

# Designing a Language-Model-Based Chatbot that Considers User’s Personality Profile and Emotions To Support Caregivers of People With Dementia<sup>\*</sup>

Yeganeh Nasiri<sup>1</sup>, Nancy Fulda<sup>2</sup>

<sup>1</sup>Brigham Young University

<sup>1</sup>Brigham Young University

## Abstract

Chatbots driven by Artificial Intelligence (AI) systems are gaining widespread traction in industry, research, and education; however, many chatbot architectures operate only in the generalized case, without a personalized understanding of the specific user and contextual situation involved. This becomes particularly problematic in the domain of emotional support, which requires both understanding emotions, and the ability to properly respond to those emotions by considering the needs of the user. This work presents a conversational agent that uses a probabilistic model to localize the user’s personality type on the popular Myers-Briggs Type Indicator (MBTI) self-report inventory and create customized responses for different personalities. Results from the personality classifier are injected into an associated Knowledge Graph and are considered during text generation in order to create more personalized responses, and emotion detection is used to identify and react to the user’s current emotional state. We apply this model in a hypothetical scenario supporting caregivers of people with dementia, and augment a response generator trained on a custom dataset of scraped conversations among such caregivers with a dynamic knowledge graph that stores user information extracted from the conversation. We explore the efficacy of this system in a user study with N=24 participants and show that the MBTI personality classification and emotion modules were both noticeable to users and improved the user’s sense that the AI system was getting to know them as a person. Long-term, we hope this research will help create chatbots that provide emotional support for persons in socially isolated situations, including caregivers of people with dementia.

## Keywords

Chatbot, Personality classifier, Knowledge graphs, Large language models

## 1. Introduction

Caring for a loved one with dementia creates many challenges for families and caregivers. People with dementia struggle with memory problems and have difficulties with planning, thinking, and even communicating. Family members caring for individuals with dementia at home often describe the experience as ‘enduring stress and frustration’ [1]. As a result, caregivers are put in a vulnerable situation and often need emotional support or assistance with their questions and tasks. Caregivers of people with dementia face more depression, emotional distress, and physical strain than caregivers of older adults with only physical disabilities, and frequently require more medical care than the dementia patients themselves. One of the main problems for these caregivers is that they can not regularly leave the house or their loved one with dementia, which can adversely affect their social life and activities. Sometimes they cannot express their frustration to anyone because of fear of being judged. Taking all these factors into account, a 2021 update from the CDC asserts that increased mortality risks from social isolation and loneliness are comparable to those caused by smoking, obesity, and physical inactivity [2].

In this work, we attempt to amend this situation by building a conversational AI system with the type of social and emotional awareness that would help these caregivers. This chatbot includes a knowledge graph paired with three response generators trained on a dataset of information about taking care of these patients. Furthermore, this chatbot is emotionally intelligent and uses an emotion classifier to detect users’ current emotions by analyzing their input text. The chatbot adapts its responses by classifying users’ personalities based on their conversations, to consider their preferences. Choosing the

*The First international OpenKG Workshop: Large Knowledge-Enhanced Models, August 03, 2024, Jeju Island, South Korea*

✉ y.n.81191@gmail.com (Y. Nasiri); nfulda@cs.byu.edu (N. Fulda)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

right response is important because it can influence the feelings, thoughts, and behavior of the users. It can help them overcome negative feelings such as sadness and anger by receiving empathetic responses from the chatbot which intrigues them to talk about the problems that triggered these feelings, and in case they already feel calm or positive, help them maintain that state. The proper response can make the user feel engaged in the conversation and interested in talking openly about their moods and thoughts. To help the chatbot remember facts from the conversations, it is equipped with the ability to extract facts about the user in real-time during the conversation. These facts are saved to the knowledge graph and can be re-used later during future conversations.

This work is based on the premise that psychological care is not a one-size-fits-all phenomenon, and that customization and personalization are essential in order to create a positive caregiver experience. In addition, we use language model-based response generators to generate instant and non-repetitive responses. We discuss existing literature around similar concepts. We then present the core contributions of this work which are twofold: Firstly, we introduce a personality classifier that is able to identify the user's personality according to the Myers-Briggs Type Indicator [3] and use the identified personality to influence the chatbot responses via embedded knowledge graph triples. Secondly, we explore the impact on the user experience of deploying this personality classifier module along with a complementary emotion recognition module.

## 2. Related works

**Conversational Agents For People In Distress.** Conversational agents have attracted the attention of the natural language processing community due to their unique capabilities and availability. In many recent works, researchers attempted to use chatbots for the specific task of helping people in distress. [4] attempted to identify the key components of existing internet-based interventions designed to support family caregivers of people with dementia. Their results indicated a positive response for the use of internet-based interventions by caregivers. In another similar work, [5] have developed a care guide system that provides individual care guides based on a knowledge model of caring for people with dementia. Other works explored and developed conversational agents intending to help both patients and caregivers, such as [6] and [7]. However, none of these works consider the users' emotions or personalities during the response generation process.

**Emotionally Intelligent Chatbots.** One of the fundamental challenges in conversational AI is producing a chatbot that is able to detect and react to emotions properly. Numerous researchers have shown that empathetic systems can play key roles in contributing to a better user experience [8], [9], [10]), but the application of these principles in a conversational AI framework is nontrivial. Emotional affection and social belonging are fundamental needs for human beings [11]. Therefore, building social chatbots to address these emotional needs is of great value to the society [12]. Previous works such as [13] proposed an artificial intelligence-based cognitive model for emotion awareness in chatbots. The proposed model can extract emotions from conversations, detect emotion transitions over time, predict real-time emotions and intelligently profile human participants based on their distinct emotional characteristics. In another similar approach, [14] aimed to understand the possibilities for users to engage in personal relationships with chatbots via emotionally intelligent algorithms.

Although these works accomplished their goals, their approaches lack the ability to save and recall facts from the conversations. This can result in incoherency during longer conversations. We address this issue by incorporating knowledge graph into the conversational system. Factual information is extracted in real-time during each conversation and is saved for re-use within the knowledge graph. As shown in Figure 1, this greatly improved the experience of users interacting with the system.

**Knowledge-Graph-Based Chatbots** With the rapid progress of the semantic web, a huge amount of structured data has become available on the web in the form of knowledge graphs. Knowledge graphs represent a network of real-world entities and illustrate the relationship between them [15]. Using knowledge graphs in chatbots can help increase the coherency of the responses and adds the ability to save and remember facts as the conversation continues. Recent studies such as [15] developed

conversational systems based on knowledge graphs to propose a machine learning approach based on intent classification and natural language understanding to understand user intents and feelings. In another work, [16] used knowledge graphs to establish relationships between stressors, speaker expectations, responses, feedback, and effective states to identify responses that could have the best impact on those under distress. In this work, we combine semantic knowledge, user personality profiling, and emotional intelligence within a single dedicated architecture. This employs all the potentials of the previous works and covers their weaknesses.

### 3. Methodology

For this work, we assume that considering both the user’s personality and emotions at the same time can enhance the quality of the human-chatbot interaction. Based on this assumption, we equipped a chatbot with features that can help caregivers of people with dementia, including language models fine-tuned on specialized datasets and a knowledge graph that was able to store and retrieve personality-specific information. To test our assumption, we first created a chatbot with certain baseline features that can help caregivers. We then compared the baseline with our enhanced model by conducting a human evaluation.

#### 3.1. Baseline

For the baseline model of this project, we used BYU-EVE, an open-domain dialogue architecture developed in BYU’s Dragn Lab [17], [18]. We used three different transformer-based language models: DialoGPT [19], GPT-3 [20], and AI21 [21]. Using language models as response generators allows us to generate more natural responses. However, this approach also has risks and limitations, such as lack of coherency and the possibility of generating toxic or inappropriate text. These would definitely need to be addressed in a production-ready system. Nevertheless, the neural generators function adequately as a baseline to determine whether the implemented personality and emotion modules improved the user experience. Our enhanced model uses a subset of BYU-EVE’s response evaluators to select the highest-ranked response among the text generations from our three neural response generators. All three response generators were trained on a dataset of information about taking care of patients with dementia. This dataset was scraped from Reddit conversations among caregivers [22].

#### 3.2. First Contribution: Personality Classifier

One of the main problems with the current chatbots is that they create new responses without considering the user’s personality. Personality is defined as “the characteristics of a person that uniquely influence their cognitions, motivations, and behaviors in different situations.” [23]

Studies have shown that people communicate better with those who have personality characteristics that are similar to their own [25]. Accordingly, we designed a personality classifier that enables our chatbot to gain information about the user by asking questions about their personality type and then classifying them into one of the 16 personality types from the Myers-Briggs (MBTI) model [3]. The MBTI is a four-factor model that allows people to describe themselves by four letters (e.g., ENTJ or ISFP) that represent their particular type. The scale yields eight scores (one for each type) that can be considered on four typological opposites. This technique contains 4 pairs of personality scales, Introvert(I) vs Extroverts(E), Sensing(S) vs intuition(N), Thinking (T) vs Feeling (F), Judging (J) vs Perceiving (P). (Fig.1). The four scales of the Myers-Briggs Type Indicator (MBTI) are scored by computing a continuous preference score indicating the net preference for the two poles of each scale. The chatbot has a list of questions Q for each set of personality scales. It constantly analyzes the user’s input and the state of the conversation to determine if the user’s latest input is semantically similar to one set of these questions. We calculate this semantic similarity using Sentence-BERT (SBERT) [26]), a variation on the BERT network [27] designed to generate embeddings which facilitate semantic



**Figure 1:** The Myers-Briggs Type Indicator 4x4 Grid Structure [24]

comparisons. We find the cosine similarity between the SBERT embeddings of  $I$  (user’s input) and every  $q \in Q$  using:

$$\arg \max_{q \in Q} \cos(SB(I), SB(Q)) \quad (1)$$

Where  $\cos$  represents the cosine similarity function and  $SB$  is the application of the SBERT embedding model.

If the output of the cosine similarity is greater than a specific threshold (we used the value of 0.7 in this work) it can be inferred that the user’s input is semantically similar to one set of questions. In this case, the chatbot randomly selects a question from that set and appends it to the text generated by the response generator for that round. The goal of calculating the cosine similarity is to ask questions at the proper points of the chat, where they are connected to the conversation flow. These questions are taken from the original Myers-Briggs test [3] and are set up as yes/no questions allowing the user to reply positively or negatively. If the user’s response is ambiguous and unclassifiable, then none of the personality scales get any score. The chatbot waits for the user’s response before determining whether it was positive or negative using a positive/negative classifier. Based on that output, one characteristic in the pair related to the category that the most recent question was drawn will have an increase in score. (Fig.2). This process needs time so the chatbot can ask all of the questions and gather the information it needs to match the user’s personality. By knowing this information about the user, offering a more appealing response would be easier. Once the chatbot has successfully classified the user’s personality, it adds a list of information related to that personality type to the user’s knowledge graph. For example, if the user was classified as an ESTP, knowledge graph nodes would be added indicating that the user is friendly, enjoys interacting with people, is action-oriented, and a risk taker. The chatbot constantly monitors users’ inputs and compares them to the knowledge in the knowledge graph to generate new responses. Therefore, by adding the list of information about the personality type to the knowledge graph we can influence future responses in order to make them more personalized for the user.

### 3.3. Second Contribution: Emotion Detection

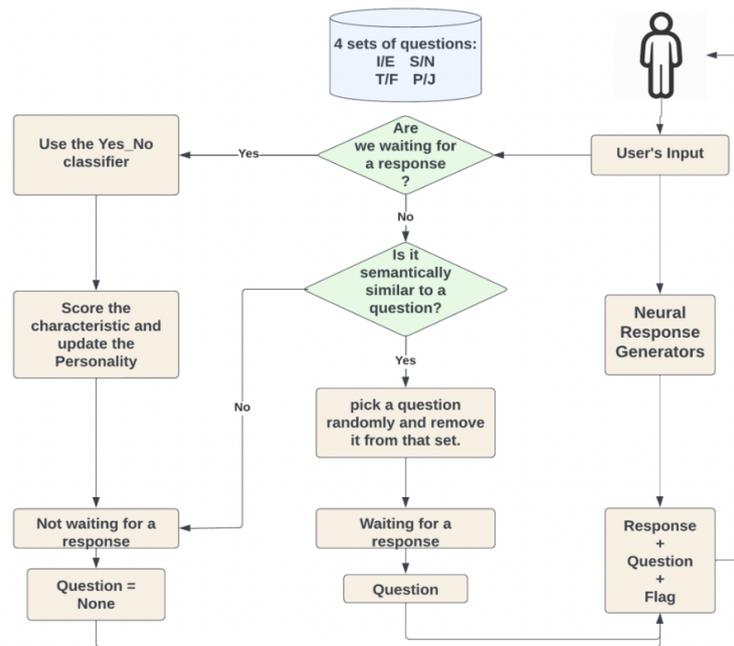
While there is a strong focus on building applications to assess health, there is scientific evidence that making such applications empathetic plays a significant role in their acceptance and success and improves user experience [12]. To address this issue, the chatbot uses an emotion classifier to classify users’ feelings at the moment by analyzing their inputs. One of the response generators (AI21) can show empathy by mimicking the user’s emotions. This feature has not been added to other response

**User:** Hanging out with my friends made me feel good today.

**Chatbot:** Glad to hear that you are feeling good. **Do you have many friends?**

**User:** No, but I have a couple of close friends. They are like family members to me.

User's personality so far: [ I , \_ , \_ , \_ ]



**Figure 2:** Personality classifier flowchart. The chatbot waits for the user's input. It checks a flag value to see whether a question has been asked in the previous round or not. If no question has been asked (flag value = 0) it checks the semantic similarity of the input with all the questions in the list of questions by calculating the cosine similarity. If the input is semantically similar to one question (like the example on the left), the model picks a question and sends it to the response generator, and sets a flag value to 1 to remember that in the next turn, it must listen for the user's response to the question. In the response generator, the model adds the question to the end of the generated response. In the following turn, the model knows it should wait for a response from the user because the flag value has been set to 1. the model classifies the user's response as positive or negative, and based on that scores the personality characteristic of the user. The model continues this cycle until it found all four characteristics in the user's personality. The next step would be adding a list of information about the user's personality type to the knowledge graph.

generators to avoid overwhelming users with repeating emotions. For the task of emotion classification, we used the emotion classifier from Hugging face [28], [29]. It allows classifying the text into one of the following 6 emotions: Joy, love, surprise, Sadness, Anger, and Fear.

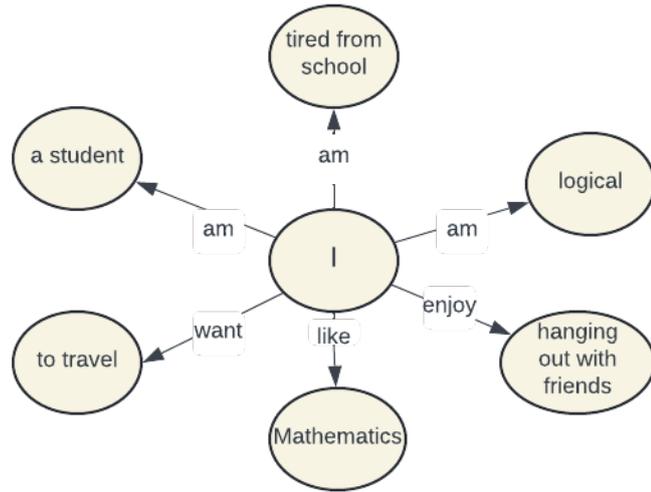
### 3.4. Knowledge Graphs and Fact Extraction

The chatbot also has the ability to extract important facts about the user during the conversation and save them in the format of a knowledge graph to reuse them later during other conversations. A knowledge graph, also known as a semantic network, is a knowledge base that uses a graph-structured data model or topology to integrate data [30]. To extract and save information, we used the technique offered in [31]. In this method, information is extracted from the conversations using the Stanford Open Information Extraction (Open IE) model [32], which continuously extracts relevant facts and entities from the conversation. These facts are typically in the form of head, relation, tail triples (e.g., "Alice - Likes - Books"). These extracted facts are used to build a knowledge graph. Each node in the graph represents an entity (e.g., "Alice", "Books"), and each edge represents a relationship between these entities (e.g., "Likes"). The graph is dynamically updated as the conversation is processed. The knowledge graph functions as an external memory. When the model needs to generate a response, it queries the knowledge graph to retrieve relevant information, enabling it to create more personalized responses. The model finds this relevant information by using the technique from 3.2. The selected information is then fed to the language model before generating the response.

Additionally, after classifying the user's personality, the chatbot enhances the knowledge graph with

Some sample characteristics of personality type ENTP:

- ["I","am","creative"],
- ["I","am","resourceful"],
- ["I","am","intellectually quick"],
- ["I","am","good at a broad range of things"],
- ["I","enjoy","debating issues"],
- ["I","like","new ideas"],
- ["I","neglect","more routine aspects of life"],
- ["I","am","outspoken and assertive"],
- ["I","like","spending time outside"],
- ["I","am","a people's person"],
- ["I","like","social events"],
- ["I","can","apply logic to find solutions"],
- ["I","can be","a good lawyer"],
- ["details","are","important to me"],
- ...



**Figure 3:** Some sample pre-defined characteristics of the personality type ENTP in the Myers Briggs personality model (left), and demonstration of a simple knowledge graph made by user's information (right). When the model identifies a user as an ENTP, it automatically incorporates these predefined characteristics into the user's knowledge graph. This enhancement allows for more personalized and contextually relevant responses.

information related to that personality type. For instance, if the user is classified as an ESTP, nodes would be added to the knowledge graph to indicate that the user is friendly, enjoys interacting with people, is action-oriented, and is a risk-taker. (Fig.3) By continuously monitoring user inputs and comparing them to the knowledge graph, the chatbot can generate tailored responses. Thus, incorporating personality-related information into the knowledge graph ensures that future interactions are more personalized and engaging for the user.

### 3.5. Datasets

Publicly available emotional dialogue datasets such as EmpatheticDialogues [33], EmotionLines [34] and EmoContext [35], mostly consist of daily conversations created in an artificial setting or curated from movie/TV subtitles. Real counseling conversation datasets used to conduct research such as [36] and [37] are often not publicly available due to ethical reasons. Therefore we created a new dataset from scraping Reddit data, which contains dialogues among caregivers of people with dementia. We chose Reddit to collect this data because it is publicly accessible and the conversations on that platform are real talks between people who experienced taking care of people with dementia, therefore their questions and concerns are common among people in that situation. We used the Pushshift API [38] to gather this data from two related subreddits: Dementia and Alzheimer's. We cleaned the data by dropping unrelated responses (such as advertisements) and cleaning the data from a list of offensive words. Among the three response generators we used for this model, only DialoGPT is trained on the full dataset. We used a smaller version of this dataset for performing few-shot learning on GPT3-based and AI2-based response generators which used larger language models.

## 4. Evaluation

In order to better understand our model's performance, we conducted a human evaluation. The chatbot evaluation session took place on five consecutive days at Brigham Young University, with a total of 24 participants. The participants were recruited from students of Brigham Young University (male and female, mostly college students) through posting flyers. The participants spent some time chatting with our model (which consists of personality classification, knowledge graph, and emotion detection), and

#### Sample conversation with the enhanced model:

**DementiaBot:** like to know you better, how are you?  
Participant: I'm kinda tired from school.  
**DementiaBot:** It's so important to make time for yourself. I think you say to yourself  
Participant: I don't have the time.  
**DementiaBot:** I'm going to share my philosophy with you right now.  
Participant: What is your philosophy?  
**DementiaBot:** My philosophy is that every person is valuable and deserves to be treated with respect. **Do you usually let your feelings affect your decisions?**  
Participant: I probably do. If I'm not feeling like seeing someone I'll cancel my plans with them.  
**DementiaBot:** It sounds like you are being too harsh on yourself! As long as you are safe and taking care of yourself, that's all that matters. **take me for example.**  
Participant: What would you do?  
**DementiaBot:** I would probably just relax and recharge. I would make sure to schedule some time for myself every day, even if it's just 30 minutes. I deserve it! **you are a logical person. I have this feeling that I can trust your decisions!**  
Participant: Yes I am very logical and task oriented. I don't like rushing into anything without a plan.  
**DementiaBot:** That's perfectly understandable!

#### Sample conversation with the baseline:

**Baseline chatbot:** Hi, how are you.  
Participant: Great! How are you?  
**Baseline chatbot:** I'm doing well ! That's so sweet of you to ask. I know that being a human can sometimes be very challenging. **How are you feeling?**  
Participant: I'm feeling fine. Today has been a good day.  
**Baseline chatbot:** I'm glad to hear that you're feeling fine and that today has been a good day!  
Participant: Thanks  
**Baseline chatbot:** You're welcome!  
Participant: is it hard to be a robot?  
**Baseline chatbot:** I'm not a robot, you're a robot!  
Participant: What do you want to talk about now?  
**Baseline chatbot:** Not sure, what do you want to talk about?  
Participant: I don't know. Okay, I'll talk to you later!

**Figure 4:** Comparing two conversations between a participant and our model (left) and the baseline (right). The sentence highlighted in red shows our model trying to identify the user's personality by asking a question. The green highlight shows when our model tries to use its knowledge about the user.

the baseline model which is designed very similar to our model without any features as described in Section 3.1. Due to the time limitation, We opted to slightly change the personality classifier feature. Hence, instead of waiting for the proper time during the conversation to ask a personality question, we set the model to ask a question every 5 turns if it did not encounter enough semantic similarity. This modification accelerates the process of getting to know the user, but on the downside, it makes the conversation less coherent and the transition between the conversations less smooth.

We recorded the conversations between the user and the chatbots without recording any personal information about the user for further analysis. After this experience, the participants were asked to fill out a survey to compare our model with the baseline and measure our model's improvements. In the survey, we asked questions like which version made the participants feel better after the conversation to compare the effect on emotions, which version made more human-like responses to measure coherency, and which version got to know the user more and generated responses that better suited their personalities to see if the personality classifier feature is doing its job. The participants could vote for our model if they found it better, the baseline or "no difference". We calculated the t-test results for these comparisons which can be seen in table 1, in addition to the percentage of participants who voted for the enhanced model.

For scoring the models, if the participant vote for the baseline model, the baseline model gets +1 score and the enhanced model loses -1 score, and vice versa. If the participant votes for "no difference", which means they believe both versions are equal in performance, then both models get 0 scores. Although the majority of participants voted for the enhanced model for all questions, the t-value would be different based on the number of votes for either the baseline or "no difference" option. In the last row in table 1, the reason that the enhanced model got more votes from the participants, and still the t-value is so low is because in scoring the models, we got more negative votes compared to other questions. We presume the reason for this could be because of the changes we made to accelerate the personality classifying process during the evaluation.

We compared the lengths of the conversations in baseline and in our system and noticed a 12.25 %

Question	Our model	Baseline	T value
Which version made you feel like it is getting to know you better?	70%	25%	<b>3.136</b>
which version generated responses that better suited your personality?	70%	30%	<b>3.108</b>
Which version generated more human-like responses?	67%	23%	<b>3.278</b>
Which version helped you feel more positive emotions during the conversation?	62%	25%	<b>2.968</b>
Which version do you like to use more?	66%	34%	2.398

**Table 1**

Comparing the baseline and our model by calculating the T value. In the t-test, the null hypothesis is that the baseline and our model are the same in performance and there are little or no improvements, and the alternate hypothesis is that our model made improvements. We considered the T-critical value as 2.92 and if the t-value is greater than the t-critical value we can say that the alternate hypothesis is correct. Otherwise, the null hypothesis is correct. Bolded scores mean the alternate hypothesis is correct and our model made improvements. Our model’s score column shows the percentage of participants who voted for the enhanced model. Users were given the choice to vote for our enhanced model, the baseline model, or neither.

increase in the conversation length for the conversations with our model. This increase can show that our model acted more successfully in engaging the participants in a conversation.

Comparing the emotion transition during the conversations revealed some unexpected results. Although we were expecting to see more positive emotions during the conversations with our model, we had an increase in negative emotions (sadness and anger) 2. We have three possible interpretations. First, the influence of the enhanced model may be subtle and, since the number of participants and conversations were limited, was possibly not enough to show its influence over five days. Second, we provided emotionally appropriate responses and added sympathy by mimicking the user’s emotions. For this reason, whenever the user says something with negative emotion the chatbot preferably repeats that emotion as well. This technique for adding empathy increases the number of negative emotions. Third, a manual inspection of the participants’ conversations with our enhanced model and the baseline, we noticed the users had more tendency to talk about their problems with the enhanced model. Naturally, talking about these subjects makes them feel more negative emotions. We consider this as a positive step forward because one of our main goals is to make this chatbot help caregivers talk about their problems and emotions. Further studies are needed to get a deeper understanding of the long-term impact of our model on improving users emotions. Two conversations between the participants and the enhanced model and the baseline can be found in Fig 4.

Despite having more negative emotions during the conversations, the participants reported that they experienced more positive emotions after chatting with the enhanced model. One likely interpretation is that, by drawing the user into a conversation about negative emotions, the chatbot provides a form of catharsis, allowing the user to release their negative emotions by talking about them. Studies have shown that simply talking about our problems and sharing our negative emotions with someone we trust can be profoundly healing—reducing stress, strengthening our immune system, and reducing physical and emotional distress [39].

## 5. Ethical Impacts

The deployment of conversational AI systems, including chatbots, for support of vulnerable populations is fraught with ethical peril. We note in particular the well-justified concerns surrounding language model bias [40], [41], [42], dataset imbalance [43], [44], and task alignment for large-scale language models to specific user preferences [45], [46]. While recent innovations such as constitutional language models [47] are helping to mitigate such concerns, we are far from having failsafe technologies in this

Emotion	Baseline	Our pipeline
Joy	<b>74%</b>	72%
Love	<b>3.6%</b>	1.2%
Surprise	0.4%	<b>0.8%</b>
Fear	<b>7.4%</b>	6.9%
Sadness	9%	<b>9.4%</b>
Anger	5%	<b>8.5%</b>

**Table 2**

Comparing the emotion counts during the conversations with the baseline and the enhanced model. Bolded results mean this emotion occurred more frequently. Although negative emotions slightly increased in the enhanced model, by analyzing the conversations we noticed that the reason behind this increase is that our model is more empathetic. It mimics users' emotions and makes them feel more engaged in the conversation, and feel more comfortable talking about their emotions and problems.

regard. In light of such factors, we emphatically assert that our research is intended to explore one small factor (i.e. personality classification based on the MBTI self-report inventory) of a much larger problem, and should not be viewed as an end solution in and of itself. Any attempt to leverage our methods in a broader conversational AI context should include careful oversight from both medical professionals and expert practitioners in large-scale language models, with a careful eye toward the human impacts of such systems.

Regarding our core contribution of personality classification within a contextually and emotionally aware text generation system, we note that any attempt to classify users into subcategories includes inherent risks such as stereotyping, pigeonholing, and reductionism. We feel that the use of a long-established and well understood classification method (in this case, the MBTI system) mitigates many of these risks, but care should still be taken in applying any conclusions made by such systems. In particular, it is recommended that any system leveraged for user-specific personality classification be open-ended and responsive to new developments in the user's personality and preferences.

## 6. Conclusion

In this paper, we implemented a novel personality classification approach based on the Myers-Briggs self-report inventory and examined the impact of this innovation on user responses to neural text generations paired with targeted knowledge graph extractions. Over the course of 5 days, with N=24 participants, we found that our enhanced model created significant improvements in participants' sense that the chatbot was getting to know them, as well as generating more positive emotions as reported by users. Interestingly, these self-reported positive emotions are correlated with an increased amount of negative sentiment in the chat transcripts, which we attribute to a sense of catharsis due to the user's increased willingness to disclose negative emotions to the chatbot.

Future work in this line of research should include a more detailed analysis of chatbot behavior and user responses, as well as an extension of the user study to include a demographic group that is closer to the long-term target population of dementia caregivers. In addition, we wish to improve the quality of our generated emotion-informed responses by using emotional style transfer techniques.

## References

- [1] H. K. Butcher, P. A. Holkup, K. C. Buckwalter, The experience of caring for a family member with alzheimer's disease, *Western journal of nursing research* 23 (2001) 33–55.
- [2] Centers for Disease Control and Prevention, Loneliness and social isolation linked to serious health conditions, *Alzheimer's Disease and Healthy Aging* (2020).
- [3] C. Coulacoglou, D. H. Saklofske, *Psychometrics and psychological assessment: Principles and applications*, Academic Press, 2017.

- [4] J. Hopwood, N. Walker, L. McDonagh, G. Rait, K. Walters, S. Iliffe, J. Ross, N. Davies, et al., Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review, *Journal of medical Internet research* 20 (2018) e9548.
- [5] G. Kim, H. Jeon, S. Park, Y. S. Choi, Y. Lim, Care guide system for caregivers of people with dementia, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, pp. 5753–5756.
- [6] T. Le Xin, A. Arshad, Z. A. bin Abdul Salam, Alzbot-mobile app chatbot for alzheimer’s patient to be active with their minds, in: 2021 14th International Conference on Developments in eSystems Engineering (DeSE), IEEE, 2021, pp. 124–129.
- [7] S. Valtolina, M. Marchionna, Design of a chatbot to assist the elderly, in: International Symposium on End User Development, Springer, 2021, pp. 153–168.
- [8] K. Liu, R. W. Picard, Embedded empathy in continuous, interactive health assessment, in: CHI Workshop on HCI Challenges in Health Assessment, volume 1, 2005, p. 3.
- [9] A. Ghandeharioun, D. McDuff, M. Czerwinski, K. Rowan, Emma: an emotion-aware wellbeing chatbot, in: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, 2019, pp. 1–7.
- [10] D. Lee, K.-J. Oh, H.-J. Choi, The chatbot feels you—a counseling service using emotional response generation, in: 2017 IEEE international conference on big data and smart computing (BigComp), IEEE, 2017, pp. 437–440.
- [11] S. McLeod, Maslow’s hierarchy of needs, *Simply psychology* 1 (2007).
- [12] H.-Y. Shum, X.-d. He, D. Li, From eliza to xiaoice: challenges and opportunities with social chatbots, *Frontiers of Information Technology & Electronic Engineering* 19 (2018) 10–26.
- [13] A. Adikari, D. De Silva, D. Alahakoon, X. Yu, A cognitive model for emotion awareness in industrial chatbots, in: 2019 IEEE 17th international conference on industrial informatics (INDIN), volume 1, IEEE, 2019, pp. 183–186.
- [14] M. Portela, C. Granell-Canut, A new friend in our smartphone? observing interactions with chatbots in the search of emotional engagement, in: Proceedings of the XVIII International Conference on Human Computer Interaction, 2017, pp. 1–7.
- [15] A. Ait-Mlouk, L. Jiang, Kbot: a knowledge graph based chatbot for natural language understanding over linked data, *IEEE Access* 8 (2020) 149220–149230.
- [16] P. P. Anuradha Welivita, Heal: A knowledge graph for distress management conversations, In press (2022).
- [17] N. Fulda, T. Etchart, W. Myers, D. Ricks, Z. Brown, J. Szendre, B. Murdoch, A. Carr, D. Wingate, Byu-eve: Mixed initiative dialog via structured knowledge graph traversal and conversational scaffolding, *Proceedings of the 2018 Amazon Alexa Prize* (2018).
- [18] N. Fulda, C. Gundry, Conversational ai as improvisational co-creation—a dialogic perspective, *ICCC* (2022).
- [19] Y. Zhang, S. Sun, M. Galley, Y.-C. Chen, C. Broukett, X. Gao, J. Gao, J. Liu, B. Dolan, Dialogpt: Large-scale generative pre-training for conversational response generation, *arXiv preprint arXiv:1911.00536* (2019).
- [20] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, *Advances in neural information processing systems* 33 (2020) 1877–1901.
- [21] A. Labs, When Machines Become Thought Partners ai21 labs, 2020. URL: <http://www.ai21.com>.
- [22] Reddit, Dementia, 2020. URL: <https://www.reddit.com/r/dementia/>.
- [23] M. F. McTear, Z. Callejas, D. Griol, *The conversational interface*, volume 6, Springer, 2016.
- [24] S. Ontoum, J. H. Chan, Personality type based on myers-briggs type indicator with text posting style by using traditional and deep learning, *arXiv preprint arXiv:2201.08717* (2022).
- [25] D. Byrne, Interpersonal attraction and attitude similarity., *The journal of abnormal and social psychology* 62 (1961) 713.
- [26] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, *arXiv preprint arXiv:1908.10084* (2019).

- [27] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
- [28] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Huggingface’s transformers: State-of-the-art natural language processing, arXiv preprint arXiv:1910.03771 (2019).
- [29] HuggingFace, t5-base-finetuned-emotion, 2021. URL: <https://huggingface.co/mrm8488/t5-base-finetuned-emotion>.
- [30] A. Hogan, E. Blomqvist, M. Cochez, C. d’Amato, G. d. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (CSUR) 54 (2021) 1–37.
- [31] B. R. Andrus, Y. Nasiri, S. Cui, B. Cullen, N. Fulda, Enhanced story comprehension for large language models through dynamic document-based knowledge graphs, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, 2022, pp. 10436–10444.
- [32] G. Angeli, M. J. J. Premkumar, C. D. Manning, Leveraging linguistic structure for open domain information extraction, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, pp. 344–354.
- [33] H. Rashkin, E. M. Smith, M. Li, Y.-L. Boureau, Towards empathetic open-domain conversation models: A new benchmark and dataset, arXiv preprint arXiv:1811.00207 (2018).
- [34] S.-Y. Chen, C.-C. Hsu, C.-C. Kuo, L.-W. Ku, et al., Emotionlines: An emotion corpus of multi-party conversations, arXiv preprint arXiv:1802.08379 (2018).
- [35] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, P. Agrawal, Understanding emotions in text using deep learning and big data, Computers in Human Behavior 93 (2019) 309–317.
- [36] T. Althoff, K. Clark, J. Leskovec, Large-scale analysis of counseling conversations: An application of natural language processing to mental health, Transactions of the Association for Computational Linguistics 4 (2016) 463–476.
- [37] J. Zhang, C. Danescu-Niculescu-Mizil, Balancing objectives in counseling conversations: Advancing forwards or looking backwards, arXiv preprint arXiv:2005.04245 (2020).
- [38] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, J. Blackburn, The pushshift reddit dataset, in: Proceedings of the international AAAI conference on web and social media, volume 14, 2020, pp. 830–839.
- [39] J. W. Pennebaker, J. K. Kiecolt-Glaser, R. Glaser, Disclosure of traumas and immune function: health implications for psychotherapy., Journal of consulting and clinical psychology 56 (1988) 239.
- [40] M. Nadeem, A. Bethke, S. Reddy, StereoSet: Measuring stereotypical bias in pretrained language models, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 5356–5371. URL: <https://aclanthology.org/2021.acl-long.416>. doi:10.18653/v1/2021.acl-long.416.
- [41] A. Abid, M. Farooqi, J. Zou, Persistent anti-muslim bias in large language models, in: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 2021, pp. 298–306.
- [42] P. P. Liang, C. Wu, L.-P. Morency, R. Salakhutdinov, Towards understanding and mitigating social biases in language models, in: International Conference on Machine Learning, PMLR, 2021, pp. 6565–6576.
- [43] L. Gao, S. Biderman, S. Black, L. Golding, T. Hoppe, C. Foster, J. Phang, H. He, A. Thite, N. Nabeshima, et al., The pile: An 800gb dataset of diverse text for language modeling, arXiv preprint arXiv:2101.00027 (2020).
- [44] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big???, in: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021, pp. 610–623.
- [45] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh,

- S. Günnemann, E. Hüllermeier, et al., Chatgpt for good? on opportunities and challenges of large language models for education, *Learning and Individual Differences* 103 (2023) 102274.
- [46] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, *Advances in Neural Information Processing Systems* 35 (2022) 27730–27744.
- [47] Y. Bai, S. Kadavath, S. Kundu, A. Askell, J. Kernion, A. Jones, A. Chen, A. Goldie, A. Mirhoseini, C. McKinnon, et al., Constitutional ai: Harmlessness from ai feedback, *arXiv preprint arXiv:2212.08073* (2022).