



Capstone

By Blake Weiss




Artificial Intelligence

- What is Artificial Intelligence and what is its effect on society
 - AI is meant to be a copy of human intelligence its goal is to allow machines/computers to achieve tasks that otherwise a human would do.
 - AI is going to transform the world with its effect on everyday life with devices like alexa, siri, google, etc...
 - Also Artificial intelligences effect in the workplace with AI expected to replace up to 80% of the current workforce in the next 15 years.
- Examples of Artificial Intelligence in everyday life
 - Healthcare: AI helps identify issues on scans and develop treatment plans for patients.
 - Finance: AI is utilized to do algorithmic black box trading it is also used to detect fraud.
 - Automotive: AI is integrated with self driving vehicles to ensure a safe trip.
 - Military: AI is being used to fly drones, maintain computer systems, utilized in fighter jets and to make and integrate with cybersecurity tools.



The Mental Health Chatbot Project

- Overview
 - Purpose: To provide people with a outlet that is ready to be talked to and can be emotionally connected with the user and help diagnose problems before the user is able to meet with a human therapist.
 - Use cases: Offer instant support to the user where otherwise may be inaccessible or limited with traditional appointment meetings with therapists.
- Challenges
 - Ethical considerations of whether or not it is fair to rely on a AI in a high tension environment whether that be in relation to suicide or other such issues that are similar.
 - Building a stable model that does not hallucinate at all or generate bad responses to the user who needs mental health support.
 - Having enough time to fine tune thousands of responses before the project was due.



Data Collection and Preparation

- Data Sources and Types
 - Health Journals, Therapy handbooks and exclusive datasets.
 - I chose to use TXT files for easy processing.
- Data Diversity
 - Using multiple data sources is extremely important because it leads to more diverse responses from the AI and more accuracy.
 - Creating Intent based categories to classify data.
 - greetings, conversational connectors, therapy responses, emotional responses, ending statements/goodbyes and misc responses/non classifiable conversational responses.
 - These 6 intent based categories I created to organize the data were extremely useful in making sure that there was enough information available for the AI to train on.



Model Building

- Model architecture
 - I utilized LSTM and Sequence 2 Sequence models integrated with attention mechanisms for contextual awareness in conversation.
 - This tool called neutron I found is a great way to visualize model architecture and weights. This will be shown on the next slide.
- Custom algorithm
 - My program is a combination of many different types of Algorithms.
 - I utilized top k, temperature and reranking of response candidates to enhance the responses correctness and coherence while also using these functions to improve upon the generative capabilities of the program.
 - I will explain how the math works on the next slide.

DENSE

Weights
kernel (256x1193)
bias (1193)

LSTM

Weights
kernel (456x1024)
recurrent_kernel (256x1024)
bias (1024)

ATTENTION

Weights
kernel (256x1)
bias (1)

Weights
kernel (256x256)
bias (256)

Weights
kernel (256x256)
bias (256)

LSTM

Weights
kernel (200x1024)
recurrent_kernel (256x1024)
bias (1024)

EMBEDDING

Weights
embeddings (1193x200)

$[(1,2,3,4)]$
 $[1,2,3,4,5,6]$
 $[(1,2,3,4)]$

Preprocessed sum num words

$$U' = \frac{1}{4} \sum_{i=1}^3 \left(\frac{\text{confidence}}{\pi \text{ predictvalue}} \right) = U' + \alpha$$

Confidence pretend value = 2

28.274

1.909859317

num words

$[This, is, the, day]$
11

num words

$$\frac{4}{4} = 1 \leftarrow \text{predict value}$$

num #'s

$$\frac{12}{4} = 3$$

$\frac{\text{num character sum}}{\text{num #'s}} = \text{predictable outcome word}$

$$\sqrt{\pi 3^2} = \text{generative outcome word} = 28.274$$

$= 30.19395932$

algrthmic chance outta int 50



Preprocessing

- Preprocessing Data the first step
 - Preprocessing utilizes the data that is provided for training and removes all unwanted characters and punctuation that is not needed.
 - Preprocessing also puts tags at the beginning and end of the sentence so that the program recognizes the sentence it puts <start> at the beginning and <end> at the ending of the sentence.
 - Preprocessing allows for the data to be cleaned before it is inputted into the AI for training.
- Basic Example of Preprocessing steps
 - Tokenization "I do not like green eggs and ham" ["I", "do", "not", "like", etc...]
 - Stop word removal get rid of common words like "the", "or", "and", "a" Stemming/Lemmatization Shorten words to their root by removing endings when applicable like "ing", "ed", "ion" and so on
 - Punctuation removal "hello how are you?" turns into "hello how are you"



Training the Model

- Dataset
 - There are 598 data points within the dataset I created. Which is a very small dataset in comparison to other AI models.
 - The data is organized in the TXT file into inputs and expected outputs for training.
- Training the Model
 - I decided against using cross validation and a validation dataset as I preferred for my system to be more simplistic.
 - I made sure the loss value was always decreasing if it wasn't the model had to be retrained.
 - I focused more on the performance of the model when training also focused on how correct the response generation was.



Post Processing

- Post Processing the last step
 - Post Processing refines the raw output that an AI spits out. It is a series of operations or techniques that get applied to the raw output to make the results of a AI's responses more suitable to its needs and requirements.
 - My Post Processing takes the generated token output by the model and turns the tokens into human readable text using the tokenizer's decode section of the method.
 - The post processing then removes the <start> and <end> tags from the sentence these tags were added during preprocessing; these tags get replaced with an empty string. This erases the tags from the output the user would see. Then lastly the post processing takes any trailing whitespace after the removal of the tags and removes the extra trailing whitespace.
 - All of this that is done in Post Processing ensures that the response the user sees is clean and makes sense.



Evaluation and Fine Tuning

- Performance and Metrics
 - Accuracy, confidence, response time, rewards, sentiment analysis, and intent classification these are the many factors that are taken into account when evaluating model performance

root - INFO - 2024-03-08 23:11:22, User Input: Hello, Bot Response: Im always here to provide answers What can I assist you with, Response Time: 9.563585, Reward: 1

root - INFO - 2024-03-08 23:11:40, User Input: Nothing much Goodbye, Bot Response: Im here if you need me for anything else, Response Time: 6.651177, Reward: 1

- These two examples show how the performance metrics are used within the AI model.
- Human Feedback
 - I used RLHF (Reinforcement Learning Human Feedback) to improve the performance of the responses from the model.
 - The AI was also issued rewards based on sentiment and the accuracy of responses that were generated from the model.



Deployment and Monitoring

- How I Monitored Consistent Performance from the model
 - Accessing an AI's performance many things are taken into account:
 - accuracy of responses, confidence, response time, rewards for responses, sentiment analysis and intent classification.
 - These performances indicate how well an AI is performing in response to the responses it is giving out.
 - I decided a Log file was needed so that I could see the ins and outs of how the computer and AI were processing the data being fed to it.
 - The hyperparameters which represent performance of certain parts of the program and can be changed and adjusted when needed would be displayed in the log file.
 - Hyperparameters can be adjusted if the AI is not performing to the quality of responses I desire.

Model in Action

- Here are some videos of the AI model in action and displaying its abilities.

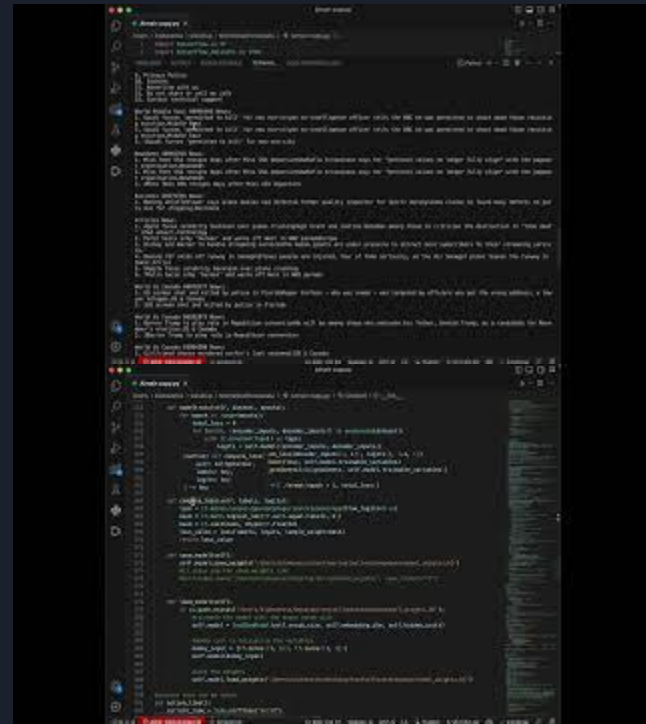


Pt 1

Pt 2

Pt 3

Pt 4





Thanks for Watching

Any Questions? Please feel free to ask!

Link to my code:

<https://github.com/BJW333/JOSH>