A Systematic Literature Review: A Comparison Of Available Approaches In Chatbot And Dialogue Manager Development

Arya Jayavardhana[1], Samuel Ady Sanjaya[2*]

[1,2] Department of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang Selatan, Banten, Indonesia.
*Corresponding Author:
Email: samuel.ady@umn.ac.id

Abstract.

The present study reviewed a number of articles chosen from a screening and selecting process on the various different methods that can be used in the context of chatbot development and dialogue managers. Since chatbots have seen a significant rise in popularity and have played an important role in helping humans complete daily tasks, this systematic literature review (SLR) aims to act as a guidance for future research. During the process of analyzing and extracting data from the 13 articles chosen, it has been identified that Artificial Neural Network (ANN), Ensemble Learning, Recurrent Neural Network (RNN), and Long-Short Term Memory (LSTM) is among some of the most popular algorithms used for developing a chatbot. Where all of these algorithms are suitable for each unique use case where it offers different advantages when implemented. Other than that, dialogue managers lean more towards the field of Deep Reinforcement Learning (DRL), where Deep Q-Networks (DQN) and its variants such as Double Deep-Q Networks (DDQN) and DDQN with Personalized Experience Replay (DDQN-PER) is commonly used. All these variants have different averages on episodic reward and dialogue length, along with different training time needed which indicates the computational power needed. This SLR aims to identify the methods that can be used and identify the best proven method to be applied in future research.

Keywords: *ANN, chatbot, dialogue manager, DQN, LSTM and RNN.*

I. INTRODUCTION

In today's modern environment, chatbots and conversational agents play an important role in helping humans complete daily tasks. This technology has helped in various fields such as education. e-commerce, health, or political field [1]. Typically, chatbots or conversational agents are created to be efficient in dealing with users or customers. With the help of natural language technologies, these systems can understand and process human inputs in order to convert them into understandable formats for computers. Initial discussions with users are usually done to gain more context of the situation before choosing an appropriate response to give back as an output [2]. However, these chatbots and intelligent agents are only trained to provide accurate existing information. These agents are unable to generate or create new information to give to the users since they work by learning from past experiences [3]. Essentially, chatbots and conversational agents can also be defined as dialogue systems that have the proficiency to accomplish human-computer interactions. With these systems, human-computer interactions are achieved smoothly by efficient understanding of natural language content. Apart from that, the integration of artificial intelligence (AI), machine learning (ML), and deep learning (DL) techniques has helped the development of these systems, giving them the ability to deliver more precise and in-depth outputs [4]. Nonetheless, a major drawback that exists in this approach is the lack of ability to create conversational agents that can think, consider, and understand user emotions.

This however, can be solved along with the progression and development of AI technologies [3].Conversational agents are divided into three main types, which are text-based agents, embodied agents, and voice-based virtual agents. These mediums help computers interact with humans in various different ways making a more versatile platform [1]. Voice-based agents are the type of agents where the communication between humans and computers are done with voice inputs. These systems can capture sound given by the users, understand, analyze, and provide feedback in the form of a voice [5]. Embodied agents on the other hand, is the type of agent that has a graphical or tangible interface. These agents can be a combination of both voice-based and text-based agents, with an addition of physical existence [6]. Text-based agents are mediums where users interact with the system via textual interactions. Textual agents will begin the conversation by gathering relevant information that can help the system understand more about the context of the users. However, textual agents do have its own limitations of language. This is why most textual agents are built upon the English language considering its ease of use and its adoption as a global language. To achieve this, Natural Language Processing (NLP) plays an important part as textual agents are able to evaluate an input by checking for data from the knowledge base and compute a decision based on the input before sending it to the user [7].

In the case of chatbots, they are conversational agents that falls into the category of text-based agents. Chatbots are also divided into three main types which are rule-based chatbots, intellectually independent chatbot, and artificial intelligence powered chatbots [3]. Rule-based chatbots are also typically known as decision-tree bots [8]. This is because these systems usually are based on a specific series of well-defined rules. They will be trained based on these rules and once they're ready, these bots will be able to provide the specific output defined in the training process [9]. There are a lot of frameworks that provide the ability to create such chatbots, these frameworks include Google DialogueFlow and IBM Watson [10]. However, the main drawback of a rule-based chatbot is the lack of ability to understand user emotions [3]. Intellectually independent chatbots are created using ML and DL techniques. With the help of neural networks, these systems have the power to think and learn from histories giving them the ability to conduct self-learning processes. ML and DL techniques have helped these systems learn about various attributes and hence giving it the ability to provide a better solution for the context given. Lastly, artificial intelligence-powered chatbots can be defined as an extension of rule-based chatbots that have been given the power of AI. With the help of Natural Language Processing (NLP), can understand user input and provide instant and easy information for the users as output [3].

In the midst of all these systems, there's a module that is often forgotten. a dialogue manager (DM). In these system processes, a DM module can be interpreted as a brain for the whole operation. In the context of a goal-oriented system, a DM is used to achieve a goal in the shortest time possible. A DM module is divided into two main parts, an agent and a state tracker. An agent functions as the system which decides what policies lie on that particular system. This can be achieved by pairing a state with an action. The agent itself consists of an intent and slot where it is made up of structured data. Lastly, a state tracker is used to monitor the whole conversation and encode any relevant history that can be of use in the future [11].The diversity of chatbots and its categories has opened various ways of approach that can be done in the context of chatbots and dialogue managers. The uniqueness of each approach can present different challenges that have to be faced. This requires a comprehensive review on the different approaches that can be done along with its advantages. Therefore, this systematic literature review is done to help identify and analyze the different approaches available. Other than that, this review can also serve as a guidance or reference for future research to highlight the technologies and techniques that can be done on this topic.

II. METHODS

To be able to achieve a more comprehensive and reliable review, this study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) as a guidance to conduct this systematic literature review (SLR) on chatbots and the implementation of dialogue managers. The PRISMA framework has been followed since it is a commonly used approach for conducting a SLR [12]. In the context of this SLR, which is focused on the various approaches that can be done when developing a chatbot and its dialogue manager, PRISMA helps ensure a rigorous process by screening and selecting relevant articles. Other than that, the PRISMA framework also has a detailed reporting checklist which ensures a clean and transparent manner [13]. This research strategy is also selected to be used in consideration of recommendation from Kitchenham et al [14].

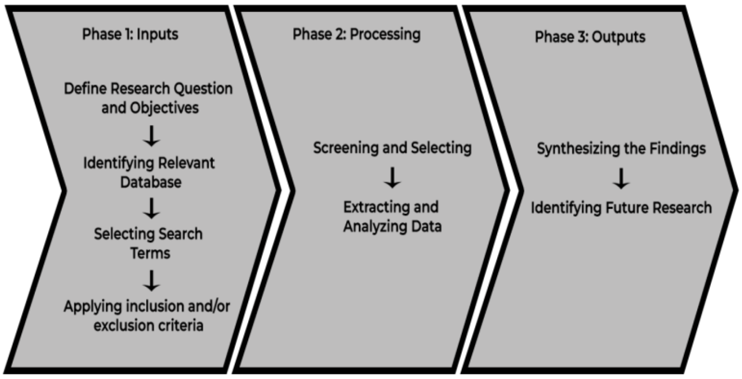

Fig 1. Research Roadmap

Following the research strategy, there will be eight sequential steps that is involved in doing this review, below is the explanation for each steps [13]:

a. Step 1 - Define research questions and objectives:

Step 1 will be defining the research questions and objectives that will help guide this review along the way. Below are the research questions and objectives for this SLR:

Research questions:

- What is the major approach that can be done when developing a chatbot and a dialogue manager?
- What is the best proven approach that can be done when developing a chatbot and a dialogue manager?
- What are the key challenges and limitations of existing dialogue manager implementations in chatbots, and how can they be addressed or improved?

Objectives:

- To provide a comprehensive review of the major approach that can be done when developing a chatbot and a dialogue manager.
- To specify the best proven approach that can be done when developing a chatbot and a dialogue manager.

b. Step 2 - Identifying relevant databases:

Once research questions and objectives are defined, step 2 will be identifying relevant databases that can be a source of data for this SLR. The databases that are chosen to be used for this SLR are Google Scholar and Scopus databases where it will be accessed through an application named Publish or Perish (POP).

c. Step 3 - Selecting search terms:

Through POP, the search term selected that can help gather relevant and updated literature are based on the selected research questions and objectives, which are “chatbot” AND “Deep learning” OR “Machine Learning”, “chatbot” AND “Dialogue Manager”, and “chatbot” AND “Dialogue Manager” AND “Deep Reinforcement Learning”.

d. Step 4 - Applying inclusion and/or exclusion criteria:

In this SLR, the inclusion criteria that is applied is the literature chosen to be used for this SLR were published between 2018 to 2023. Other than that, the chosen literature is only chosen if it were available in English or Bahasa Indonesia. On the other hand, the exclusion criteria is that any literature that does not use machine learning or deep learning techniques to develop a chatbot will be excluded.

e. Step 5 - Screening and selecting:

Next step would be screening and selecting the literature that has been collected. The literature chosen will be based on its relevance to the research questions and objectives defined in this SLR. Full-text articles will be reviewed and literature that does not have any relevancy towards the research questions and objectives will not be selected.

f. Step 6 - Extracting and analyzing data:

After the literature has been screened and selected, it would then be analyzed and any data relevant to the research questions and objectives of this SLR will be extracted. This includes things such as methods used, implementation process, and evaluation process.

g. Step 7 - Synthesizing the findings:

The extracted information from those literatures will be later presented in a narrative and table format in this SLR. This will give both an explanation and visual representation of the data gathered.

h. Step 8 - Identifying future research:

Lastly, based on the analysis done in this SLR, potential future research will be identified and presented as a conclusion for this paper.

III. RESULT AND DISCUSSION

By following the steps that have been defined previously, over 600 articles have been identified and through the screening and selecting process, where 13 articles have been chosen to be eligible and relevant in answering the research questions and objectives that have been defined previously. The summary of those 13 articles is shown in Table 1. From these 13 selected articles, we will conduct a comprehensive analysis of the algorithms employed, the case studies taken, their applications, advantages in comparison, and information regarding recent studies in this field. In the following section, we will provide a detailed summary of each of these articles. The table summarizes different algorithms used in chatbot development, highlighting their strengths. Artificial Neural Network (ANN) is versatile, suitable for various chatbots like tourism, quizzes, and eHealth. Recent research confirms its reliability in education.

Table. 1. Overall comparison between algorithms

No.	Algorithm	Suited use case	Commonly used in	Advantages	Recent study
[15], [16], [17], [18]	Artificial Neural Network (ANN)	Well suited for regression, classification, image processing, and character recognition.	Tourism chatbot, Quiz generation chatbot, and eHealth chatbot.	• Tolerant to missing input values or corrupted cells • Have the ability to self-organize to fit changes in information	A recent study in a chatbot development using ANN for higher-educational institutions has shown that ANN models are dependable with accurate results having an average RMSE value of 0.124 from 3 different ANN models.
[19], [20], [21]	Ensemble Learning	Well suited to help improve the performance of existing algorithms such as GRU, RNN, or LSTM.	Fashion chatbot, Diabetes treatment chatbot	• Can boost weak models • Can be well utilized by using growing computational power	A recent study shows that with meta-algorithms such as ensemble learning, it has potential to provide better generalization and map strong correlations via conversation in a humane sense.
[22], [23]	Recurrent Neural Network (RNN)	Well suited for sequence prediction and natural language generation, therefore is capable of text generation where prediction of values based on attributes can be done.	Emotion recognition chatbot, Technical support chatbot	• Fit for contextual input sequences • Works well for modeling temporal structure	A recent study shows that developing a chatbot using RNN can suit Seq2Seq problem statements better since it is built over a knowledge domain.

No.	Algorithm	Suited use case	Commonly used in	Advantages	Recent study
[24], [25], [26], [27]	Long-Short Term Memory (LSTM)	Well suited for time series forecasting, and can be used as a conversational agent,	Twitter chatbot, anime chatbot	• Handles long-term dependencies better • More memory capacity than RNN	A recent study on the variant of recurrent neural network model in context learning chatbot shows that LSTM is the best choice when the state of dialogue and state of conversation needs to be kept and tracked.

Ensemble Learning enhances existing algorithms for fashion and diabetes chatbots, offering better performance. It's adaptable with growing computing power. Recurrent Neural Network (RNN) is great for text generation, suitable for emotion recognition and tech support chatbots. Recent studies show its compatibility with Seq2Seq problems. Long-Short Term Memory (LSTM) is perfect for time series forecasting and chatbots like Twitter and anime. It handles long-term dependencies well. Recent research confirms its effectiveness in context learning chatbots. In summary, these algorithms offer specific advantages for different chatbot types. Researchers and developers can use this information to choose the best algorithm for their needs.In the context of dialogue management within the domain of Deep Reinforcement Learning (DRL), a comparative analysis was conducted to assess the performance of various DRL techniques, each paired with a Support Vector Machine (SVM) classifier. The Table 2 summarizes the findings, including the average dialogue length, the average episodic reward, and the training time in hours for different DRL algorithms. These algorithms play a vital role in shaping the efficiency and effectiveness of dialogue managers in conversational AI systems.

Table 2. Overall comparison between DRL techniques for dialogue manager

No.	Algorithm (paired with SVM)	Average dialogue length	Average episodic reward	Training time in hours
[28]	Deep Q-Networks (DQN)	673.45 ± 564.02	−6.89 ± 5.62	71.97
	Double Deep Q-Networks (DDQN)	791.65 ± 529.12	−8.51 ± 5.40	93.65
	DDQN with Prioritized Experience Replay (DDQN-PER)	1342.3 ± 915.20	−13.51 ± 9.15	52.56
	DDQN with Prioritized Experience Replay (DDQN-PER) (Stand-alone)	1039.07 ± 933.62	−11.26 ± 9.17	112.12

These algorithms exhibit notable differences in performance across several key metrics. For instance, Deep Q-Networks (DQN) demonstrate a moderate average dialogue length of 673.45, coupled with a negative average episodic reward of -6.89, and a training time of 71.97 hours. Meanwhile, the Double Deep Q-Networks (DDQN) show a slightly longer average dialogue length at 791.65, a more negative average episodic reward of -8.51, and a longer training time of 93.65 hours. Interestingly, the DDQN with Prioritized Experience Replay (DDQN-PER) emerges as a standout performer in terms of dialogue length, with an average of 1342.3, although it exhibits a more negative average episodic reward of -13.51. However, this superior dialogue length comes at the cost of a shorter training time, requiring only 52.56 hours. On the other hand, the standalone DDQN-PER exhibits an average dialogue length of 1039.07, a slightly better average episodic reward of -11.26, but a notably longer training time of 112.12 hours.These comparative results highlight the trade-offs between dialogue length, reward, and training time associated with these DRL techniques in dialogue management. Researchers and practitioners should consider these factors carefully when selecting an algorithm to suit the specific requirements of their conversational AI systems. Furthermore, these findings underscore the importance of a nuanced evaluation of DRL techniques for dialogue management to achieve optimal performance in real-world applications.The research mind map on

Fig 2 provides an overview of key elements in the field of Chatbot and Dialogue Manager Development, highlighting the associated algorithms, application areas, and evaluation metrics:

Machine Learning and Deep Learning serve as the foundational concepts for chatbot and dialogue manager development, with various algorithms and techniques contributing to their effectiveness. Artificial Neural Network (ANN), as seen in references [15], [16], [17], [18], plays a significant role in chatbot development. ANN is a versatile choice, adaptable to various applications. Recurrent Neural Network (RNN), as referenced in [22], [23], is particularly well-suited for text-based chatbots, excelling in natural language understanding. Ensemble Learning ([19], [20], [21]) and Long-Short Term Memory (LSTM) ([24],[25],[26],[27]) algorithms offer enhanced performance, making them valuable choices for chatbot development in specific domains. Dialogue Manager is the core component in chatbot development, overseeing conversations and interactions. It is applied in various domains, including Health ([15],[16],[17],[18]), Tourism ([16]), Education ([17],[18], [27]), Fashion ([21]), Contextual Chatbots like Twitter Bots ([24]), Agriculture ([25],), and Anime ([26]). Deep Reinforcement Learning introduces an additional layer of sophistication to chatbots, with algorithms like Deep Q-Network (DQN), Double Deep Q-Network (DDQN), and DDQN with Personalized Experience Replay (DDQN-PER) contributing to chatbot development in specific applications, such as Tourism, Education, Fashion, and others. Evaluation metrics are crucial for assessing chatbot performance. These include measures like MRR, RMSE, Average Dialogue Length, Accuracy, F1-Score, Average Episodic Reward, BLEU, and Recall. These metrics help gauge the effectiveness and quality of chatbot interactions. This mind map offers a comprehensive visual representation of the interconnected elements in chatbot and dialogue manager development, including the algorithms used, their applications, and the essential evaluation metrics employed to assess performance.

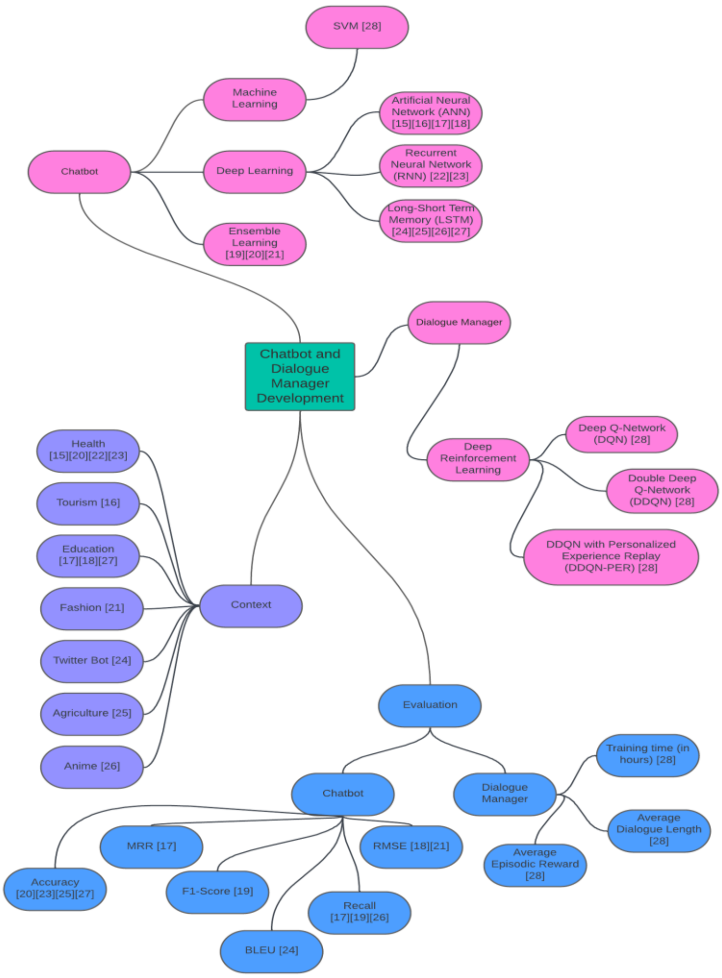

Fig 2. Research Mindmap

(RQ1) What is the major approach that can be done when developing a chatbot and a dialogue manager?

When developing a chatbot, there's a few approaches/algorithms that can be used and implemented. From table 1, it can be concluded that Artificial Neural Network (ANN), Ensemble Learning, Recurrent Neural Network (RNN), and Long-Short Term Memory (LSTM) are among some of the most popular algorithms to be used. Each of these algorithms have its own preferred use case along with advantages of its own. As can be seen in table 1, ANN is commonly known for its adaptability for tasks such as regression, classification, image processing, and character recognition. It can be found implemented in a few chatbot applications such as tourism chatbots, quiz generation chatbots, and eHealth chatbots [15], [17]. A key strength that ANN has is in its ability in handling missing input values or data corruption effectively [16]. A recent study even showed ANN model reliability in the context of highly educational institutions, showing promising, high, and accurate results with an average of 0.124 RMSE value from three distinct ANN models [18].Another notable approach that can be taken is Ensemble Learning, which is usually used to help enhance the performance of existing ML or DL algorithms such as GRU, RNN and LSTM [18].

This approach is usually found with unique purposes, such as fashion chatbots or diabetes treatment chatbot[20], [22]. Ensemble Learning can help fortify weaker models and as computational power continues to grow, its use case can become even more common [19]. A recent study has shown that the use of meta-algorithms such as ensemble learning can help chatbots achieve superior generalization ability and help map strong correlations better in a humane and conversational context [20]. Other than that, RNN is also another popular algorithm that is used in the development of chatbots. With its distinctive ability for sequence prediction and natural language generation, it excels in chatbot applications where text generation and attribute prediction is needed. This also makes RNN an optimal choice for chatbots including emotion recognition and technical support chatbots [22], [23]. RNN is particularly known to be suitable in handling contextual input sequences and modeling temporal structure [22]. A recent study has showcased the efficiency of RNN in addressing Sequence-2-Sequence (Seq2Seq) problems, especially when chatbots are developed with a specific knowledge domains [23]. Lastly, from an extended branch of RNN algorithms, the LSTM algorithms also offer a unique set of capabilities in a chatbot development context. LSTM has proved itself to be a strong algorithm to use when handling time-series forecasting problems and as a conversational agent, such as twitter and anime chatbot [24], [26].

LSTM excels when it comes to handling long-term dependencies and can hold more memory capacity than the traditional RNN models [25]. In a recent study, it has been showcased that LSTM can be particularly useful when the preservation and tracking of dialogue and conversation states are needed [27]. On the other hand, algorithms used for dialogue managers have focused more in the realm on Deep Reinforcement Learning (DRL). First, the standard Deep Q-Networks (DQN) has reported an average of 673.45 dialogue length with an average episodic reward of -6.89. It should be also noted that the training time for DQN is the second lowest among the variants of DQN which is around 71.97 hours. Next, Double Deep Q-Networks (DDQN) has seemed to outperform DQN slightly with an average dialogue length of 791.65 and a slightly improved average episodic reward of -8.51. However, it should be noted that the training time for DDQN is significantly higher at 93.65 hours which indicates the need for additional computational resources to achieve its performance. There is also a variant called DDQN with Prioritized Experience Replay (DDQN-PER) where it shows significant improvement with an average dialogue length of 1342.3, -13.51 average episodic reward, and only 52.56 hours of training time. Additionally, there is also a test where DDQN-PER is not paired with any algorithms and it has maintained a good balance between performance and training time [28].

(RQ2) What is the best proven approach that can be done when developing a chatbot and a dialogue manager?

In the context of chatbot development, the most effective approach will depend on the precise use case and objectives given. However, among all the popular algorithms used, Long-Short Term Memory (LSTM) still emerges as the best proven approach that can be used. LSTM exceptional capability for handling time-series and being conversational agents [24], [26]. Its ability to handle long-term dependencies and having a larger memory capacity has distinguished it from other approaches [24]. This is also further proven when the tracking and preserving of dialogue and conversation states is needed [27]. Therefore making LSTM stand as a reliable choice, offering a more versatile solution across various different conversational needs.As for dialogue managers, the development of this module leans more towards the landscape of Deep Reinforcement Learning (DRL). DDQN with Prioritized Experience Replay (DDQN-PER) has shown the best performance as it demonstrates impressive advancements in dialogue lengths and average episodic reward. It also has the best and most efficient training time making it a more computational friendly choice. Standard Deep Q-Networks (DQN) also has showcased reasonably good performance with an average dialogue length of 673.45 and average episodic reward of -6.89. The training time needed for this variant is also relatively efficient clocking in only around 71.97 hours [28]. Nonetheless, the preferred variant can depend on the project scope and problems statements given. Standard DQN can still persist as a more practical and optimal choice for more compact projects, even when DDQN-PER may appear better on paper.

(RQ3) What are the key challenges and limitations of existing dialogue manager implementations in chatbots, and how can they be addressed or improved?

One significant challenge and limitation that exists in existing dialogue manager implementation for chatbots revolves around the complexity and inefficiency when handling a large state and action set. When these state and action pairs grow too big, it can lead to inefficient computation and slow down the dialogue management in the process [29]. This can surely affect the ability of the chatbot to make quick and informed decisions to give to the users, where this may affect user experience. To help address this issue, implementing Deep Reinforcement Learning (DRL) techniques as opposed to using regular Reinforcement Learning (RL) can help strengthen the ability of the dialogue manager to learn more optimal policies and navigate through big state-action pairs more effectively making user experience more seamless [11].DRL itself is the combination of RL with Deep Neural Network (DNN), where DNN will act as an approximation function to help the system calculate the optimal Q-Value. This helps the dialogue manager generalize all state-action pairs, where this obviously benefits the agent since by having the ability to generalize all state-action pairs, an agent may accurately predict the Q-Value of a previously unseen state-action pairs more effectively. Other than that, an improvement has been made where an agent that is using DQN can now calculate the Q-Value for all state-action pairs in a single pass, in contrast to the previous method where neural networks have to make multiple passes in order to calculate all Q-Values for all available state-action pairs [30].

(RQ4) How do machine learning and NLP techniques influence the effectiveness of a chatbot and how can these techniques be optimized for different use cases?

For starters, the effectiveness of a chatbot is hugely influenced by machine learning and NLP techniques. Obviously, chatbots with AI powered systems will have better capabilities that regular rule-based chatbot [3]. The effectiveness of the machine learning techniques for chatbots of different use cases may vary depending on the algorithms that excel in that specific use case. For example, Artificial Neural Networks (ANN) have the versatility to do tasks such as classification, image processing, and character recognition which makes it better for chatbots in the field of tourism or quiz generation [15], [17].

Meanwhile, Recurrent Neural Network (RNN) and Long-Short Term Memory (LSTM) demonstrate a better capability in handling sequential prediction with better natural language generation which makes it better in text generation and as a conversational agent where there may be a more contextually rich conversation [22],[23],[24],[26]. To be able to optimize these techniques for different use cases, it is imperative that the algorithms are tailored to that specific chatbot functionality. Machine learning models can be fine-tuned to be able to adapt to various different fields and can achieve a high accuracy [31]. Other than that, using ensemble learning techniques can help make a more robust model that is content aware in a dialogue management system. By combining models via ensemble learning, the chatbot developed can have more generalization over a humane conversation and help map strong correlation efficiently [20]. These understanding can help optimize and harness the true potential of each algorithm to enhance the performance and effectiveness of algorithms for different use cases.

IV. CONCLUSION

To conclude everything, the review done in this SLR has shown some of the most popular approaches when it comes to developing a chatbot and its dialogue manager. Based on the 17 articles chosen to be analyzed in this review, ANN, Ensemble Learning, RNN, and LSTM are among some of the most popular algorithms used. Among these algorithms, LSTM still excels in being the most optimal choice as it is a more versatile option. As for dialogue managers, DDQN-PER has shown to have the best performance on paper, while standard DQN still showed promising results for more compact projects. In the future, a combination of DQN with LSTM as a dialogue system can be tested to see and evaluate its performance as a conversational agent.

V. ACKNOWLEDGMENTS

This research received support from the Institution of Research and Community Services at Universitas Multimedia Nusantara. We extend our appreciation to our colleagues at the Big Data Laboratory within the Information Systems Department at Universitas Multimedia Nusantara, whose valuable input and expertise greatly enriched the research.

REFERENCES

[1] M. Allouch, A. Azaria, and R. Azoulay, "Conversational Agents: Goals, Technologies, Vision and Challenges," *Sensors*, vol. 21, no. 24, p. 8448, Dec. 2021, doi: 10.3390/s21248448.

[2] B. Mott, J. Lester, and K. Branting, "Conversational Agents," 2004. doi: 10.1201/9780203507223.ch10.

[3] A. Joseph and N. G. Eapen, "Conversational Agents and Chatbots," in *Machine Learning and Deep Learning in Natural Language Processing*, Boca Raton: CRC Press, 2023, pp. 17–52. doi: 10.1201/9781003296126-4.

[4] H. N. Io and C. B. Lee, "Chatbots and conversational agents: A bibliometric analysis," in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Dec. 2017, pp. 215–219. doi: 10.1109/IEEM.2017.8289883.

[5] A. S. Miner, A. Milstein, S. Schueller, R. Hegde, C. Mangurian, and E. Linos, "Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health," ***JAMA Intern. Med.,*** vol. 176, no. 5, p. 619, May 2016, doi: 10.1001/jamainternmed.2016.0400.

[6] J. Cassell, "Embodied conversational agents: Representation and intelligence in user interfaces," *AI Mag.*, vol. 22, no. 4, pp. 67–83, 2001.

[7] D.W.Massaro, M. M. Cohen, S. Daniel, and R. A. Cole, "Developing and Evaluating conversational Agents," in *Human Performance and Ergonomics*, Elsevier, 1999, pp. 173–194. doi: 10.1016/B978-012322735-5/50008-7.

[8] B. Liu and C. Mei, "Lifelong Knowledge Learning in Rule-based Dialogue Systems," Nov. 2020, doi: 10.48550/arXiv.2011.09811.

[9] J. Singh, M. H. Joesph, and K. B. A. Jabbar, "Rule-based chabot for student enquiries," ***J. Phys. Conf. Ser***., vol. 1228, no. 1, p. 012060, May 2019, doi: 10.1088/1742-6596/1228/1/012060.

[10] S. A. Thorat and V. Jadhav, "A Review on Implementation Issues of Rule-based Chatbot Systems," ***SSRN Electron. J.***, 2020, doi: 10.2139/ssrn.3567047.

[11] B. E. Nishimoto and A. H. Reali Costa, "Dialogue Management with Deep Reinforcement Learning: Balancing Exploration and Exploitation," in *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, Oct. 2019, pp. 449–454. doi: 10.1109/BRACIS.2019.00085.

[12] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *Syst. Rev.*, vol. 10, no. 1, p. 89, Dec. 2021, doi: 10.1186/s13643-021-01626-4.

[13] J. Yang, Y.-L. Chen, L. Y. Por, and C. S. Ku, "A Systematic Literature Review of Information Security in Chatbots," *Appl. Sci.*, vol. 13, no. 11, p. 6355, May 2023, doi: 10.3390/app13116355.

[14] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Inf. Softw. Technol.*, vol. 55, no. 12, pp. 2049–2075, Dec. 2013, doi: 10.1016/j.infsof.2013.07.010.

[15] S. Ayanouz, B. A. Abdelhakim, and M. Benhmed, "A Smart Chatbot Architecture based NLP and Machine Learning for Health Care Assistance," in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, Mar. 2020, pp. 1–6. doi: 10.1145/3386723.3387897.

[16] S. Dey and D. Shukla, "Analytical study on use of AI techniques in tourism sector for smarter customer experience management," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Mar. 2020, pp. 1–5. doi: 10.1109/ICCSEA49143.2020.9132925.

[17] A. S. Sreelakshmi, S. B. Abhinaya, A. Nair, and S. Jaya Nirmala, "A Question Answering and Quiz Generation Chatbot for Education," in *2019 Grace Hopper Celebration India (GHCI)*, Nov. 2019, pp. 1–6. doi: 10.1109/GHCI47972.2019.9071832.

[18] N. I. Mohd Rahim, N. A. Iahad, A. F. Yusof, and M. A. Al-Sharafi, "AI-Based Chatbots Adoption Model for Higher-Education Institutions: A Hybrid PLS-SEM-Neural Network Modelling Approach," *Sustainability*, vol. 14, no. 19, p. 12726, Oct. 2022, doi: 10.3390/su141912726.

[19] H. Cuayáhuitl *et al.*, "Ensemble-based deep reinforcement learning for chatbots," *Neurocomputing*, vol. 366, pp. 118–130, Nov. 2019, doi: 10.1016/j.neucom.2019.08.007.

[20] M. Bali, S. Mohanty, D. S. Chatterjee, M. Sarma, and R. Puravankara, "Diabot: A Predictive Medical Chatbot using Ensemble Learning," ***Int. J. Recent Technol. Eng.***, vol. 8, no. 2, pp. 6334–6340, Jul. 2019, doi: 10.35940/ijrte.B2196.078219.

[21] Sapna *et al.*, "Recommendence and Fashionsence," in *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, Jan. 2019, pp. 256–259. doi: 10.1145/3297001.3297035.

[22] F. Patel, R. Thakore, I. Nandwani, and S. K. Bharti, "Combating Depression in Students using an Intelligent ChatBot: A Cognitive Behavioral Therapy," in *2019 IEEE 16th India Council International Conference (INDICON)*, Dec. 2019, pp. 1–4. doi: 10.1109/INDICON47234.2019.9030346.

[23] M.-C. Lee, S.-Y. Chiang, S.-C. Yeh, and T.-F. Wen, "Study on emotion recognition and companion Chatbot using deep neural network," *Multimed. Tools Appl.*, vol. 79, no. 27–28, pp. 19629–19657, Jul. 2020, doi: 10.1007/s11042-020-08841-6.

[24] Y. M. Çetinkaya, İ. H. Toroslu, and H. Davulcu, "Developing a Twitter bot that can join a discussion using state-of-the-art architectures," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, p. 51, Dec. 2020, doi: 10.1007/s13278-020-00665-4.

[25] B. Arora, D. S. Chaudhary, M. Satsangi, M. Yadav, L. Singh, and P. S. Sudhish, "Agribot: A Natural Language Generative Neural Networks Engine for Agricultural Applications," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, Feb. 2020, pp. 28–33. doi: 10.1109/IC3A48958.2020.233263.

[26] Y.-T. Wan, C.-C. Chiu, K.-W. Liang, and P.-C. Chang, "Midoriko Chatbot: LSTM-Based Emotional 3D Avatar," in *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, Oct. 2019, pp. 937–940. doi: 10.1109/GCCE46687.2019.9015303.

[27] K. Pathak and A. Arya, "A Metaphorical Study Of Variants Of Recurrent Neural Network Models For A Context Learning Chatbot," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, Nov. 2019, pp. 768–772. doi: 10.1109/ISCON47742.2019.9036167.

[28] T. Saha, D. Gupta, S. Saha, and P. Bhattacharyya, "Reinforcement Learning Based Dialogue Management Strategy," 2018, pp. 359–372. doi: 10.1007/978-3-030-04182-3_32.

[29] G. Xu, H. Lee, M.-W. Koo, and J. Seo, "Optimizing Policy via Deep Reinforcement Learning for Dialogue Management," in *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jan. 2018, pp. 582–589. doi: 10.1109/BigComp.2018.00101.

[30] A. Tsantekidis, N. Passalis, and A. Tefas, "Deep reinforcement learning," in *Deep Learning for Robot Perception and Cognition*, Elsevier, 2022, pp. 117–129. doi: 10.1016/B978-0-32-385787-1.00011-7.

[31] P. A. Olujimi and A. Ade-Ibijola, "NLP techniques for automating responses to customer queries: a systematic review," *Discov. Artif. Intell.*, vol. 3, no. 1, p. 20, May 2023, doi: 10.1007/s44163-023-00065-5.