Mental Health Dialogue System for Emotional Well-being

Subramanya N **Venkata Bramara Parthiv Dupakuntla** **Shashank Srigiri**
snagabhushan@umass.edu vdupakuntla@umass.edu ssrigiri@umass.edu

1 Problem statement

Mental health is the 'unseen' state of a human being that defines one's emotional state of well-being. In this rapidly changing world, mental health and its effects have become an important aspect of research among both the medical, and scientific communities. It is well known that a person who is troubled by any sort of disease, whether it is physical or mental, does not often choose to come forward and ask for help. This is very prominent in the case of mental health issues such as 'depression', 'anxiety', 'stress', 'PTSD' [Post Traumatic Stress Disorder], etc. Based on these premises, we find our motivation in building a mental health dialogue system so that people can cross the above mentioned barrier and seek help and advice. To do so, we explore the realms of Conversational AI, particularly from an emotional stand point and present a novel T5 architecture along with the baseline that successfully incorporates the emotions a counsellor or a therapist needs to express towards their clients. We evaluate these models on both automated evaluation metrics and human evaluation metrics and present our findings.

2 What you proposed vs. what you accomplished

After numerous discussions with the professor and the internal reviews we had, we made the following modifications to our proposed model:

- *Reinforcement Learning Approach*: We proposed an RL based approach to overcome the problems caused by SEQ2SEQ models. We realised that this reward based approach is not feasible due to the lack of token level rewards, and because we only get rewards at the sequence level.
- *Using BERT as the backbone*: Owing to the professor's lectures on Text Generation models such as GPT2 and T5, we realised that BERT could not be used to generate text without significant modifications and left-to-right language models such as GPT2 and conditional Language Models such as T5 would lead to a better approach for this problem
- *Encoder-Decoder Architecture*: We proposed to use an encoder-decoder architecture which simulates the virtual agents talking to each other and we reward or penalize based on the dialogue generated, in a similar style of Alpha-Go. However, we implemented a novel T5 architecture that will be explained later in the below sections.
- We have built and trained two T5 models with one being the baseline, and one being our novel approach.
- We have substantial evaluation results to show that the novel T5 architecture we built outperformed the baseline model.
- *Change in Evaluation Metrics*: We have also changed the evaluation metrics with respect to our proposed evaluation metric, the Facebook FAISS (Johnson et al., 2017). After an in-depth literature review, we designed a detailed error analysis framework using both the automated evaluation metrics such as the held-out perplexity, BLEU (Papineni et al., 2002) Score, as well as the human evaluation metrics, explained later in the below sections.

3 Related work

All the related work categorically falls into two types of queries being used, contextual queries and non contextual queries. In contextual queries, we have a long-term dependency somewhere in the past and the patient is referring to that specific context and asking a query related to that. Non contextual queries are the type of queries in which even though the chat is continuous, the patient might ask another query independent and unrelated to the context of the conversation. It is very important to understand what the category a query falls under in order to build a better dialogue system.

Prior studies (Wang et al., 2012) and (Coulibaly et al., 2014) have done this by using handcrafted features. Such features can be Bag of words, some kind of word patterns, lexicon patterns, etc. Biyani et al. ((Coulibaly et al., 2014)) were able to show that using words or lexicons that are subjective or using concrete context-based words like medicine names, other side effects, and the medication process is among the most helpful features in building an online support system. (Khanpour and Caragea, 2018) proposed a CNN + LSTM implemented model where they are using embeddings acquired by passing them through CNN and later passing that through the LSTM for the other downstream task. This usage of CNN + LSTM helped them in achieving higher, better emotional chatbot models. CNN is being used to extract highlevel features that can carry the gist of the sentence and LSTM represents the function that returns the response. It is important to understand the developments in statistical dialogue system. Empirically put, building a statistical dialogue system falls under the two following categories:

One way is that we can simulate a supervised learning approach where we give a huge amount of training data and the models learn based on these inputs. Over time, depending on how well the model is trained, it then starts giving similar responses. Researchers have used various methods to approach the Dialogue Generation problem. (Ritter et al., 2011) approach this response generation problem as a statistical machine translation (SMT) problem. There are neural response generation models which have gained a lot of attention recently. In addition to these, (Sutskever et al., 2014), proposed an LSTM sequence-to-sequence (SEQ2SEQ) model using Maximum Likelihood Estimation (MLE). There have been various approaches that were studied in the past from retrieval models to generative models(Wu et al., 2019; Cai et al., 2019; Weston et al., 2018). However, the major issues in most of the models were that the responses were very generic and repetitive (Vinyals and Le, 2015).

Because of this, an effort was made in order to make the chat more informative by keeping the information of the context (Sordoni et al., 2015) (Serban et al., 2016), along with generating diverse and contextualized responses. A GAN-based approach was introduced by Wang and Wan, 2018 which controlled the sentiment of the generated text. (Rashkin et al., 2019) introduced the EmpatheticDialogues dataset and fine-tuned the baselines to generate emotional responses. (Santhanam and Shaikh, 2019) finetuned the GPT2 (Radford et al., 2019). (Lee et al., 2020) used reinforcement learning for predicting the response along with sentiment look-ahead.

The other and optimal approach is similar to the one we would follow in this project, which is task-oriented. In this project, we introduce an end-to-end emotional chatbot which not only tries to generate cohesive responses but also tries to take empathy into account by using 2 different losses. Our model tries to minimize the loss over time and learns to give appropriate answers.

4 Dataset

We have explored the mental health dataset we collected from the two volumes of anonymous Counselling and Psychotherapy sessions. We have chosen to utlise the EmpatheticDialogues dataset over this because of the following reasons:

- It has 1649 conversations where there are several instances of male and female clients, and also instances with multiple clients in the same therapy session.
- After formatting into a single structure of conversations starting with the client and having an equal number of client and therapist responses, and selecting conversations that have at least 20 dialogues, we had our final dataset of 823 conversations.
- In these 823 conversations, the dialogues between the client and the therapist are natural and hence included a lot of redundant words such as a sentence - ”I'll I'll do this but but I get what you are trying to tell me so so umm umm I'll I'll take your leave”. Which made it unfeasible to remove all these redundancies and create a homogeneous dataset.

Empathetic Dialogues dataset:

- We observed that the Empathetic Dialoguess(Rashkin et al., 2019) dataset will be the ideal fit for our problem scenario as it contains several emotions that a human can display.
- It is a crowdsourced dataset using 810 MTurk workers via the ParlAI platform.
- The process of collection involves 2 steps. Firstly, Pair of workers are asked to select an emotion word and describe a situation where they felt that way. Then, they had a conversation with each other about the situations.
- Each worker had to contribute one situation description, and one pair of conversation.
- In one conversation, they had to contribute as a speaker about the situation they contributed and in another as a listener where situation is contributed by the speaker.

Brief Statistics of the data:

- It comprises of 24,850 conversations gathered from 810 participants.
- Our train/validation/test split is according to 80%/10%/10% which finally amounts to 19533 / 2770 / 2547 conversations, respectively.

4.1 Data preprocessing

The data is present in the following format. We have 3 columns, namely

- Context
- ID
- Utterance sentence

Each ID represents a whole conversation and each row with an ID represents a single utterance. Each ID is present multiple number times representing an input sentence or 1 turn.
When the patient speaks a sentence for the first time, that is passed to the model, and the model generates the response. However, the next time we need the context of the previous sentences, and due to which we append the utterance1+utterance2 and then we combine them into a single string and then pass that through the model. And similarly, for the 3rd utterance, we pass the utterance1+utterance2+utterance3 to the model.
One other important thing is that the maximum sequence length that our model can take is 50. If the sentence length is less than 50 then we add ¡PAD¿ tokens to make the length 50. In between the different utterance sentences, we append the ¡SEP¿ token to notify that they are different utterances. If the sentence length is more than 50 then we split the combined sentence into multiple inputs and then operate on that.

5 Baselines

The model that we use in our project is called the T5 model. T5 stands for "Text-to-Text Transfer Transformer" and was recently proposed by google in the paper "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". This paper proposes a unified framework to solve many NLP tasks using the text-to-text format.

The Input/Output for the T5 model is as follows: We basically corrupt the input and then we ask the model to predict what was corrupted. It is basically set up as a generation problem rather than as a classification problem.

Input/Output for T5: We start with an unlabelled arbitrary sentence. For example, consider the original text.

- Original Text Ex: "We don't see things as they are, we see them as we are."
- Input Text Ex: "We don't see things $< X >$, we see them as $< Y >$"
- Target Text Ex: $< X >$ as they are $< Y >$ we are $< Z >$

We then mask out approx 15% of the tokens in the sequence. Unlike BERT, T5 prefers to mask the contiguous spans of text. Ex: "as they are" We then replace these spans with special tokens. Ex: $< X >, < Y >$. In the above example, two different spans were masked out with 2 different lengths. $< X >$ is of length 3 and $< Y >$ is of length 2. The input text is fed into the transformer encoder and the target output is what the decoder is supposed to produce. This is basically the same as filling the special tokens.

Context	**Conv_id**	**Utterance**
sentimental	hit:0_conv:1	I remember going to see the fireworks with my best friend. It was the first time we ever spent time alone together. Although there was a lot of people_comma_ we felt like the only people in the world.
sentimental	hit:0_conv:1	Was this a friend you were in love with_comma_ or just a best friend?
sentimental	hit:0_conv:1	This was a best friend. I miss her
terrified	hit:2_conv:5	Job interviews always make me sweat bullets_comma_ makes me uncomfortable in general to be looked at under a microscope like that.
terrified	hit:2_conv:5	Don't be nervous. Just be prepared.
terrified	hit:2_conv:5	I feel like getting prepared and then having a curve ball thrown at you throws you off.

Table 1: Examples of Empathetic Dialogues Dataset

Context	***Input***	***Output***
sentimental	I remember going to see the fireworks with my best friend. It was the first time we ever spent time alone together. Although there was a lot of people, we felt like the only people in the world.	Was this a friend you were in love with, or just a best friend?
sentimental	I remember going to see the fireworks with my best friend. It was the first time we ever spent time alone together. Although there was a lot of people, we felt like the only people in the world. Was this a friend you were in love with, or just a best friend?	This was a best friend. I miss her
terrified	Job interviews always make me sweat bullets, makes me uncomfortable in general to be looked at under a microscope like that.	Don't be nervous. Just be prepared.
terrified	Job interviews always make me sweat bullets, makes me uncomfortable in general to be looked at under a microscope like that. Don't be nervous. Just be prepared.	I feel like getting prepared and then having a curve ball thrown at you throws you off.

Table 2: Examples of dataset after Preprocessing

Architecture:

- Left: The left most architecture uses full masking in encoder and uses causal masking decoder. We also have encoder-decoder attention in this architecture.

- Middle: A language model (LM) consisting of one Transformer stack. The input that is fed is the concatenation of input and target along with using the causal making entirely.

- Last: Adding a prefic to a LM allows fully-visible masking on the input.

The way we calculate loss is like the standard LM loss, i.e, for every token, we compute the cross-entropy loss and then we backprop after predicting the entire sequence. Now, after backprop and training, the pertained model is ready and we want to fine-tune the model on the actual task that we care about.We can use transfer learning and use this pertained model to solve some other downstream tasks of our interest.

6 Approach

Our approach is based on the assumption that an empathetic conversational agent should mirror the emotion of the speaker (Carr et al., 2003). We propose to use a mode that favors sentiment understanding and empathetic response generation using the sentiment of each dialogue context. It is based on the Text-to-Text Transformer (T5) (Raffel et al., 2020) and we extend it with a sentiment analysis model and weighted loss during training, in order to apply sentiment understanding and en-

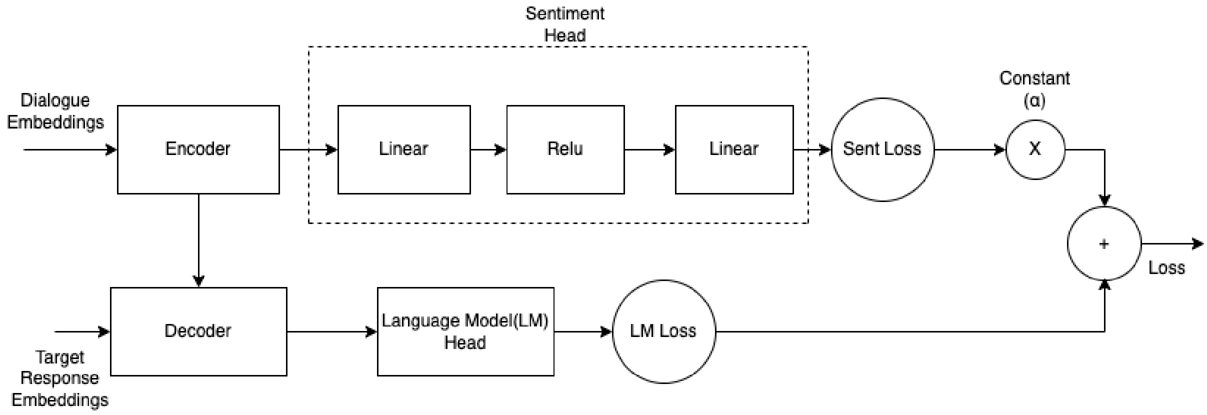

Figure 1: Novel T5 Architecture

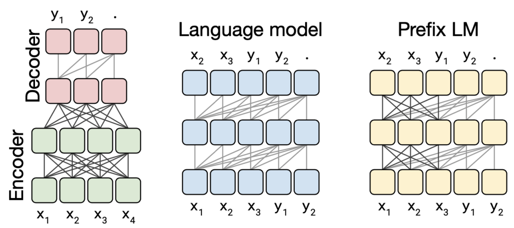

Figure 2: T5 Architecture (Raffel et al., 2020)

force empathetic response generation.

- *Language Modeling*: To optimize the response language modeling objective we use the contextualized representation of the gold response and we apply language modeling by predicting the reply tokens using the cross-entropy loss. We denote that loss as L_{lm}.
- *Sentiment Understanding*: To optimize the sentiment understanding objective, we pass the contextualized representation of the dialogue context through the 2-layer sentiment classifier and we apply sentiment classification using cross-entropy loss. We denote that loss as L_{sent}. In this way, the model learns to predict the sentimental state of the dialogue, and specifically that of one of the speakers, using the sentiment labels we created.

The model is penalized, when the sentiment representation of the generated response is different from that of the dialogue context, not only favoring sentiment understanding but also promoting empathetic response generation. Our final fine-tuning loss function is the weighted sum of the mentioned losses: $Loss : L_{lm} + \alpha * L_{sent}$ (Where α is a constant)

6.1 Implementation Details:

For our baseline, we use the T5-base model from the HuggingFace library which has 12 layers, 768 hidden-states, 3072 feed-forward hidden-states and 12 heads.
To build our novel T5 model, we use a 300-d dimensional space for the sentiment representations obtained from the 2-layer classifier on top of the baseline.
Finally, the baseline and the Novel T5 model have approximately 222M and 223M parameters respectively.
During training, we set to 0.8. After experimenting with various empirically selected value pairs for the parameters , we found that the selected values yield the slightly best PPL for the validation set.

We use the Adam optimizer, setting the learning rate equal to 2e5 and the weight decay to 1e6. We also use a batch size of 8. All hyperparameters were manually tuned and the set with the best validation perplexity was chosen. All models were trained in a single Tesla K80 GPU provided by Google Colab. During inference time, we use top-p (nucleus) sampling method (Holtzman et al., 2020) with top-k filtering (Fan et al., 2018), by setting threshold probability *p* to 0.9 and *topk* to 10. We also add a length penalty equal to 0.6 and we set the maximum length of the generated response to be equal to 40.
Table 3 illustrates the corresponding models and the files used in our project.

6.2 Evaluation Framework

There are 2 types of evaluation metrics we utilize to check the model's performance across various

File Name	Descriptions
.\experiments\train_novel_T5.py	Training Novel-T5
.\experiments\test_novel_T5.py	Test Novel-T5
.\experiments\train_T5.py	Train T5
.\experiments\test_T5.py	Test T5
.\experiments\core\data\empdataset.py	Get Empathetic Dialogue
.\experiments\data\empatheticdialogues	Dataset

Table 3: File Path and Descriptions

Model	BLEU1	BLEU2	BLEU3	BLEU4	Avg BLEU	PPL
T5 (Baseline)	0.26	0.23	**0.21**	0.17	0.2175	13.5
Novel T5 (Ours)	**0.28**	**0.24**	0.20	**0.19**	**0.2275**	**13.9**

Table 4: Automated Evaluation Metric Results

Excellent	Good	Fair	Poor	Very Poor
5	4	3	2	1

Table 5: Likert Scale of Quality

Model	Emotional Appropriateness	Relevance	Readability
T5 (Baseline)	3.62	3.66	4.64
Novel T5	**3.76**	**3.72**	**4.78**

Table 6: Examples where our model outperformed the baseline

Input	T5 (Baseline) Output	Novel-T5 Output	Gold Response
I was bowling last week and I accidentally used the person's custom ball that was next to me!	Oh no! That's so embarrassing!	Oh no! What happened?	oh, how he reacted next
I found an injured bird in my yard today, it was only a baby too!	Oh no! What happened?	I am sorry to hear that. Did you get it to come home?	How terrifying. I hope you were able to find and kill it.
I woke up needing to use the bathroom at midnight. As soon as I sit down a giant bug flew into my face! I screamed so loud that I woke up my husband and he though we were being burgled!	I am sure he was scared.	That is horrible. I hope you were able to sleep well.	Did you help it?
I am terrified that my wife is going to leave me. Why would she leave you? Because she is talking to other men behind my back	I am sure she will.	That is a very scary thought.	oh no, i'm so sorry to hear that. I know your heart is broken. Did you try to talk to her about it?
I was going through some boxes the other day. I found some old pictures of my kids I thought were gone	Oh wow, that's so sad. I hope you find them again soon.	I bet they were a lot of fun.	That is exciting! I love having pictures to look back on.
I love my husband! He has been so sweet this week!Aww! That is so cute! Has he done something special this week that made him extra sweet? He knew I had a rough day today so he brought home my favourite wine and Taco Bell! It's the little things :)	I love Taco Bell! I love Taco Bell!	I bet! I'm glad he did something special!	That is so sweet of him. Its definitely the little things.
I'm feeling a little down right now. The weather is perfect, but I miss not going to the beach with my husband and dog and catching the sunset. Those were great times! I bet! Im sorry you are feeling down. Yeah, we would take a bottle of wine with us, watch the surfers and just kick back. Wish those days would return.	I'm sorry to hear that. I'm glad you are feeling better.	I bet! I miss the sunsets too.	Times like that would be so awesome! I hope they come back for you
Its so quiet in my house, my husband is out of town for work. Hey I like alone time! Just gotta find some cool stuff to do! True, I'm trying but keep hearing creepy noises. Guess I need to turn the TV up louder haha	I'm sorry to hear that. I'm sorry.	I'm sure you will find some fun things to do!	oh no!! The ghosts are coming!! Yeah turn the TV up LOUD :D
I slipped and feel at work the other day. Several people saw it. I wanted to just crawl in a hole and hide afterwards.	That's scary. Did you get hurt?	That's scary. Did you get hurt?	Oh dear that's embarrassing! I hope you didn't hurt yourself.

Table 7: Examples where our model outperformed the baseline

aspects. They are:

Automated Evaluation Metrics: We report the perplexity (PPL) of the actual (gold) response as in (Li et al., 2016) Moreover, we report BLEU scores, BLEU1 (1-gram), BLEU2 (2-gram), BLEU3 (3-gram), and BLEU4 (4-gram) between the model and the gold response as shown in Table 4.

Human Evaluation Metrics: From the 2547 conversations in the test set, we get 5259 sentences and we have randomly sampled 500 sentences for evaluation purposes. We have used the Likert Scale for quality as shown in Table 5 and measured the sentences across three metrics, ”Emotional Appropriateness”, ”Relevance”, ”Readability”.
The results of the evaluation are presented as in Table 6:

We have also presented a detailed Table 7 where we illustrate nine such examples where Novel T5 out performed the baseline T5.

7 Error analysis

It is very important to understand that our model enacts the behaviour of a counsellor or a therapist. This implies that the emotional elements we try to infuse into our model and expect to be shown are ”Empathy”, ”Self-assurance”, and ”Acknowledgment”. It is clearly observed that the Novel T5 gives significant prominence to the attributes of emotion such as described above when compared to the baseline T5.

After evaluating the baseline outputs with respect to the human evaluation attributes we defined in the Evaluation Framework section, the following are the inferences that can be made about the baseline model:

- The T5 model could not correctly respond to questions being asked. It was either repeating the question or repeating the sentence “I’m so sorry to hear that”
- The T5 model lacks relevancy with respect to the input when compared to the reference sentence at times.
- The T5 Model sometimes repeats the same sentence. For example, some outputs were “I’m glad you are happy. I’m glad you are happy”, “I hate politics too. I hate politics too.”, “I’m so sorry to hear that. I’m so sorry to hear that.”
- We personally feel that some responses from the T5 model were better than the references in regards to the elements of ‘empathy’, ‘self-assurance’, and ‘acknowledgement’ present in the sentences. We are not trying to check the correctness of the model but are trying to check if the model gathered enough emotions to respond with, which is exactly what we need from a counsellor.
- At times the responses lacked relevancy. We felt that in particular input sentences where the discussion was about a favourite band or about a specific kind of flies at home, the model responded in context with the emotion but lacked the relevant information in its output.

Examples where the baseline lacked relevancy:
Example 1:

- Input: Listening to old artists like they Might Be Giants is my jam
- Output: I love old artists.
- Reference: Wow I haven’t thought of them in years. Sometimes it is good to have bands you can always go back to

Example 2:

- Input: I found out yesterday that I have a nest of yellow jacket bees in my front yard. They can be so aggressive.
- Output: I bet you are scared.
- Reference: I hate bees and wasps so much. Maybe you should call pest control?

8 Contributions of group members

We have equally contributed towards the project. The contribution is as follows:

- Subramanya: Performed data preprocessing, trained models, and writing.
- Parthiv: Performed error evaluation, trained models, and writing.
- Shashank: Performed data preprocessing, trained models, and writing.

9 Conclusion

We believe that through this project, we gained a tremendous amount of practical knowledge on the key topics that were taught in class such as Text generation models, and Evaluation Metrics. After implementing our models, we observed several key takeaways especially from the realms of Conversational AI such as how chatbots work, how text generation models can inculcate emotions and how to generalize chatbots and make them task independent as well. We believe that there is a lot of scope in future for emotion specific chat bots and if we could further continue this project, we believe that we can develop various enhanced architectures using prompt tuning where we specify the type of emotion the conversation should carry throughout the dialogue and finally develop a complete dialogue system that can respond with any emotional way the user wants it to.

References

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences*, 100(9):5497–5502.

Coulibaly, S., Roulin, A., Balog, S., Biyani, M. V., Foster, E. J., Rowan, S. J., Fiore, G. L., and Weder, C. (2014). Reinforcement of optically healable supramolecular polymers with cellulose nanocrystals. *Macromolecules*, 47(1):152–160.

Fan, A., Lewis, M., and Dauphin, Y. (2018). Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 889–898, Melbourne, Australia. Association for Computational Linguistics.

Holtzman, A., Buys, J., Du, L., Forbes, M., and Choi, Y. (2020). The curious case of neural text degeneration.

Johnson, J., Douze, M., and Jégou, H. (2017). Billion-scale similarity search with gpus. *arXiv preprint arXiv:1702.08734*.

Khanpour, H. and Caragea, C. (2018). Fine-grained emotion detection in health-related online posts. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1160–1166, Brussels, Belgium. Association for Computational Linguistics.

Lee, K., Lee, K., Shin, J., and Lee, H. (2020). Network randomization: A simple technique for generalization in deep reinforcement learning.

Li, J., Galley, M., Brockett, C., Gao, J., and Dolan, B. (2016). A diversity-promoting objective function for neural conversation models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119, San Diego, California. Association for Computational Linguistics.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, page 311–318, USA. Association for Computational Linguistics.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer.

Rashkin, H., Smith, E. M., Li, M., and Boureau, Y.-L. (2019). Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.

Ritter, A., Cherry, C., and Dolan, W. B. (2011). Data-driven response generation in social media. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 583–593, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Santhanam, S. and Shaikh, S. (2019). Emotional neural language generation grounded in situational contexts.

Serban, I. V., Sordoni, A., Bengio, Y., Courville, A., and Pineau, J. (2016). Building end-to-end dialogue systems using generative hierarchical neural network models.

Sordoni, A., Bengio, Y., Vahabi, H., Lioma, C., Simonsen, J. G., and Nie, J.-Y. (2015). A hierarchical recurrent encoder-decoder for generative context-aware query suggestion.

Subramanya N, Shashank Srigiri, V. B. P. D. (2021). Mental Health Dialogue System for Emotional Well-being.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Vinyals, O. and Le, Q. (2015). A neural conversational model.

Wang, Y.-C., Kraut, R., and Levine, J. M. (2012). To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 833–842.

Zaranis, E., Paraskevopoulos, G., Katsamanis, A., and Potamianos, A. (2021). Empbot: A t5-based empathetic chatbot focusing on sentiments.