

TherapyBot: a chatbot for mental well-being using transformers

Deepak Dharrao¹, Shilpa Gite²

¹Department of Computer Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India

²Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India

Article Info

Article history:

Received Jul 15, 2023

Revised Aug 4, 2023

Accepted Nov 13, 2023

Keywords:

Chatbot

Conversational artificial intelligence

Deep learning

Mental well-being

Natural language processing

Psychology

Transformers

ABSTRACT

The field of natural language processing (NLP) and conversational artificial intelligence (AI) has one ingenious application in the psychological space. Depression and anxiety are two major issues that the world is facing, with close to 41% of adults reporting these symptoms in the United States alone, as of December 2020. It has also been observed that most of the people are not open about it. As a result, it is critical to address this issue on a global scale. Developed countries reportedly have 9 psychiatrists per 100,000 people. One way to mitigate this is the use of chatbots. We propose a transformer-based methodology to build a therapy bot that has been trained on a combination of open-domain conversations from a publicly available dataset and therapist-client conversations from a self-constructed dataset. This end-to-end data-driven model shows quality performance in conversations and adds value by aiding in the case of mental health issues. The proposed architecture is proven to be effective in its usability in the psychological space for both single-turn and multi-turn dialogue. The performance of the proposed system shows a loss is 0.29 and perplexity is 1.34, both metrics keep gradually decreasing which means an improvement in the performance of the chatbot system.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Deepak Dharrao

Department of Computer Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University)

Lavale, Mulshi, Pune, Maharashtra 412115, India

Email: deepakdharrao@gmail.com

1. INTRODUCTION

The chatbot system is more popular nowadays because of its ability to converse and interact with humans efficiently. It interacts using written, spoken, and visual languages. Chatbot tools are used widely in various sectors like education, marketing, medical for improving user satisfaction through timely responses. These tools can very effectively be used to solve queries of individuals with mental disorders. In situations like mental disorders, chatbots can be useful tools for individuals who are hesitating to discuss disputes related to mental health and to take advice from other human beings with direct contact.

One of the fundamental goals of artificial intelligence (AI), especially natural language processing (NLP), has been to create intelligent conversation systems that react as naturally as a person to a user question, whether for a specific task or a general one. The sector of NLP has witnessed stellar growth in the past few decades. One subdomain that stands out is conversational AI, which has seen some breakthroughs [1]. One ingenious application that has been put forth is its use in the psychological space. One of the fundamental goals of AI, especially NLP, has been to create intelligent conversation systems that react as

naturally as a person to a user question, whether for a specific task or a general one. This is known as conversational AI, and it encompasses a wide range of technologies such as question-answering systems [2], [3]. Domain-specific/open-domain chatbots, and so on. Therapies for mental well-being in the form of chatbots are progressing a lot. A machine may help, listen to, and counsel a person in an unbiased manner without any judgment is the main reason for its progress [4].

2. BACKGROUND OF CONVERSATIONAL ARTIFICIAL INTELLIGENCE

Significant volumes of data on conversations for model training have been made available over the last decade, encouraging findings such as increasing customer satisfaction, conversion, and improving marketing performance. Furthermore, recent breakthroughs in deep learning, a subset of AI that uses neural networks to create intelligent systems; reinforcement learning, another subset of AI in which an agent learns to perform a specific task by interacting with its environment by being rewarded or penalized based on its actions; and multi-task learning, yet another subset of AI in which multiple learning tasks are undertaken concurrently, have aided conversational agents in evolving at an incredible rate. Many prominent industrial conversation systems have been constructed utilizing an amalgamation of supervised, unsupervised, and reinforcement learning [5]. Despite their extensive usage, they face a number of issues, such as failing to understand the user's sentiment [6], losing track of the dialogue [7], providing boring non-contextual replies [8], or just stumbling with modern-day lingo [9], [10].

Recent breakthroughs in NLP have provided us with language models to mitigate some of these challenges. The transformer architecture has become tremendously popular owing to the fact that it consistently outperforms other language models such as recurrent neural networks (RNNs) [11]. This methodology, which utilizes fully connected neural network layers and the concept of self-attention, helps retain longer conversation histories, leading to consistent, contextual, and improved conversation. Researchers have also demonstrated how unsupervised pre-training of huge language models on a vast corpus of data leads to improved performance when fine-tuned on specific tasks. This can be observed clearly when we look at OpenAI's GPT series: GPT, GPT-2, and GPT-3 which is the best language model the world has seen yet, with its ability to cater to any language task, be it question-answering, reading comprehension, text summarization, text generation or conversation modeling [12], [13].

2.1. Mental health: a growing concern

One of the medical terms known as mental disease is often acknowledged as a mental health issue. These types of concerns consist of a wide range of broad range of complications that influence human thoughts behaviors and emotions. There are numerous types of issues like addictive behaviors of human beings, nervous situations, sadness, and problems related to diets are all symptoms of mental disease. In various surveys, it is noted that many people from different age groups face these issues these days [14].

However, when any symptoms related to mental sickness possess consistent stress and widely impact on working ability of any human being, such a health problem becomes a mental health disorder. A mental sickness can have a negative impact on anyone's happiness and create complications in their day-to-day life, such as doing routine work at home, kitchen while studying or teaching at school or college, working at the office or it also impacts personal life. In most situations, symptoms can be treated with a mix of medicines, and talk therapy i.e., psychotherapy. Depression and loneliness are one of the significant issues that our community is facing today. It is also observed that most people are not open about it. Hence, it is imperative to address this issue quickly on a global level. Unfortunately, the hospitality services are insufficient to solve this grave problem. In developed regions, there are around nine psychiatrists per 100,000 people [15]. The situation gets worse in developing countries.

2.2. Conversational artificial intelligence in psychological space

Although there is still a lot to explore when we talk about conversational AI in a psychological space, its prospect is now visible as a guide for the prevention, remedy, and observation-up/relapse prevention of psychological troubles and mental disorders. They could be used in the future for suicide prevention. In the remedy of psychological issues, chatbots might offer tools that individuals should work with on their person. After the crowning glory of classical psychotherapy, chatbots are probably the next step to stabilize intervention effects, facilitate the transfer of the healing content material into daily life, and decrease the probability of relapse. Studies show that people find it difficult to open up to a therapist or a friend or colleague. Many do not even have access to a therapist. In such cases, the intervention of conversational AI is necessary [16].

Multiple mental health chatbots, such as Woebot and Replika, witness good results [17], [18]. They help people start that initial conversation about their issues, which then becomes a regular activity. They

provide a safe space for those who are not comfortable discussing their thoughts with another individual. Therefore, virtual therapy given by a chatbot could improve access to psychological treatment and is more straightforward for those who are hesitant to talk with a therapist.

This project aims to create an open-domain generative model for conversational AI agents leveraging a transformer-based architecture. The agent must be able to comprehend the user input statements and generate close-to-human responses. The AI agent is expected to operate in the domain of mental health by providing psychotherapy [19], [20].

2.3. Dataset description

The Facebook dataset is an open-source dataset consisting of many open-domain conversations of about 5-6 sentences. They are between two individuals, which make up 58,881 input-output pairs. Certain sample open-domain conversations from this dataset are shown in Figure 1.

```
[
  "I felt afraid. Still i remember it",
  "What were you afraid of? ",
  "Two years ago_comma_ i was admitted in hospital. I was ill",
  "Oh No_comma_ well glad to see you are here still! ",
  "Luckily"
],
[
  "We have a new manager at work and it isnt going well. ",
  "what is your job position?",
  "I do sales work_comma_ but he always lies to us and takes our bonus money. ",
  "That's very bad"
],
```

Figure 1. Sample conversations from the Facebook dataset

On the other hand, the CounselChat dataset [21], [22] includes 2,130 question-answer pairs from conversations that occur between a therapist and their client, see Figure 2. These have been scraped from counselchat.com and cover over 31 different topics ranging from ‘depression’ to ‘substance abuse’ to ‘military issues’, see Figure 3. The questions are relatively short in this CounselChat dataset, but most of the responses are tremendously long in terms of the number of words as shown in Figure 4. Thus, making it infeasible for us to use them. We also observe that there are a greater number of responses as compared to questions. This implies that there are multiple responses to each question in our dataset. This helps us create a more adaptive conversational model.

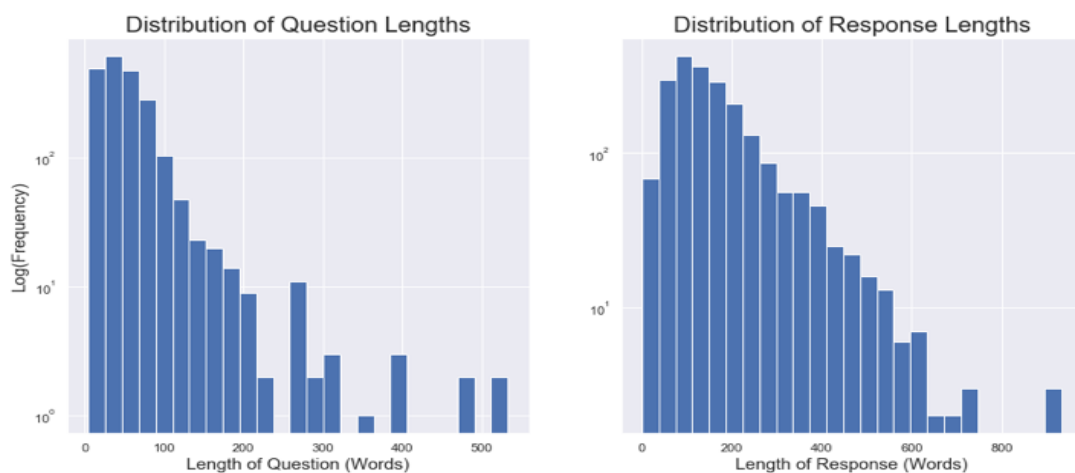


Figure 2. Distribution of question and response length from CounselChat

Furthermore, certain topics occur very frequently in our data. These are depression, anxiety, and relationships/intimacy from the list of 31 topics. This corpus, which is an amalgamation of two datasets has been pre-processed by performing the steps: i) Extract conversation pairs from the datasets into a list of

“questions” and “answers” simultaneously; ii) Replace words like “he’s” or “they’d” with “he is” and “they would” simultaneously; iii) Remove special characters; iv) Tokenize the data, by breaking a sentence into multiple words aka tokens. Add ‘start and end’ tokens to showcase the beginning and end of every sentence; v) Encode the tokenized sentences by converting each word to a number/vector in n-dimensional space; vi) Filter out the sentences those having more than 60 tokens; and vii) Pad the final tokenized sentences to 60 tokens.

In this research to make model training feasible given the hardware constraints we have limited the maximum length of the sentence to 60 words. At the end of pre-processing, we obtain a dataset of 20,096 input-output pairs with a vocab size of 8515.

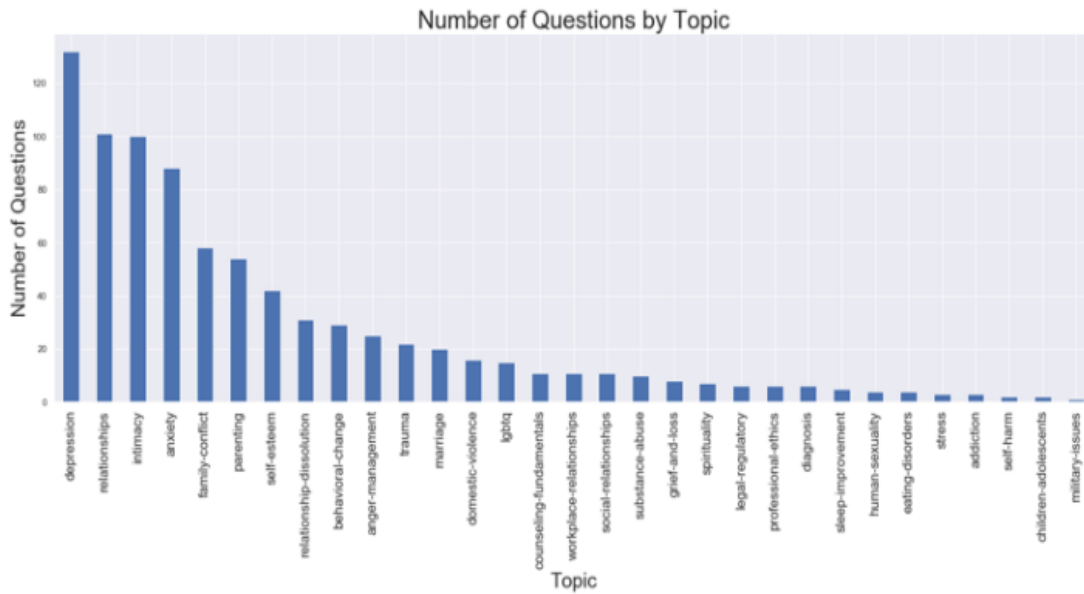


Figure 3. CounselChat number of questions by topic

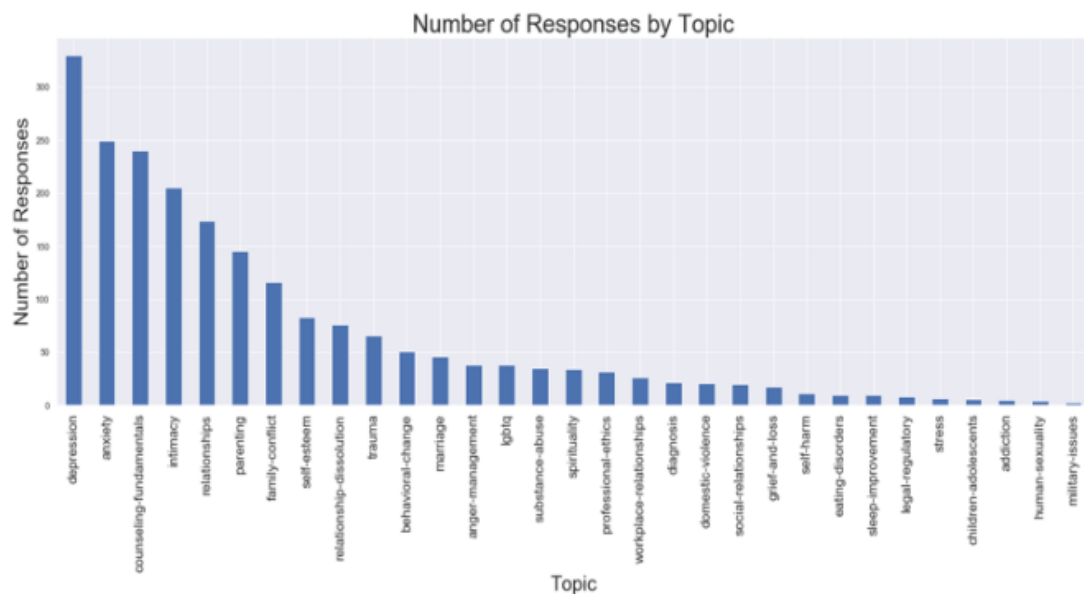


Figure 4. CounselChat number responses by topic

3. SYSTEM DESIGN

In this section, the dataset collection process, model training, and prediction tasks are explained. The system architecture as shown in Figure 5 the system design can be described in three major parts first one is dataset collection from open domain dataset, model training, and prediction from trained model.

3.1. Dataset collection

In the model training phase, the dataset comprises open domain conversations present in the “Facebook dataset” and domain-specific therapist-client question answers scraped from counselchat.com, to create the “CounselChat dataset”. Due to an absolute lack of high-quality mental health-oriented conversational data, we train the model on an open-domain dataset, followed by domain-specific data [23]. This ensures that the model can engage in day-to-day conversation by enhancing its knowledge, meanwhile providing therapy, and talking about sensitive topics when required.

3.2. Model training

The dataset from open domain i.e., Facebook and CounselChat fetched for training the proposed model. The system is trained and saves the logs and model weights. The predefined questions and responses on various topics are used to train the proposed system.

3.3. Prediction

The proposed system was trained with the help of open-domain datasets. Then this trained model is used for the prediction of responses to the newly added questions during the testing phase of research work. The responses are generated to work as the AI-assisted therapist. The implemented model could generate the responses that were validated by the experts.

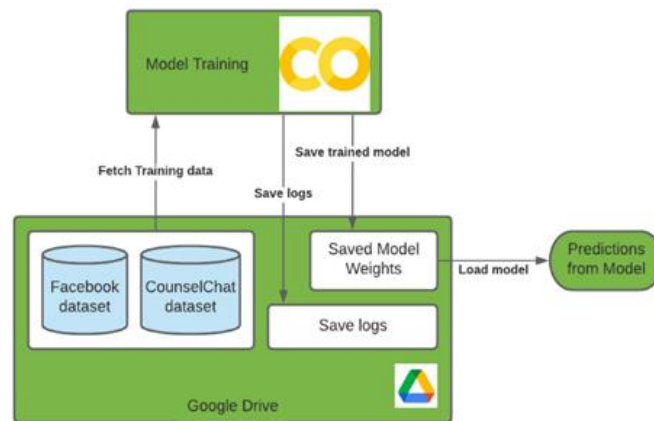


Figure 5. System architecture

4. MODEL ARCHITECTURE

The transformer model [24] was built using TensorFlow 2.0. Previously, the concept of encoder-decoder with a base model as RNN/long short-term memory (LSTM) was used for most NLP tasks but it was not efficient for understanding the long-term context. So, the concept of a transformer was introduced [25]. The transformer’s architecture as shown in Figure 6 is quite like the encoder decoder but the base model used here is a transformer. Every transformer has 6 layers of encoders and 6 layers of decoders [26]. Each encoder in the system has a self-attention layer and a feed-forward neural network [27]. The words must pass through all these layers of the encoder and then to the decoder. While the model is dealing with a word, the self-attention layer permits it to observe auxiliary positions in the input sequence for better encoding of that word. It utilizes a neural network architecture entirely based on a self-attention mechanism due to which it can work parallel and reduce the number of computations per layer [28], [29]. It works with variable-sized inputs along with blocks of self-attention layers rather than using RNNs or convolutional neural networks (CNNs) like most conversational models. It is also quite good at capturing long-term context since it consists of two parts: encoder and decoder. Every word from the sentence is embedded into the vector of size 512 before passing to the first encoder using Bag of Words and Word2Vec. This embedding happens only at the bottommost encoder. The size of the vector is a hyperparameter which is the length of the longest sentence in the dataset.

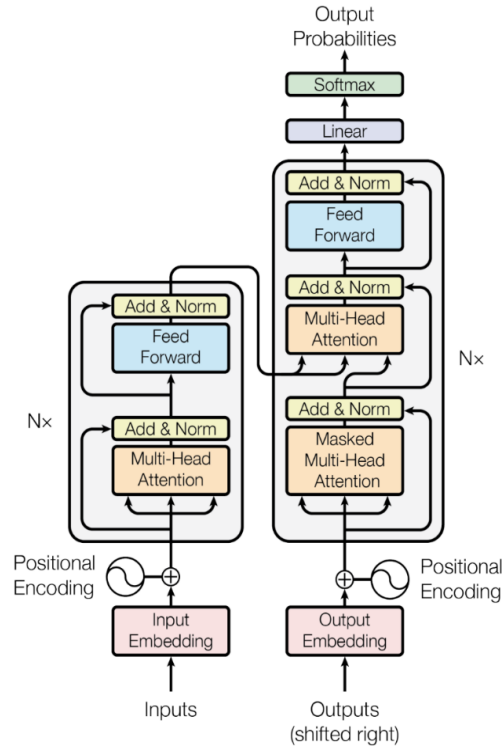


Figure 6. Transformer architecture

After completing this embedding, positional encoding is also done for each word by attaching a vector to each input word so that it understands the position of the word in sequence as well. This architecture ensures that given the clear lack of mental health data, we can leverage open-domain data and then proceed to fine-tune our model on domain-relevant data [24]. We also inculcate language-check libraries into our workflow to fix grammatical errors in the response.

4.1. Attention

A detailed discussion on the first layer i.e. self-attention and how to calculate attention is provided in this section [24]. The first phase in computing self-attention is to produce three vectors from every input passed in the encoder. So, for every word, we create a query vector, key vector, and value vector. These vectors are formed by multiplying the embedding by three matrices. New vectors are smaller in dimension (64) compared to embedding dimension (512). We assign weight vectors for each query, key, and value vector at the start [30]. Then we multiply embedding the vector of the 1st word with the weight vector to get the 1st query vector. A similar process happens for every word of the sentence. We finish up forming a query, key, and value projection for every word in the input sentence, just to give attention to the required words.

The second step is to calculate the dot product of the query vector and key vector to get a score. The score defines how much attention to place on supplementary parts of the input sequence as we encode a word from the given sentence at a particular position. The third step is to divide the scores by the square root of the dimension of the key vector, and then we hand over the calculated result through a SoftMax function. SoftMax normalizes the scores so that they are positive and add up to one. The scaled dot product attention mechanism as shown in Figure 7(a) used in this case can be described as (1).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Where, Q is the matrix that comprises the input query, representing a vector corresponding to a word in the provided sequence. K represents entire keys, i.e. vector notations of cumulative words in the sequence. V are those values, which once more represent the vector of all the words in the sequence.

The attention mechanism is presented in Figure 7 with scaled dot-product and multi-head attention techniques. Figure 7(a) presents the scaled dot-product attention mechanism. As shown in Figure 7(b) for

encoder-decoder multi-head attention modules containing several attention layers working parallel. Here, V has a similar word sequence as Q. Although, for the attention module that consists of the encoder and the decoder sequences, V differs from the sequence signified by Q. The multi-head attention is made up of four parts: i) linear layers which then divide into multiple heads, ii) scaled dot-product attention, iii) concatenation of all these heads, and iv) final linear layer. Here each multi-head attention block accepts Q, K, and V as the inputs.

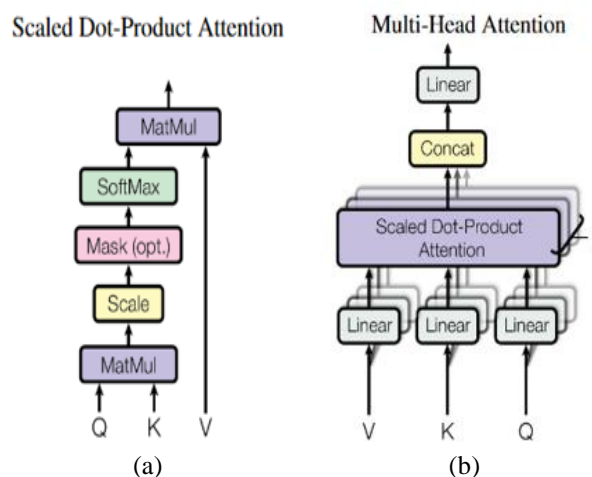


Figure 7. Attention mechanism (a) scaled dot-product attention and (b) multi-head attention

5. PROPOSED TRANSFORMER

As transformers gained a lot of attention in the technical world, we tried to implement them with some amendments. In this section, the description of the proposed transformer is provided which comprises four phases as depicted in Figure 8. It contains masking, positional encoding, encoder, and decoder.

5.1. Masking

These models are auto-regressive in nature i.e.; they make predictions one step at a time by using the outputs until that point [31]. During training, we use teacher-forcing. Hence, the correct output is passed to the upcoming time step irrespective of what was predicted at the present time step. As the transformer forecasts every word, self-attention permits it to consider the words that came before it in the input sequence to forecast the next word. A look-ahead mask is used by the model to prevent it from peaking at the expected output.

5.2. Positional encoding

A positional encoding vector is added later to the initial embedding of the input sequences for each word. To provide a sense of order to the model positional encoding was added [32], [33]. This is added to the input and output embedding since the model does not use any RNN layers, so this helps grasp the relative position between the words in a sentence. The proposed transformer is shown in Figure 9.

5.3. Encoder

As shown in Figure 9(a), every transformer has an encoding component and a decoding component. An encoder Figure 9(a) comprises input embedding, positional encoding, and “x” encoder layers. They are *responsible for analyzing and representing the input sequence in a way the model can understand*.

5.4. Decoder

As shown in Figure 9(b) i.e., another half of the transformer comprises output embedding, positional encoding, and “x” decoder layers. The Encoder and decoder layers are made up of multi-head attention and dense layers. Without the transformer decoder, there would be no way to generate the output sequence. Without an encoder, the transformer decoder would miss important contextual information, resulting in lower-quality output. Combining an encoder and transformer decoder is key to the effectiveness of the transformer architecture in NLP tasks.

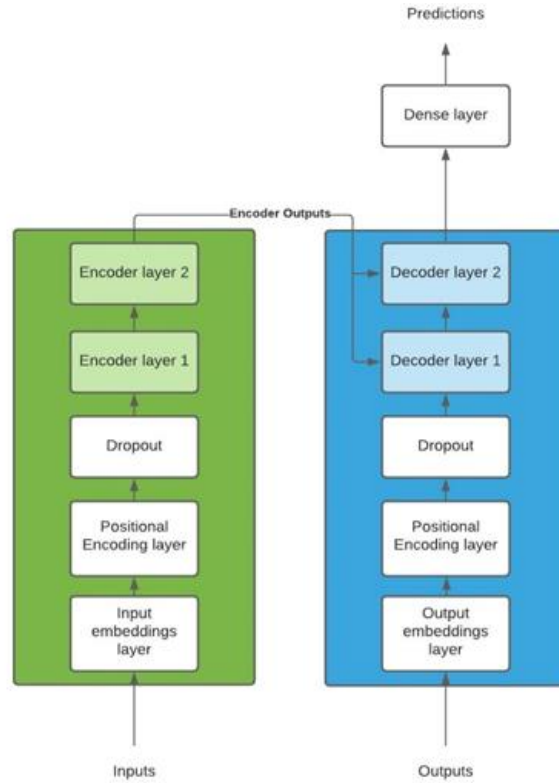


Figure 8. Proposed transformer structure

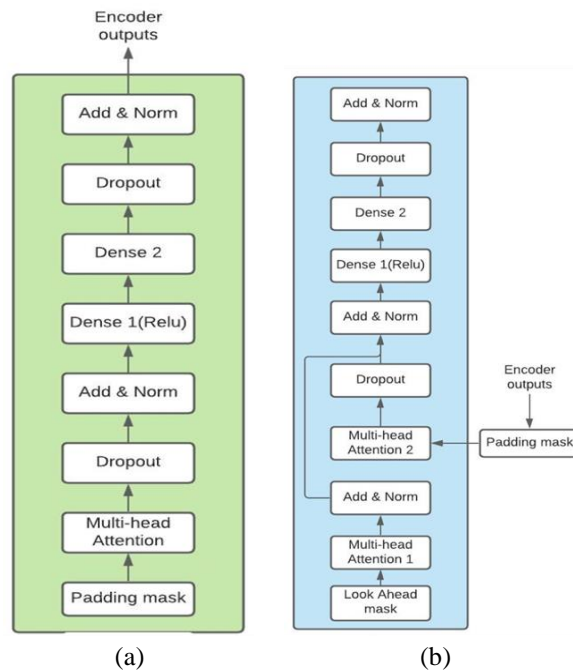


Figure 9. Proposed transformer component (a) encoder layer and (b) decoder layer

6. EXPERIMENTAL DETAILS

The model is then initialized with the following parameters number of the encoder and decoder layers=2, model depth or $d_{model}=256$, number of attention heads=8, number of units=512, dropout=0. We train the model over 12 epochs with a batch size of 64 on the two datasets collectively. The loss function used is

“sparse categorical cross entropy” [34] and the optimizer is “Adam”. We further use a customized learning rate as seen in Figure 10. Here we observed that the learning rate gradually increased in a linear manner from 0.0000 to 0.0010 for training steps from 0 to 3000 training steps and the learning rate slightly decreased for training steps above 3,000.

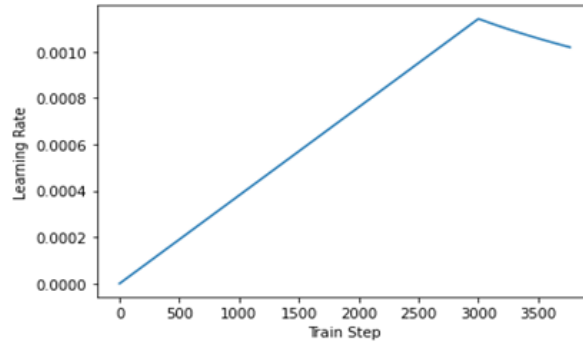


Figure 10. Custom learning rate

6.1. Performance analysis

The performance of this transformer model when trained collectively on both datasets was analyzed using two metrics as shown in Figures 11 and 12. As shown in Figure 11 the result for loss is 0.29. It helps us find the similarity between the output generated by the model and the expected output present in our data. The lower the perplexity, the better the model is said to perform the results are shown in Figure 12. Here the result for perplexity is 1.34. The obtained results from both these metrics keep steadily decreasing, hence indicating an improvement in our chatbot’s performance.

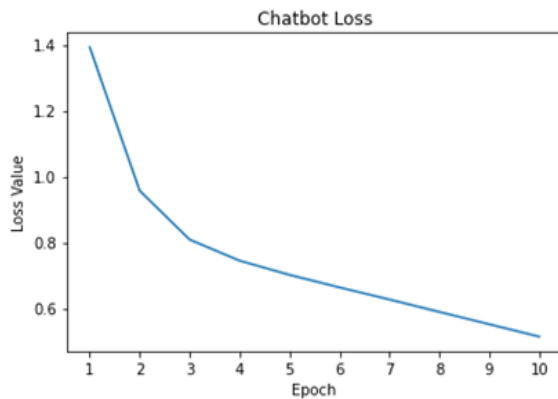


Figure 11. Loss over epochs

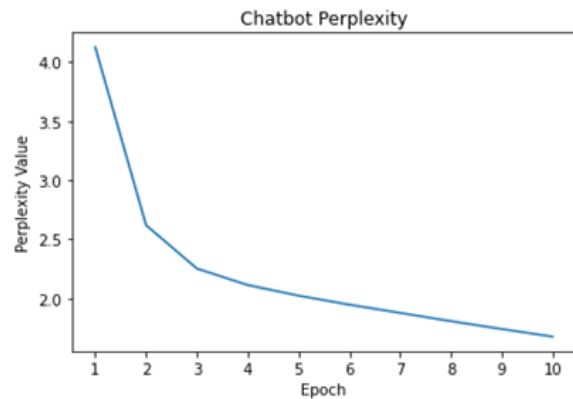


Figure 12. Perplexity over epochs

6.2. Human evaluation

We also did some manual checks by observing the outputs of this chatbot as shown in the Figure 13 screenshot captured from our implementation results. By observing the results obtained from our proposed model we can say that these types of chat-bots are very effective for starting smooth conversations with the users. We can observe that our system starts conversations with very simple question-answers, and it helps to establish a good environment for further discussion and user can easily get support from the automated system like their close friends. Though the approach gives the desired accuracy, limitations could be stated as it works only for the English language, and the accuracy may vary depending on the dataset, pre-processing, training samples, and other language-related parameters. Dataset may play a very important role here as the core learning is solely dependent on it. Anger-based evaluation [35] or stress detection using social media posts [36] can also be seen as an extension to similar problems based on the availability of good datasets.

```

User: i feel stressed out
Bot: i am sorry to hear you are feeling that way
User: i have a lot of work to finish
Bot: i bet you are going to have a good time !
User: i doubt that
Bot: i am sorry to hear that . that is holding you back ?

Input: Is it okay to feel sad all the time?
Output: i am so sorry to hear that . it is a very normal to feel better now .

Input: I want to kill myself.
Output: why do you say that ?

Input: I feel lonely.
Output: why are you scared

Input: My pet died. I can't stop crying.
Output: i am sorry to hear that . how are you feeling now ?

Input: I had a huge fight with my friend today. I hate him.
Output: oh that is bad . what happened ?

Input: oh that is bad . what happened ?
Output: i had to work

Input: i had to work
Output: what did you do ?

Input: what did you do ?
Output: i did not tell them because i was very upset about it .

Input: My pet died. I can't stop crying now.
Output: i am sorry to hear that . i hope you are okay .

Input: I had a fight with my friend today.
Output: oh no ! what happened ?

Input: oh no ! what happened ?
Output: he ran away

```

Figure 13. Sample outputs of chatbot

7. CONCLUSION

In this paper, we have proposed a chatbot system specially for Mental Well-Being. The results obtained conclude that conversational agents can aid people in starting that initial conversation about their issues, which then becomes a regular activity, and they feel a safe space discussing their thoughts. Therefore, providing virtual therapy and hence improving access to psychological treatment. We have used Transformers to obtain a great chatbot that can track context over time and does not produce bland responses like “I don’t know”. Furthermore, we ensure that the responses are grammatically correct. Due to a lack of high-quality conversational data related to mental health, we use two different datasets for training, one is a vast open-domain dataset by Facebook and the other is mental health QA data obtained from Counsel Chat. The results obtained from the final model have a loss value of 0.29 and a perplexity of 1.34. The obtained results from both these metrics keep steadily decreasing, hence indicating an improvement in our chatbot’s performance.

In the future researchers can explore many new avenues for this project such as using a bigger model to obtain better results, as seen in countless research experiments lately. Reinforcement learning can be incorporated for continuous user feedback integration to improve model performance. Integration of sentiment analysis using multi-task learning to improve the responses is also a promising field. Generative adversarial networks can also be explored for building chatbots. Transfer learning on state-of-the-art models can be carried out and benchmarked against the proposed approach. Other emotions detection such as anger and happiness can also be detected. It can also lead to mental stress detection using texts on social media. Overall there are many such avenues for future research work.

ACKNOWLEDGMENTS

The authors thank Symbiosis International (Deemed University), Pune, India.





REFERENCES

- [1] M. Aleedy, H. Shaiba, and M. Bezbradica, "Generating and analyzing Chatbot responses using natural language processing," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 9, pp. 60–68, 2019, doi: 10.14569/ijacsa.2019.0100910.
- [2] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on freebase from question-answer pairs," *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1533–1544, 2013.
- [3] J. Bao, N. Duan, M. Zhou, and T. Zhao, "Knowledge-based question answering as machine translation," in *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 967–976, doi: 10.3115/v1/p14-1091.
- [4] J. Cahn, "Chatbot: architecture, design, and development," University of Pennsylvania, 2017.
- [5] Y. L. Hsueh and T. L. Chou, "A task-oriented chatbot based on LSTM and reinforcement learning," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 1, pp. 1–27, Jan. 2022, doi: 10.1145/3529649.
- [6] X. Han and M. Zhou, "Designing effective interview chatbots: Automatic chatbot profiling and design suggestion generation for chatbot debugging," in *Conference on Human Factors in Computing Systems - Proceedings*, New York, NY, USA: ACM, May 2021, pp. 1–15, doi: 10.1145/3411764.3445569.
- [7] B. Galitsky, "A social promotion chatbot," in *Developing Enterprise Chatbots*, Cham: Springer International Publishing, 2019, pp. 427–463, doi: 10.1007/978-3-030-04299-8_12.
- [8] G. Caldarini, S. Jaf, and K. McGarry, "A literature survey of recent advances in chatbots," *Information (Switzerland)*, vol. 13, no. 1, pp. 1–22, Jan. 2022, doi: 10.3390/info13010041.
- [9] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI blog*, 2018.
- [10] Y. Zhang *et al.*, "Dialogpt: large-scale generative pre-training for conversational response generation," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 270–278, 2020, doi: 10.18653/v1/2020.acl-demos.30.
- [11] A. K. Kushwaha and A. K. Kar, "MarkBot – a language model-driven chatbot for interactive marketing in post-modern world," *Information Systems Frontiers*, pp. 1–18, Aug. 2021, doi: 10.1007/s10796-021-10184-y.
- [12] W. D. Heaven, "How to make a chatbot that isn't racist or sexist," *Ethics of Data and Analytics*, pp. 60–62, 2022, doi: 10.1201/9781003278290-10.
- [13] N. Shi, Q. Zeng, and R. Lee, "The design and implementation of language learning chatbot with xai using ontology and transfer learning," *arXiv preprint arXiv:2009.13984*, 2020, pp. 305–323, 2020, doi: 10.5121/csit.2020.101124.
- [14] V. More, V. Koli, V. Vijaykumar, and V. Jethani, "Chatbot for mental well-being," *ITM Web of Conferences*, vol. 40, p. 03019, Aug. 2021, doi: 10.1051/itmconf/20214003019.
- [15] S. Almanzar, C. L. Katz, and B. Harry, "Treatment of mentally III offenders in nine developing Latin American countries," *Journal of the American Academy of Psychiatry and the Law*, vol. 43, no. 3, pp. 340–349, 2015.
- [16] R. Meadows, C. Hine, and E. Suddaby, "Conversational agents and the making of mental health recovery," *Digital Health*, vol. 6, pp. 1–11, Jan. 2020, doi: 10.1177/2055207620966170.
- [17] J. J. Prochaska *et al.*, "A therapeutic relational agent for reducing problematic substance use (Woebot): Development and usability study," *Journal of Medical Internet Research*, vol. 23, no. 3, p. e24850, Mar. 2021, doi: 10.2196/24850.
- [18] L. M. Possati, "Psychoanalyzing artificial intelligence: the case of replika," *AI and Society*, vol. 38, no. 4, pp. 1725–1738, 2023, doi: 10.1007/s00146-021-01379-7.
- [19] R. Fulmer, A. Joerin, B. Gentile, L. Lakerink, and M. Rauws, "Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: randomized controlled trial," *JMIR Mental Health*, vol. 5, no. 4, p. e64, Dec. 2018, doi: 10.2196/mental.9782.
- [20] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," *JMIR Mental Health*, vol. 4, no. 2, pp. 1–11, Jun. 2017, doi: 10.2196/mental.7785.
- [21] A. Welivita and P. Pu, "Curating a large-scale motivational interviewing dataset using peer support forums," *Proceedings - International Conference on Computational Linguistics, COLING*, vol. 29, no. 1, pp. 3315–3330, 2022.
- [22] N. Bertagnolli, "Counsel chat: bootstrapping high-quality therapy data," May 2020. Accessed: Nov. 06, 2023. [Online]. Available: <https://towardsdatascience.com/counsel-chat-bootstrapping-high-quality-therapy-data-971b419f33da>
- [23] A. A. Abd-Alrazaq, M. Alajlani, N. Ali, K. Denecke, B. M. Bewick, and M. Househ, "Perceptions and opinions of patients about mental health chatbots: scoping review," *Journal of Medical Internet Research*, vol. 23, no. 1, pp. 1–15, Jan. 2021, doi: 10.2196/17828.
- [24] Vasvani A. *et al.*, "Attention is all you need," *Advances in neural information processing systems*, pp. 5998–6008, 2017.
- [25] J. Chen, O. Agbodike, and L. Wang, "Memory-based deep neural attention (mDNA) for cognitive multi-turn response retrieval in task-oriented chatbots," *Applied Sciences (Switzerland)*, vol. 10, no. 17, pp. 1–11, Aug. 2020, doi: 10.3390/app10175819.
- [26] R. Csaky, "Deep learning based chatbot models," pp. 1–67, Aug. 2019, Available: <http://arxiv.org/abs/1908.08835>
- [27] H. Zhang, Y. Lan, L. Pang, J. Guo, and X. Cheng, "ReCoSa: detecting the relevant contexts with self-attention for multi-turn dialogue generation," *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 3721–3730, 2020, doi: 10.18653/v1/p19-1362.
- [28] Y. Ding, M. Jia, Q. Miao, and Y. Cao, "A novel time-frequency Transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings," *Mechanical Systems and Signal Processing*, vol. 168, Apr. 2022, doi: 10.1016/j.ymssp.2021.108616.
- [29] R. Pedro and A. L. Oliveira, "Assessing the impact of attention and self-attention mechanisms on the classification of skin lesions," *Proceedings of the International Joint Conference on Neural Networks*, 2022, doi: 10.1109/IJCNN55064.2022.9892274.
- [30] J. Zhang, J. Zhang, S. Ma, J. Yang, and G. Gui, "Chatbot design method using hybrid word vector expression model based on real telemarketing data," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 4, pp. 1400–1418, Apr. 2020, doi: 10.3837/TIIS.2020.04.001.
- [31] E. V. Epure and R. Hennequin, "Probing pre-trained auto-regressive language models for named entity typing and recognition," *2022 Language Resources and Evaluation Conference, LREC 2022*, pp. 1408–1417, 2022.
- [32] R. Goel, S. Susan, S. Vashisht, and A. Dhanda, "Emotion-aware transformer encoder for empathetic dialogue generation," in *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2021*, IEEE, Sep. 2021, pp. 1–6, doi: 10.1109/ACIIW52867.2021.9666315.
- [33] L. H. Baniata, S. Kang, and I. K. E. Ampomah, "A reverse positional encoding multi-head attention-based neural machine translation model for arabic dialects," *Mathematics*, vol. 10, no. 19, pp. 1–25, Oct. 2022, doi: 10.3390/math10193666.





- [34] J. J. Bird, "Improving customer service chatbots with attention-based transfer learning," pp. 1–9, 2021, Available: <http://arxiv.org/abs/2111.14621>
- [35] A. Surana *et al.*, "An audio-based anger detection algorithm using a hybrid artificial neural network and fuzzy logic model," *Multimedia Tools and Applications*, Oct. 2023, doi: 10.1007/s11042-023-16815-7.
- [36] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, "Machine learning driven mental stress detection on reddit posts using natural language processing," *Human-Centric Intelligent Systems*, vol. 3, no. 2, pp. 80–91, Mar. 2023, doi: 10.1007/s44230-023-00020-8.

BIOGRAPHIES OF AUTHORS



Dr. Deepak Dharrao     has completed her Ph.D. in Computer Science and Engineering with Specialization in Computer Vision. Currently working as an Assistant Professor in the Department of CSE-IT at Symbiosis Institute of Technology, Lavale, Symbiosis International University. He has 15 years of experience in research and academics. He has authored or co-authored over 30 technical articles in refereed journals and conference proceedings. Dr. Deepak is a member of IEEE, ACM, and a Life Member of ISTE professional organizations. His research interests include the areas of image and real-time video processing, data science, artificial intelligence, and machine learning. He can be contacted at email: deepakdharrao@gmail.com.



Dr. Shilpa Gite     associate professor, Symbiosis Institute of Technology, Symbiosis Centre for Applied AI (SCAAI), SIU, Pune. She has 17 years of experience in research and academics. She has authored or co-authored over 60 technical articles in refereed journals and conference proceedings. Dr. Shilpa is a Senior member of IEEE organizations. Her research interests include the areas of deep learning, computer vision, NLP, artificial intelligence, and machine learning. She can be contacted at email: shilpa.gite@sitpune.edu.in.