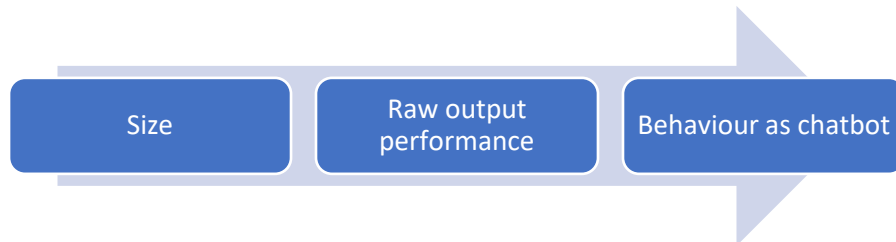


FBS-2841: Select the top 3 LLM models to test for the chatbot implementation and performing experiment.

### Model Shortlisted

1. GPT-NeoX 20B (MIT)
2. GPT-J 6B (Apache 2.0)
3. OpenLLaMa 7B (Apache 2.0)
4. Falcon 40B (Apache 2.0)

### Framework for Evaluation



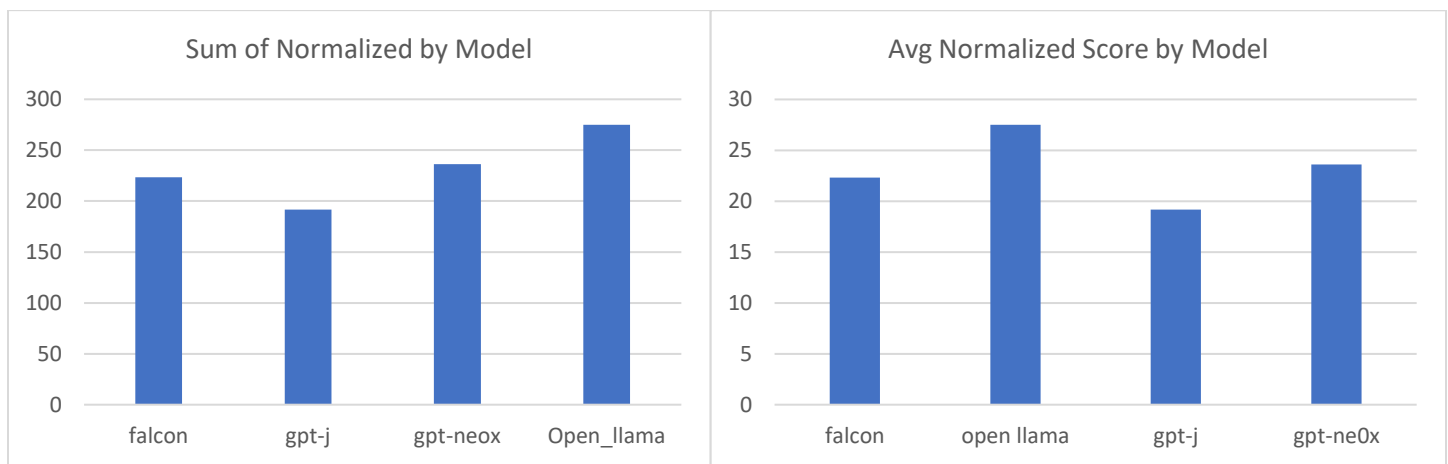
### Performance – I: Raw output performance

- How the model performance without any filter?
- How raw model output relevant to ford specific software without any tweaking?
- Quantify the output to our use case.

#### Approach: Creating Custom Score for comparison.

1. Given list of known ford software
2. Standard Prompt is given to all Models.
3. Output generated is used to check similarity with given list of ford specific software.
4. Get cosine similarity scores for the output.
5. Repeat this for 10 queries and take an average to determine the score for evaluation.
6. Compare models score for their performance without any filter.
7. Select the best model.

**Results:** Open Llama is recommended on basis of its performance as its average normalized similarity score with respect to ford specific tech stack is 27.495750818011%. Also, it is based on Red Pajama dataset which is relevant to our use case.



## Open Llama 7B v2 open instruct.

\*General information about selected model

Source: [VMware/open-llama-7b-v2-open-instruct · Hugging Face](https://huggingface.co/VMware/open-llama-7b-v2-open-instruct)

Instruction-tuned version of the fully trained Open Llama 7B v2 model. The model is open for COMMERCIAL USE. This model performs better on code compared to v1 due to the improvements made on the base model by the openlm-research team. The instruction model is trained on an improved instruction tuning dataset compared to v1

- **Model :** Open-llama-v2
- **Size:** 7B parameters
- **Dataset:** Open-instruct
- **NOTE:** The model was trained using the Alpaca prompt template

### License

- CC BY-SA-3.0 (Commercially Viable!)
- Base Language Model (openlm-research/open\_llama\_v2\_7b) is under apache-2.0
- Fine-Tuning Dataset (VMware/open-instruct) is under cc-by-sa-3.0

### Datasets used for Fine-Tuning.

- **Open-instruct.**
  - Open-instruct-v1
  - Mosaic/Dolly-HHRLHF + filtered OASST1 - cc by 3.0
- **Subset of COT SUBMIX (FROM FLAN V2) Zeroshot examples**
  - ESNLI - MIT
  - ECQA - CDLA 1.0 - Sharing
  - Strategy - MIT
  - CREAK - MIT
  - gsmk8 - MIT
  - aqua - MIT
  - qasc - Apache 2.0

### Standard prompts used.

1. What are the technologies to build chatbot?
2. What technologies are needed to create a web based dashboard for a cloud computing project?
3. What technologies are used for creating a blog?
4. What technologies can be used for reporting on the performance of sales?
5. What technology recommendation for AI chat bot?
6. What technology recommendation for big data analytics?
7. What technology recommendation for building an app that can run in multiple clouds?
8. What technology recommendation for building an inventory management system?
9. What technology recommendation for cloud Machine Learning deployment?
10. What technology recommendation for creating an application to manage cloud computing?
11. What technology recommendation for dashboard for my Hr data?
12. What technology recommendation for dashboard for my sales?
13. What technology recommendation for data warehousing?
14. What technology recommendation for Natural Language Processing?
15. What technology recommendation for project management?
16. What type of database should be used for a web application?
17. What type of technology would be best for developing mobile application?

\*Citizen developer prompts llm generated.

Output Table

model	prompt	output	temp	Similarity score	Normalized
falcon	What technology recommendation for dashboard for my sales?	['Qlik Sense','Tableau','Power BI','Excel','Google Sheets','Python','R','SQL','MySQL','PostgreSQL','Knex']	0	13.91591	47.03845
falcon	What technology recommendation for dashboard for my sales?	["python", "qlik sense", "power bi", "tableau"]	0	3.413944	10.78511
falcon	What technology recommendation for big data analytics?	["hadoop", "python"]	0	11.82625	39.82483
falcon	What technology recommendation for project management?	["gantt chart"]	0	0	-1
falcon	What technology recommendation for building an app that can run in multiple clouds?	["big data", "hive"]	0	28.96827	99
falcon	What type of database should be used for a web application?	["relational", "NoSQL"]	0	0	-1
falcon	What technology recommendation for dashboard for my Hr data?	["qlikview", "power bi" ]	0	4.828045	15.66667
falcon	What technology recommendation for project management?	["jira", "agile development"]	0	0	-1
falcon	What technology recommendation for data warehousing?	["hive", "kafka"]	0	0	-1
falcon	What technology recommendation for dashboard for my Hr data?	["QlikView","Power Bi","Tableau","Power BI","QlikView" ]	0	4.638636	15.01282
Open_llama	What technology recommendation for dashboard for my sales?	["nodejs","javascript","python","java"]	0	8.36242	27.86751
Open_llama	What technology recommendation for dashboard for my sales?	["sql","excel","access","pyhton", "qlik sense"]	0	13.65577	46.14045
Open_llama	What technology recommendation for cloud Machine Learning deployment?	["AWS", "Azure", "Google Cloud Platform"]	0	22.43873	76.45967
Open_llama	What technology recommendation for big data analytics?	["Hadoop", "Hive", "Spark"]	0	4.828045	15.66667
Open_llama	What technology recommendation for building an app that can run in multiple clouds?	["AWS","Azure","Aegon"]	0	4.828045	15.66667
Open_llama	What technology	["python","pytorch","numpy","	0	5.288859	17.25742

	reccomendation for AI chat bot?	matplotlib","gym","sklearn","seful","replit","jovo","ai"]			
Open_llama	What technology reccomendation for Natural Language Processing?	["python", "spacy", "sklearn", "numpy"]	0	4.18121	13.43376
Open_llama	What technology reccomendation for project management?	["mpp", "ms project", "trello"]	0	0	-1
Open_llama	What technology reccomendation for project management?	["jira", "Confluence"]	0	0	-1
Open_llama	What technology reccomendation for data warehousing?	["sap", "sql server", "oracle", "ms access", "sas"]	0	18.96419	64.46537
gpt-j	What technology reccomendation for building an inventory management system?	["python", "sqlite", "python"]	0	7.479576	24.81989
gpt-j	What type of technology would be best for developing mobile application?	["python", "java", "react", "nodejs"]	0	12.54363	42.30127
gpt-j	What technology reccomendation for cloud Machine Learning deployment?	["python", "apache spark"]	0	9.656091	32.33333
gpt-j	What technology reccomendation for Natural Language Processing?	["tensorflow", "pytorch"]	0	0	-1
gpt-j	What technology reccomendation for big data analytics?	["mongodb", "mongoose", "python"]	0	4.828045	15.66667
gpt-j	What technology reccomendation for dashboard for my sales?	["python", "sqlite", "pandas", "kibana"]	0	4.18121	13.43376
gpt-j	What technology reccomendation for AI chat bot?	["python", "botfuel"]	0	5.913124	19.41241
gpt-j	What technology reccomendation for data warehousing?	["python", "hadoop", "mysql"]	0	9.656091	32.33333
gpt-j	What technology reccomendation for building an app that can run in multiple clouds?	["aws", "aws cloudformation"]	0	0	-1
gpt-j	What technology reccomendation for project management?	["python", "sqlite", "redis", "trello"]	0	4.18121	13.43376
gpt-neox	What technologies are needed to create an web based dashboard for a cloud computing project?	["python", "qlik sense", "amazon aws", "microsoft azure"]	0	12.64279	42.64358
gpt-neox	What technology	["amazon aws", "microsoft	0	12.54363	42.30127

	reccomendation for creating an application to manage cloud computing?	azure"]			
gpt-neox	What technology reccomendation for cloud Machine Learning deployment?	["AWS machine learning"]	0	0	-1
gpt-neox	What technology reccomendation for big data analytics?	["apache spark","kafka"]	0	4.828045	15.66667
gpt-neox	What are the technologies to build chatbot?	["python","node js"]	0	14.48414	49
gpt-neox	'What technology reccomendation for building an app that can run in multiple clouds?	["AWS", "Azure", "Google Cloud"]	0	20.90605	71.16878
gpt-neox	What technologies can be used for reporting on the performance of sales?	["hadoop", "spark"]	0	5.913124	19.41241
gpt-neox	What type of technology would be best for developing mobile application?	["ios development", "mobile development"]	0	0	-1
gpt-neox	What technologies are used for creating a blog?	["django-cms", "django"]		0	-1
gpt-neox	What technology reccomendation for project management?	["pymultipy"]		0	-1