# Personalized Chatbots with Emotion-Aware Responses
## Project Proposal

**Mayank Mayank (301563549)**
mma266@sfu.ca

**Aryaman Bahuguna (301563474)**
aba214@sfu.ca

## Abstract

Human-computer interactions have gained significant traction in this era, with the ability of AI-driven chatbots to tailor their responses based on user input. This makes them essential for enhancing user satisfaction. Our project aims to develop a simple chatbot that adapts its responses based on the user's emotional tone detected through text input. This system can help automate tasks across various sectors such as customer service, education, therapy, etc., where empathetic interactions are crucial.

Our approach involves a two-stage pipeline:

1. Emotion Detection
2. Emotiona Sensitive Response Generation

To classify the user input into one of several emotional categories, a transformer-based model (e.g., BERT) can be trained on emotion-labeled datasets such as *GoEmotions*. Next, a generative language model (e.g., GPT-2) can be trained to produce responses that align with the detected emotion. The model's evaluation will be conducted using both quantitative metrics (BLEU, ROUGE, sentiment alignment scores) and qualitative assessments (human feedback).

The project utilizes open-source NLP frameworks (e.g., Hugging Face Transformers, PyTorch, TensorFlow) and freely available computing resources (Google Colab, Kaggle) to ensure easy and stable usage. Additionally, an interactive interface setup using Streamlit or Gradio deployment will allow users to engage with the chatbot and assess its efficiency.

The project effectively promotes and contributes to the novel area of emotionally aware AI by combining both sentiment analysis and conditional response generation. The ability to generate empathetic and contextually aware responses makes our chatbot a useful tool for emotional intelligence applications, thus providing the potential to improve the user experience across multiple domains.

## 1 Introduction and Motivation

Recent progress in Natural Language Processing (NLP) has produced chatbots capable of generating coherent responses. However, most current conversational agents do not adapt to the user's emotional state. Emotion-aware systems are particularly relevant in customer service, healthcare (e.g., therapy bots), and education (Shum et al., 2018). By detecting and responding to user emotions, a chatbot can achieve more empathetic and context-aware interactions, thus improving user satisfaction and engagement.

**Problem Statement:** We aim to build a chatbot that, given a user query, first identifies the user's emotional tone and then generates a response reflecting sensitivity to that emotion.

**Why This Approach?** Our goal is to combine robust emotion detection (using a classification model) with an effective generative pipeline for personalized, emotionally aligned replies. This project is both technically feasible and highly impactful, providing valuable real-world applications in mental health support, user engagement, and general conversational AI systems.

## 2 Related Work

Transformer-based language models have yielded state-of-the-art results in text classification and generation tasks. Devlin et al. (2019) introduced BERT, which was refined into RoBERTa for improved performance in various NLP benchmarks (Liu et al., 2019). For conversational text generation, GPT-2 (Radford et al., 2019) and T5 (Raffel et al., 2020) have shown strong performance.

Several datasets, such as GoEmotions (Demszky et al., 2020), support fine-grained emotion recognition. Our project builds on this prior work by integrating a dedicated emotion classifier with a conditional text generation system. We will compare different pretrained variants (e.g., BERT vs. RoBERTa for detection and GPT-2 vs. T5 for generation) to identify the most effective pipeline for emotionally aware dialogue.

## 3 Approach

**Emotion Detection:** We will fine-tune a RoBERTa-based classifier (and optionally compare with BERT) on the GoEmotions dataset, which contains 58k Reddit comments with 27 emotion labels (Demszky et al., 2020). We will measure performance using F1 scores across these emotion labels.

**Conditional Response Generation:** We plan to adapt GPT-2 (or T5 as a variant) to generate text conditioned on the detected emotion. Specifically, we will append a special token indicating the emotion category to the dialogue history before feeding it into the generative model. This ensures the response is tailored not just to the context but also to the user's emotional state.

**Model Variants:** We will systematically compare:

- BERT vs. RoBERTa for emotion classification
- GPT-2 vs. T5 for response generation

## 4 Experimental Setup

**Datasets:** The GoEmotions dataset (Demszky et al., 2020) will be our primary source for emotion annotations. For conversational contexts, we may leverage DailyDialog or similar publicly available dialogue corpora.

**Implementation:** We will use the Hugging Face Transformers library (Wolf et al., 2020). The workflow involves:

1. Fine-tune the emotion classifier.
2. Use that classifier to label user queries in a conversation.
3. Inject the predicted emotion label into the generative model for conditional response generation.

**Evaluation:** We will measure:

- **Emotion Classification**: Accuracy, F1 score on GoEmotions.
- **Response Quality**: BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004).
- **Emotional Appropriateness**: Human feedback via short user studies or crowd-sourced annotations (asking participants to rate how well the response matches the intended emotion).

**Baselines:** To demonstrate effectiveness, our baseline systems will include:

- A purely context-based chatbot without emotion conditioning.
- Variants of the pipeline where we switch emotion classifiers (BERT vs. RoBERTa) and generators (GPT-2 vs. T5).

## 5 Timeline and Work Breakdown

**By Milestone (March 19):**

- Collect and preprocess data (GoEmotions + optional dialogue dataset).
- Fine-tune RoBERTa for emotion detection and validate initial performance.
- Begin integrating the emotion detection output into GPT-2 for preliminary testing.

**Post-Milestone to Final:**

- Finish pipeline integration with GPT-2 or T5.
- Thoroughly evaluate the models on automated metrics and gather human feedback.
- Compare all baselines and variants.
- Finalize write-up, including error analysis and discussion.

**Team Member Responsibilities.**

- **Mayank Mayank:** Data preprocessing, emotion detection experiments and Human feedback collection, analysis.
- **Aryaman Bahuguna:** Conditional generative model fine-tuning, evaluation scripts and overall system integration.

## References


Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Polisi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 4040–4054.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 4171–4186.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*, pages 311–318.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Heung-Yeung Shum, Xiaodong He, and Di Li. 2018. From ELIZA to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1):10–26.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clément Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Mo Funtowicz, and Jamie Brew. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP)*, pages 38–45.