Human-Robot Interaction on Elderly Companion Robot Development Using Dual Intent Entity Transformer

Naufal Haidar Rauf
Department of Informatics Engineering
Faculty of Computer Science
Universitas Dian Nuswantoro
Semarang, Indonesia
111201912342@mhs.dinus.ac.id

Heru Agus Santoso
Department of Informatics Engineering
Faculty of Computer Science
Universitas Dian Nuswantoro
Semarang, Indonesia
heru.agus.santoso@dsn.dinus.ac.id

***Abstract*—The growing elderly population and the increasing demand for nursing home care have led to a need to improve the quality of life for residents. A popular solution is developing a companion robot to assist residents with various tasks and provide human-like interaction. In this paper, we present the prototype of a companion robot currently website-based, equipped with a Dual Intent Entity Transformer to understand what the user wants. The results of the study show the prototype is capable of classifying user messages according to their appropriate intent with precision, recall, F-score, and accuracy on average 91%, 91%, 90%, and 87% respectively using privately owned dataset. Future works such as expanding the robot's insight and tuning the model would improve the service provided by the companion robot.**

***Keywords*—companion robot, nursing home, robot development**

I. Introduction

Social interaction is fundamental for exchanging information, perceiving and responding to verbal, facial, and gesture cues [1], which is especially important for the elderly living in nursing homes where caregivers face challenges in providing the best care due to the steady increase of elderly population every year [2], deteriorating health of elderly, financial insecurity, loss of friend or family member [3], this present as a challenge for the caregiver in providing the best care for the resident living in nursing home.

Society 5.0 is defined as a society led by technological and scientific innovation [4]. This society aims to build a human-centered society that can easily provide products and services that meet a variety of underlying needs, reduce economic and social disparities, and thereby significantly improve the quality of life [5]. The paucity of caretakers is evident in a field study being conducted at the Harapan Ibu Nursing Home in Semarang. Caretakers at the Harapan Ibu Nursing Home in Semarang became overworked due to the resident-to-staff ratio of 10:1, which led to several issues, including residents forgetting to take their medication, forgetting to pray, and a lack of spiritual and entertainment fulfillment. A typical solution to these problems is to create a companion robot to aid caregivers in providing vital care. In line with the Society 5.0 goal, efforts are underway to develop a companion robot to assist the daily life of the elderly [6]–[8].

The current technological capability of companion robot development requires components such as a user interface, knowledge base, and NLU module. A component of the aforementioned NLU module called an intent classifier serves as a link between user interaction within the system and their intentions [9]. Any companion robot must have an intent classifier since users speak in a variety of ways even when their words or intentions are the same. For instance, saying "Give me information about diabetes medication" or "Hey, did you know the medicine to treat diabetes?" can communicate the desire for medical information. As such, the DIET Classifier has been developed by RASA to bridge the gap between varying message patterns and similar intentions [10]. The very component that will be exposed first is the user interface. Therefore, component usage must be simple for the user to assist users in making decisions [11], especially for users with visual impairment such as blindness [12]. In general, the development of the software of companion robots is as long as the development of the hardware, if not longer. Typically require several algorithms. This further increases the development time as researchers need to familiarize themselves with each algorithm. Therefore, a solution to develop a chatbot in a companion robot with a platform that provides the algorithm for the chatbot to reduce complexity is necessary.

In this paper, we propose a prototype of a companion robot i.e., Sekar, that is capable of understanding user messages with a Dual Intent Entity Transformer, in addressing problems faced by caregivers at Harapan Ibu Nursing Home. To limit the scope of the research, we focused on the chatbot as the software of the companion robot, built using a website-based framework and using RASA as a framework to develop the model. The key contribution of this paper is building a website-based companion robot prototype. The rest of the paper is organized as follows. Section II presents an overview of relevant research on companion robots. Section III describes the main technologies of the proposed companion robot. Finally, Section IV describes the implementation and results of the proposed method, and Section V summarizes the work and outlines future work.

II. Related Works

A conversational agent/chatbot must be able to discern user messages and respond appropriately to suit the user's needs. This skill, namely intent prediction/intent classification, serves three functions. The first goal is to determine user intent so that the chatbot can deliver answers and adapt prior responses. The second goal of intent categorization is to learn and replicate human agent behavior so that the conversational agent can distinguish distinct intent and when to provide spontaneous requests for more information or specifics. The final goal of intent categorization, specifically the trained model, is to be used for data analysis and conversational answer generation [13].

Any task-oriented conversational system must have intent detection [14]. To grasp the user's present objective, the system must utilize its intent classifier to categorize the user's utterance (given in various natural languages) into one of several predetermined classifications, i.e., intents. For example, in the e-banking domain, intents can be a *failed top-up* or *lost credit card*. Getting the intent wrong results in the first point of failure for a conversational agent [15].

The user interface is another key feature of a conversational agent. A good conversational agent should have a user interface that is easy to use and fulfills the needs of the user. There are various approaches to creating user interfaces. User-centered design (UCD) is one of the methods to create it [16]. The goal of UCD is to create user interfaces that promote user satisfaction by matching the design with the user's expectations and goals [17] and offering greater information retention [18]. The breakthrough in the Artificial Intelligence and Machine Learning department has resulted in a user-friendly Voice User Interface (VUI) and an accessible ICT solution. A case study of Voice User Interface encourages and assists the elderly in using Internet of Things (IoT) devices in their homes [19].

III. METHOD

A. DIET (Dual Intent Entity Transformer) Classifier

To properly engage with the user in chatbot systems, we must first comprehend what the user wants. Intent classification (i.e. categorizing the inquiry) can propose plausible limits for obtaining an answer. One case example would be "Apa itu tadarusan?" (English: What does Quran recitation mean?) and "Mau nanya tentang pengajian" (English: Wanna ask about Quran recitation). If the chatbot understands those questions about Quran recitation in Indonesian, presented in formal and casual style, the chatbot will likely be less confused and serve the relevant answer.

The proposed chatbot uses a DIET Classifier as its intent classification. DIET (Dual Intent Entity Transformer) is a new multitask architecture for entity classification and recognition with the ability to incorporate trained word embeddings from the model language and combine them with sparse words in plug-and-play mode. A DIET that does not include trained embeddings will improve difficult NLU data sets. Furthermore, incorporating a pre-trained embedding of words and sentences from the language model enhances the overall accuracy of all tasks [20]. The representation of DIET architecture shown in Fig. 1 has several important parts. The phrase "mainkan lagu Metallica one" has the intention of playing a song from Metallica titled "One" and the entity name is "lagu_lagu" (English: songs) with the weight of the feed-forward layer distributed to the token.

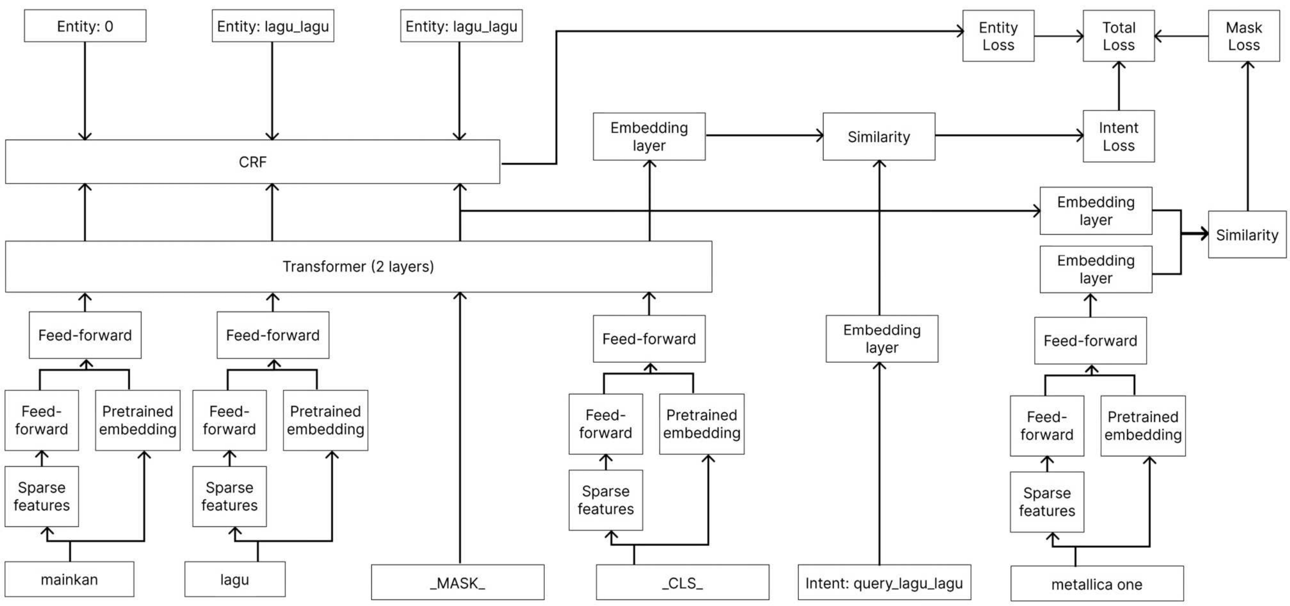

Fig. 1. Architecture of DIET Classifier with an example query "mainkan lagu Metallica one" (English: play Metallica one song)

B. Architecture Overview

The proposed approach shown in Fig. 2 takes advantage of the RASA Platform, a ready-to-use conversational AI question-and-answer action with initial knowledge of different domains and present it to the user with a style that mimics a chatting app.

framework using Python as its programming language and JSON to store the knowledge [21] Fig. 3 depicts its architecture. The proposed companion robot would engage in

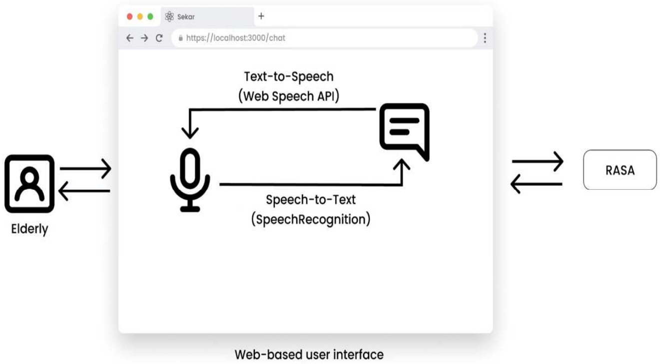

Fig. 2. Architecture of Sekar

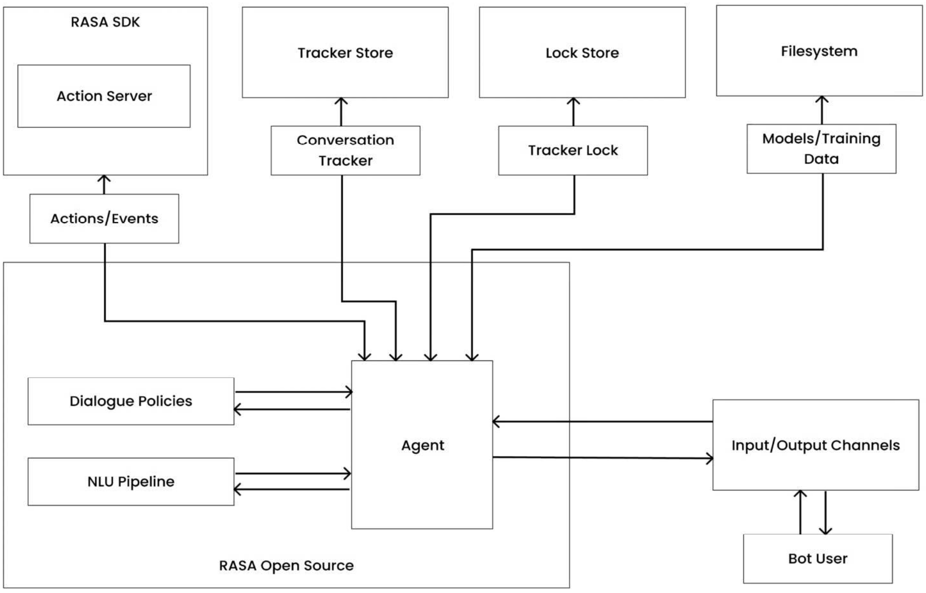

Fig. 3. RASA Architecture

The website-based user interface is developed using React.js [22], a JavaScript library that provides a live preview feature for a better developer experience, a vast library to offers such as SpeechRecognition [23] for developing speech-to-text feature, and Tailwind CSS [24] for shaping up component of the interface (i.e. button, text, etc).

The conversation with the chatbot went as follows:

- The user will ask questions to the companion robot, whether by typing the question or using dictation to ask verbally. Speech-to-text function is handled using SpeechRecognition while text-to-speech is handled using Web Speech API [25]. These functions are present in the proposed companion robot to help the disabled elderly interact with the robot.
- After typing/dictating the question, the user must click the "Send" button, indicated with a paper airplane icon,
- Lastly, the robot will respond with the appropriate answer and the cycle continues

RASA Platform architecture represented in Fig. 3 has several key components. NLU Pipeline is the part responsible for intent classification, entity extraction, and response retrieval. Dialogue Policies is the name of the dialogue management part that decides the next action that should be taken in the conversation based on the context provided [26].

To build the NLU model used in the NLU Pipeline, RASA provides support for developing a knowledge base simply using a .json file. Setting up the pipeline for the NLU model is done through the config.yml file. This pipeline covers intent classifier, entity extractor, and policies. Configuration of our pipeline, visualized in the chart, shown in Fig. 4.

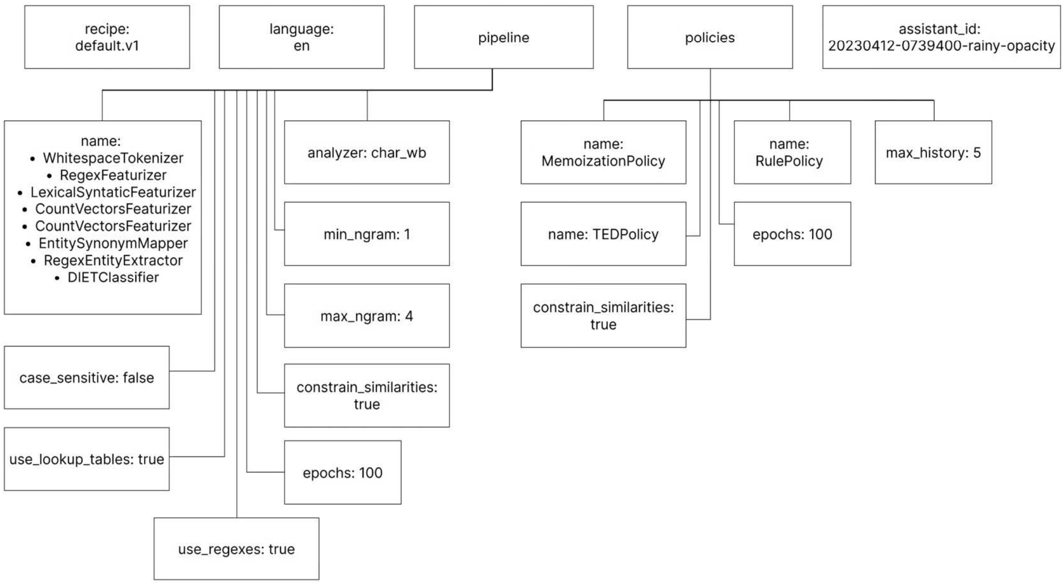

Fig. 4. NLU pipeline of the proposed system

IV. Implementation and Evaluation

A. Design Implementation and Demonstration Environment

In general, developing a chatbot takes a long time. We chose to set the chatbot in website-based form to shorten the development process because web-based applications are easy to create and install. The proposed chatbot also incorporates a RASA Open Source third-party conversational AI system. Similar conversational AI solutions, such as Google Nest and Alexa, have limits in terms of study, user interface design, and scalability.

The proposed chatbot for companion robot has the following features:

- Scalable knowledge base: Adding a new knowledge base or removing a knowledge base is as easy as manipulating the .json file, then executing the "rasa train" command on the terminal to retrain the model,
- Real-time: User can have a conversation with a robot in real-time,
- Multi-platform: Since it's web-based, it theoretically can be operated from different devices, whether smartphone, laptop, tablet, etc., but current development is only optimized on desktops.

As for the demonstration, we are using the Windows 10 operating system and Google Chrome browser. Hardware used are listed below:

- Intel NUC 11 Performance Kit | NUC11PAHi7 with 4 cores 8 threads processor, 8 GB DDR4 RAM, and 256 GB SSD storage,
- External microphone for audio input that is required in speech-to-text function,
- External speaker/headphone for audio output that is required in text-to-speech function,
- 24-inch computer monitor with Full HD resolution to display the visual,
- Keyboard for sending messages/queries in typed form

B. Intent Classification Task

Performing tests on intent classification tasks are done programmatically using terminal commands. For this test, we use "rasa data split nlu" to split the data for training and testing purposes. Then, we use "rasa test nlu --cross-validation" to perform a cross-validation test that will utilize both training and testing data.

Setting up stories/scenarios to be tested is done through the test_stories.yml file. Each domain (i.e., religious, entertainment, healthcare, personalization) will have 2 scenarios/test paths. Fig. 5 shows a snippet of scenarios being tested on our model.

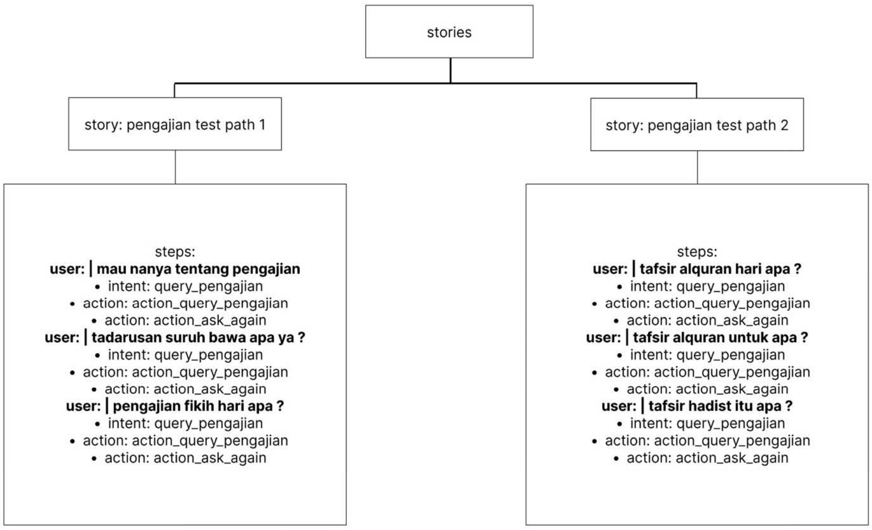

Fig. 5: Snippet of stories for testing intent classification

Performing intent classification tests to determine the model capability using datasets that are privately owned. This dataset contains different variations of question and request sentences in Indonesian and the path and actions that will be taken by the model. Measuring the intent classification commonly uses precision, recall, F-score, and accuracy. Each of these measurements uses a combination of different values.

The precision of the model is calculated by dividing the true positives value and the combined value of true positives and false positives, shown in (1). True positives are the total value of expected positive intent correctly predicted as positive by the model and false positives are the total value of expected positive intent incorrectly predicted by the model.

$$Precision = \frac{(true\ positives)}{(true\ positives + false\ positives)} \quad (1)$$

The recall is calculated by dividing the true positives value and the combined value of true positives and false negatives, shown in (2). False negatives are the total value of expected negative intent that is incorrectly predicted.

$$Recall = \frac{(true\ positives)}{(true\ positives + false\ negatives)} \quad (2)$$

F-measure, also known as F-score or F_1-score is calculated using the formula shown in (3). This measurement relies on precision and recall, where precision requires true positives and false positive values while recall requires true positives and false negative values.

$$F - measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (3)$$

The accuracy of the model is calculated using the formula shown in (4). This measurement combined four values: true positives, true negatives, false positives, and false negatives.

$$Accuracy = \frac{(true\ positives + true\ negatives)}{(true\ positives + true\ negatives + false\ positives\ +\ false\ negatives)} \quad (4)$$

Using those equations results in each query getting precision, recall, and F-score. Accuracy is the exception since its value is applied to all queries. To get the average score for each measurement, all scores from each query on certain measurements are added and then divided by how many the queries are being tested. In this proposed system we tested 5 queries; "query_solat", "query_obat", "query_lagu_lagu", "query_penyakit", and "query_pengajian".

Precision on average is 91%, meaning that intent classification is correct 91% of the time. Recall on average is 91%, meaning it correctly classifies 91% of all intents. and the F-score on average is 90%. Accuracy for the intent classification task is 87%, meaning the intent classifier correctly predicts 87% intent from the total predictions. Some of the challenges faced during the intent classification test are a limited reference to the test conducted in previous studies, abstracted test results from RASA as in the breakdown on how much the 4 values (total true positives, true negatives, false positives, and false negatives) from each parameter, and sometimes there is an error that didn't even hint the root cause of the error. To overcome these challenges, we present the results as-is, generated from RASA.

More work is needed to improve the model such as expanding the user's question example, expanding the knowledge base entries, rewriting test paths, etc.

C. User Interface Implementation

Any chatbot must have an efficient user interface because it directly affects the user's satisfaction and experience. Some of the user interface implementation and dialogue between the robot and user are shown in Fig. 6 and Fig. 7 along with additional explanation and the translated conversation in English.

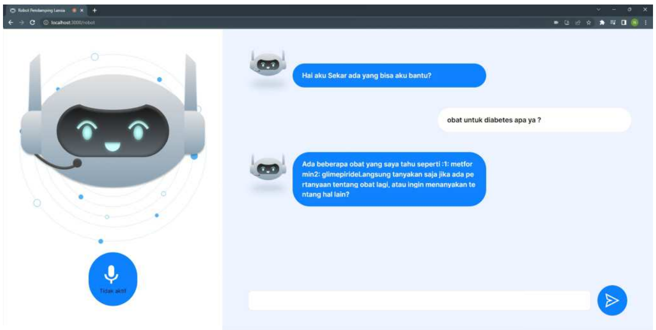

Fig. 6: Asking for diabetes medication

Starting the conversation, Sekar greets the user with "Hi, I'm Sekar. Can I help you?" to which the user replies "What is the medication for diabetes?". After Sekar understand that the user asking about diabetes medication, it responds with "There are several medicines that I'm aware of such as 1:metformin, 2: glimepiride. Feel free to ask if you have other question about medication, or would you like to ask about another thing?". For context, metformin and glimepiride are medications to control high blood sugar in type 2 diabetes patients. Glimepiride may be used with other diabetes medications [27], [28].

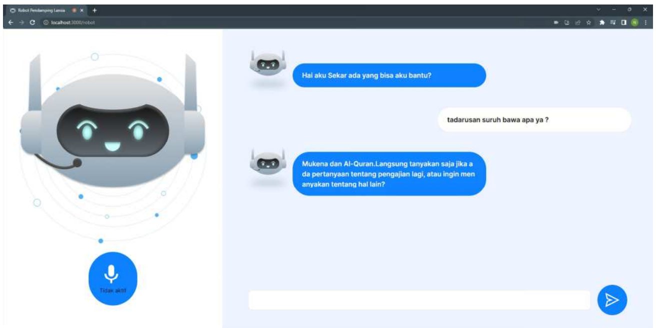

Fig. 7: Asking for what to bring during Quran recitation

In another instance, Sekar greets the user again as in Fig. 6 explanation, and asks if there is anything the user needs help with. This behavior is expected since Sekar has been programmed to greet every time the webpage is rendered, be it the moment this prototype fired up or when the user refreshes the webpage. This time, a user asked "What are the items to bring for Quran recitation?" to which Sekar responded with: "Mukena and Quran. Feel free to ask if you have other questions about Quran recitation, or would you like to ask about another thing?" Mukena is a type of hijab well-known in Indonesia that is almost exclusively worn by Muslim women for conducting prayer [29].

Some of the challenges faced in the user interface development are the layout of the conversation page can not adapt outside the resolution of the computer monitor previously stated and the robot logo that was supposed to represent facial expression could not be integrated with the website-based framework at the moment. To overcome these challenges, we choose to keep the user interface present on a current computer monitor and disable the face expression feature until a solution to integrate it has been found.

The current iteration of the user interface still requires some improvement such as adding a function to automatically send a message after the microphone is inactive for a few seconds and increasing flexibility to adapt on different resolution display devices.

V. Summary

This study suggested the development of a companion robot chatbot prototype using React.js for the user interface and RASA for the NLU model. The suggested chatbot makes use of the user-friendly RASA conversational AI, which incorporates DIET (Dual Intent Entity Transformer), a RASA-developed intent classifier. Analysis of the results of the intent classification test demonstrates that RASA is highly confident in its interpretation of user messages. The suggested solution offers a user-friendly interface for interacting with the robot as well.

Future work on the suggested system will address flaws in the existing implementation and incorporate improvements including expanding the knowledge base, adding a facial expression to indicate the robot's current attitude, improving the chatbot's layout responsiveness, and deploying the chatbot online.

References

[1] C. D. Frith and U. Frith, "Social Cognition in Humans," *Curr. Biol.*, vol. 17, no. 16, pp. R724–R732, Aug. 2007, doi: 10.1016/j.cub.2007.05.068.

[2] BPS-Statistics Indonesia, "Elderly Statistics 2022." BPS-Statistics Indonesia, Dec. 27, 2022. Accessed: Apr. 07, 2023. [Online]. Available: https://www.bps.go.id/publication/2022/12/27/3752f1d1d9b41aa69be4c65c/statistik-penduduk-lanjut-usia-2022.html

[3] BKKBN-National Population and Family Planning Board, "Healthy, Active, and Dignified Elders," BKKBN. Accessed: Apr. 07, 2023. [Online]. Available: https://www.bkkbn.go.id/berita-lansia-sehat-aktif-dan-bermartabat

[4] "Report on The 5th Science and Technology Basic Plan." Council for Science, Technology and Innovation Cabinet Office, Government of Japan, Dec. 18, 2015. Accessed: Apr. 08, 2023. [Online]. Available: https://www8.cao.go.jp/cstp/kihonkeikaku/5basicplan_en.pdf

[5] K. Fukuda, "Science, technology and innovation ecosystem transformation toward society 5.0," *Int. J. Prod. Econ.*, vol. 220, p. 107460, Feb. 2020, doi: 10.1016/j.ijpe.2019.07.033.

[6] S. B. Lee and S. Hun Yoo, "Design of the companion robot interaction for supporting major tasks of the elderly," in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, Jeju: IEEE, Jun. 2017, pp. 655–659. doi: 10.1109/URAI.2017.7992695.

[7] K. Chen, V. W. Lou, K. C. Tan, M. Wai, and L. Chan, "Effects of a Humanoid Companion Robot on Dementia Symptoms and Caregiver Distress for Residents in Long-Term Care," *J. Am. Med. Dir. Assoc.*, vol. 21, no. 11, pp. 1724-1728.e3, Nov. 2020, doi: 10.1016/j.jamda.2020.05.036.

[8] H.-M. Gross, A. Scheidig, S. Muller, B. Schutz, C. Fricke, and S. Meyer, "Living with a Mobile Companion Robot in your Own Apartment - Final Implementation and Results of a 20-Weeks Field Study with 20 Seniors," in *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada: IEEE, May 2019, pp. 2253–2259. doi: 10.1109/ICRA.2019.8793693.

[9] H. Sayedi, "Intent Recognition in NLP." Accessed: Jun. 11, 2023. [Online]. Available: https://www.taus.net/resources/blog/intent-recognition-in-nlp

[10] "Introducing DIET: state-of-the-art architecture that outperforms fine-tuning BERT and is 6X faster to train," Rasa. Accessed: Apr. 27, 2023. [Online]. Available: https://rasa.com/blog/introducing-dual-intent-and-entity-transformer-diet-state-of-the-art-performance-on-a-lightweight-architecture/

[11] S. B. Barnes, "User Friendly: A Short History of the Graphical User Interface," 1996.

[12] A. Khan and S. Khusro, "Blind-friendly user interfaces – a pilot study on improving the accessibility of touchscreen interfaces," *Multimed. Tools Appl.*, vol. 78, no. 13, pp. 17495–17519, Jul. 2019, doi: 10.1007/s11042-018-7094-y.

[13] C. Qu, L. Yang, W. B. Croft, Y. Zhang, J. R. Trippas, and M. Qiu, "User Intent Prediction in Information-seeking Conversations," in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, Glasgow Scotland UK: ACM, Mar. 2019, pp. 25–33. doi: 10.1145/3295750.3298924.

[14] X. Liu, A. Eshghi, P. Swietojanski, and V. Rieser, "Benchmarking Natural Language Understanding Services for building Conversational Agents." arXiv, Mar. 26, 2019. Accessed: Jun. 07, 2023. [Online]. Available: http://arxiv.org/abs/1903.05566

[15] I. Casanueva, T. Temčinas, D. Gerz, M. Henderson, and I. Vulić, "Efficient Intent Detection with Dual Sentence Encoders." arXiv, Mar. 10, 2020. Accessed: Jun. 05, 2023. [Online]. Available: http://arxiv.org/abs/2003.04807

[16] D. Saffer, *Designing for Interaction: Creating Innovative Applications and Devices*. New Riders, 2010.

[17] L. M. Hasani, D. I. Sensuse, Kautsarina, and R. R. Suryono, "User-Centered Design of e-Learning User Interfaces: A Survey of the Practices," in *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*, Yogyakarta, Indonesia: IEEE, Sep. 2020, pp. 1–7. doi: 10.1109/IC2IE50715.2020.9274623.

[18] R. G. Bias and D. J. Mayhew, *Cost-Justifying Usability: An Update for the Internet Age*. Elsevier, 2005.

[19] J. Kowalski *et al.*, "Older Adults and Voice Interaction: A Pilot Study with Google Home," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk: ACM, May 2019, pp. 1–6. doi: 10.1145/3290607.3312973.

[20] T. Bunk, D. Varshneya, V. Vlasov, and A. Nichol, "DIET: Lightweight Language Understanding for Dialogue Systems." arXiv, May 11, 2020. Accessed: May 08, 2023. [Online]. Available: http://arxiv.org/abs/2004.09936

[21] A. Karajgi, "Building a Chatbot with Rasa," Medium. Accessed: Apr. 12, 2023. [Online]. Available: https://towardsdatascience.com/building-a-chatbot-with-rasa-3f03ecc5b324

[22] "React." Accessed: Apr. 18, 2023. [Online]. Available: https://react.dev/

[23] "Learn How To Build Web Apps With Speech Recognition." Accessed: Apr. 18, 2023. [Online]. Available: https://webspeechrecognition.com/

[24] "Tailwind CSS - Rapidly build modern websites without ever leaving your HTML." Accessed: Apr. 18, 2023. [Online]. Available: https://tailwindcss.com/

[25] "Web Speech API - Web APIs | MDN." Accessed: Apr. 18, 2023. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API

[26] "Rasa Architecture Overview." Accessed: May 08, 2023. [Online]. Available: https://rasa.com/docs/rasa/arch-overview/

[27] "Glimepiride Oral: Uses, Side Effects, Interactions, Pictures, Warnings & Dosing - WebMD." Accessed: Jun. 05, 2023. [Online]. Available: https://www.webmd.com/drugs/2/drug-12271/glimepiride-oral/details

[28] "Metformin Oral: Uses, Side Effects, Interactions, Pictures, Warnings & Dosing - WebMD." Accessed: Jun. 05, 2023. [Online]. Available: https://www.webmd.com/drugs/2/drug-11285-7061/metformin-oral/metformin-oral/details

[29] M. Muslih, M. A. Mun'im, and K. Mahbubah, "Mukena in Madurese Social Life," *Islam. J. Studi Keislam.*, vol. 15, no. 2, Art. no. 2, Mar. 2021, doi: 10.15642/islamica.2021.15.2.223-241.