# COMP5313WA 2024 Chatbot Contest

EXECUTIVE SUMMARY of Group No. 3

Students Names:

1212312 Murudkar, Priyamvada pmurudk1@lakeheadu.ca

1222198 Nagarajan Thiruppathi, Thanesh tnagaraj@lakeheadu.ca

1207090 Nagpal(Nagpal), Gurveen gnagpal1@lakeheadu.ca

1214653 Neen, Jumaina Rahman jneen@lakeheadu.ca

1216604 Pala, Gowthami Sai Sree Pooja gpala@lakeheadu.ca

1217458 Parikh, Vaibhav vparikh5@lakeheadu.ca

1228123 Patadia, Chintan Ashwin cpatadia@lakeheadu.ca

1215654 Patel, Foram fpatel10@lakeheadu.ca

# Overview - The Quick Pitch

This is a brief section that describes what your comparison task in designing the selected chatbots.

The main focus of the comparison task is evaluating the effectiveness of the chosen chatbots in terms of first aid recommendations. In particular, we evaluate the ability of ChatGPT and a Neural Network with BioBERT to comprehend user inquiries, interpret medical data, and provide precise and useful advice in emergency scenarios. In the end, we hope to determine which chatbot model provides better performance and usability for the specific application of first aid recommendations by comparing the advantages and disadvantages of each method.

# The Problem

Here is where you describe the problem that you are solving.

* What are the conversational chatbots that you are focusing on?
* Review some existing solutions that you have seen for such conversational chatbots?
* What are the pros and cons of each existing solution?

Ther are some already existing solutions for conversational chatbots other than that we are focusing on they are:

* **Seq2Seq model:**

Theis model is mostly based on Recurrent Neural Network (RNN) or Transformer, which are widely used for generating sequence to sequence output by encoding the input sequence which make them more suitable for dialogue generation and human like conversation.

* **Transformer based model:**

Although we have model like GPT and Bert which originated from transformer there are also other model which is based on transformers like T5 which is good for summarization of text and Transformer-XL which improves the long-range dependencies and generate coherent text.

* **LSTM (Long Short-Term Memory):**

LSTM is also a type of RNN which is a improved version of RNN which can store the memory of the data for a period of time, it also determines with is important enough to be stored and which should be deleted. It is a huge leap in sentence generation and holding conversation with a human.

* Covi-Bio Bert:

This model is a lot similar to our model which is Bio-Bert but Covi-Bio Bert is specifically pre-trained on data and queries related to covid and covid related information to provide information with much accuracy.

The pros and cons of existing solution?

* **Seq2Seq model:**

**Pros:**

Flexible architecture: it can handle various type of conversational tasks, including dialog generation and text summarization

Long-range dependency: Recurrent Seq2Seq model like those based on LSTM can capture long range dependencies in sequential data, which is crucial for generating coherent responses in conversation.

**Cons:**

Exposed to bias: They are exposed to bias, where they can generate unrealistic responses due to the unbalanced or bias in the dataset that it has been trained on.

Limited context understanding: Traditionally, the model may struggle with understanding context contextual information, leading to generate irrelevant responses.

* **Transformer:**

**Pros:**

Transformer based model utilize self-attention mechanism, enables parallel processing of input sequence and effective training on Moden hardware

Scalability: They are scalable to a large datasets and vocabulary size, making them suitable for training on extensive conversational corpora.

**Cons:**

Limited interpretability: They may lack in interpretability, making it challenging to understand the underlying reason behind their generated responses.

Data inefficiency: It requires a large amount of data and computational resources to train on.

* **LSTM**

**Pros:**

**Handling Long Sequences:** LSTMs are well-suited for processing sequences of data with long-range dependencies. They can capture information from earlier time steps and remember it for a more extended period, making them effective for tasks like natural language processing (NLP) and time series analysis.

**Avoiding Vanishing Gradient Problem:**  LSTMs address the vanishing gradient problem, which is a common issue in training deep networks, particularly RNNs. The architecture of LSTMs includes gating mechanisms (such as the forget gate) that allow them to control the flow of information and gradients through the network, preventing the gradients from becoming too small during training.

**Cons:**

**Overfitting:** Like other deep learning models, LSTMs are susceptible to overfitting when there is insufficient training data. Regularization techniques like dropout can help mitigate this issue.

**Long Training Times:** Training deep LSTM models on large datasets can be time-consuming and may require powerful hardware, such as GPUs or TPUs.

We are focusing on the following chatbots:

**Neural Net using BioBERT:** BioBERT is a customized version of BERT (Bidirectional Encoder Representations from Transformers) that is used in neural nets for biomedical text mining applications. When combined with a neural network, it can comprehend and react to questions on medical and healthcare subjects, with an emphasis on biological and biomedical ideas, in the context of a conversational chatbot. This enables more precise and specialized answers in the healthcare field.

**ChatGPT:** Built on the GPT (Generative Pre-trained Transformer) architecture, ChatGPT is an AI model for conversational applications created by OpenAI. It is intended to comprehend natural language text responses and provide responses that are human-like. ChatGPT has the ability to have many kinds of discussions, exchange information, respond to queries, and even help with a variety of activities.

* Who is the target application?

The purpose of these conversational chatbots is to offer advice on first aid. This entails providing direction and support to people in dire circumstances where they could need emergency medical help.

**ChatGPT for First Aid:** ChatGPT can help users by giving detailed instructions on how to provide first aid in a variety of situations. Before expert medical assistance arrives, it can assist users in assessing the situation, identifying injuries or symptoms, and suggesting the best course of action.

**Using BioBERT with Neural Nets for First Aid:** The system can comprehend and handle medical queries related to first aid with an emphasis on biomedical terminology and knowledge by utilizing BioBERT within a conversational chatbot framework. This makes it possible to provide recommendations that are more precise and thorough and that are suited to certain illnesses or injuries.

Example Situation: Let's say someone comes across someone who is having a severe allergic reaction. They can communicate with the conversational chatbot in this scenario, explain their symptoms, and ask for advice. Neural Net utilizing BioBERT and ChatGPT can both give comprehensive directions on getting emergency medical services. Conversational chatbots can offer valuable assistance in providing accurate and timely first aid recommendations to users, potentially helping to save lives in emergency situations.

* Review some existing solutions that you have seen for such conversational chatbots?
* What are the pros and cons of each existing solution?

# The Solution

Here you describe how your design and experimentation including the comparison produces ideas and prototype for effective conversational chatbots.

* How did model each chatbot ?

**NeuralNet Chatbot:** The BioBERT model with neural net is used to simulate the BioBERT chatbot's question-answering capabilities. This model is specifically intended to comprehend and reply to queries pertaining to medical and healthcare themes. It has been refined using biomedical text data. The chatbot matches user inquiries with preset patterns and responses kept in a JSON file by using cosine similarity between sentence embeddings. Using a sentence transformer model, the method encodes user questions and predetermined patterns into embeddings. The most suitable response is then found by computing cosine similarity scores.

**ChatGPT:** We used an API integration to refine a pre-trained GPT (Generative Pre-trained Transformer) model that we used to model our ChatGPT chatbot. Through this interface, we were able to harness the power of large-scale language models that have been trained on a variety of text corpora, allowing ChatGPT to comprehend user inquiries and produce conversational, human-like responses. We tried to adjust the model's replies to the context of giving first aid recommendations by fine-tuning it on particular conversational datasets. This will make sure that the model can interact with users and provide pertinent advice in emergency scenarios.

* Have tested your chatbots?

Yes, tests on both chatbots have produced positive outcomes. The NeuralNet chatbot can accurately respond to a range of first aid scenarios and is skilled at managing inquiries within its specific medical specialty. It is important to note that the NeuralNet dataset is smaller than ChatGPT, which restricts the range of medical issues it can cover. However, ChatGPT shows flexibility in comprehending and answering a variety of user inquiries, including complicated ones. Even with its broad reach, ChatGPT can offer thorough answers to several questions about different topics.

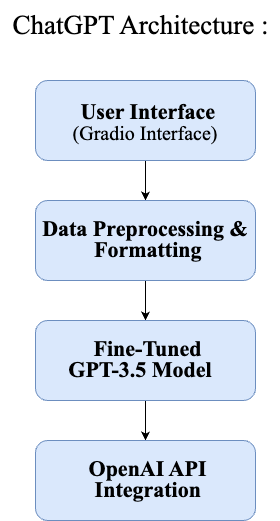
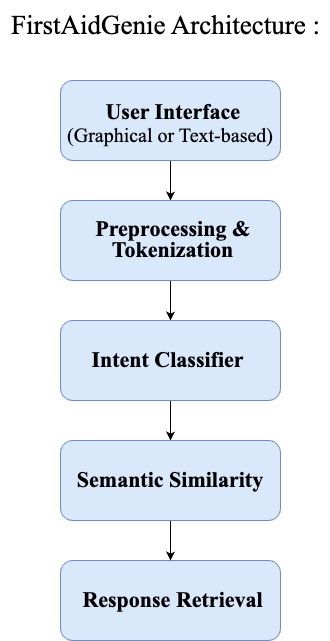
* Plans for improvements?

We can update the dataset to increase the chatbots' efficacy in offering medical advice and first-aid recommendations. First, in order to cover a wider range of first aid scenarios and healthcare-related questions, it is necessary to increase the coverage of intents by finding gaps in the existing dataset and adding new intents. This entails investigating typical first-aid scenarios and medical emergencies to guarantee thorough coverage of pertinent subjects. Secondly, more robust and diverse training examples for each intent can be produced by using data augmentation approaches. Enhancing the dataset with synonyms, word variations, and paraphrases of preexisting patterns might help the chatbots respond to a greater variety of user queries.

# Highlights on the Chatbots Capabilities

Describe both chatbots design and features/architecture/model and use some graphs, figures and images to illustrate your chatbots.

The architecture of both approaches and their explanation is below.



1. **"**FirstAidGenie**"** is based on advanced NLP and deep learning models, including BioBERT and SentenceTransformer.

**Input Processing:**

* User Input Reception: Gradio is used to develop a graphical user interface (GUI) or a text-based interface to receive user queries.
* Preprocessing: Preprocessing operations like tokenization and embedding are performed to prepare the queries for more processing and analysis.

**Intent Classification:**

* Intent Matching: A JSON file containing predefined intent patterns is compared to the preprocessed queries.
* Semantic Similarity Computation: SentenceTransformer, which encodes sentences into dense embeddings, is used to compute the semantic similarity between user queries and intent patterns.

**Response Generation:**

* Intent Resolution: Based on the calculated similarity scores, the chatbot determines the most pertinent intent related to the user inquiry.
* Response Retrieval: From the intents JSON file, the chatbot retrieves pre-programmed responses that match the specified intent.
* Response Selection: Predetermined parameters, including the most significant similarity score or other relevance metrics, are used to select the most appropriate response.

**Model Implementation:**

* A neural network model has been used to encode user queries and intent patterns into embeddings and calculate similarity scores between them. For tasks involving the computation of semantic similarity with medical questions, the BioBERT model is utilized.

**User Interface:**

* Text-Based Interface: Using a text-based interface, users can converse with the chatbot by submitting questions and getting conversational answers.
* Gradio Interface: As an alternative, users can communicate with the chatbot via a graphical user interface (GUI) made with Gradio. This tool offers an aesthetically pleasing and user-friendly text input and response display interface.
* Through the Gradio interface, users can ask the chatbot medical-related questions, and the chatbot will respond with appropriate responses based on the trained neural network and BioBERT embeddings.

**Overall,** this design allows the chatbot to comprehend user inquiries, recognize necessary intents, and deliver precise and contextually relevant answers, improving user experience and usability.

1. Chatbotis based on OpenAI's GPT-3.5 model.

**Data Processing:**

* Conversion to ChatGPT Format: First, the dataset must be transformed into a format that can be used to train a GPT-3.5 model. This entails putting intentions, trends, and reactions in an orderly fashion.

**Fine Tuning the Model:**

* Data Formatting for Fine-Tuning: To produce a training file appropriate for fine-tuning the GPT-3.5 model, further formatting is applied to the dataset. This entails arranging user inquiries and related answers in a way that complies with OpenAI's training specifications.
* Creation of the Fine-Tuning task: Next, a fine-tuning task for the GPT-3.5 model is produced using the formatted data. To train the model for first aid assistance and medical queries, it must be trained on the dataset that has been provided.

**ChatGPT Response Generation:**

* Generating Responses: After the model is adjusted, it can produce reactions in response to user cues. To provide contextually relevant and educational answers to user inquiries on first aid procedures and medical help, the chatbot uses the trained GPT-3.5 model.

**User Interface:**

* Gradio Interface: Gradio is used to design the chatbot's interface, offering a user-friendly setting for communication. Enhancing the overall user experience, users can input medical queries and receive conversational responses.

**OpenAI Integration:**

* Using the OpenAI API, the chatbot can utilize the GPT-3.5 model for response creation and fine-tuning. Thanks to this connectivity, the trained model and the chatbot interface may communicate effortlessly.

**Overall,** this architecture makes creating a chatbot skilled at offering medical advice and first-aid support easier. Utilizing the GPT-3.5 model's capabilities and integration with an intuitive interface, the chatbot provides precise and beneficial solutions to individuals needing medical advice.

# Walkthrough the Neural Network and BioBERT-based Chatbot Settings an Example of Use

Discuss the main steps necessary for the particular chatbot to be set to work.

* What is your design model?

**Design Model:**

* An architecture based on neural networks is the basis of the design paradigm used to develop chatbots. To be more precise, it makes use of neural network models that have already been trained for natural language processing tasks like computing semantic similarity and answering questions. The architecture of the chatbot involves encoding user questions and intent patterns into embeddings, utilizing neural network-based models to calculate similarity scores, and choosing relevant responses based on these scores.
* What resources will you need? (APIs, Libraries, Datasets, etc)

**Resources needed:**

**Libraries:** TensorFlow, TFLearn(a high-level deep learning library built on top of TensorFlow), Transformers, SentenceTransformers, NLTK, Gradio, and other libraries are used in the code.  
**Pretrained Models:** The code uses BioBERT and MiniLM models pre-trained on biomedical text data.  
**Data:** The JSON-formatted intents data includes patterns and answers for different intents.

* What application you chose?

**Chosen Application:**

* Its contribution to improving healthcare delivery is emphasized, especially in situations where prompt medical attention is essential. The answer also touches on the possibility of ongoing enhancement and growth of the chatbot's functionalities, such as enhancing its natural language comprehension and adding features for tailored recommendations. It emphasizes how essential chatbots are to equipping people to handle critical situations and medical emergencies.
* Its contribution to improving healthcare delivery is emphasized, especially in situations where prompt medical attention is essential. The answer also touches on the possibility of ongoing enhancement and growth of the chatbot's functionalities, such as enhancing its natural language comprehension and adding features for tailored recommendations. It emphasizes how essential chatbots are to equipping people to handle critical situations and medical emergencies.
* The IPYNB Prototyping

To facilitate interactive execution and prototyping in a Colab environment, the code is organized in a Jupyter Notebook (.ipynb).

* Additional Notes
* The chatbot design uses pre-trained neural network models—BioBERT in particular—to perform tasks related to answering questions and computing semantic similarity.  
  It employs vector space representation of user queries and intent patterns using pre-trained models' embeddings.
* By choosing the intent that has the highest similarity score to the input query, the chatbot seeks to offer appropriate answers to user requests.
* Users can communicate with the chatbot by submitting questions using the Gradio interface, and they will get real-time answers.

# Walkthrough of the ChatGPT-based Chatbot Settings an Example of Use

Discuss the main steps necessary for the ChatGPT set to work.

* What is your design model?

**Design Model:**

* The design approach focuses on fine-tuning OpenAI's GPT-3.5 model to provide a chatbot designed expressly to offer advice and first-aid support. Preprocessed and prepared to meet the needs of the GPT-3.5 model, the data is then used to fine-tune the model, resulting in replies that are useful in medical contexts and contextually relevant.
* What resources will you need? (APIs, Libraries, Datasets, etc)

**Resources needed:**

**Libraries:** To connect with the OpenAI GPT-3.5 model and generate a graphical user interface, the code uses the openai and gradio libraries, respectively.  
**Datasets:** A JSON file containing intents and related patterns/responses is needed to train and fine-tune the chatbot.  
**API Password:** It takes access to OpenAI's GPT-3.5 model API to adjust and produce results.

* What application you chose?

**Chosen Application:**

* The chosen application is a first-aid chatbot that answers users' medical questions and suggests first-aid techniques. It converses with people and provides advice based on their medical needs by leveraging GPT-3.5's conversational features.
* The IPYNB Prototyping

The code is organized and written in a Jupyter Notebook environment (firstaidgpt.ipynb). Each piece of the code is devoted to a particular task, such as data processing, model optimization, and Gradio-assisted GPT-3.5 model interface to create an intuitive user interface.

* Additional Notes
* The algorithm preprocesses the data to properly prepare the data for fine-tuning the GPT-3.5 model.  
  It uses OpenAI's API to adjust the model and produce responses to user commands.
* Gradio is used to construct the chatbot interface, which lets users submit medical questions and get conversational answers.
* The chatbot uses GPT-3.5's capabilities to deliver contextually relevant recommendations and provide helpful and informative responses for first-aid support.

# Comparison and Analytics

Compare the two chatbots and take position. Use relevant measures to do the analytics on comparisons

* What is your arguments and counter arguments?

**Arguments:**

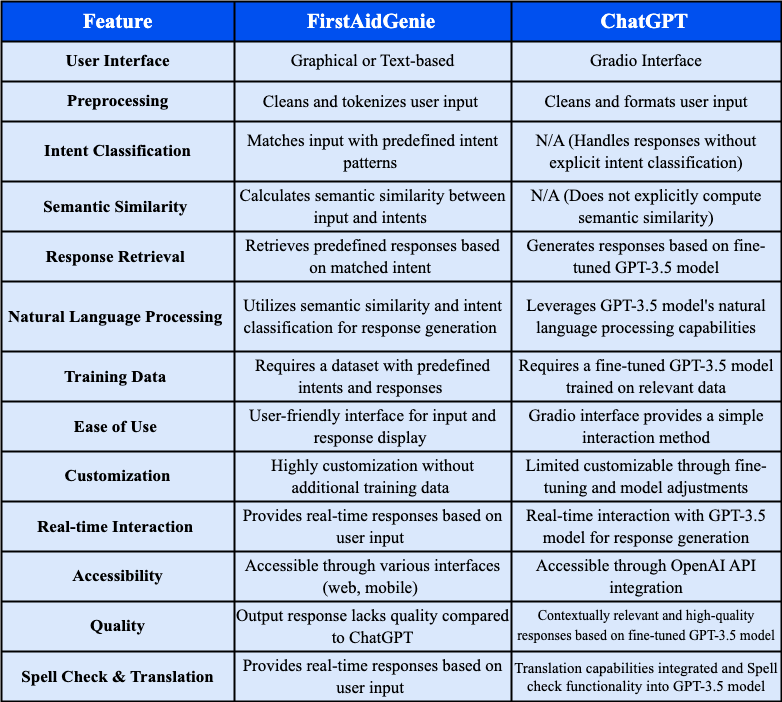
* **Effectiveness**: Using cutting-edge technology to supply precise and beneficial answers to user inquiries, both chatbots use advanced natural language processing (NLP) models to assist in medical settings.
* **Accessibility**: The FirstAidGenie chatbot has an easy-to-use UI that makes it simple for users to ask questions and get answers. On the other hand, the GPT-3.5-based chatbot functions via API integration and would need more technical know-how to be set up and utilized.

**Counterarguments:**

* **Accuracy**: Although both chatbots try to provide accurate answers, there may be times when the generated responses are not exact or need to sufficiently address the user's query, which could result in confusion or false information.
* What is your position?

The point of view adopted is that both chatbots can significantly contribute to providing medical help. The **FirstAidGenie** chatbot excels in user accessibility and usability, while the GPT-3.5-based chatbot demonstrates the power of sophisticated natural language processing models in producing contextually relevant responses. However, combining the two methods is the best way to establish a comprehensive and user-friendly medical help platform.

* What kind of measures you used for the comparison?



# Public Github Link:

# References

[1] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2019, September 10).

BioBERT: A pre-trained biomedical language representation model for biomedical text mining.

OUP Academic. https://academic.oup.com/bioinformatics/article/36/4/1234/5566506?login=false

[2] Bai, Y., & Zhou, X. (2020). Automatic Detecting for Health-related Twitter Data with BioBERT

(pp. 63–69). https://aclanthology.org/2020.smm4h-1.10.pdf

[3] Das, B., & Nirmala, S. J. (2022, October 1). Improving Healthcare Question Answering System by

Identifying Suitable Answers. IEEE Xplore.

https://doi.org/10.1109/MysuruCon55714.2022.9972435

[4] Julius, F. G., & Nayak, A. K. (n.d.). Advancing COVID-19 Inquiry Responses Through Transfer

Learning-Based Question Entailment Methodology. Research Gate.

https://www.researchgate.net/publication/378275113\_Advancing\_COVID-19\_Inquiry\_Responses\_

Through\_Transfer\_Learning-Based\_Question\_Entailment\_Methodology

[5] Devika, N., & Anoop, V. S. (n.d.). Biomedical Named Entity Recognition from Malaria Literature

using BioBERT. IEEE Explore.

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10335049>

[6] Zhang, Z., & Chen, A. L. P. (2022, November 3). Biomedical named entity recognition with the

combined feature attention and fully-shared multi-task learning - BMC bioinformatics. BioMed

Central. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04994-3

[7] Sharaf, S., & Anoop, V. S. (n.d.). An Analysis on Large Language Models in Healthcare: A Case

Study of BioBERT. https://arxiv.org/. https://arxiv.org/pdf/2310.05421

[8] Yu, X., Hu, W., Lu, S., & Sun, X. (n.d.). BioBERT based named entity recognition in electronic

medical record. IEEE Explore.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8965108

[9] Xie , W., Han, H., Xu, J., Zhao, C., & Li, T. (n.d.). Extracting Drug-Drug Interactions from

Biomedical Texts Using BioBERT with Improved Focal Loss. IEEE Explore.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10246572

[10] Xygi, E., Andriopoulos, A., & Koutsomitropoulos, D. (n.d.). Question Answering Chatbots for

Biomedical Research using Transformers. IEEE Explore.

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10066979

[11] Kant Singh, S., Kumar, S., & Singh Mehra, P. (n.d.). Chat GPT & Google Bard AI: A Review.

IEEE Explore. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10263706

[12] Roumeliotis, K. I., & Tselikas, N. D. (2023, May 26). Chatgpt and open-AI models: A preliminary

review. MDPI. https://www.mdpi.com/1999-5903/15/6/192

[13] Jhajaria, S., & Kaur, D. (n.d.). Study and Comparative Analysis of ChatGPT, GPT and DAll-E2.

IEEE Explore. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10307823&tag=1

[14] Brady D. Lund, Ting Wang, Nishith Reddy Mannuru, & Bing Nie. (n.d.). ChatGPT and a New Academic

Reality: AI-Written Research Papers and the Ethics of the Large Language Models in Scholarly

Publishing. https://arxiv.org/pdf/2303.13367

[15] Neumann, Michael & Rauschenberger, Maria & Schön, Eva-Maria. (2023). "We Need To Talk About

ChatGPT": The Future of AI and Higher Education. 10.25968/opus-2467.