

COMP5313WA 2024 Chatbot Contest

EXECUTIVE SUMMARY of Group No. 3

Students Names:

1212312 Murudkar, Priyamvada pmurudk1@lakeheadu.ca
1222198 Nagarajan Thirupathi, 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?

We are focusing on the following chatbots:

Neural Net using BioBERT: BioBERT is a customized version of BERT (Bidirectional Encoder Representations from Transformers) used in neural nets for biomedical text mining applications. Combined with a neural network, it can comprehend and react to questions on medical and healthcare subjects, emphasizing biological and biomedical ideas, in 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 Open AI. It is intended to comprehend natural language text responses and provide human-like responses. ChatGPT can have many kinds of discussions, exchange information, respond to queries, and even help with various 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 providing first aid in various 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, emphasizing 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 with a severe allergic reaction. In this scenario, they can communicate with the conversational chatbot, explain their symptoms, and ask for advice. Neural Net, utilizing BioBERT and ChatGPT, can provide comprehensive directions for 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 emergencies.

- Review some existing solutions that you have seen for such conversational chatbots.
- There are some already existing solutions for conversational chatbots other than that we are focusing on; they are:
- **Seq2Seq model:**
This model is based chiefly on Recurrent Neural Network (RNN) or Transformer, which are widely used for generating sequence-to-sequence output by encoding the input sequence, which makes them more suitable for dialogue generation and human-like conversation.
- **Transformer based model:**
Although we have models like GPT and Bert, which originated from transformers, there is also another model which is based on transformers, like T5, which is suitable for the summarization of text, and Transformer-XL, which improves the long-range dependencies and generates coherent text.
- **LSTM (Long Short-Term Memory):**
LSTM is also a type of RNN, an improved version of RNN that can store the data's memory for a period of time; it also determines which is important enough to be stored and which should be deleted. It is a massive leap in sentence generation and holding a conversation with a human.
- **Covi-Bio Bert:**
This model is a lot similar to our model, which is Bio-Bert. Still, Covi-Bio Bert is specifically pre-trained on data and queries related to covid COVID and COVID-related information to provide accurate information.
- What are the pros and cons of each existing solution?
- **Seq2Seq model:**
Pros:
Flexible architecture: it can handle various conversational tasks, including dialogue generation and text summarization.
Long-range dependency: Recurrent Seq2Seq models, 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 imbalance or bias in the dataset it has been trained on.
Limited context understanding: Traditionally, the model may need to help understand context contextual information, leading to generating irrelevant responses.
- **Transformer:**
Pros:
The transformer-based model utilizes a self-attention mechanism, enabling parallel input sequence processing and practical training on Modern hardware.
Scalability: They are scalable to large datasets and vocabulary sizes, making them suitable for training on extensive conversational corpora.

Cons:
Limited interpretability: They may lack interpretability, making understanding the underlying reason behind their generated responses challenging.
Data inefficiency: It requires a lot 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.

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 simulates the BioBERT chatbot's question-answering capabilities. This model specifically intends to comprehend and reply to queries about 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 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 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 conversational datasets. This will ensure that the model can interact with users and provide pertinent advice in emergencies.

- Have tested your chatbots?

Yes, tests on both chatbots have produced positive outcomes. The Neural Net 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 Neural Net dataset is more minor than ChatGPT, which restricts the range of medical issues it can cover. However, ChatGPT shows flexibility in comprehending and answering various user inquiries, including complicated ones. Even with its broad reach, ChatGPT can answer several questions about different topics thoroughly.

- Plans for improvements?

We can update the dataset to increase the chatbots' efficacy in offering medical advice and first-aid recommendations. First, to cover a broader 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

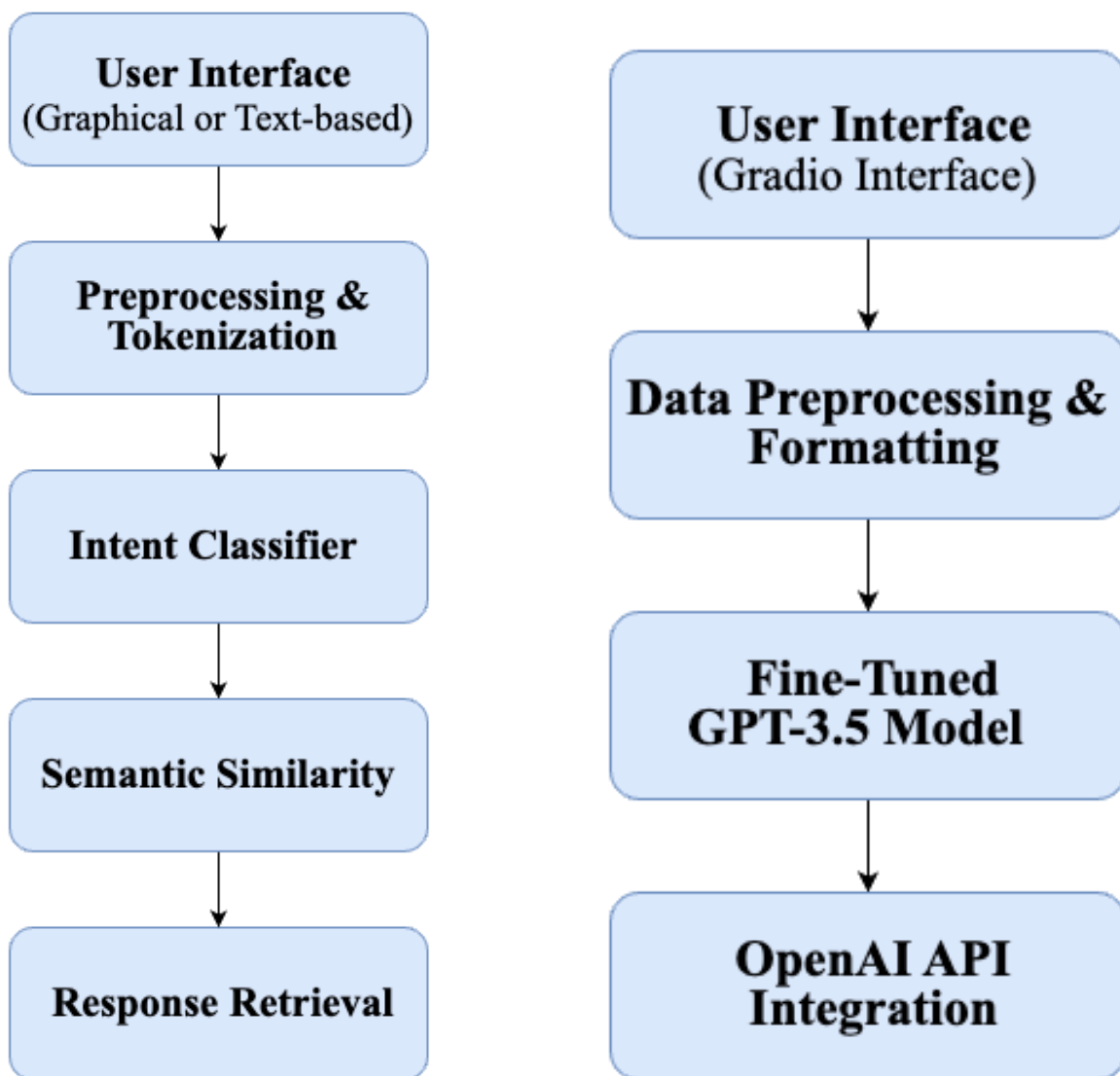
entails investigating typical first-aid scenarios and medical emergencies to guarantee thorough coverage of pertinent subjects. Secondly, more robust and diverse training examples for each intent can be produced by using data augmentation approaches. Enhancing the dataset with synonyms, word variations, and paraphrases of preexisting patterns might help the chatbots respond to more 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.

FirstAidGenie Architecture : ChatGPT Architecture :



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. Tokenization and lemmatization of user queries are performed first using spaCy. By taking this step, you can be sure that the input is ready for additional analysis.

Intent Classification:

- A Sentence Transformer model is used to encode the input query into a semantic space. Next, the cosine similarity between the encoded query and the patterns linked to various intents is calculated. This aids in categorizing the user's intention according to the pattern that is most comparable.

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.
- After classifying an intent, the chatbot chooses a response from the response pool associated with that intent. The response that best fits the intended patterns and the encoded query with the highest similarity score is chosen.

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.

Response Post-processing:

After the chosen response, it is translated back into the language initially detected (if relevant) and lemmatized using spaCy. The response is improved for readability and coherence in this step.

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.

Evaluation:

Metrics like as the BLEU score, which gauges the degree of similarity between generated and reference responses, can be used to assess the chatbot's responses. This aids in evaluating the outputs of the chatbot in terms of quality and relevancy.

Overall, the architecture shows how deep learning models and language processing techniques can be integrated for conversational applications by allowing the chatbot to comprehend user queries, classify intents, retrieve relevant responses, and provide seamless interaction through an interface.

2. Chatbot is 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. The design concept calls for the integration of several parts to produce a conversational chatbot that can comprehend user inquiries and offer appropriate replies. For intent categorization and response creation, it integrates deep learning models with natural language processing approaches. Tokenization, semantic similarity calculation, response selection, language translation, and a user interface are some of the components of the architecture.

- 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, spaCy, 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:

The selected application is a First Aid Genie chatbot, which responds to user inquiries to deliver first aid knowledge. When users enter their symptoms or worries, the chatbot provides relevant first aid guidance or information in response. The chatbot makes use of deep learning models and sophisticated natural language processing techniques to comprehend user intent and provide pertinent responses.

- The IPYNB Prototyping

Jupyter Notebook (IPYNB) implements the code, making prototyping and experimentation simple. Real-time development and testing of the chatbot code can be done on an interactive platform provided by the notebook. It makes it easier to quickly iterate and improve the functionality and design of the chatbot.

- Additional Notes
- The chatbot's architecture makes use of BioBERT to answer questions, which helps it comprehend inquiries about medicine.
- Intent categorization and answer retrieval are aided by semantic similarity computation using Sentence Transformers.
- To guarantee a flawless user experience, the chatbot's responses are post-processed for coherence and readability.
- The chatbot is deployed using Gradio and has an easy-to-use interface that makes it possible for people to communicate with it.
- Evaluation measures that measure the quality of the chatbot's responses and help it perform better over time include the BLEU score.

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?

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

- Measures Applied: Comparison of Keyword Detection Efficiency and Processing Time
- Comparison Measure: Efficiency in finding keywords and processing speed.
- Justification: Biobert's proficiency at identifying keywords stems from their specific training in biomedical content. It works exceptionally well for deciphering medical searches and finding pertinent keywords. However, because of its intricate architecture and high processing demands, it could take longer than other models to answer queries.

Chatbot is based on OpenAI's GPT-3.5 model.

- Measures Applied: Spell Checking, Personalization, History Handling, and Language Understanding.
- Comparison Measure: Knowledge of Language, Customization, Spelling Check, and History Management
- Justification Being a big language model, ChatGPT exhibits great language understanding abilities and can offer customized responses depending on the conversation's context. It has functions like spell-checking to guarantee precise answers. Furthermore, it can retain past discussions, allowing the dialogue to

flow smoothly. Nonetheless, there might be situations in which the bot's decision-making in subsequent contacts is impacted by the conversation history stored, which could result in unfavorable outcomes.

- Feature Analysis of both chatbots is shown below:

Feature	FirstAidGenie	ChatGPT
User Interface	Graphical or Text-based	Gradio Interface
Preprocessing	Cleans and tokenizes user input	Cleans and formats user input
Intent Classification	Matches input with predefined intent patterns	N/A (Handles responses without explicit intent classification)
Semantic Similarity	Calculates semantic similarity between input and intents	N/A (Does not explicitly compute semantic similarity)
Response Retrieval	Retrieves predefined responses based on matched intent	Generates responses based on fine-tuned GPT-3.5 model
Natural Language Processing	Utilizes semantic similarity and intent classification for response generation	Leverages GPT-3.5 model's natural language processing capabilities
Training Data	Requires a dataset with predefined intents and responses	Requires a fine-tuned GPT-3.5 model trained on relevant data
Ease of Use	User-friendly interface for input and response display	Gradio interface provides a simple interaction method
Customization	Highly customization without additional training data	Limited customizable through fine-tuning and model adjustments
Real-time Interaction	Provides real-time responses based on user input	Real-time interaction with GPT-3.5 model for response generation
Accessibility	Accessible through various interfaces (web, mobile)	Accessible through OpenAI API integration
Quality	Output response lacks quality compared to ChatGPT	Contextually relevant and high-quality responses based on fine-tuned GPT-3.5 model
Spell Check & Translation	Provides real-time responses based on user input	Translation capabilities integrated and Spell check functionality into GPT-3.5 model

Public Github Link:

<https://github.com/vaibhav720/Conversational-Contest---BioBert-vs-ChatGPT-First-Aid>

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] Jhalaria, 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.