Plagiarism Checker X - Report

Originality Assessment

1%

Overall Similarity

Date: Jan 9, 2026 (04:05 AM)
Matches: 37 / 3805 words
Sources: 3

Remarks: Low similarity detected, consider making necessary changes if needed.

Verify Report:
Scan this QR Code

Journal of Mathematics Instruction, Social Research and Opinion Vol. 5, No. 1, March 2026, pp. 15 – 25, https://doi.org/10.58421/misro.v5i1.984 ISSN 2962-7842 15 Journal homepage: https://journal-gehu.com/index.php/misro A Natural Language Processing-Based Chatbot as a Medium for Consultation and Education on Direct-Contact Infectious Diseases Serlina1, Eva Darnila2, Rini Meiyanti3 1,2,3Universitas Malikussaleh, Lhokseumawe, Indonesia Article Info ABSTRACT Article history: Received 2025-12-07 Revised 2026-01-02 Accepted 2026-01-03 Direct-contact infectious diseases such as influenza, diphtheria, tuberculosis (TB), scabies, varicella, impetigo, herpes simplex, and HIV remain public health threats. Limited access to accurate information encourages the development of chatbots as educational media. This study aims to design and build an NLP-based chatbot named SerMediCare to provide consultation and education on infectious diseases. The Research and Development (R&D) method with an iterative approach was used, including needs analysis, data collection from journals and medical books, and interviews with healthcare workers; system design; model training; and implementation on a web platform. The dataset was prepared in JSON format, including patterns, responses, and tags, and trained with a Transformer-based model to accurately recognize user intent. Evaluation results show that SerMediCare achieves 86% accuracy, indicating its ability to provide relevant responses to user queries. Black box testing confirmed that all features function properly. This chatbot is expected to be an effective digital tool for improving health literacy and facilitating public access to reliable information about infectious diseases. Keywords: Chatbot NLP Disease Education AI Corresponding Author: Serlina Faculty of Engineering, School of Informatics Engineering, Universitas Malikussaleh, Indonesia Email: serlina.210170195@mhs.unimal.ac.id 1. INTRODUCTION The rapid advancement of information technology has led to the development of various digital innovations, including chatbots powered by Natural Language Processing [1]. This technology enables systems to understand and respond to human language more naturally and interactively [2], allowing information and educational services to be delivered quickly, automatically, and efficiently. As a result, NLP-based chatbots have gained widespread adoption across multiple fields. In the healthcare sector, NLP-based chatbots have been effective [3] in providing accessible information without requiring direct interaction with medical professionals [4].

 These systems allow users to obtain accurate answers in a language they can easily understand, thus supporting improvements in public health literacy [5]. Previous studies also indicate that NLP significantly enhances user experience and satisfaction when accessing health-related information [6]. In Indonesia, infectious diseases transmitted through direct contact remain a persistent public health issue [7]. Conditions such as influenza, diphtheria, tuberculosis (TBC), scabies, chickenpox (varicella) [8], impetigo, herpes simplex, and human immunodeficiency virus (HIV) continue to show notable transmission rates, particularly in densely populated areas and regions with limited healthcare access [9]. This situation highlights the need for innovative solutions to strengthen public awareness and education. Direct-contact infectious diseases spread through physical touch or droplets expelled during coughing or sneezing [10]. Epidemiological studies emphasize that the transmission process involves multiple contributing factors, including the causative agent, the host, and environmental conditions [11]. Understanding these mechanisms is essential for strengthening preventive efforts, especially by providing accurate and easily understandable information to the community. NLP-based chatbots offer an innovative solution for delivering information on symptoms, risk factors, prevention strategies, and early management of infectious diseases [12]. With the ability to process natural language, these chatbots can interpret conversational context and respond to various forms of questions [13], including those expressed in different languages, expressions, or dialects—an important feature in Indonesia’s linguistically diverse population. Beyond educational purposes, NLP-based chatbots can reduce healthcare workers' workload by providing essential information commonly sought by the public [14]. These systems can be continuously updated to incorporate the latest developments in health information [15], including evolving epidemiological conditions. NLP-based chatbots serve as adaptive, sustainable tools for health communication [16]. Given these challenges and needs, this study aims to develop an NLP-based chatbot as a medium for consultation and education on direct-contact infectious diseases. The system is designed to deliver accurate and accessible information on illnesses such as influenza, diphtheria, tuberculosis, scabies, varicella, impetigo, herpes simplex, and HIV. It is expected that this chatbot will improve public understanding while contributing to reducing infectious disease transmission in Indonesia. 2. METHOD 2.1 Research Method This study employs the 3 Research and Development (R&D) method to develop an NLP-based chatbot designed [17] as a consultation and educational tool for direct-contact infectious diseases. This method was selected to ensure the resulting chatbot meets user needs while providing accurate, relevant information. The research stages include needs analysis, data collection, system design, development, testing, and evaluation, all conducted iteratively to achieve an optimal level of quality.

https://doi.org/10.58421/misro.v5i1.984 17 2.2 Data Gathering Procedure The data used in this study were collected through several structured approaches to ensure the chatbot received accurate, medically reliable information. Literature reviews were conducted by examining scientific journals and medical books on directcontact infectious diseases, including influenza, diphtheria, tuberculosis, scabies, varicella, impetigo, herpes simplex, and HIV. These sources provided comprehensive information on disease characteristics, including symptoms, transmission mechanisms, risk factors, prevention methods, and general treatment guidelines. In addition, consultations with medical professionals were conducted to obtain expert insights on critical information to convey to users, including frequently overlooked symptoms and practical preventive advice. This combination of data sources ensures that the dataset reflects both scientific evidence and real-world clinical experience. The collected information was then organized into a structured dataset suitable for Natural Language Processing applications. Data preparation involved compiling pairs of questions and answers (Q&A) based on common user inquiries and medically validated responses [18]. Medical insights obtained from consultations with healthcare professionals at RSUD Muda Sedia Aceh Tamiang were also incorporated to ensure clinical accuracy. The dataset was further refined through preprocessing steps to ensure clarity, consistency, and relevance before being processed by the NLP model. This approach yields a dataset that provides a solid foundation for training the chatbot to deliver accurate, context-aware, and user-friendly responses. 2.3 Data Analysis The data analysis stage focuses on examining the collected dataset to identify patterns, key concepts, and thematic structures related to direct-contact infectious diseases [19]. Each question–answer pair was evaluated to ensure medical accuracy, relevance, and alignment with the chatbot's educational objectives. The analysis also involved categorizing information into thematic groups, including symptoms, transmission modes, prevention strategies, and early treatment recommendations. This categorization supports the systematic organization of knowledge, enabling the NLP model to generate precise and contextually appropriate responses. To prepare the dataset for Natural Language Processing, linguistic and semantic analysis was conducted to assess variations in user language patterns [20], including synonyms, informal phrases, and common misunderstandings associated with infectious diseases. This step helps identify potential ambiguities, redundancies, and language inconsistencies that could affect the model's performance. By refining the dataset through semantic normalization and clarity enhancement, the analysis ensures the chatbot can interpret diverse user inputs and deliver accurate, easily understood answers. This process forms the foundation for building a robust and reliable NLP model.

https://doi.org/10.58421/misro.v5i1.984 3. RESULTS AND DISCUSSION 3.1. Long Short-Term Memory (LSTM) Computation This section explains the mathematical computations that occur in each main component of an LSTM, presented in the following order: the cell state update, the forget gate, the input gate, and the output gate. The step-by-step description aims to clarify how the LSTM manages and stores important information [21] over long time horizons through interacting gates and states. The mathematical formulas for these computations are given below. A. Cell State (Ct) Previous Cell state (C_{t-1}) : [0.4, –0.3] Forget gate (f_t) : [0.562, 0.635] Input gate (i_t) : [0.55, 0.622] Candidate cell state ($\tilde{c}_t$) : [0.245, 0.422] Computation Steps a. First Element: Forget : 0.562 × 0.4 = 0.2248 Input : 0.55 × 0.245 = 0.1348 Sum : 0.2248 + 0.1348 = 0.3596 b. Second Element: Forget: 0.635 × (–0.3) = –0.1905 Input: 0.622 × 0.422 = 0.2625 Jumlah: –0.1905 + 0.2625 = 0.072 Resulting Cell State (C_t) = [0.3596, 0.072] From the computation results, the updated cell state is obtained as C_t = [0.3596, 0.072]. This value is produced by combining retained information from the forget gate with newly added information from the input gate. The first component is calculated by preserving a portion of the previous memory (0.562 × 0.4) and adding new information (0.55 × 0.245), while the second component follows the same principle for the corresponding element. In this way, the cell state serves as a long-term memory repository that is selectively updated, maintaining a balance between past information and newly introduced signals. This mechanism ensures that only relevant information is preserved, thereby supporting accuracy and consistency in sequential learning processes. B. Forget State The sigmoid activation yields the forget gate output f_t = [0.562, 0.635]. These values represent the degree of retention applied to the previous cell state. The first value, 0.562, indicates that approximately 56.2% of the information from the first element of the previous cell state is preserved, while the rest is discarded. Similarly, the second value, 0.635, signifies that about 63.5% of the information from the second element remains relevant and is carried forward to the following computation stage within the LSTM unit. In essence, the forget gate

https://doi.org/10.58421/misro.v5i1.984 plays a critical role in regulating the flow of long-term memory by removing information deemed unnecessary while retaining information that contributes to subsequent prediction or learning processes. C. Input State For the input gate, the sigmoid activation produces i_t = [0.55, 0.622]. These values indicate that approximately 55% of the new information will be incorporated into the first element of the cell state, while about 62.2% will be incorporated into the second element. The input gate plays a crucial role in the memory-updating process, regulating how much of the current input contributes to the cell state. Higher activation values correspond to a greater amount of new information being stored, meaning the input gate acts as a filter that determines the relevance of incoming data. These values also reflect the model’s level of confidence in the new information. Thus, the input gate helps maintain a balance between previously stored memory and newly acquired information 3.2. Model Evaluation The evaluation assessed the performance of the SerMediCare chatbot model using two main metrics during training: accuracy and loss. Figure 1. Evaluation Model Figure 1 presents the evaluation results of the NLP-based chatbot model training using two primary metrics, namely accuracy and loss, on the training data over 200 epochs. 3.3. Deployment Deployment in the SerMediCare application is the process of deploying the Natural Language Processing (NLP)-based chatbot into a website platform using the Flask framework. This process includes integrating all system components, such as the user interface, chatbot backend, and the database that supports the question-and-answer functionality. With this deployment, users can access the application online without downloading or installing additional applications. This makes it easier for users to conduct consultations and obtain educational information regarding infectious diseases transmitted through direct contact, including influenza, diphtheria, tuberculosis, scabies, varicella, impetigo, herpes simplex, and HIV. With proper deployment, the

https://doi.org/10.58421/misro.v5i1.984 system can operate in real time, remain responsive, and be ready for use by the community at any time. a. Home Page Figure 2. Home Page The Home Page displays the main interface of the SerMediCare website, which is the deployed result of the Natural Language Processing-based chatbot for consultation and education on infectious diseases transmitted through direct contact. b. Start Chatbot Page Figure 3. Start Chatbot Page On the Start Chatbot

Page, users can begin interacting with the chatbot to consult and obtain educational information regarding various infectious diseases transmitted through direct contact. The interface displays a real-time conversation area between the user and the chatbot. In the example shown, the user types the message “hello,” and the chatbot responds with a greeting, “How are you? Is there anything I can help you with?”. Furthermore, when the user asks “what is HIV?”, the chatbot provides an informative response, explaining that

https://doi.org/10.58421/misro.v5i1.984 HIV is a virus that attacks the human immune system, particularly CD4 cells. If left untreated, HIV can progress into AIDS (Acquired Immunodeficiency Syndrome). The SerMediCare chatbot accurately and quickly responds to user input based on embedded data, indicating that the system is functioning well and effectively. The simple, responsive user interface also facilitates interaction, while the chatbot’s ability to understand the context of questions demonstrates that the integration of the NLP model and dataset has been successfully implemented. c. Information Page Figure 3. Start Chatbot Page The Information Page of the SerMediCare chatbot application highlights six main features designed to provide information and education related to infectious diseases transmitted through direct contact. These six features are presented in an informative and accessible manner to help users understand the symptoms, causes, transmission, prevention, and treatment of various diseases. The first feature, Disease Consultation, allows users to consult the chatbot directly for basic information on infectious diseases, including symptoms and treatment methods. This feature serves as the main entry point for users to explore information according to their needs. The second feature, Prevention and Education, provides users with knowledge about appropriate preventive measures and relevant health education to reduce the risk of disease transmission. The third feature is the Interactive Feature, which enables users to engage in direct question-and-answer interactions with the chatbot. Through this feature, users can obtain personal and instant information based on the questions they ask. The fourth feature is Real-Time Service, demonstrating the chatbot’s ability to respond quickly and directly to user questions, eliminating the need for long waiting times to receive answers to their concerns or needs. The fifth feature, Personalized Experience, allows the chatbot to provide responses tailored to previous conversation history or user questions, making the interaction more relevant and personalized. The sixth feature is Access for All, ensuring that all chatbot services can be accessed by anyone, anywhere, at any time. This aims to make health information accessible to all layers of society, including those living in areas with limited access to healthcare professionals. With these six features, the SerMediCare chatbot is expected to serve as an effective and adaptive medium for

 consultation and education, helping communities understand and manage infectious diseases transmitted through direct contact. 3.4 Discussion on Chatbot as a Medium for Health Consultation and Education The chatbot system addresses directly transmitted infectious diseases, including influenza, diphtheria, tuberculosis, scabies, varicella, impetigo, herpes simplex, and HIV. It was developed using a supervised learning approach with a categorized Q&A dataset structured into tags, patterns, and responses. The classification model was trained and saved as a .h5 file, then integrated into a web interface using the Flask framework. Figure 4. Chatbot as a Medium for Consultation and Education This section presents a practical example of user-chatbot interaction on the topic of herpes simplex. When the user asks, “What is herpes simplex?”, the chatbot provides a comprehensive explanation: it is an infection caused by the herpes simplex virus (HSV) that affects the oral (oral herpes) or genital (genital herpes) areas. It highlights that the disease is highly contagious through direct contact with bodily fluids and provides information on common symptoms such as itching, pain, and fluid-filled blisters, emphasizing the importance of treatment for faster recovery. When the user follows up with “What symptoms usually appear in herpes simplex patients?”, the chatbot delivers a detailed response, describing clinical symptoms including painful skin blisters, pain, fever, and swollen lymph nodes. These responses demonstrate the chatbot’s ability to understand user intent through natural language processing (NLP) and provide accurate, relevant, and educational information on the disease. In addition to the interaction example shown above, Figure 4 illustrates the chatbot interface implemented on a web-based platform. The interface is designed to be user-friendly and straightforward, allowing users to submit health-related questions and receive responses in real time easily. This visual representation supports the practical implementation of the chatbot as a consultation medium accessible without medical expertise.

https://doi.org/10.58421/misro.v5i1.984 The findings indicate that the chatbot can deliver structured, informative responses on disease definitions, symptoms, and transmission. This supports the role of natural language processing (NLP) in enabling the system to understand user intent and provide relevant answers. The conversational format shown in Figure 4 also helps users engage more actively with the system, which is important for health education purposes. These results are consistent with previous studies that reported chatbot-based health systems as effective tools for improving access to medical information[22]. Prior research has highlighted that chatbots can reduce barriers to health consultation by offering quick responses and minimizing user hesitation 23], especially when discussing sensitive health topics. Compared to static health information platforms, the proposed chatbot provides a more interactive and personalized learning experience. However, the performance of the chatbot is influenced by the quality of the training dataset and predefined question–answer patterns[24]. When user input closely matches the trained data, the chatbot produces accurate and relevant responses, as demonstrated in the interaction example. Similar limitations have been identified in earlier studies, suggesting that continuous dataset expansion and model improvement are necessary to enhance system robustness[25]. Overall, the chatbot demonstrated in Figure 4 shows potential as a supportive medium for consultation and health education. With further development, such as expanding disease coverage and improving language understanding, the chatbot can complement existing digital health information systems. 4. CONCLUSION Based on the 3 research and development of the SerMediCare chatbot, several conclusions can be drawn. The NLP-based chatbot was successfully designed and developed as a medium for consultation and education on directly transmitted infectious diseases, including influenza, diphtheria, tuberculosis, scabies, varicella, impetigo, herpes simplex, and HIV, with NLP technology enabling it to understand and respond to user questions contextually. Black Box testing demonstrated that the chatbot provides accurate and relevant responses to various user inquiries. SerMediCare has been successfully deployed on a web platform using the Flask framework, allowing the public to access health consultation and educational services easily and quickly, without direct contact with healthcare professionals. Model performance evaluation showed significant improvements in accuracy and reductions in loss during training, indicating the system's growing ability to understand user question patterns. The chatbot achieved 86% accuracy, demonstrating strong performance in classifying questions and delivering appropriate responses. Beyond serving as a source of basic information on infectious diseases, the chatbot enhances public health literacy, reduces misinformation, and alleviates the educational burden on healthcare providers.

REFERENCES

[1] F. Aslam, "The Impact of Artificial Intelligence on Chatbot Technology: A Study on the Current Advancements and Leading Innovations," 2023. [Online]. Available: www.ajpojournals.org [2] Z. Wang et al., "Interactive Natural Language Processing," May 2023, [Online]. Available: http://arxiv.org/abs/2305.13246

[3] H. Mendapara, S. Digole, M. Thakur, and A. Dange, "AI Based Healthcare Chatbot System by Using Natural Language Processing," International Journal of Scientific Research and Engineering Development, vol. 4, 2021, [Online]. Available: www.ijsred.com [4] C. Bulla, C. Parushetti, A. Teli, S. Aski, and S. Koppad, "A Review of AI Based Medical Assistant Chatbot," 2020. [5] P. J. Fitzpatrick, "Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners," 2023, Frontiers Media SA. doi: 10.3389/fdgth.2023.1264780. [6] K. Nawab, G. Ramsey, and R. Schreiber, "Natural Language Processing to Extract Meaningful Information from Patient Experience

Feedback," Appl Clin Inform, vol. 11, no. 2, pp. 242–252, Mar. 2020, doi: 10.1055/s-0040-1708049. [7] S. Fauziyah et al., "How should indonesia consider its neglected tropical diseases in the COVID-19 era? Hopes and challenges (review)," Biomed Rep, vol. 14, no. 6, Jun. 2021, doi: 10.3892/br.2021.1429. [8] J. A. Guzman-Cottrill et al., "SHEA practice update: Infection prevention and control (IPC) in residential facilities for pediatric patients and their families," Infect Control Hosp Epidemiol, vol. 46, no. 1, pp. 3–26, Jan. 2025, doi: 10.1017/ice.2024.124. [9] L. Chen, T. Chen, T. Lan, C. Chen, and J. Pan, "The Contributions of Population Distribution, Healthcare Resourcing, and Transportation Infrastructure to Spatial Accessibility of Health Care," Inquiry (United States), vol. 60, Jan. 2023, doi: 10.1177/00469580221146041. [10] K. Randall, E. T. Ewing, L. C. Marr, J. L. Jimenez, and L. Bourouiba, "How did we get here: What are droplets and aerosols and how far do they go? A historical perspective on the transmission of respiratory infectious diseases," Oct. 12, 2021, Royal Society Publishing. doi: 10.1098/rsfs.2021.0049. [11] R. M. . Merrill, Introduction to epidemiology. Jones and Bartlett Publishers, 2024. [12] S. Chakraborty et al., "An AI-Based Medical Chatbot Model for Infectious Disease Prediction," IEEE Access, vol. 10, pp. 128469–128483, 2022, doi: 10.1109/ACCESS.2022.3227208. [13] M. Aleedy, H. Shaiba, and M. Bezbradica, "Generating and Analyzing Chatbot Responses using Natural Language Processing," 2019. [Online]. Available: www.ijacsa.thesai.org [14] I. Odo and H. Ajose-Adeogun, "Quick Response Code Cross Current International Journal of Medical and Biosciences Abbreviated Key Title: Cross Current Int J Med Biosci Enhancing Healthcare Communication Efficiency Through NLP-Driven Chatbots: Impact on Patient Satisfaction and Practitioner Workload," Cross Current Int J Med Biosci, vol. 4, no. 5, p. 76, 2022, doi: 10.36344/ccijmb.2022.v04i05.001. [15] O. Baclic, M. Tunis, K. Young, C. Doan, and H. Swerdfeger, "Challenges and opportunities for public health made possible by advances in natural language processing," Canada Communicable Disease Report, pp. 161–168, Jun. 2020, doi: 10.14745/ccdr.v46i06a02. [16] Roshni Ojha, "From 2 Algorithms to Conversations: The Influence of Natural Language Processing on Chatbot Innovation,"

2024. [17] S. Marchese, "Tesi Di Laurea Magistrale In Management Engineering-Ingegneria Gestionale Influence Of Nlp On Organizational Design And Development Of Dynamic Capabilities," 2024. [18] W. U. Hasan, "Design And Evolution Of Conversational Ai For Healthcare: From Structured Data Collection To Culturally Sensitive And Adaptive Support For Chronic Disease Management And Adrd Care" 2025. [19] L. Judijanto, H. Hermansyah, K. P. Ningsih, D. Anurogo, and M. Firdaus, "The Role of Big Data Technology in Predicting and Managing the Spread of Infectious Diseases," Journal of World Future Medicine, Health and Nursing, vol. 2, no. 2, pp. 216–227, Mar. 2024, doi: 10.70177/health.v2i2.757. [20] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," Multimed Tools Appl, vol. 82, no. 3, pp. 3713–3744, Jan. 2023, doi: 10.1007/s11042-022-13428-4. [21] J. Abonyi, R. Károly, and G. Dörgö, "Event-tree based sequence mining using LSTM Deep-learning model," Complexity, vol. 2021, 2021, doi: 10.1155/2021/7887159. [22] Z. Xiao, Q. V. Liao, M. Zhou, T. Grandison, and Y. Li, "Powering 1 an AI Chatbot with Expert Sourcing to Support Credible Health Information Access," in International Conference on Intelligent User Interfaces, Proceedings IUI, Association for Computing Machinery, Mar. 2023, pp. 2–18. doi: 10.1145/3581641.3584031. [23] Y. C. Tseng, W. Jarupreechachan, and T. H. Lee, "Understanding the Benefits and Design of Chatbots 3 to Meet the Healthcare Needs of Migrant Workers," Proc ACM Hum Comput Interact, vol. 7, no. CSCW2, Oct. 2023, doi: 10.1145/3610106.

 [24] N. A. N. M. Isa, S. N. A. Jawaddi, and A. Ismail, "Experimental Evaluation of Machine Learn ing Models for Goal-oriented Customer Service Chatbot with Pipeline Architecture," Sep. 2024, [Online]. Available: http://arxiv.org/abs/2409.18568 [25] I. Basharat and S. Shahid, "AI-enabled chatbots healthcare systems: an ethical perspective on trust and reliability," J Health Organ Manag, 2024, doi: 10.1108/JHOM-10-2023-0302.

Sources

1	https://juji.io/publications/ INTERNET <1%
2	https://multiarticlesjournal.com/abstract_more_data/304/from-algorithms-to-conversations-the-influence-of-natural-language-processing-on-chatbot-innovation INTERNET <1%
3	https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000492 INTERNET <1%

EXCLUDE CUSTOM MATCHES	OFF
EXCLUDE QUOTES	OFF
EXCLUDE BIBLIOGRAPHY	ON