intent-oriented question answering for institutional information services [2], [3]. In Indonesia, conversational agents have also been implemented for university admission and academic information services, indicating a continuing need for information channels that remain accessible beyond conventional office hours [4], [5], [6].

The communication channel and response-control mechanism are equally important in the implementation of academic chatbots. WhatsApp-based university and library chatbots have been reported as practical service interfaces because students can access them through a familiar messaging environment [7], [8]. Other student-support systems have investigated multilingual interaction [9], web-based university assistants [10], neural-network-based academic information centers [11], Dialog flow-oriented educational agents [12], and WhatsApp–Dialog flow integration [13]. Cloud-based messaging has also been applied to emergency communication [14], while university chatbots have been evaluated in operational contexts in which response reliability is important [15]. More recently, retrieval-augmented academic-service chatbots have been investigated as an alternative approach for flexible institutional knowledge access [16].

*Corresponding author.

Email addresses: arifgaming2124@gmail.com (Syahputra)

DOI: 10.15294/rji.v4i2.62866

Nevertheless, systematic reviews of educational chatbots indicate that evaluation methodology, domain grounding, response reliability, and user-centered assessment remain recurring concerns [17], [18].

For Indonesian-language services, the system must accommodate different ways of expressing academic information needs and distinguish supported requests from unsupported inputs. Classification performance can be sensitive to the amount and distribution of labeled examples [19]. Research on new-intent discovery also indicates that a fixed intent inventory does not eliminate the possibility of previously unseen or poorly represented information needs [20]. Therefore, evaluating an academic chatbot requires more than determining whether an intent label can be predicted; it is also necessary to examine whether uncertain questions are handled appropriately and whether the final response remains consistent with verified institutional information. Table 1 compares approaches and highlights the study's service-evaluation gap.

Table 1. Comparison of representative chatbot studies and the research gap addressed in this study

No	Context	Approach	Evaluation / Remaining Gap
[1]	Higher-education administration and learning support	Text classification and entity recognition	Demonstrated chatbot feasibility but did not focus on WhatsApp-based controlled institutional responses.
[7]	University admission through WhatsApp	WhatsApp chatbot for admission information	Demonstrated messaging-channel feasibility; evaluation emphasis differs from multi-layer service verification.
[11]	Academic information center	Deep feed-forward neural network	Reported model performance but did not separate intent agreement, response conformity, and end-to-end integration.
[12]	Educational information management	Dialog flow-based educational chatbot	Demonstrated Dialog flow applicability, but response governance and fallback-risk analysis remained limited.
[16]	Prospective-student academic services	Retrieval-Augmented Generation	Provides flexible knowledge retrieval, but response governance differs from a bounded verified-response service.
This study	Current-student academic services through WhatsApp	Dialog flow ES, controlled response layer, and MySQL logging	Separates informal-query recognition from verified response delivery and evaluates functional, intent, response, and integration layers independently.

At Universitas Bina Darma, Student Care/PPM handles recurring inquiries related to KRS, SISFO, tuition payments, scholarships, academic letters, examinations, academic schedules, and other student services. The operational problem is twofold. First, students frequently express the same information need using informal Indonesian wording, abbreviations, or different sentence structures. Second, the institution requires responses to remain consistent with verified service procedures. Consequently, an academic chatbot that correctly detects an intent but subsequently provides an incomplete or mismatched response would not fully resolve the information-service problem.

Accordingly, this study develops and evaluates a natural language processing-based academic chatbot integrated with WhatsApp. The work does not propose a new NLP algorithm. Instead, its contribution is positioned as applied system engineering, consisting of four elements: (1) a bounded service architecture that separates recognition of informal student queries from controlled institutional responses; (2) confidence-based routing to fallback handling for uncertain inputs; (3) persistent logging of messages, detected intents, and confidence values for monitoring; and (4) a layered evaluation framework that separately examines functional correctness, intent agreement, response conformity, and end-to-end integration. The research question is whether this architecture can reliably deliver academic information within the evaluated scope and which failure modes remain after technical integration.

METHODS

Research design and data

This study employed an engineering research approach because its primary outcome is a deployable software artifact accompanied by an operational evaluation. The research workflow consisted of problem identification, needs assessment, construction of the institutional service knowledge base, intent and entity configuration, WhatsApp and backend integration, conversation logging, and multi-layer system testing.

Primary data for the needs assessment were obtained from 100 active students through an online questionnaire. The student population used in determining the sample size consisted of approximately 6,000 active students. The sample size was calculated using Yamane's simplified formula [21], which is algebraically equivalent to the formulation commonly referred to as Slovin's formula in Indonesian applied research, as shown in Equation (1).

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

Using $N = 6,000$ and $e = 0.10$:

$$n = \frac{6000}{1 + 6000(0.10)^2} = 98.36 \approx 100 \text{ respondents}$$

where n is the sample size, $N = 6,000$ is the population size, and $e = 0.10$ is the selected margin of error; the sample was rounded to 100 respondents. A 10% margin of error was selected because the questionnaire functioned as an exploratory needs assessment rather than as an inferential evaluation of chatbot performance. The questionnaire was intended to prioritize frequently requested academic-service topics and capture representative student expressions, rather than to estimate population parameters with high statistical precision. Consequently, the 100 respondents should not be interpreted as providing exhaustive coverage of all linguistic variations used by the entire student population. Less frequently expressed information needs may therefore be underrepresented. The questionnaire identified five dominant information categories, as presented in Table 2.

Table 2. Distribution of student information needs

Topic	Respondents	Share
KRS & SISFO problems	33	33%
Scholarships & KIP	19	19%
Tuition payment	18	18%
Recommendation letters & internships	17	17%
Academic schedules	13	13%
Total	100	100%

Secondary information was obtained from verified institutional service materials, including Student Care/PPM information, academic calendars, service forms, procedures, and other institutional information used to construct controlled chatbot responses. Questionnaire data captured students' information needs and language, while institutional materials provided authoritative chatbot answers.

Data processing and NLP configuration

Questionnaire text was cleaned and normalized by removing unnecessary punctuation, standardizing capitalization, and normalizing common informal expressions while preserving academic terms and abbreviations that carry specific meanings, including KRS, KHS, SISFO, UKT, KIP, and PPM. The questionnaire therefore contributed observed forms of student language and topic priorities, while verified institutional sources were separately used to define controlled responses.

An audit of the final Dialog flow export identified 2,572 training-phrases across 127 user-says files, together with 129 custom intents and 22 entities. The figure of 2,572 entries must not be interpreted as 2,572 independent questionnaire responses. During iterative system development, phrase-level provenance metadata was not retained in a form that could reliably identify every final entry as an original respondent expression, a normalized expression, or an editor-added variant. Therefore, this study reports the final audited Dialog flow inventory and explicitly acknowledges the absence of complete phrase-level provenance as a reproducibility limitation rather than claiming an undocumented automatic data-augmentation procedure. Dialog flow ES was configured using intents, entities, training phrases, and fallback handling. Intent matching produces a detected intent and a confidence score [22]. In the implemented system, the backend applied an operational confidence threshold of 0.70. When the detected supported intent met the threshold, the backend returned the corresponding controlled institutional

response. Inputs with insufficient confidence or unsupported information needs were routed to failure or clarification handling. The 0.70 threshold was adopted as an operational safety rule rather than as a statistically optimized decision boundary. No dedicated threshold-calibration experiment was performed; therefore, the value should not be interpreted as an empirically optimal threshold.

System architecture

Figure 1 presents the integrated system architecture. The system architecture separates the communication layer, language-understanding layer, application-logic layer, and data-storage layer. Incoming WhatsApp messages are received through Baileys and forwarded to the Node.js backend. The backend sends the message to Dialog flow ES for intent and entity recognition. The detected intent and confidence score are subsequently evaluated before the backend selects either the corresponding controlled response or fallback handling. Conversation data, detected intents, confidence values, and message timestamps are stored in MySQL for monitoring and subsequent evaluation. Figure 2 illustrates the message-processing flow and confidence-based response routing.

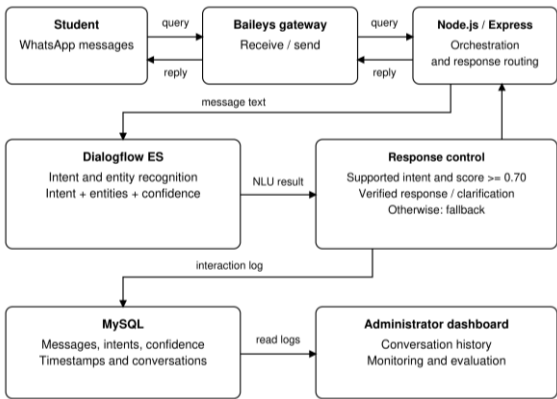


Figure 1. Integrated architecture of the academic conversational system

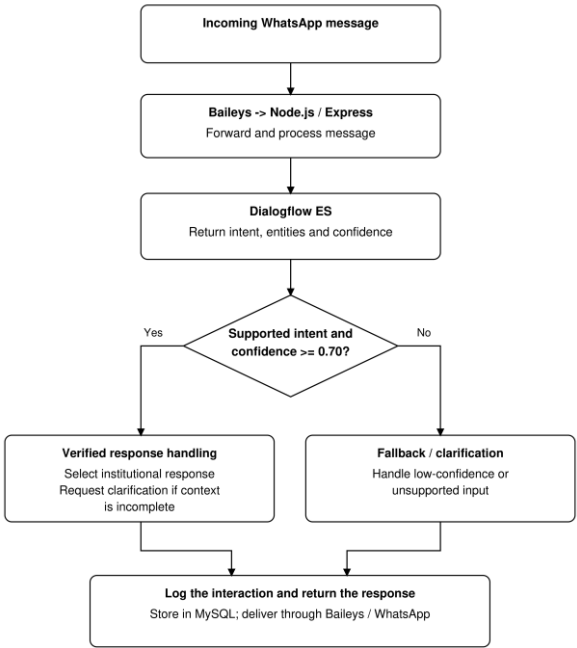


Figure 2. Message-processing flow and confidence-based response routing

Figure 2 details the message-processing sequence and confidence-based routing. The separation between recognition and response delivery enables the language-understanding to result to be analyzed independently from the quality of the institutional answer. It also allows low-confidence decisions and fallback behavior to be inspected through stored conversation logs. Figure 3 presents the database entity relationship supporting conversation logging and monitoring.

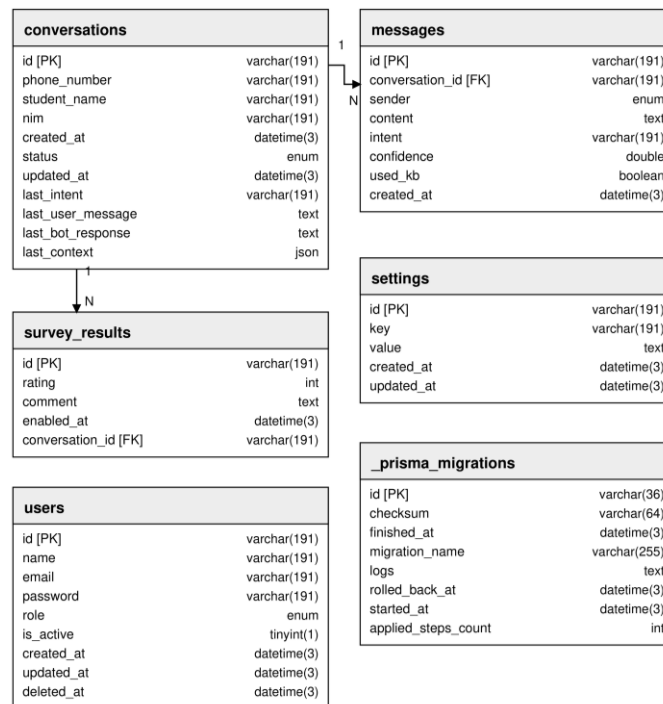


Figure 3. Database entity relationship supporting conversation logging and monitoring

Figure 3 presents the database entities and their relationships. The simplified data model contains only entities directly relevant to the implemented system and reported evaluation. Conversation records store session-level information, while message records contain message content, detected intent, confidence, and timestamps. User and configuration information supports system administration and monitoring.

Testing procedure

Four evaluation layers were employed. First, black-box testing evaluated whether 13 core system functions produced the expected externally observable behavior. Second, intent testing compared the detected intent with the expected intent for 24 test utterances. Third, response testing evaluated 15 final chatbot answers to determine whether the responses addressed the requested service and remained consistent with the configured institutional information. Fourth, integration testing evaluated six end-to-end scenarios involving WhatsApp/Baileys, Node.js, Dialog flow ES, MySQL, and the administrator dashboard.

The 24 intent cases were purposively selected as a functional sample covering the major service families represented in the needs assessment and operational knowledge base, together with an out-of-scope input. The test set was not designed to provide one evaluation example for each of the 129 deployed intents. Therefore, the reported intent-agreement percentage represents performance on the evaluated sample and should not be interpreted as exhaustive accuracy across the complete Dialog flow agent. A broader stratified evaluation covering every intent family would be required to support an agent-wide generalization claim. For intent evaluation, a test case was considered correct when the detected intent exactly matched the ground-truth label. A different production label was also classified as a semantic match when it represented the same academic service need and required the same operational action. Semantic matches are explicitly identified in the results rather than being presented as exact matches. This criterion prevents production-specific naming differences from being automatically classified as functional failures; however, it also introduces researcher judgment because independent inter-rater assessment was not performed.

Response conformity was assessed by comparing the returned response against the configured institutional service information. Consequently, this evaluation represents engineering verification and should not be interpreted as a blinded human-user assessment of response usefulness. The evaluation rates were calculated using the formulas presented in Equations (2)–(4):

$$IA = \frac{C_i}{T_i} \times 100\% \quad (2)$$

IA denotes intent agreement; C_i is the number of correct intent cases, including accepted semantic matches, and T_i is the total number of intent cases.

$$RC = \frac{C_r}{T_r} \times 100\% \quad (3)$$

RC denotes response conformity; C_r is the number of conforming responses, and T_r is the total number of response cases.

$$IS = \frac{S_g}{T_g} \times 100\% \quad (4)$$

IS denotes integration success; S_g is the number of successful integration scenarios, and T_g is the total number of integration cases. No stress/load testing, live usability study, or formal confidence-threshold calibration was conducted in the present study.

RESULT AND DISCUSSION

System implementation

The implemented chatbot provides an end-to-end interaction path from a student's WhatsApp message to a logged institutional response. Baileys operates as the WhatsApp communication gateway, Node.js coordinates application processing and routing, Dialog flow ES provides intent, entity, and confidence outputs, and MySQL stores interaction data. Students do not require a separate application or special command syntax. Questions can be submitted through ordinary Indonesian-language WhatsApp messages within the supported academic-service domain.

Anonymized WhatsApp conversation

English translation of a recorded interaction

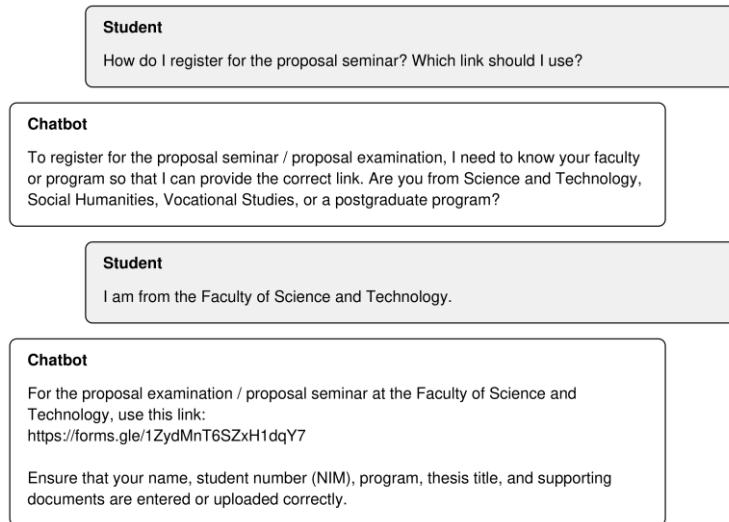


Figure 4. Example of academic-service interaction through WhatsApp

Figure 4 presents an anonymized English transcription of a recorded WhatsApp interaction. The conversation content is retained, while names, student identifiers, and telephone details are omitted. In services requiring additional context, such as faculty or study-program information, the chatbot may request clarification before returning the corresponding controlled institutional response. Dialog flow ES functions as the language-recognition component. Different linguistic expressions representing the same information need are organized into intents, while entities capture structured terms when necessary. Figure 5 presents a schematic configuration summary combining the audited inventory reported in Methods with selected intent names from the agent console.

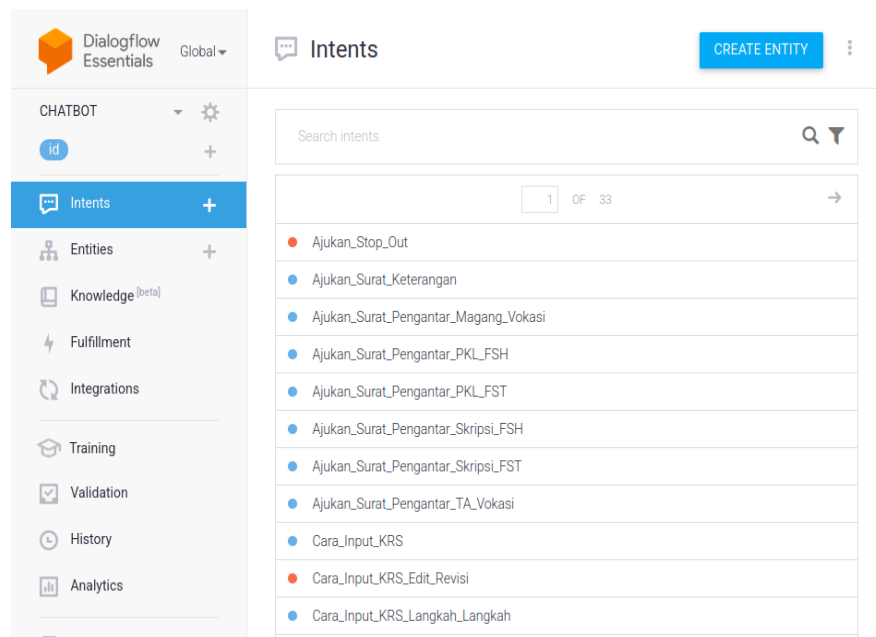


Figure 5. Dialog flow intent configuration used by the NLP component

The administrator dashboard complements the chatbot by displaying conversation history and stored evaluation information. This operational visibility enables developers or administrators to distinguish failures originating from message transport, intent recognition, confidence routing, or response mapping.

Evaluation results

Black-box testing

Table 3 presents the black-box testing results. Black-box testing verified the externally observable behavior of message reception, NLP processing, controlled responses, fallback handling, database storage, response delivery, authentication, and conversation monitoring.

Table 3. Black-box testing results

Function tested	Scenario	Input	Expected output	Observed output	Status
WhatsApp message reception	Student sends an academic question through WhatsApp	Academic question message	Message is received and forwarded to backend	Message received and forwarded to backend	Successful
Message processing	Backend receives a WhatsApp message	Student message	Backend processes the message and sends the query to Dialog flow	Message processed and sent to Dialog flow	Successful
Intent detection	Student sends a question matching a service category	Academic question	Dialog flow returns the detected intent and confidence score	Intent and confidence score obtained	Successful
Standard response	Question is recognized with the expected intent	Academic question	System returns the Standard Response for the detected intent	Standard Response returned for the detected intent	Successful
Default fallback response	Student sends an unrecognized or out-of-scope question	Out-of-context question	System returns the Default Fallback Response automatically	Default Fallback Response returned	Successful
Clarification response	Student sends a question without sufficient information	Ambiguous or incomplete question	System requests additional information	Clarification response returned	Successful

Function tested	Scenario	Input	Expected output	Observed output	Status
Conversation storage	System finishes processing a student message	Message, intent, confidence score, response	Conversation data are stored according to the designed schema	Conversation data stored in database	Successful
Response delivery	System has generated a response	Standard or fallback response	Response is delivered to the student through WhatsApp	Response received through WhatsApp	Successful
Admin login - valid	Admin logs in with valid credentials	Valid email and password	System authenticates and opens dashboard	Admin login successful and dashboard shown	Successful
Admin login - invalid	Admin logs in with invalid credentials	Invalid email and password	System rejects login and shows failure information	Login rejected and failure information shown	Successful
Dashboard	Admin opens dashboard after login	Authenticated admin	System shows dashboard information and data summary	Dashboard information and summary displayed	Successful
Conversation list	Admin opens conversation menu	Click conversation menu	System shows student conversation history	Conversation list displayed	Successful
Conversation detail	Admin selects a conversation	Click one conversation record	System shows conversation details and message history	Details and message history displayed	Successful

All 13 black-box scenarios produced the expected outputs. The results indicate that the implemented communication pipeline, fallback handling, conversation persistence, and administrative functions were operational under the tested conditions.

Intent testing

Table 4 presents the intent-testing results. Intent testing evaluated whether the system identified the intended academic-service category from the student query. Each result was compared with the ground-truth intent, and semantic equivalence was accepted when the deployed intent name differed but represented the same service meaning. This criterion is important because the final agent contains program-, faculty-, or action-specific intent variants.

Table 4. Intent-testing results

Test question	Ground-truth intent	Detected intent	Confidence	Assessment	Note
Admin, how do I fill in the KRS?	KRS_Information	KRS_Information	0.80	Correct	-
When does KRS registration open?	KRS_Information	KRS_Information	0.80	Correct	-
Why has my KRS not appeared in SISFO?	KRS_SISFO_Issue	KRS_SIAKAD_Issue	0.45	Correct (semantic)	Different name; same meaning. Below 0.70.
I cannot log in to SISFO.	SISFO_Issue	SISFO_Issue	0.86	Correct	-
How do I view my KHS?	KHS_Information	KHS_Information	1.00	Correct	-
What is the tuition-payment deadline?	Tuition_Fee_Information	Tuition_Fee_Information	0.91	Correct	-
My tuition has been paid but has not been recorded.	Tuition_Fee_Information	Tuition_Fee_Information	0.70	Correct	-
How do I apply for tuition installments?	Tuition_Fee_Installment	Tuition_Fee_Installment	0.86	Correct	-

Test question	Ground-truth intent	Detected intent	Confidence	Assessment	Note
What are the scholarship requirements?	Scholarship_Information	Scholarship_Information	0.74	Correct	-
When does KIP registration open?	KIP_Scholarship_Info	KIP_Scholarship_Info	0.82	Correct	Identical labels.
How do I request an active-student letter?	Active_Student_Certificate	Active_Student_Certificate	0.90	Correct	-
I want to request an internship introduction letter.	Internship_Cover_Letter	Submit_Vocational_Internship_Cover_Letter	1.00	Correct (semantic)	Different name; same meaning.
What are the requirements for a PKL letter?	Internship_Cover_Letter	Internship_Thesis_Cover_Letter	0.73	Correct (semantic)	Different name; same meaning.
How do I request a thesis introduction letter?	Thesis_Cover_Letter	Submit_FST_Thesis_Cover_Letter	0.76	Correct (semantic)	Different name; same meaning.
What are the Seminar Proposal requirements?	Proposal_Defense_Information	Proposal_Defense_Information	0.87	Correct	-
Where is the Seminar Proposal registration link?	Proposal_Defense_Information	Proposal_Defense_Information	0.79	Correct	-
How do I register for the comprehensive examination?	Comprehensive_Exam_Information	Register_Vocational_Comprehensive_Exam	1.00	Correct (semantic)	Different name; same meaning.
What are the Seminar Hasil requirements?	Results_Seminar_Information	Register_Results_Seminar	0.94	Correct (semantic)	Different name; same meaning.
How do I register for the thesis examination?	Master_Thesis_Defense_Information	Register_Postgraduate_Thesis_Defense	1.00	Correct (semantic)	Different name; same meaning.
Where is the PPM Online link?	PPM_Online_Information	PPM_Online_Information	0.90	Correct	-
How do I access the PPM Online form?	PPM_Online_Information	PPM_Online_Information	0.82	Correct	-
When is the final examination schedule?	Academic_Schedule	Academic_Schedule	0.72	Correct	-
How do I use general academic services?	General_Service_Information	Default Fallback Intent	0.00	Incorrect	Informasi_Layanan_Umum was not available in the agent.
I have a question unrelated to academics.	Default Fallback Intent	Default Fallback Intent	0.00	Correct	Out-of-scope academic context.

Twenty-three of the 24 test cases were classified as correct, resulting in an intent agreement of 95.83%. Several accepted cases represent semantic rather than exact label matches because the production agent contains faculty- or program-specific intents that correspond to the same service action. The low-confidence KRS/SISFO case is particularly important for interpreting the result. Although the detected intent represented a semantically equivalent service problem, its confidence value was only 0.45, which is below the operational threshold of 0.70. Consequently, semantic classification agreement and actual response routing are not equivalent concepts: a semantically acceptable prediction can still be rejected by the confidence-control rule.

The single incorrect case involved a general academic-service intent that was not available in the deployed agent and was consequently routed to fallback handling. This failure indicates an intent-coverage limitation rather than a communication-layer failure.

Response testing

Table 5 presents the response-testing cases and reported assessments. Response testing evaluated the final response returned through the Dialog flow-to-backend processing path. A response was considered conforming when it addressed the requested academic service and remained consistent with the configured institutional information.

Table 5. Response testing results

Intent	Test question	Standard response defined	Observed response	Assessment
KRS_Information	Admin, how do I fill out the KRS?	Yes	For KRS information, please follow the KRS registration schedule specified in the academic calendar. Ensure that your SISFO account is accessible and the selected courses comply with the study program regulations. If the KRS portal is not open or if you encounter issues, please provide details of the problem so you can be directed further.	Conforming
SISFO_Issue	I cannot log in to SISFO.	Yes	SISFO issues may include login failures, missing data, underplayed KRS/KHS, or system errors. Please state the issue you are experiencing so we can direct you more accurately.	Conforming
Tuition_Fee_Information	When is the deadline for tuition payment?	Yes	For schedule details and tuition payment deadlines, please check the official university announcements or the applicable academic calendar.	Non-conforming
Scholarship_Information	What are the scholarship requirements?	Yes	Scholarship information can be accessed through official university announcements or the relevant unit. Please ensure you prepare the required documents according to the specific scholarship type.	Conforming
Active_Student_Certificate	How do I request an active student certificate?	Yes	Submissions for the Active Student Certificate can be made via the following link: https://forms.gle/z1Jc8uzs7nht6Cp49 . Any missing documents or form entry errors will be communicated/sent via email. Please ensure the email address used is active. Thank you.	Conforming
Internship_Cover_Letter	Admin, I want to request an internship cover letter.	Yes	Requests for an Internship/Final Project Cover Letter can be submitted via the following link: https://sites.google.com/binadarma.ac.id/suratppmonline/home . Any missing documents or form entry errors will be communicated/sent via email. Please ensure the email address used is active. Thank you.	Conforming
Thesis_Cover_Letter	How do I apply for a thesis cover letter?	Yes	Requests for an Internship/Final Project Cover Letter can be submitted via the following link: https://sites.google.com/binadarma.ac.id/suratppmonline/home . Any missing documents or form entry errors will be communicated/sent via email. Please ensure the email address used is active. Thank you.	Conforming
Proposal_Defense_Information	What are the requirements for the Proposal Defense?	Yes	To register for the Proposal Defense/Proposal Exam, please sign up via the following link: https://sites.google.com/binadarma.ac.id/suratppmonline/home . Missing documents and the exam schedule will be notified via email. Please ensure the email address used is active. Thank you.	Conforming
Comprehensive_Exam_Information	How do I register for the comprehensive exam?	Yes	To register for the Comprehensive Exam - Faculty of Vocational Studies, please use the following link: https://forms.gle/5qUwtghEfBwDsnzQ9 . Ensure that details such as name, NIM, study program/faculty, and supporting documents are filled out correctly as required by the form.	Conforming
Results_Seminar_Information	What are the requirements for the results seminar?	Yes	For the Results Seminar/Results Exam, please specify your faculty or study program first.	Conforming

Intent	Test question	Standard response defined	Observed response	Assessment
Master_Thesis_Defense_Information	How do I register for the master's thesis defense?	Yes	To register for the Master's Thesis Defense - Postgraduate Program, please use the following link: https://forms.gle/UcKe1pJZVYK1QKpr7 . Ensure that details such as name, NIM, study program/faculty, and supporting documents are filled out correctly as required by the form.	Conforming
PPM_Online_Information	Admin, where is the link for Online PPM?	Yes	Online PPM services can be accessed via the following link: https://sites.google.com/binadarma.ac.id/suratppmonline/home . Please select the service that matches your academic needs.	Conforming
KHS_Information	How can I view my KHS?	Yes	To view the KHS, students can access SISFO using their respective accounts. If the KHS is not visible yet, ensure that the grading period is open or contact the relevant unit for further confirmation.	Conforming
Academic_Schedule	When is the final exam schedule?	Yes	For class schedules, midterms, final exams, or the academic calendar, please check the official announcements from the university or the relevant academic unit.	Conforming
General_Service_Information	How do I follow/access general academic services?	Yes	For general academic service information, please contact the student service unit or access the official academic portal.	Non-conforming

Thirteen of the 15 responses were assessed as conforming, producing a response-conformity rate of 86.67%. The two non-conforming cases were the tuition-payment deadline question (case 3) and the general academic-service question (case 15). In case 3, the answer referred the student to university announcements or the academic calendar without providing the requested deadline. In case 15, the answer directed the student to the service unit or academic portal without explaining the requested access procedure. Both responses were therefore assessed as insufficiently specific to address the question. These results demonstrate that successful message transmission and even acceptable intent recognition do not automatically guarantee a sufficiently specific final answer.

Integration testing

Table 6 presents the integration testing results. Integration testing verified the complete interaction path among WhatsApp/Baileys, the Node.js backend, Dialog flow ES, MySQL, and the administrator interface.

Table 6. Integration-testing results

Scenario	Input	Observed process	Expected	Actual result	Status
WhatsApp message delivery	Academic question	Message received by Baileys and forwarded to backend	Message received by backend	Message received by Baileys and processed by backend through <code>processIncomingMessage</code>	Successful
Backend message processing	Student message	Backend sends text to Dialog flow	Message processed by Dialog flow	Across 46 test conversations, messages were sent to Dialogflow project <code>gotaniesaku</code> and intent plus confidence were returned	Successful
Intent detection	Academic question	Dialog flow classifies the intent	Intent and confidence obtained	Classification succeeded; unknown input produced Default Fallback Intent with confidence 0, while domain intents such as <code>Kendala_KRS_SIAKAD</code> and <code>Pembayaran_UKT_Cicilan</code> were returned	Successful
Response delivery	Recognized intent	Backend receives Dialog flow result and produces response	Response sent back through WhatsApp	Response was produced, stored, and returned through WhatsApp via Baileys	Successful

Scenario	Input	Observed process	Expected	Actual result	Status
Message storage	Student message and chatbot response	Backend stores conversation data	Data stored in database	MySQL stored 46 conversations and 106 messages, including intent and confidence for each message	Successful
Conversation monitoring	Stored conversation data	Dashboard retrieves data from backend/database	Conversation displayed on dashboard	Inbox and /log-uji displayed conversation lists and message content through API with HTTP 200 and populated data	Successful

All six integration scenarios were successful. In the recorded evaluation environment, the system stored 46 conversations and 106 messages containing relevant intent and confidence information. The 100% integration success indicates that the identified deficiencies in response quality were not caused by broken connectivity among the major system components. Instead, the remaining problems originated primarily from intent coverage and response mapping.

Overall testing summary

The results of the overall evaluation are presented in Table 7.

Table 7. Overall evaluation summary

Evaluation type	Test cases	Successful	Unsuccessful	Rate
Intent testing	24	23	1	95.83%
Response testing	15	13	2	86.67%
Integration testing	6	6	0	100.00%

Table 7 summarizes the evaluation outcomes. The combined evaluation shows a consistent pattern: integration reliability is strongest (100.00%), intent recognition is also high (95.83%), and response conformity is the main remaining weakness (86.67%). Black-box testing additionally confirms that all 13 tested core functions produced their expected outputs.

Discussion

The principal finding of this study is not simply the high intent-agreement percentage, but the measurable difference between recognition quality and response quality. Intent agreement reached 95.83% on the 24-case evaluation set, whereas response conformity was lower at 86.67%. This difference is operationally important for academic-service chatbots because a technically acceptable intent prediction has limited value when it does not lead to a sufficiently specific and institutionally appropriate response. Reporting the evaluation layers separately therefore exposes service failures that would remain hidden if evaluation were based exclusively on intent accuracy or successful message delivery.

The contribution of the present work should consequently be understood as applied system engineering rather than an algorithmic advance in NLP. The admission-oriented, WhatsApp-based, and neural-classifier studies compared in Table 1 establish the feasibility of conversational academic services. A systematic review by Debets et al. found increasing use of chatbot builder platforms and a predominance of perceptual evaluation criteria in educational chatbot research [23]. The present study combines a familiar WhatsApp interface with a controlled institutional response layer, an explicit confidence-routing mechanism, persistent conversation logging, and a layered technical evaluation strategy. Its contribution therefore concerns administrative response reliability, while its lack of direct user evaluation remains a limitation relative to perceptual assessment. This architecture is especially relevant for administrative information services in which an incorrect but fluent response may be more problematic than explicitly returning fallback or requesting clarification.

The handling of informal Indonesian expressions is supported by the student needs-assessment data and the deployed training-phrase inventory. However, this study does not include a controlled baseline experiment comparing normalized and non-normalized expressions. Therefore, the results should not be interpreted as quantifying the isolated contribution of language normalization. Indonesian NLU research has demonstrated substantial linguistic variation across domains and communication styles [24], while out-of-

scope intent research shows that strong in-scope classification does not automatically guarantee reliable rejection of unsupported questions [25]. In the present study, the correct fallback for the explicitly unrelated question demonstrates desirable rejection behavior, whereas fallback for the intended general academic-service question reveals an undesirable coverage gap.

The 0.70 confidence threshold provides an operational safeguard but also creates a trade-off. The KRS/SISFO test was semantically acceptable despite receiving a confidence value of only 0.45. Under the deployed routing rule, such a case would still be rejected and sent to fallback handling. Because the threshold was not calibrated using a dedicated validation set, receiver-operating-characteristic analysis, or precision-recall optimization, it should be considered an implementation parameter rather than evidence of an optimal classification boundary. Future work should therefore evaluate multiple threshold values and explicitly measure the trade-off between accepting incorrect service intents and unnecessarily rejecting valid queries.

The response-testing failures further indicate that knowledge and response maintenance are at least as important as intent recognition. The tuition payment answer did not provide the requested deadline, while the general service answer provided only generic redirection without explaining how to access the service. In contrast, all integration scenarios were successful. This distinction provides a practical maintenance implication: improving the system does not necessarily require changes to the messaging infrastructure; instead, improvements should prioritize intent-coverage auditing, finer-grained response mapping, and systematic versioning of institutional information. This finding is consistent with dialogue-system evaluation literature emphasizing that technical or automatic measures alone cannot fully represent usefulness, appropriateness, or user experience [26]. The current engineering evaluation can demonstrate that the system functions and identify observable response defects, but it cannot establish student satisfaction, perceived usefulness, or real-world service effectiveness without direct user evaluation.

Several limitations constrain the interpretation and generalizability of the findings. First, the 24 intent cases constitute a purposive sample and cover only a subset of the 129 custom intents deployed; therefore, the 95.83% value represents agreement on the reported test set rather than agent-wide classification accuracy. Second, judgments concerning semantic equivalence and response conformity involve researcher interpretation and were not independently validated using multiple evaluators. Third, no live student usability or satisfaction study was conducted. Fourth, the 0.70 confidence threshold was not statistically calibrated. Fifth, concurrency, stress, and load behavior were not tested. Sixth, institutional service information is maintained through configured responses rather than being synchronized in real time with an academic-information-system API. Seventh, complete phrase-level provenance was not retained for the 2,572 final training-phrase entries, which limits reproducibility of the exact transformation from observed questionnaire language to the final deployed agent. Finally, the 10% sampling margin used for the exploratory questionnaire may underrepresent less frequent information needs and linguistic patterns.

CONCLUSION

This study developed and evaluated a WhatsApp-based academic chatbot for student information services at Universitas Bina Darma using Baileys, Node.js, Dialog flow ES, and MySQL. Within the reported evaluation scope, all 13 black-box scenarios produced the expected outputs, intent agreement reached 95.83% (23/24), response conformity reached 86.67% (13/15), and all six integration scenarios were successful. The findings indicate that the principal residual risk does not lie in component connectivity but in service-intent coverage and response specificity. A chatbot may correctly process and classify a student message while still returning an insufficiently specific answer. Therefore, the contribution of this work is a controlled and logged service architecture that separates language recognition from institutional response delivery and evaluates these layers independently, rather than proposing a new NLP algorithm. Future research should conduct stratified evaluation across the complete deployed intent inventory, retain phrase-level provenance during training-data development, involve independent evaluators and live student participants, calibrate the fallback threshold using empirical validation data, conduct concurrency and load testing, and integrate verified institutional information through maintained APIs. These improvements would strengthen reproducibility, user-centered validation, scalability, and the reliability of academic information delivery.

REFERENCES

- [1] [1] H. T. Hien, P.-N. Cuong, L. N. H. Nam, H. L. T. K. Nhung, and L. D. Thang, "Intelligent Assistants in Higher-Education Environments: The FIT-EBot, a Chatbot for Administrative and Learning Support," in *Proc. 9th International Symposium on Information and Communication Technology (SoICT)*, 2018, pp. 69–76, doi: 10.1145/3287921.3287937.
- [2] [2] H. A. Santoso et al., "Dinus Intelligent Assistance (DINA) Chatbot for University Admission Services," in *Proc. 2018 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2018, pp. 417–423, doi: 10.1109/ISEMANTIC.2018.8549797.
- [3] [3] B. R. Ranolija, N. Raghuwanshi, and S. Singh, "Chatbot for University Related FAQs," in *Proc. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 1525–1530, doi: 10.1109/ICACCI.2017.8126057.
- [4] [4] A. Z. Amrullah, A. S. Anas, and G. Primajati, "Implementasi Chatbot sebagai Virtual Assistant Penerimaan Mahasiswa Baru pada Universitas Bumigora," *Jurnal Bumigora Information Technology (BITE)*, vol. 4, no. 1, pp. 17–26, 2022, doi: 10.30812/bite.v4i1.1664.
- [5] [5] Munawirah, E. Mardjani, and K. Kristian, "Aplikasi Chatbot Berbasis Dialogflow dan NLP untuk Pelayanan Informasi Akademik pada Fakultas Ilmu Komputer Universitas Tomakaka," *Jurnal Ilmiah Multidisiplin Amsir*, vol. 3, no. 2, pp. 99–115, 2025. [Online]. Available: <https://stieamsir.ac.id/journal/index.php/abrj/article/view/610>
- [6] [6] D. Saientisna, I. G. A. G. A. Kadyanan, I. B. M. Mahendra, and V. G. A. Pradika, "Chatbot Pelayanan Informasi Kampus," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 2, no. 1, pp. 251–254, 2023, doi: 10.24843/JNATIA.2023.v02.i01.p29.
- [7] [7] B. B. C. Zoraya, S. Mumin, and A. Andhykac, "WhatsApp Chatbot Implementation of New Student Admission Information Service in Universitas Nahdlatul Ulama Sidoarjo," *Informatics, Electrical and Electronics Engineering (Infotron)*, vol. 5, no. 1, pp. 26–37, 2025, doi: 10.33474/infotron.v5i1.23365.
- [8] [8] M. A. Ahwan, N. Abiadhoh, A. B. Kusuma, and U. F. Alam, "Pembangunan WhatsApp Chatbot sebagai Layanan Kecerdasan Buatan di Perpustakaan UIN Walisongo Semarang," *Berkala Ilmu Perpustakaan dan Informasi*, vol. 20, no. 1, pp. 119–132, 2024, doi: 10.22146/bip.v20i1.7043.
- [9] [9] K. Ralston, Y. Chen, H. Isah, and F. H. Zulkernine, "A Voice Interactive Multilingual Student Support System Using IBM Watson," in *Proc. 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2019, pp. 1924–1929, doi: 10.1109/ICMLA.2019.00309.
- [10] [10] N. P. Patel, D. R. Parikh, D. A. Patel, and R. R. Patel, "AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT)," in *Proc. 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2019, pp. 148–150, doi: 10.1109/ICECA.2019.8822176.
- [11] [11] R. Faurina, D. Revanza, and A. Sopran, "Pengembangan Chatbot Menggunakan Deep Feed-Forward Neural Network sebagai Pusat Layanan Informasi Akademik," *Jurnal Eksplora Informatika*, vol. 11, no. 2, pp. 120–129, 2023, doi: 10.30864/eksplora.v11i2.833.
- [12] [12] M. M. A. Abdelmoiz, M. M. M. Mostafa, and T. H. A. Soliman, "Developing an Educational Chatbot for Scientific Data Management Course Using DialogFlow," *Appl. Math. Inf. Sci.*, vol. 18, no. 3, pp. 629–640, 2024, doi: 10.18576/amis/180315.
- [13] [13] B. Albab and I. Nuryasin, "Pengembangan Chatbot Berbasis WhatsApp Menggunakan Dialogflow dan NLP untuk Layanan Informasi Toko Fotokopi," *JIKO (Jurnal Informatika dan Komputer)*, vol. 9, no. 3, pp. 594–605, 2025, doi: 10.26798/jiko.v9i3.2046.
- [14] [14] O. Peña-Cáceres, A. Távora-Ramos, T. Correa-Calle, and T. More-More, "WhatsApp-Based Cloud Service Chatbot Application for Emergencies or Disasters," *J. Adv. Inf. Technol.*, vol. 15, no. 3, pp. 435–445, 2024, doi: 10.12720/jait.15.3.435-445.
- [15] [15] A. Balderas, R. F. García-Mena, M. Huerta, N. Mora, and J. M. Dodero, "Chatbot for Communicating with University Students in Emergency Situation," *Heliyon*, vol. 9, no. 9, Art. no. e19517, 2023, doi: 10.1016/j.heliyon.2023.e19517.
- [16] [16] R. W. Sugiarto, P. Sokibi, and P. Rizkiyah, "Chatbot Layanan Akademik Calon Mahasiswa UIC Menggunakan Metode Retrieval-Augmented Generation," *Jurnal Riset Rumpun Ilmu Teknik*, vol. 4, no. 3, pp. 246–249, 2025, doi: 10.55606/juritek.v4i3.6664.
- [17] [17] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots Applications in Education: A Systematic Review," *Comput. Educ. Artif. Intell.*, vol. 2, Art. no. 100033, 2021, doi: 10.1016/j.caeai.2021.100033.
- [18] [18] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI Chatbots in Education: Systematic Literature Review," *Int. J. Educ. Technol. High. Educ.*, vol. 20, Art. no. 56, 2023, doi: 10.1186/s41239-023-00426-1.
- [19] [19] I. Casanueva, T. Temčinas, D. Gerz, M. Henderson, and I. Vulić, "Efficient Intent Detection with Dual Sentence Encoders," in *Proc. 2nd Workshop on Natural Language Processing for Conversational AI*, 2020, pp. 38–45, doi: 10.18653/v1/2020.nlp4convai-1.5.
- [20] [20] H. Zhang, H. Xu, T.-E. Lin, and R. Lyu, "Discovering New Intents with Deep Aligned Clustering," *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 16, pp. 14365–14373, 2021, doi: 10.1609/aaai.v35i16.17689.
- [21] [21] T. Yamane, *Statistics: An Introductory Analysis*, 2nd ed. New York, NY, USA: Harper & Row, 1967.
- [22] [22] Google Cloud, "Intent Matching – Dialogflow ES," *Google Cloud Documentation*, 2026. [Online]. Available: <https://docs.cloud.google.com/dialogflow/es/docs/intents-matching>. Accessed: Sep. 9, 2026.
- [23] [23] T. Debets, S. K. Banihashem, D. Joosten-Ten Brinke, T. E. J. Vos, G. Maillette de Buy Wenniger, and G. Camp, "Chatbots in education: A systematic review of objectives, underlying technology and theory, evaluation criteria, and impacts," *Comput. Educ.*, vol. 234, Art. no. 105323, 2025, doi: 10.1016/j.compedu.2025.105323.
- [24] [24] B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *Proc. 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2020, pp. 843–857, doi: 10.18653/v1/2020.aac1-main.85.
- [25] [25] S. Larson et al., "An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction," in *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, 2019, pp. 1311–1316, doi: 10.18653/v1/D19-1131.
- [26] [26] J. Deriu, A. Rodrigo, A. Otegi, G. Echegoyen, S. Rosset, E. Agirre, and M. Cieliebak, "Survey on Evaluation Methods for Dialogue Systems," *Artif. Intell. Rev.*, vol. 54, pp. 755–810, 2021, doi: 10.1007/s10462-020-09866-x.