

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.

<u>ALL</u> 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).□Attach a Moodle submission receipt of the online project submission, to
each project (including multiple copies).□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.

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.

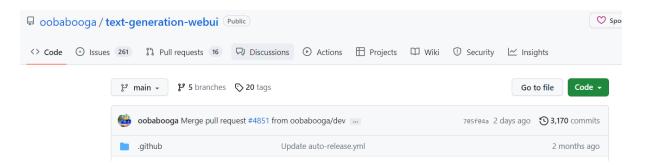


Figure 1: The github page

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.

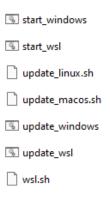


Figure 3: The batch file

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

Chat	Default	Notebook	Parameters	Model	Training	Session				
			AI How can	I help you	today?					
			You Hello.							
			AI Hi there!	What can I	assist you v	with today?				
			You Can you	pretend to	be a recept	onist?				
			AI Sure thir	ıg! May I kr	iow who's ca	alling please				
								Show co	ntrols (Ctrl+S)	
		≡							Generate	

Figure 4: The webui gui

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

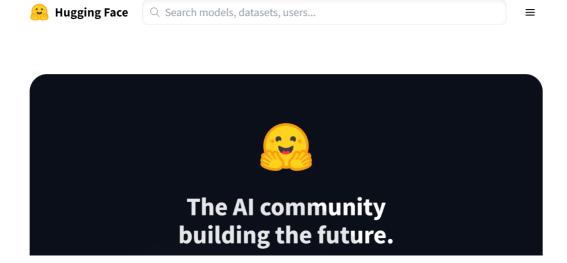


Figure 5: The huggingface website

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

Hugging Face Q S	sets 📓 Spaces 🧯 Do	cs 🚔 Solutions Pricing 💉	r≡ Log In Sign Up
• OpenAssistant/oasst-sft-4-pythia-12	2b-epoch-3.5 🖻 🕜	♡ like 346	
🖻 Text Generation 🔗 Transformers 🗘 PyTorch 🔀 E	inglish gpt_neox sft	Inference Endpoints 🗇 text-ge	neration-inference
₫ License: apache-2.0			
🖗 Model card → 🗉 Files 🥚 Community 😰		: 🕲 Train - 🕫 Deploy	 Use in Transformers
	🖉 Edit model card		
ppen-Assistant SFT-4 12B Model		Downloads last month 11,395	$\sim\sim$
his is the 4th iteration English supervised-fine-tuning (SF	-T)		
nodel of the <u>Open-Assistant</u> project. It is based on a Pythi	ia 12B	🔸 Inference API 🛈	
nat was fine-tuned on human demonstrations of assistan	it	🕫 Text Generation	Examples 🗸
onversations collected through the <u>https://open-assistar</u>	<u>nt.io/</u>		
uman feedback web app before March 25, 2023.		Your sentence here	

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: face ":" character like this: facebook/galactica-125m:main. To downloa	book/galactica-125m. To specify a branch, add it at the end after a d 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.

Downloading the model to models\TheBloke_OpenAssistant-Llama2-13B-Orca-8K-3319-AWQ	
100%	7.02k /7.02k ?iB/s
100%	883 /883 ?iB/s
100%	175 /175 174kiB/s
100%	4.77k /4.77k 9.47MiB/s
100%	151 /151 ?iB/s
100%	21.2k /21.2k ?iB/s
100%	90.0 /90.0 ?iB/s
100%	485 /485 ?iB/s
100%	746 /746 ?iB/s
100%	1.84M /1.84M 5.31MiB/s
100%	500k /500k 1.56MiB/s
1%	65.0M /7.25G 11.0MiB/s

Figure 8: downloading the model from huggingface

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

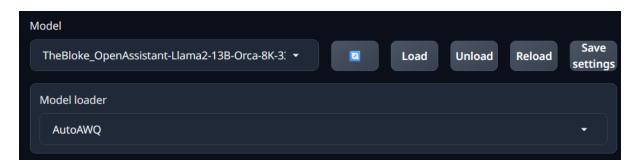


Figure 9: loading the downloaded model from huggingface

Once loaded, the model can be used.

3 Chatbot Arena

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

Chatbot Arena (battle) Chatbot Arena (side-by-side)	Single Model Leaderboard
💢 Chatbot Arena 💢 : Benchma	rking LLMs in the Wild
Blog GitHub Paper Dataset Twitter Discord	
Rules	
• Chat with two anonymous models side-by-side and vot	e for which one is better!
You can do multiple turns of conversations before votin	-
 The names of the models will be revealed after your vot names are revealed will not count towards the leaderboard 	ie. Conversations with identity keywords (e.g., ChatGPT, Bard, Vicuna) or any votes after the j.
• Click "Clear history" to start a new round.	
Leaderboard	
See Imsys/chatbot-arena-leaderboard or the 4th tab above	on this page.
Battle	
Please scroll down and start chatting. The models include	both closed-source models (e.g., ChatGPT) and open-source models (e.g., Llama).
👳 Model A	@ Madel B
Enter your prompt here and press ENTER	Send
Clear history	🖸 Regenerate 🛛 📾 Share

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.

Chatbot Arena 🔀 : Benchmarking I	LMS III the Wild		
Rules			
at with any two models side-by-side and vote! I can continue chatting for multiple rounds.			
k "Clear history" to start a new round.			
hoose two models to compare			
t-3.5-turbo-0613	•	gpt-4-turbo	
Expand to see 20+ model descriptions		Phy 4 million	
Indel A		63 Model B	
Hello		Hello	
Hello! How can I assist you today?		Hello! How can I assist you today?	
Can you pretend to be a receptionist?		Can you pretend to be a receptionist?	
Of course! I can certainly assist you with receptionist-like	tasks. How can I help you?	Of course! I'll take on the role of a virtua	al receptionist. How may I help you today?
I need to find a room		I need to find a room	
Certainly! I'd be happy to assist you in finding a room. Co some more information? Are you looking for a room in a a room for rent?		Certainly! To assist you in finding a roor 1. What type of room are you looking space)? 2. What is the location or area where	for (e.g., hotel room, meeting room, office
👈 A is better	👉 B is better	🤝 Tie	📍 Both are bad
Enter your prompt and press ENTER			Sen
🔤 Clear history	Reg	enerate	share
meters			

Ar

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.

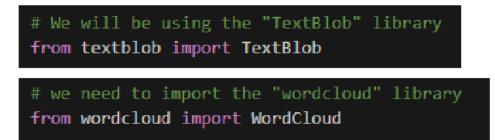


Figure 12: The libraries

References

Nothing cited for this file.