J-INTECH (Journal of Information and Technology)
Accredited Sinta 4 Ministry of Higher Education, Science and Technology Republic of Indonesia SK No. 10/C/C3/DT.05.00/2025
E-ISSN: 2580-720X || P-ISSN: 2303-1425

FAQ Chatbot for Small Businesses on the Web Using Semantic Search and Response Ranking

Abi Armansyah [1*], Ari Muzakir [2], Evi Yulianingsih [3]

[1,2,3]*Bina Darma University, Faculty of Science and Technology, Information Systems Study Program, Jl. Jenderal Ahmad Yani No.3, 9/10 Ulu, Kecamatan Seberang Ulu I,Palembang, Sumatera Selatan 30111, Indonesia*

Keywords
Cosine Similarity; FAQ Chatbot; Semantic Search; Small Business; TF-IDF; Web Application;

Corresponding Author:
aabiarmansyah12@email.com

Abstract
Small businesses often handle customer questions through manual replies via chat applications or phone calls, causing repetitive work, delayed responses, and inconsistent information delivery. This study proposes a web-based FAQ chatbot that answers user questions by performing semantic search over an Indonesian FAQ knowledge base and ranking the most relevant response. The chatbot applies a lightweight information retrieval approach using TF-IDF vectorization and cosine similarity to compute the relevance score between the user query and FAQ entries (question and tags). The system then selects the top-ranked FAQ entry and returns its associated answer, meaning the semantic matching is performed at the question-to-question level, not directly between questions and answers. The top results are ranked, and the chatbot returns the best answer along with a confidence score and the top three candidate questions to increase transparency. If the score is below a predefined threshold, the system provides a fallback response and suggests related topics rather than forcing an incorrect answer. The system is implemented as a PHP–MySQL web application with an administrator dashboard that supports secure login, FAQ CRUD management, chat logging, and usage analytics. Functional verification is conducted using black-box testing across main modules, including authentication, FAQ management, chatbot interaction, logging, and analytics dashboards. The expected contribution of this work is a practical and low-cost chatbot solution that can be deployed by small businesses to reduce repetitive customer service workload, accelerate response time, and provide measurable service insights through log-based analytics. Future improvements include expanding the knowledge base, enhancing Indonesian text normalization, and adopting embedding-based retrieval for better semantic matching.

1. Introduction

Small businesses often receive repetitive customer questions related to operating hours, services, warranty policies, payment methods, and delivery options. In many cases, these inquiries are handled manually via chat messages or phone calls, which can increase repetitive workload, slow response time, and create inconsistent information delivery when message volume grows. Prior studies in MSME contexts report that customer service handled through manual interactions can affect service responsiveness[1], and chatbot adoption is frequently proposed to reduce routine workload and improve response speed in small-business customer service processes[2].

Although chatbots are widely recognized as a solution to automate frequent inquiries, not all implementations are suitable for small businesses. Many advanced approaches require labeled training data, ongoing model maintenance, and additional infrastructure. For small businesses with limited resources, a lightweight approach that remains easy to deploy, easy to update, and controllable in terms of answer correctness is often more practical.

Recent studies show that retrieval-based FAQ chatbots using TF-IDF weighting and cosine similarity can effectively match user questions to knowledge-base items and rank the most relevant answer[3]. This approach has been applied in FAQ chatbot contexts for student admissions and organizational helpdesk scenarios, indicating that TF-IDF + cosine similarity remains a strong baseline for practical deployments with a curated FAQ database[4].

In Indonesian-language services, similar retrieval pipelines have been implemented for campus information chatbots and virtual assistants using preprocessing and TF-IDF-based relevance scoring[5]. Beyond classic TF-IDF, recent work also explores stronger semantic retrieval using Sentence-BERT (SBERT) with cosine similarity to improve matching quality, reinforcing that similarity scoring and response ranking are central mechanisms in FAQ chatbot design[6],[7].

Therefore, this study proposes a web-based FAQ chatbot for small businesses that applies TF-IDF–based semantic retrieval and response ranking over an Indonesian FAQ knowledge base[8], complemented by a confidence threshold to avoid forcing low-confidence answers. The novelty of this work lies not in the TF-IDF baseline itself, but in the practical deployment-oriented design and measurable workflow, including: (1) a transparent top-3 candidate suggestion mechanism presented to users when confidence is low, (2) an admin dashboard for continuous knowledge-base refinement (CRUD + tags) combined with log-based analytics (top queries, top matched FAQs, average score), and (3) an explainable debug view that exposes preprocessing output, TF-IDF weights, and cosine similarity ranking to support evaluation and iterative improvement. These components enable small businesses to maintain and improve the chatbot without retraining models, while providing quantitative evidence from interaction logs to monitor retrieval quality

Therefore, this study proposes a web-based FAQ chatbot for small businesses that applies TF-IDF–based semantic retrieval and response ranking over an Indonesian FAQ knowledge base[2], complemented by a confidence threshold to avoid forcing low-confidence answers. The novelty of this work lies not in the TF-IDF baseline itself, but in the practical deployment-oriented design and measurable workflow, including: (1) a transparent top-3 candidate suggestion mechanism presented to users when confidence is low, (2) an admin dashboard for continuous knowledge-base refinement (CRUD + tags) combined with log-based analytics (top queries, top matched FAQs, average score), and (3) an explainable debug view that exposes preprocessing output, TF-IDF weights, and cosine similarity ranking to support evaluation and iterative improvement. These components enable small businesses to maintain and improve the chatbot without retraining models, while providing quantitative evidence from interaction logs to monitor retrieval quality[9]. This paper contributes a lightweight, explainable, and maintainable FAQ chatbot workflow for small businesses, combining retrieval ranking, confidence control, and log-driven knowledge-base refinement in a PHP–MySQL implementation.

2. Research Method

his study applies an applied web-based system development approach to build an FAQ chatbot for small businesses, starting from requirement identification, database and interface design, implementation, and functional evaluation[10]. The system is implemented using PHP–MySQL, where the FAQ knowledge base is maintained through an admin dashboard (CRUD), while user interactions are served through a chat interface. The overall development flow follows practical deployment-oriented chatbot engineering patterns reported in recent work on retrieval-based conversational systems and RAG-based virtual assistant development in industry settings[11],[12]. TF-IDF weighting is applied to transform normalized text into numerical features, as widely used in NLP tasks to represent documents for downstream scoring and classification[13].

For the chatbot answering mechanism, a retrieval and ranking pipeline is used: Indonesian text is normalized (lowercasing, punctuation removal, tokenization, stopword removal, and light stemming), then the FAQ documents (question + tags) are transformed into TF-IDF vectors, and user queries are ranked using cosine similarity. The system returns the top answer with a similarity score, and logs every interaction for analytics. Evaluation is conducted through 11 black-box functional test cases covering login, FAQ management (CRUD), chat endpoint, logging, and dashboard statistics. To validate chatbot responses, we used 8 interaction logs generated during trial usage and inspected the top-1 match, top-3 candidates, cosine similarity score, and threshold decision (0.25). Response validity was confirmed by checking whether the returned FAQ matched the intended topic and by reviewing the frequency of low-confidence fallback cases, FAQ management, chat endpoint, logging, and dashboard statistics) and retrieval output checks using representative user queries. This method is consistent with recent studies that evaluate chatbot response quality using vectorization + cosine similarity and emphasize the importance of preprocessing and dataset curation for better retrieval performance[14],[15].

In addition to the retrieval pipeline, the study applies a continuous improvement mechanism through admin management and log-based monitoring. The FAQ knowledge base is updated by administrators via CRUD operations, while every user interaction is recorded in the database (question, matched FAQ ID, similarity score, and timestamp). These logs are then summarized in the analytics dashboard to identify frequently asked questions, the most matched FAQ items, and the proportion of low-confidence queries that trigger fallback suggestions. This feedback loop supports iterative refinement of FAQ content and tags, ensuring that the chatbot remains accurate and relevant as customer needs evolve.

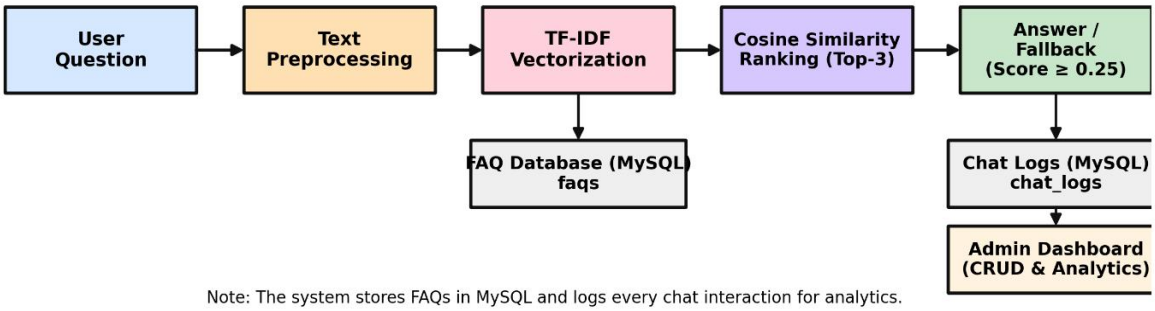

Figure 1. The proposed FAQs chatbot system

3. Result and Discussions

In the implementation results, the system successfully presents three core components that reflect the proposed workflow. First, the public chatbot interface demonstrates the TF-IDF and cosine similarity retrieval mechanism in practice: when a user submits a vague query such as “hallo,” the system does not force an unreliable answer, but instead returns a fallback message and provides top-3 related FAQ suggestions that can be clicked to guide the user; meanwhile, for a more specific query such as service duration, the chatbot returns the most relevant FAQ answer and maintains a coherent conversation flow. Second, the Admin FAQ

Management page confirms that the knowledge base can be updated dynamically through CRUD functions (create, edit, delete) with search and tag support, allowing business owners to refine content without modifying source code. Third, the Admin Statistics dashboard summarizes interaction logs and matching performance—showing total chats, today's chats, average similarity scores (overall and matched), the most frequent user questions, and the most frequently matched FAQ items—indicating that logging and analytics operate correctly and can support continuous improvement of FAQ coverage and retrieval quality.

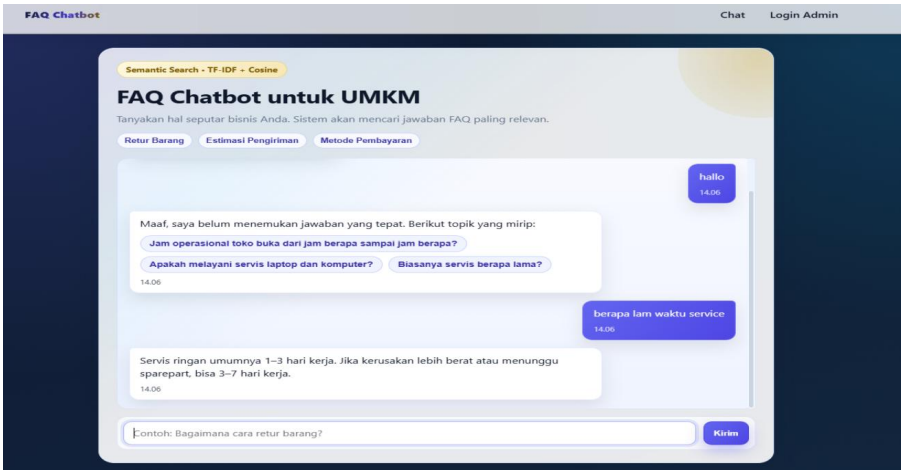

Figure 2. Web Chatbot

This figure shows the public chatbot interface where users ask questions and receive responses generated through semantic search (TF-IDF + cosine similarity). The interface supports two output behaviors: when the similarity score is low, the system returns a fallback message and provides top-3 related FAQ suggestions as clickable buttons to help users reformulate or select a closer topic; when the query is sufficiently relevant, the chatbot returns the best-matched FAQ answer directly. This design improves user experience by avoiding forced incorrect answers while still guiding users toward available information in the knowledge base.

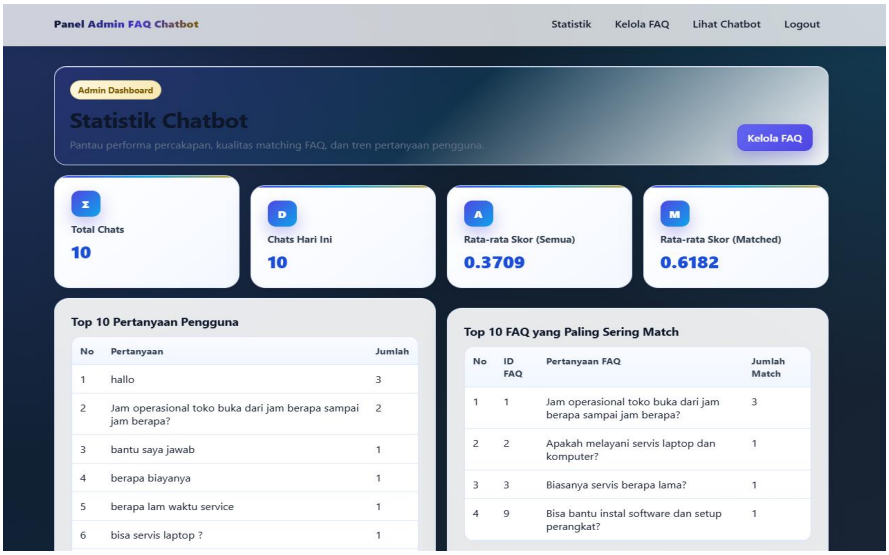

Figure 3. Dashboard Admin

This figure shows the Admin Statistics dashboard that summarizes chatbot usage and retrieval performance based on stored interaction logs. Key indicators include total chats, chats today, average similarity score (overall), and average similarity score (matched), which help evaluate how confidently the system matches

user queries to FAQs. The dashboard also lists top user questions and top matched FAQs, enabling admins to identify trending issues, detect unanswered/low-confidence topics, and prioritize FAQ improvements. Overall, this page demonstrates that the system not only answers questions but also provides data-driven feedback for continuous knowledge-base optimization.

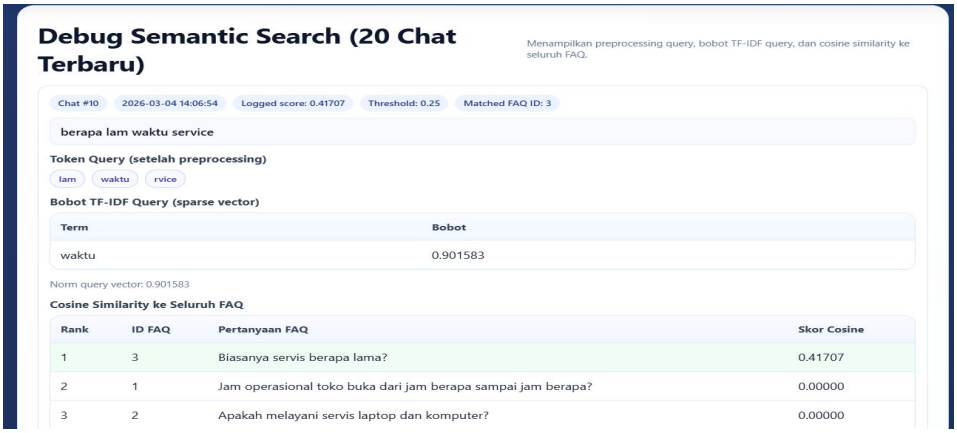

Figure 4. Debug Semantic Seacrh

Although Figure 4 is presented as a debugging view, it is included to provide traceability and transparency of the retrieval mechanism, which is essential for validating a ranking-based chatbot. The figure explicitly shows the intermediate outputs preprocessing tokens, TF-IDF query weights, and cosine similarity scores across FAQs so readers can verify that the selected answer is produced by a reproducible scoring process rather than an opaque rule. This visualization supports the scientific explanation by (1) demonstrating how input normalization affects term representation, (2) showing which terms contribute most strongly to the query vector, and (3) confirming that the chosen FAQ is the top-ranked item above the threshold. In addition, the same outputs are useful for iterative refinement (e.g., adjusting stopwords/tags or threshold) based on observed mismatches, thereby linking implementation evidence to retrieval evaluation.

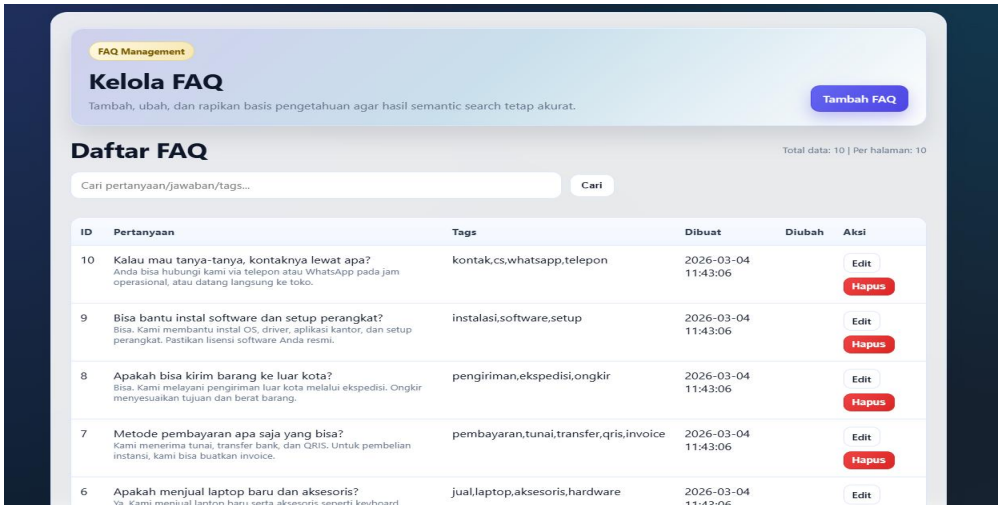

Figure 5. Admin FAQ Management

This figure presents the Admin FAQ Management page used to maintain the chatbot's knowledge base. The admin can perform CRUD operations (create, edit, delete) on FAQ items and manage supporting metadata such as tags, which are used to strengthen matching during retrieval. The table view displays the FAQ list along with timestamps (created/updated), and the search bar allows quick filtering by keyword. This module

ensures the system is maintainable and scalable, because knowledge updates can be done without changing program code, enabling continuous refinement based on user needs.

The results indicate that the proposed system works effectively as a lightweight retrieval-based FAQ chatbot. In the trial logs (N = 8 chats) with a 0.25 threshold, the system achieved an average similarity score of 0.4115 (overall) and 0.6584 for matched cases, indicating that most accepted matches were produced with relatively strong similarity. These quantitative indicators, together with the logging and analytics dashboard, support the effectiveness of the retrieval and ranking mechanism for practical FAQ automation: the admin dashboard enables continuous knowledge-base updates (CRUD), while the chatbot interface applies similarity scoring to avoid forcing low-confidence answers and instead provides top-3 related suggestions. This behavior is consistent with recent retrieval-based chatbot implementations that employ TF-IDF and cosine similarity to match user questions against FAQ data and improve the efficiency of information services, highlighting the importance of structured FAQ data and log-driven improvement[16]. In addition, recent research on retrieval-based chatbot designs for institutional information services emphasizes the value of retrieval and ranking strategies (including similarity-based retrieval and evaluation) to maintain response relevance and reliability when the knowledge base is curated and updated over time an approach that aligns with this study's logging and analytics module for identifying frequent queries and refining FAQ coverage[17].

Performance Evaluation

To validate the proposed semantic search and response ranking approach, this study evaluates the system from two perspectives: (1) functional correctness of the web application modules and (2) retrieval performance of the FAQ matching mechanism. Functional verification used 11 black-box test cases covering authentication, FAQ CRUD, chatbot request/response, logging, and analytics. A test case was marked Passed if the observed output matched the expected output (correct page redirection/message, correct database update, and correct access control), and Failed if the output deviated from the expected behavior or produced an error. All functional test cases were executed in the same deployment environment and documented in Table 1. conducted using black-box test cases covering authentication, FAQ management, chatbot interaction, logging, and analytics. For retrieval evaluation, the 8 chats shown in Table 3 represent an initial demonstration dataset collected during trial usage to verify that the retrieval pipeline (preprocessing, TF-IDF weighting, cosine ranking, and the 0.25 threshold) behaves as intended. The goal of this stage is validation and transparency, not large-scale benchmarking. Future work will expand the evaluation with a larger set of queries and broader FAQ coverage to provide stronger statistical evidence. Key indicators include Top-1 correctness, Top-3 hit rate, fallback rate (queries below the threshold), and similarity-score statistics. In this implementation, the chatbot applies a confidence threshold of 0.25, logs every interaction, and reports aggregate statistics such as total chats and average similarity scores (overall and matched) to support continuous improvement.

Tabel 1. Performance Evaluation

No	Module / Feature	Test Scenario	Expected Output	Result
1	Admin Login	Correct username & password	Redirect to admin dashboard	Passed
2	Admin Login	Wrong credentials	Error message, stay on login page	Passed
3	Session Security	Login process	Session regenerated, admin session stored	Passed
4	FAQ Create	Add new FAQ with valid data	Data saved & shown in list	Passed
5	FAQ Validation	Empty question/answer	Validation error shown	Passed
6	FAQ Edit	Update question/tags/answer	Updated content displayed	Passed
7	FAQ Delete	Delete FAQ with CSRF token	Data removed safely	Passed
8	Chat Ask Endpoint	Submit question via UI	Bot response returned (answer or fallback)	Passed
9	Fallback Mechanism	Low similarity query	Fallback + top-3 suggestions shown	Passed

10	Logging	Submit any chat query	Row inserted into chat_logs	Passed
11	Analytics	Open statistics page	Metrics and top lists displayed	Passed

Tabel 2. Retrieval Evaluation

User Query	Top-1 Matched FAQ (ID)	Top-1 FAQ Question (Short)	Cosine Score	Threshold (0.25)	Decision	Notes
Berapa lam waktu service	3	"Biasanya servis berapa lama?"	0.41707	0.25	Matched	Shown in debug screen
Hallo	–	–	<0.25 (example)	0.25	Fallback	Suggest top-3 topics
Jam operasional toko buka dari jam berapa sampai jam berapa ?	1	"Jam operasional…"	≥0.25	0.25	Matched	Typical exact FAQ match
Berapa biayanya	–	–	<0.25 (likely)	0.25	Fallback	Not in FAQ KB (pricing not stored)
Bisa servis laptop ?	2	"Melayani servis laptop & komputer?"	≥0.25	0.25	Matched	Should match service FAQ
Bantu saya jawab	–	–	<0.25 (likely)	0.25	Fallback	Non-informational query

Tabel 3. Summary Statistics

Metric	Value
Total chats	8
Chats today	8
Average similarity score (overall)	0.4115
Average similarity score (matched only)	0.6584
Threshold	0.25

As an additional note, the evaluation results confirm that the proposed approach is not only implementable but also measurable: the system can be assessed through functional testing outcomes and log-based retrieval indicators (similarity scores, matched frequency, and fallback behavior)[18],[19]. By combining a confidence threshold with transparent top-3 candidate suggestions, the chatbot reduces the risk of incorrect answers while still guiding users toward relevant topics. This makes the solution suitable for small-business deployment, where knowledge-base updates are frequent and operational simplicity is essential, and it provides a clear basis for future refinement using richer FAQ coverage and stronger semantic retrieval methods. Future work will extend the evaluation using a larger query set and more systematic metrics, following recent directions on structured evaluation frameworks for retrieval-augmented generation chatbots[20].

4. Conclusions and Future Works

This study developed a web-based FAQ chatbot for small businesses that applies semantic search and response ranking using TF-IDF and cosine similarity over an Indonesian FAQ knowledge base. The implemented system successfully provides relevant answers for specific queries and applies a confidence-threshold mechanism to prevent low-quality responses by returning fallback messages with top-3 related

question suggestions. The admin dashboard enables secure login, FAQ CRUD management, and log-based analytics, allowing continuous improvement of FAQ coverage based on real user interaction patterns. Overall, the results indicate that the proposed approach is practical, lightweight, and maintainable for small-business customer service automation.

5. References

[1] J. Cordero and L. Barba-guaman, "Use of chatbots for customer service in MSMEs," vol. 22, no. 1, pp. 185–197, 2026, doi: 10.1108/ACI-06-2022-0148.

[2] K. Marcineková, A. J. Sujová, and R. ˇDurica, "Implementing AI Chatbots in Customer Service Optimization — A Case Study in Micro-Enterprise," vol. 16, no. 1078, pp. 1–18, 2025.https://doi.org/10.3390/info16121078

[3] S. Adam and E. Lulianthy, "Frequenly Ask Question (FAQ) Chatbot for New Student Admission System Using Natural Language Processing at Politeknik Aisyiyah Pontianak," *JTKSI*, vol. 04, no. 03, 2021.

[4] G. H. Setiawan and I. M. B. Adnyana, "Improving Helpdesk Chatbot Performance with Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity Models," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 252–257, 2023.https://doi.org/10.30871/jaic.v7i2.6527

[5] B. S. Pramono, M. A. N. Fathihah, M. A. Dzikri, and I. M. C. D. Noto, "Penerapan TF-IDF dan Cosine Similarity dalam Chatbot Virtual Asisten : Studi Kasus Layanan Informasi Pendaftaran Sekolah Menengah Pertama," vol. 4, no. 1, pp. 150–158, 2025.

[6] D. Kristanto, R. A. Ramadhani, and A. B. Setiawan, "Pengembangan Chatbot Layanan Informasi Kampus Menggunakan TF-IDF," vol. 22, no. 2, pp. 103–115, 2025, doi: 10.33364/algoritma/v.22-2.2350.

[7] R. M. Holis, P. Eko, P. Utomo, and B. F. Hutabarat, "Semantic FAQ Chatbot Using SBERT (Sentence-BERT) and Cosine Similarity for Academic Services," vol. 5, no. 2, pp. 915–922, 2025.https://doi.org/10.47709/brilliance.v5i2.7027

[8] D. Steybe *et al.*, "Evaluation of a context-aware chatbot using retrieval-augmented generation for answering clinical questions on medication-related osteonecrosis of the jaw," *J. Cranio-Maxillo-Facial Surg.*, vol. 53, no. 4, pp. 355–360, 2025, doi: 10.1016/j.jcms.2024.12.009.

[9] X. Lin, X. Wang, B. Shao, and J. Taylor, "How Chatbots Augment Human Intelligence in Customer Services : A Mixed-Methods Study How Chatbots Augment Human Intelligence in Customer Services : A Mixed-Methods Study ABSTRACT," *J. Manag. Inf. Syst.*, vol. 41, no. 4, pp. 1016–1041, 2025, doi: 10.1080/07421222.2024.2415773.

[10] E. Aprianto, D. Mahdiana, and A. Wibowo, "Optimizing Bag of Words and Word2Vec with Vocabulary Pruning and TF- IDF Weighted Embeddings for Accurate Chatbot Responses in Indonesian Treasury Services," vol. 7, no. 1, pp. 587–605, 2026. https://doi.org/10.52436/1.jutif.2026.7.1.5370

[11] Asmaidin and C. B. Santoso, "Evaluasi Metode Retrieval pada Chatbot Domain Khusus Berbasis Retrieval-Augmented Generation," *JSAI J. Sci. Appl. Informatics*, vol. 09, no. 1, pp. 105–111, 2026. https://doi.org/10.36085/jsai.v9i1.9897

[12] D. Baur, J. Ansorg, C. Heyde, P. Med, and A. Voelker, "Development and Evaluation of a Retrieval-Augmented Generation Chatbot for Orthopedic and Trauma Surgery Patient Education : Mixed-Methods Study," *JMIR AI*, vol. 4, 2025, https://doi.org/10.2196/75262.

[13] R. Fernando, Y. D. Proboningrum, and S. D. Supriati, "NLP Implementation For AI Generated Text Detection (ChatGPT) Using Naive Bayes Method," no. 10, pp. 292–302, 2025. https://doi.org/10.32664/j-intech.v13i02.2026

[14] R. Yang, M. Fu, C. Tantithamthavorn, C. Arora, L. Vandenhurk, and J. Chua, "The Journal of Systems & Software RAGVA : Engineering retrieval augmented generation-based virtual assistants," *J. Syst. Softw.*,

vol. 226, no. January, p. 112436, 2025, doi: 10.1016/j.jss.2025.112436.

[15] M. S. Ibrahim, A. D. Sallibi, A. A. Hadi, and A. A. Nafea, "Passer Journal of Basic and Applied Sciences A Chatbot for Frequently Asked Questions using TF-IDF and Query Expansion Techniques," vol. 7, no. 1, pp. 506–512, 2025, doi: 10.24271/PSR.2025.487515.1805.

[16] G. Pratista, V. C. Mawardi, and I. Lewenusa, "Pemanfaatan Chatbot Retrieval-Based dan Analisis Sentimen Untuk Meningkatkan Layanan Informasi Interaktif di Radio Untar," *Com. Commun. Inf. Technol. J.*, vol. 3, no. 2, 2025, doi: 10.47467/comit.v3i2.8424.

[17] M. Yusuf, R. S. Perdana, and P. P. Adikara, "Analisis Perbandingan Kinerja Metode Retrieval Cosine Similarity dan BM25 pada Chatbot berbasis Retrieval-Augmented-Generation," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 1–6, 2026.

[18] S. U. Singh and A. S. Namin, "A survey on chatbots and large language models : Testing and evaluation techniques," *Nat. Lang. Process. J.*, vol. 10, no. January, p. 100128, 2025, doi: 10.1016/j.nlp.2025.100128.

[19] I. Justice, Y. M. Kobara, J. Owolabi, A. A. Akpan, and O. F. Offodil, "Conversational and generative artificial intelligence and human – chatbot interaction in education and research," vol. 32, pp. 1251–1281, 2025, doi: 10.1111/itor.13522.

[20] R. M. Amodia, C. Tirnauca, and M. Zorrilla, "SoftwareX RAGBOT CLI : a Python library for running and evaluating retrieval-augmented generation chatbots," *SoftwareX*, vol. 32, no. August, p. 102458, 2025, doi: 10.1016/j.softx.2025.102458.