

Configuration Manual

MSc Research Project
MSCDAD JAN23A I

Talha Masood
Student ID: 20164467

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing



Student Name:	Talha Masood
Student ID:	20164467
Programme:	MSCDAD JAN23A I
Year:	2023
Module:	MSc Research Project
Supervisor:	Jorge Basilio
Submission Due Date:	14/12/2023
Project Title:	Configuration Manual
Word Count:	Approx 300
Page Count:	7

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	29th January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Talha Masood
20164467

1 Introduction

This is the configuration manual for the project "An Overview of the use of Small Scale AI and LLM Models in the context of Receptionist Chatbots".

The main websites to be used are Github and Huggingface. The main programming language to be used is Python.

The libraries to be used are TextBlob and WordCloud. The Python version used in the virtual environment is Python 3.10.13.

2 Oobabooga

The WebUi is available for download on its github repository. The image in figure 1 shows the main page. It can be downloaded as a zip file.

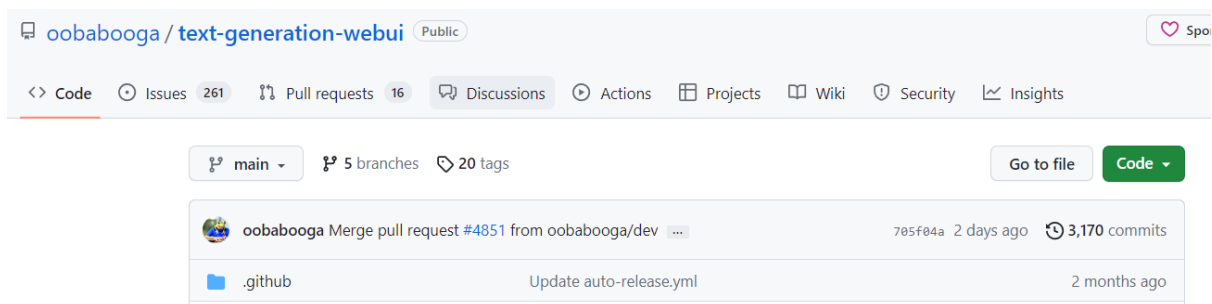


Figure 1: The github page

After installation, it needs to be extracted. The zip file is shown in figure 2.

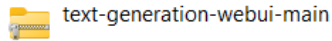


Figure 2: The zip file

After extracting, run the batch file. The file name is start-windows. It is shown in figure 3.

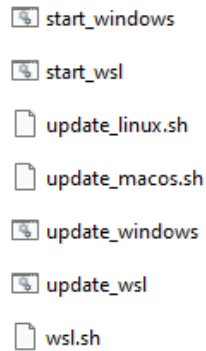


Figure 3: The batch file

Running this will install the webui. After installation, the webui interface will show as shown in figure 4.

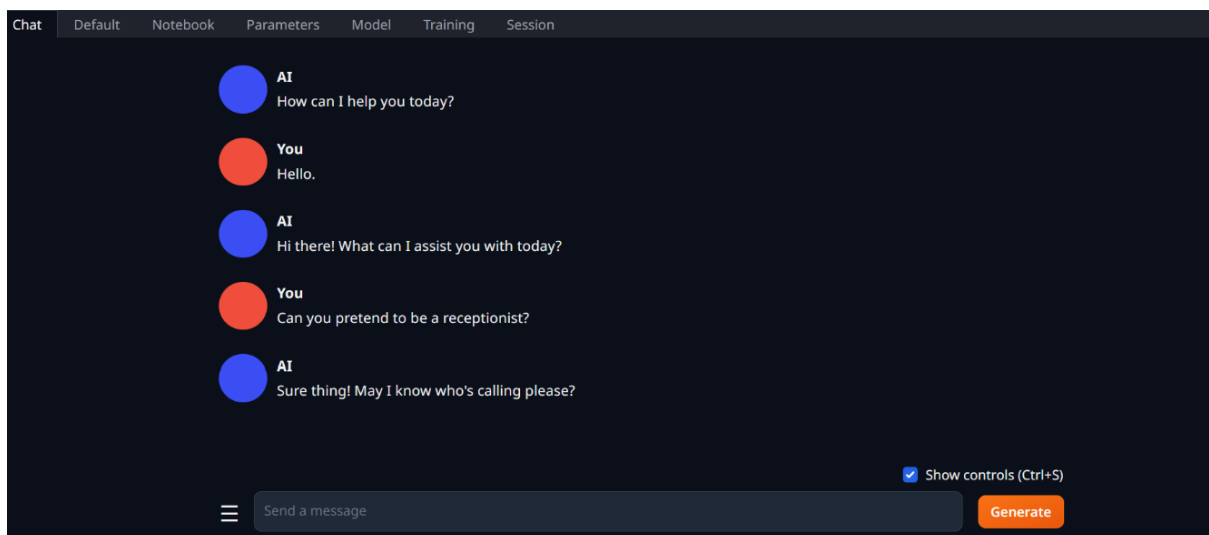


Figure 4: The webui gui

After the webui is loaded the next step is to get models from the huggingface website. The main page is shown in figure 5.

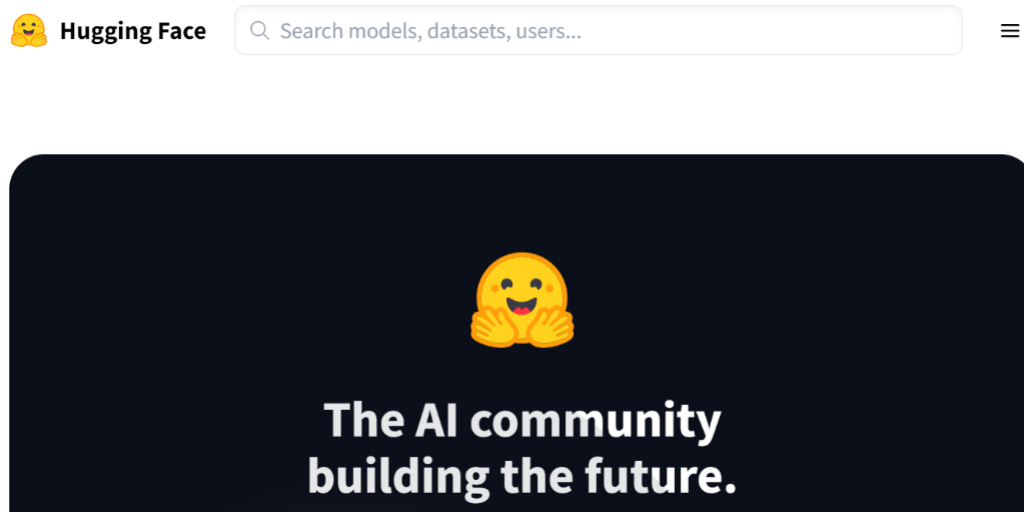


Figure 5: The huggingface website

A model can be downloaded from its page. An example is shown in figure 6.

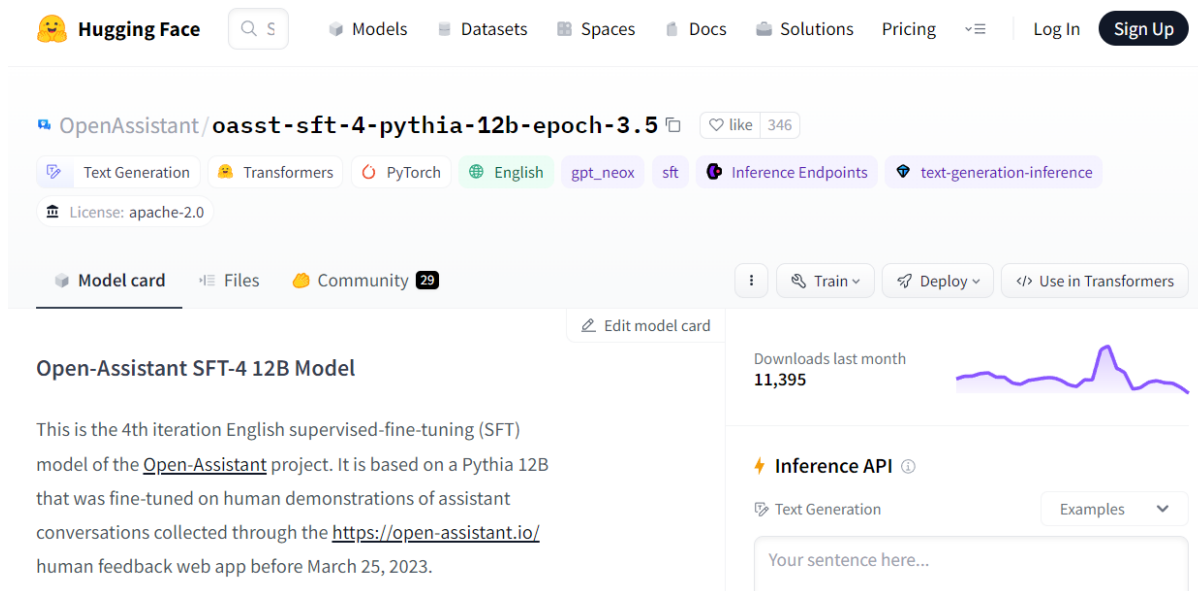


Figure 6: page of a model on huggingface

The link can be copied and pasted in the model tab of the webui GUI. The way to do this is shown in figure 7.

Download model or LoRA

Enter the Hugging Face username/model path, for instance: facebook/galactica-125m. To specify a branch, add it at the end after a ":" character like this: facebook/galactica-125m:main. To download a single file, enter its name in the second box.

OpenAssistant/oasst-sft-4-pythia-12b-epoch-3.5

File name (for GGUF models)

Download Get file list

Figure 7: downloading the model from huggingface

The download process begins and we can see the progress on the command line as shown in figure 8.

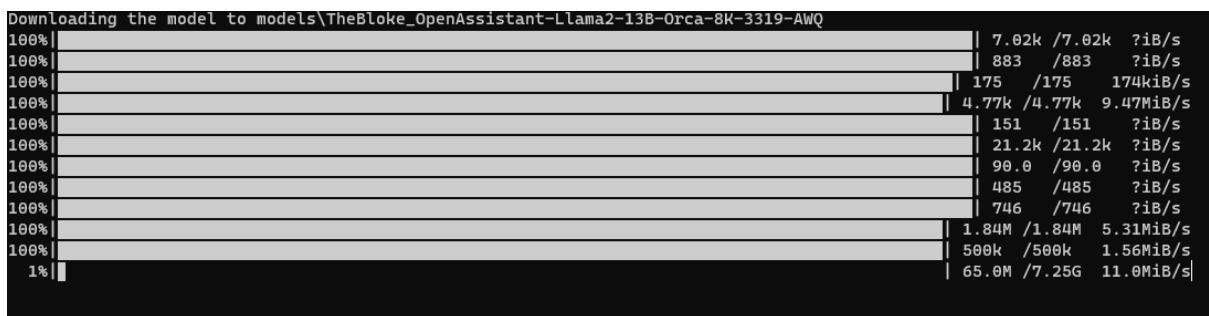


Figure 8: downloading the model from huggingface

After the model is downloaded, it must be loaded as shown in figure 9.

Model

TheBloke_OpenAssistant-Llama2-13B-Orca-8K-3: Load Unload Reload Save settings

Model loader

AutoAWQ

Figure 9: loading the downloaded model from huggingface

Once loaded, the model can be used.

3 Chatbot Arena

Chatbot arena can be used from the website as shown in figure 10.

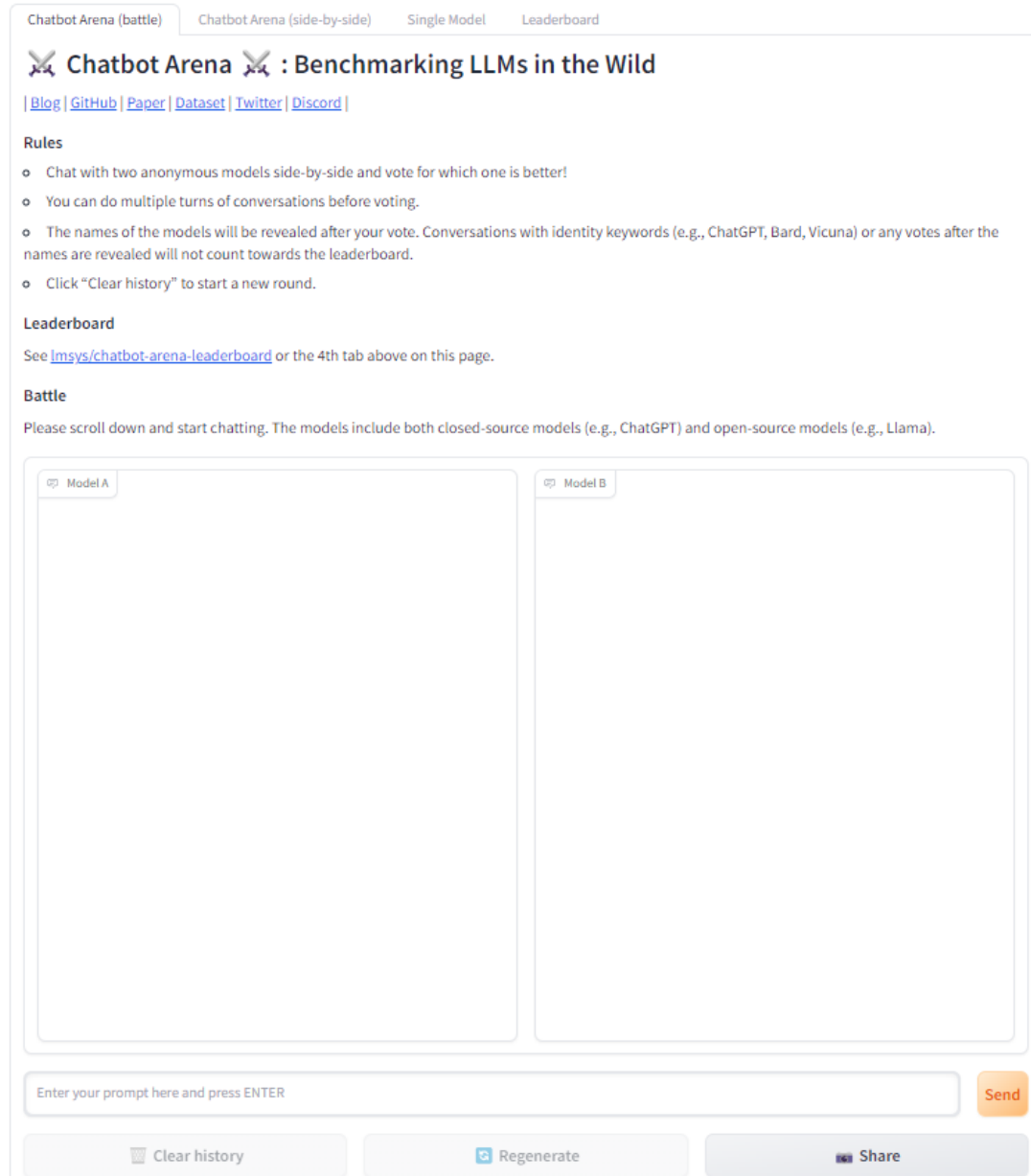


Figure 10: The chatbot arena page

Speed may vary depending on the model used as well as the time of day. The models can be chosen from a drop down menu at the top.

Two conversations can be had at a time as shown in figure 11.

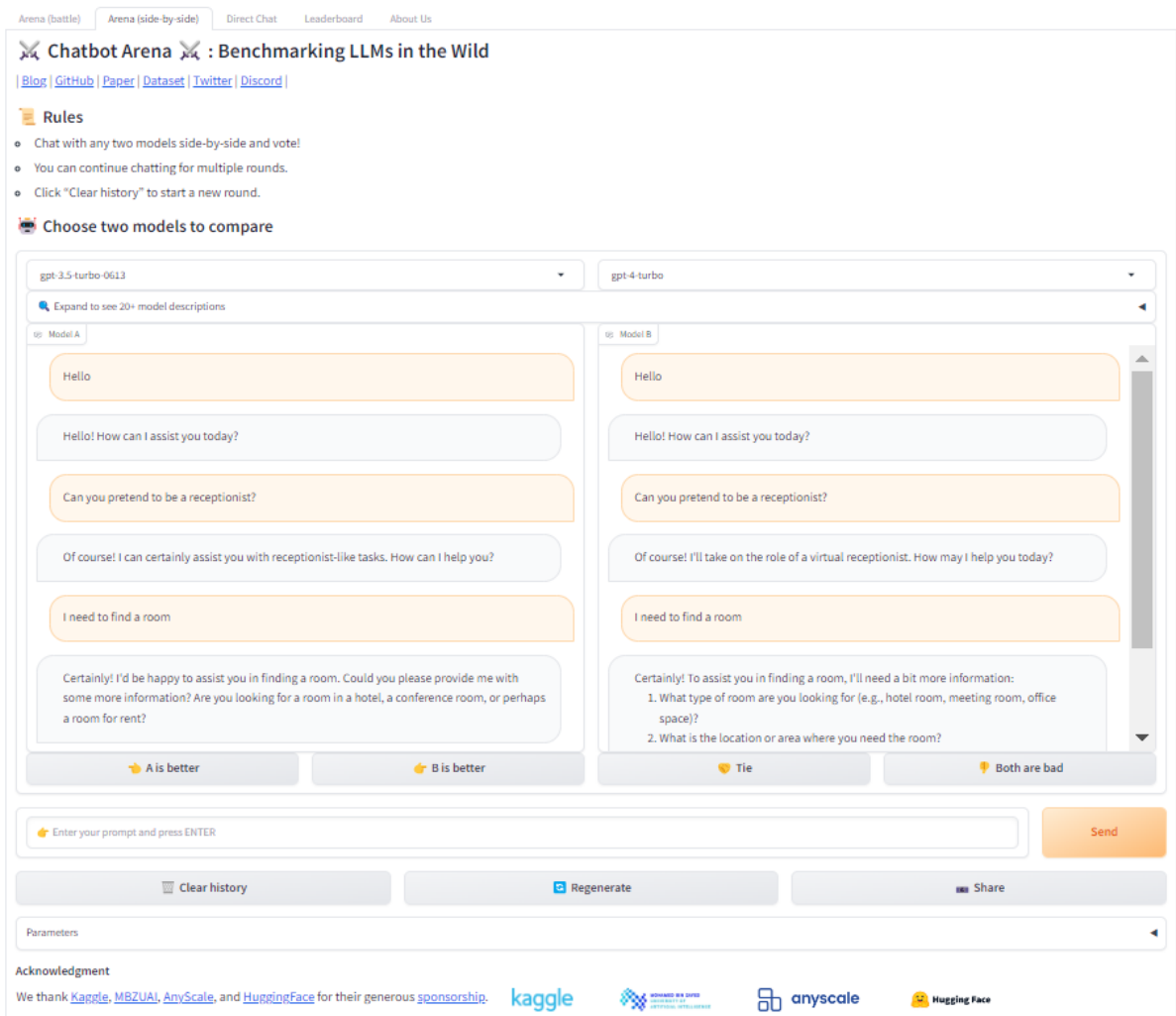


Figure 11: The chatbot arena page

4 Wordcloud and Sentiment analysis libraries

Sentiment analysis and wordcloud are done by the libraries shown in the figure below.

```
# We will be using the "TextBlob" library
from textblob import TextBlob

# we need to import the "wordcloud" library
from wordcloud import WordCloud
```

Figure 12: The libraries

References

Nothing cited for this file.