CLAMAC: Design of a Continuous Learning Aspect-Analysis Model for development of Multidomain Adaptive Chat Companions

Rajiv Khobragade[1]Pranay Deepak Saraf[2]Supriya Thakur[3]

[1] Research Scholar, Department of Computer Science & Engineering
G. H. Raisoni College of Engineering,Nagpur, India
rajiv.madhav59@gmail.com
[2]Assistant Professor,Department of Computer Science & Engineering
G. H. Raisoni College of Engineering,Nagpur, India
pranay.saraf@raisoni.net
3Supervisor,Corporate Segments Team,
Willis Tower Watson, Mumbai, India
supriya.thakur@willistowerswatson.com

Abstract: Chatbots (or robo advisors) assist users to quickly solve trivial issues with minimum support staff. Chatbots are often designed via use of rule-based models, which allows them to identify solutions to fixed question sets. But such bots are static in nature, and their use often affects customer retention when used under real-time scenarios. To overcome this issue, dynamic Machine Learning (ML) based bots were designed, which use a large training corpus with multiple domains for assisting users. But these models have limited aspect-aware capabilities, due to which a single classification error reduces the usability of entire chat sessions. Moreover, these models train the chatbot on multiple domains, which reduces its accuracy & increases delay needed to classify user inputs. To overcome these limitations, this text proposes design of a novel continuous learning aspect-analysis model for development of multidomain adaptive chat companions. The proposed model initially uses an aspect-based classification layer, that assists in segregating user input into multiple aspect types. Based on this segregation, multiple domain-specific deep learning models are trained, which assist in improving usability of the chat sessions. The aspect-based model is built using Genetic Algorithm (GA) based adaptive clustering model, while a MobileVNet2 Model is used to train multiple deep learning methods that assist in user query classifications. Due to use of GA based clustering, the model is capable ofidentifying aspects with an accuracy of 98.5%, which allows the Convolutional Neural Network (CNN) to effectively select the correct training model for solving user queries. The model also uses an incremental learning layer, that assists in continuously improving its performance w.r.t. intelligent user feedback analysis. The proposed model was deployed on a wide variety of domains including Medical Assistance, Hospitality Management, E-Commerce Websites, and Online Food Ordering websites. Upon evaluation of the proposed model on these applications, an accuracy of 97.6% was observed, along with a precision of 95.4%, recall of 96.5%, and delay of less than 1 second for different input queries. Due to such a high-performance, the proposed model is capable of deployment for a wide variety of chat companion scenarios.

Keywords: Chatbot, Machine, Learning, Genetic, Algorithm, MobileVNet2, Aspect, CNN, Clustering, Model

1. Introduction

Chatbot design is a task of multiple domain that entails modelling of corpus collection techniques, pre-processing, query evaluation, text matching, sentiment analysis, clustering, classification, and post processing models. Design of a typical chatbot model [1] is depicted in figure 1, wherein user input is processed via entity evaluation, intent classification, dialog-based modelling, multiple social & professional channels, etc. to estimate query-based responses.

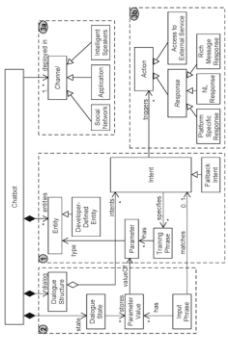

Figure 1. Design of typical chatbot model

These responses are aggregated via an action layer, that assists in selecting input-specific responses, which are presented to the users. The model uses a combination of multiple deep learning models, that work in collaboration to improve chatbot's real-time performance. Efficiency of these models largely depends upon quality of training corpus, due to which these chatbots require large multidomain training datasets. Similar models [2, 3] along with their nuances, advantages, limitations, and future research scopes are discussed in the next section of this text. Based on this discussion it can be noticed that the existing models are either static or have limited aspect-aware capabilities, due to which a single classification error reduces the usability of entire chat sessions. To overcome these issues, section 3 proposes design of a novel continuous learning aspect-analysis model for development of multidomain adaptive chat companions. The proposed CLAMAC model uses a combination of Genetic Algorithm with dynamic clustering for aspect-based analysis, and MobileVNet2 based CNN for identification of input-based responses. The proposed model was tested under different applications, and its efficiency was evaluated in section 4, where it was compared in terms of accuracy, precision, recall, and response time metrics. Finally, this text concludes with some interesting observations about the proposed model and recommends methods to further improve its performance.

2. Literature review

Chatbot models are increasing rapidly due to availability of simplified interfaces. For instance, work in [4, 5, 6] propose use of QANet (Question Answering Network) model, retrieval-polished (RP) model, and trust-based response model, which assists in estimation of context-specific features to answer user queries. But these models work in single domain environments, and cannot be scaled due to their internal design characteristics. To design a scalable chatbot, work in [7] proposes combination of Modelling Language for Artificial Intelligence, Sentiment Analysis&automatic Natural Language Generation, which assists in improving multidomain answering capabilities via effective feature augmentation analysis. This model is useful for Sports, Entertainment, E-Commerce, and other real-time applications. Extensions to this model are discussed in [8, 9, 10], which propose use of Business Transformation chatbots, knowledge base bots (KBot), and bioinspired optimization bots, that assist in improving scalability via effective feature augmentation process. Performance of these models is further improved via use of Machine Learning Bots (MLBs) [11], cognitive behavior therapy (CBT) based bots [12], manually tuned bots [13], ad Utterance-to-Utterance Interactive Matching Network (UIMN) bots [14], that assist in extensive feature analysis. These bots can be trained for multiple domains, but are highly complex to deploy due to their inherent corpus-based large-feature set extraction characteristics. To simplify this process, work in [15, 16] propose use of Seq2Seq (S2S), and Modular Cognitive Agent (MCA) based chatbots, which assist in reducing feature complexity via use of variance-based feature selection processes. These models are further extended via use of Robustly optimized Bidirectional Encoder Representations from Transformers (RBERT) [17], and Natural Answer Generation (NAG) models [18], which assist in aspect-based analysis. But these models use single CNNs for multidomain learning, which limits their real-time performance under linked user queries. To get over these limitations, the succeeding sectionsuggests a design of a continuouslylearning aspect-analysis Model for development of Multidomain adaptive chat companions. The proposed model was tested on a wide variety of datasets and compared with various state-of-the-art methods for showcasing its utility under different scenarios.

3. Design of Continuous Learning Aspect-Analysis Model for development of Multidomain Adaptive Chat Companions

In contingent to the literature survey, it was perceived that the existing chatbot models have limited aspect-aware capabilities, due to which a single classification error reduces the usability of entire chat sessions. These models work on single classifier-based multidomain training process, which limits their accuracy when multidomain queries are asked during user sessions. To integrate multidomain queries while maintaining high efficiency, this section proposes design of a novel continuous learning aspect-analysis model for development of adaptive chat companions. Design of the suggested model is depicted in figure 2, wherein it can be noticed that an aspect-based classification layer for segregation of user input into multiple aspect types is connected with multiple, domain-specific deep learning models. This combination assists in improving usability of the chat sessions due to high accuracy for real-time user queries. The aspect-based model is built using Genetic Algorithm (GA) based adaptive clustering model, while a MobileVNet2 Model is used to train multiple deep learning methods that assist in user query classifications. The GA Model uses Natural Language Processing (NLP) based operations to identify different aspects via Bidirectional Encoder Representations from Transformers (BERT) process.

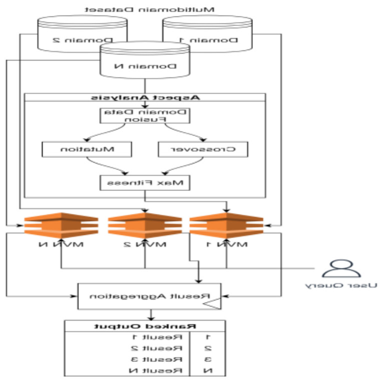

Figure 2. Complete flowchart of the proposed model. The model works via the following process,

- Initialize GA Parameters,
  - Number of iterations (N_i)
  - Number of solutions (N_s)
  - Learning rate (L_r)
  - Maximum number of BERT Embeddings ($E(Max)$)
- Initially mark all solutions as ‘to be mutated’
- For each iteration in 1 to N_i, perform the following tasks,
  - For each solution in 1 to N_s, perform the following tasks,
    - If the solution is marked as ‘not to be mutated’, then go to the next solution for checking process.
    - Else, generate a new solution via the following process,
      - Stochastically estimate maximum number of BERT Embeddings via equation 1 as follows,

$N(Emb) = STOCH\left(L_r * E(Max), E(Max)\right) \dots (1)$Where, $STOCH$ represents a Markovian stochastic process used to generate a number between the given range of inputs.

- Process training set input queries for each set of BERT embeddings, and estimate accuracy of aspect evaluation via equation 2, $Acc = \frac{N_{CA}}{N_T} \dots (2)$

Where, N_{CA}&N_T represents correctly classified aspects and total aspects used for classification by the BERT process.

- Estimate solution fitness via equation 3,

$f = \frac{\sum_{i=1}^{N(Emb)} Acc_i}{N(Emb)} \dots (3)$At the end of each iteration, estimate iteration fitness threshold via equation 4 as follows, $f_{th} = \sum_{i=1}^{N_s} f_i * \frac{L_r}{N_s} \dots (4)$

- At the end of each iteration, mark the solutions as ‘to be mutated’, where $f_i \leq f_{th}$, else mark others as ‘not to be mutated’
- Upon completion of all iterations, select BERT configuration with maximum fitness levels Due to selection of highest accuracy BERT Model, the GA Method is capable of identifying input queries with high accuracy, which assists in selection of proper MobileVNet2 Models. Depending upon number of domains used for analysis, individual corpuses are collected and a particular MobileVNet2 Model is selected for answering queries. Internal layers for the used MobileVNet2 Model are depicted in figure 3, wherein Dilated convolutions, Depth wise convolutions, Pointwise convolutions, Depth wise separable convolutions, and Global average pooling operations can be observed. Here each input query is processed via Word2Vec feature estimation which is trained for 150528 different corpus words. These features are converted into a size of 224x224x3, which assists in effective convolution-based feature representation for classification process. The convolutional features (CF) are extracted via equation 5, wherein leaky Rectilinear Unit (LReLU) is used for dynamic feature sizing for different input representations.

$$CF = Leaky_{ReLU}\left(\frac{r}{2}+a,\frac{c}{2}+b\right) * \sum_{a=0}^{\frac{r}{2}}\sum_{b=0}^{\frac{c}{2}} W2V(i-a,j-b) \ldots (5)$$

Where, r, c represents row & column size for dilated convolutions, a, b represents padding sizes, while $W2V$ represents Word2Vec features for the given input texts.

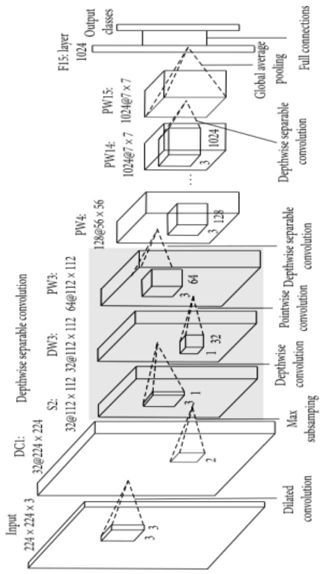

Figure 3. Design of the used MobileVNet2 Model for query response processThese features are activated via a Leaky ReLU Model, which is represented via equation 6 as follows,

$$LReLU(x,y) = 0.1 * CF, when\ Var(CF) \geq 0$$

$else, 1, when\ Var(CF) \leq 0$ ... (6)The leaky RELU activation layer used in this case removes 10% of all low variance features, which assists in feature selection process. This process is continued for multiple layers, which assists in identification of large input set features. The extracted features are classified using a Fully Connected Neural Network (FCNN), which uses Soft Max based activation layer, and is represented via equation 7 as follows,

$R(Out) = SoftMax\left(\sum_{i=1}^{N_f} b + CF_i * w_i\right)$ ... (7)Where, w & b represents weights and bias values for different input feature sets, and are tuned via hyperparameter tuning process, while $R(Out)$ represents output response for given input query, which is selected from the pre-set answer classes stored in respective datasets. Based on this process, multiple models are trained on different datasets, and their efficiency is evaluated in terms of accuracy, precision, recall, and response delay metrics. These metrics are evaluated for different datasets, and their performance is discussed in the next section of this text.

3. Results & performance evaluation

To evaluate performance of the proposed method, corpuses from standard datasets were used for training different MobileVNet2 models. These datasets were collected from the following sources,Microsoft Research WikiQA Corpus, available at https://www.microsoft.com/en-us/download/details.aspx?id=52419

- Question-Answer Dataset from Wikipedia, available at http://www.cs.cmu.edu/~ark/QA-data/
- Question Answering Collections from TREC, available at https://trec.nist.gov/data/qa.html
- Yahoo! Language Dataset, available at https://webscope.sandbox.yahoo.com/catalog.php?datatype=l
- ConvAI3 Dataset, available at https://convai.io/data/

A collection of 1.5 million records were used to train 5 different MobileVNet2 Models for Medical Assistance, Hospitality Management, E-Commerce Websites, and Online Food Ordering applications. The dataset was divided in a ratio of 70:30, wherein 70% entries were used for training, while remaining 30% were used for testing different performance metrics. These metrics include Accuracy (A), Precision (P), Recall (R), & Response Time (RT) of the proposed model, and was compared with QA Net [4], UI MN [14], and RB ERT [17] methods. Based on these configurations, accuracy of chat responsew.r.t. Number of Evaluations (NE) is tabulated in table 1 as follows,

NE	**A (%)**	**A (%)**	**A (%)**	**A (%)**

	QA Net [4]	UI MN [14]	RB ERT [17]	CLA MAC
2.93k	73.57	65.87	84.33	96.95
5.85k	73.98	66.24	84.81	97.37
9k	74.22	66.44	85.08	97.67
11.93k	74.44	66.65	85.34	97.89
14.85k	74.56	66.75	85.47	98.00
18k	74.60	66.78	85.51	98.03
20.93k	74.61	66.79	85.52	98.04
23.85k	74.61	66.80	85.53	98.04
27k	74.61	66.80	85.53	98.05
29.25k	74.62	66.80	85.54	98.05
33.75k	74.62	66.80	85.54	98.05
36k	74.62	66.80	85.54	98.06
38.93k	74.63	66.81	85.55	98.06
41.85k	74.63	66.81	85.55	98.06
45k	74.78	66.95	85.73	98.22

Table 1. Accuracy of different chatbot models

From these results and figure 4,it can be perceived that the suggested model is 23.8% better than QA Net [4], 31.5% better than UI MN [14], and 10.5% better than RB ERT [17] in terms of multiple domain accuracy performance.

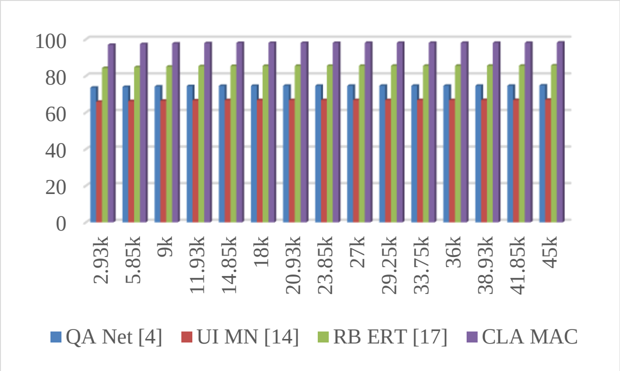

Figure 4. Accuracy of different chatbot models

This is due to use of MobileVNet2 with GA based BERT Modelling, which combines different high-performance classification &aspect-based selection to achieve better performance with lower error rates when compared to standard models. Similar observations are made for precision (P) values, and can be observed from the following table 2,

NE	P (%) QA Net [4]	P (%) UI MN [14]	P (%) RB ERT [17]	P (%) CLA MAC
2.93k	80.79	71.35	84.42	85.68
5.85k	81.24	71.75	84.90	86.05
9k	81.49	71.97	85.17	86.31
11.93k	81.75	72.20	85.43	86.51
14.85k	81.88	72.31	85.57	86.60
18k	81.92	72.35	85.61	86.63
20.93k	81.92	72.36	85.61	86.64
23.85k	81.93	72.36	85.62	86.65
27k	81.93	72.36	85.62	86.65
29.25k	81.93	72.37	85.62	86.65
33.75k	81.93	72.38	85.62	86.65
36k	81.93	72.38	85.62	86.66
38.93k	81.94	72.38	85.63	86.66
41.85k	81.95	72.38	85.64	86.75
45k	82.12	72.53	85.81	86.91

Table 2. Average precision values for different chatbot models

From these results and figure 5,it can be perceived that the suggested model is 3.5% better than QA Net [4], 14.5% better than UI MN [14], and 1.4% better than RB ERT [17] in terms of multiple domain precision performance. This is due to use of MobileVNet2 with GA based BERT Modelling, which combines different high-performance classification &aspect-based selection to achieve better performance with lower error rates when compared to standard models.

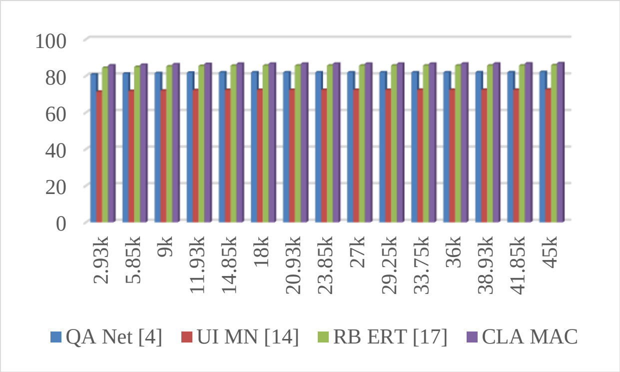

Figure 5. Average precision values for different chatbot models

Similar observations are made for recall (R) values, and can be observed from the following table 3,

NE	R (%) QA Net [4]	R (%) UI MN [14]	R (%) RB ERT [17]	R (%) CLA MAC
2.93k	79.78	70.47	83.38	91.12
5.85k	80.23	70.87	83.85	91.52
9k	80.48	71.08	84.10	91.80
11.93k	80.72	71.30	84.36	92.01
14.85k	80.84	71.41	84.49	92.10
18k	80.88	71.45	84.54	92.13
20.93k	80.89	71.46	84.54	92.14
23.85k	80.90	71.46	84.55	92.15
27k	80.90	71.46	84.56	92.15
29.25k	80.91	71.47	84.56	92.15
33.75k	80.91	71.48	84.56	92.15
36k	80.91	71.48	84.56	92.16
38.93k	80.92	71.48	84.57	92.16
41.85k	80.93	71.48	84.57	92.26
45k	81.09	71.63	84.74	92.44

Table 3. Average recall values for different chatbot models

From these results and figure 6,it can be perceived that the suggested modelis 10.5% better than QA Net [4], 19.5% better than UI MN [14], and 8.3% better than RB ERT [17] in terms of multiple domain recall performance.

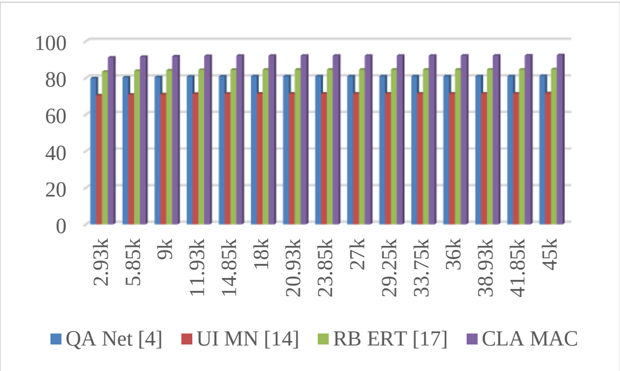

Figure 6. Average recall values for different chatbot models

This is due to use of MobileVNet2 with GA based BERT Modelling, which combines different high-performance classification &aspect-based selection to achieve better performance with lower error rates when compared to standard models. Similar observations are made for delay (D) values, and can be observed from the following table 4,

NE	D (ms) QA Net [4]	D (ms) UI MN [14]	D (ms) RB ERT [17]	D (ms) CLA MAC
2.93k	37.38	42.32	35.76	27.20
5.85k	37.16	42.08	35.56	27.06
9k	37.04	41.96	35.44	26.98
11.93k	36.92	41.82	35.34	26.88
14.85k	36.88	41.74	35.30	26.86
18k	36.88	41.72	35.28	27.30
20.93k	36.88	41.72	35.28	27.60
23.85k	36.88	41.72	35.28	28.00

27k	36.88	41.72	35.28	28.56
29.25k	36.88	41.72	35.28	28.52
33.75k	36.88	41.72	35.28	28.72
36k	36.88	41.72	35.28	29.04
38.93k	36.88	41.72	35.28	29.10
41.85k	36.88	41.72	35.28	29.14
45k	36.81	41.63	35.21	29.31

Table 4. Average delay (D) values for different chatbot models

In contingent to this assessment and figure 7, it is noticed that the proposed model is 15.5% faster than QA Net [4], 20.5% faster than UI MN [14], and 14.5% faster than RB ERT [17], under different types of user queries.

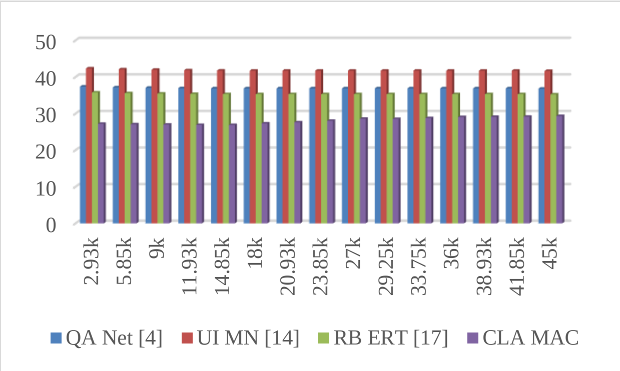

Figure 7. Average delay (D) values for different chatbot models

Due to which, the proposed model is capable of being deployed for a wide variety of real-time high-speed applications. This is due to use of MobileVNet2 with GA based BERT Modelling, which combines different high-performance classification &aspect-based selection to achieve better performance with lower error rates when compared to standard models. Due to this high-performance, the proposed model is useful for a wide variety of real-time chat bot design deployments under multidomain application scenarios.

4. Conclusion and future work

The proposed model initially uses a GA based BERT training method, which is capable of effective aspect classification under multiple datasets. The BERT model is followed by multiple MobileVNet2 based CNN blocks, which assist in deploying a domain specific query resolution process. The model is trained on multiple datasets, and it was observed that the proposed model is 23.8% better than QA Net, 31.5% better than UI MN, and 10.5% better than RBERT in terms of multiple domain accuracy, it was also observed that the proposed model is 3.5% better than QA Net, 14.5% better than UI MN, and 1.4% better than RBERT in terms of precision, while the proposed model is 10.5% better than QA Net, 19.5% better than UI MN, and 8.3% better than RBERT in terms of multiple domain recall performance. Apart from this, the proposed model was observed to be 15.5% faster than QA Net, 20.5% faster than UI MN, and 14.5% faster than RBERT, under different types of user queries, which makes it useful for a wide variety of chatbot design applications.

5. References

[1] L. M. Sánchez-Adame, S. Mendoza, J. Urquiza, J. Rodríguez and A. Meneses-Viveros, "Towards a Set of Heuristics for Evaluating Chatbots," in IEEE Latin America Transactions, vol. 19, no. 12, pp. 2037-2045, Dec. 2021, doi: 10.1109/TLA.2021.9480145.

[2] G. Daniel, J. Cabot, L. Deruelle and M. Derras, "Xatkit: A Multimodal Low-Code Chatbot Development Framework," in IEEE Access, vol. 8, pp. 15332-15346, 2020, doi: 10.1109/ACCESS.2020.2966919.

[3] G. A. Santos, G. G. de Andrade, G. R. S. Silva, F. C. M. Duarte, J. P. J. D. Costa and R. T. de Sousa, "A Conversation-Driven Approach for Chatbot Management," in IEEE Access, vol. 10, pp. 8474-8486, 2022, doi: 10.1109/ACCESS.2022.3143323.

[4] E. H. -K. Wu, C. -H. Lin, Y. -Y. Ou, C. -Z. Liu, W. -K. Wang and C. -Y. Chao, "Advantages and Constraints of a Hybrid Model K-12 E-Learning Assistant Chatbot," in IEEE Access, vol. 8, pp. 77788-77801, 2020, doi: 10.1109/ACCESS.2020.2988252.

[5] L. Zhang, Y. Yang, J. Zhou, C. Chen and L. He, "Retrieval-Polished Response Generation for Chatbot," in IEEE Access, vol. 8, pp. 123882-123890, 2020, doi: 10.1109/ACCESS.2020.3004152.

[6] B. Srivastava, F. Rossi, S. Usmani and M. Bernagozzi, "Personalized Chatbot Trustworthiness Ratings," in IEEE Transactions on Technology and Society, vol. 1, no. 4, pp. 184-192, Dec. 2020, doi: 10.1109/TTS.2020.3023919.

[7] R.Bharathi, T.Abirami,” Energy efficient compressive sensing with predictive model for IoT based medical data transmission”, Journal of Ambient Intelligence and Humanized Computing, November 2020, https://doi.org/10.1007/s12652-020-02670-z
[8] A. Miklosik, N. Evans and A. M. A. Qureshi, "The Use of Chatbots in Digital Business Transformation: A Systematic Literature Review," in IEEE Access, vol. 9, pp. 106530-106539, 2021, doi: 10.1109/ACCESS.2021.3100885.
[9] A. Ait-Mlouk and L. Jiang, "KBot: A Knowledge Graph Based ChatBot for Natural Language Understanding Over Linked Data," in IEEE Access, vol. 8, pp. 149220-149230, 2020, doi: 10.1109/ACCESS.2020.3016142.
[10] R. Rajkumar and V. Ganapathy, "Bio-Inspiring Learning Style Chatbot Inventory Using Brain Computing Interface to Increase the Efficiency of E-Learning," in IEEE Access, vol. 8, pp. 67377-67395, 2020, doi: 10.1109/ACCESS.2020.2984591.
[11] R. Ren, M. Zapata, J. W. Castro, O. Dieste and S. T. Acuña, "Experimentation for Chatbot Usability Evaluation: A Secondary Study," in IEEE Access, vol. 10, pp. 12430-12464, 2022, doi: 10.1109/ACCESS.2022.3145323.
[12] K. Denecke, S. Vaaheesan and A. Arulnathan, "A Mental Health Chatbot for Regulating Emotions (SERMO) - Concept and Usability Test," in IEEE Transactions on Emerging Topics in Computing, vol. 9, no. 3, pp. 1170-1182, 1 July-Sept. 2021, doi: 10.1109/TETC.2020.2974478.
[13] C. B. Nordheim, A. Følstad and C. A. Bjørkli, "An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study," in Interacting with Computers, vol. 31, no. 3, pp. 317-335, May 2019, doi: 10.1093/iwc/iwz022.
[14] J. -C. Gu, Z. -H. Ling and Q. Liu, "Utterance-to-Utterance Interactive Matching Network for Multi-Turn Response Selection in Retrieval-Based Chatbots," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 369-379, 2020, doi: 10.1109/TASLP.2019.2955290.
[15] K. Palasundram, N. MohdSharef, K. A. Kasmiran and A. Azman, "SEQ2SEQ++: A Multitasking-Based Seq2seq Model to Generate Meaningful and Relevant Answers," in IEEE Access, vol. 9, pp. 164949-164975, 2021, doi: 10.1109/ACCESS.2021.3133495.
[16] D. Carlander-Reuterfelt, Á. Carrera, C. A. Iglesias, Ó. Araque, J. F. Sánchez Rada and S. Muñoz, "JAICOB: A Data Science Chatbot," in IEEE Access, vol. 8, pp. 180672-180680, 2020, doi: 10.1109/ACCESS.2020.3024795.
[17] B. Kim, J. Seo and M. -W. Koo, "Randomly Wired Network Based on RoBERTa and Dialog History Attention for Response Selection," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 2437-2442, 2021, doi: 10.1109/TASLP.2021.3077119.
[18] K. Palasundram, N. MohdSharef, K. A. Kasmiran and A. Azman, "Enhancements to the Sequence-to-Sequence-Based Natural Answer Generation Models," in IEEE Access, vol. 8, pp. 45738-45752, 2020, doi: 10.1109/ACCESS.2020.2978551.