FBS-2842: Review available LLMs and gather their Open-Source Licenses

Sources
1. [GitHub - Hannibal046/Awesome-LLM: Awesome-LLM: a curated list of Large Language Model](#)
2. [The Rise of Open-Source LLMs and Its List | by Manikanth | Jul, 2023 | Medium](#)

Explored base LLM along with instruction finetuned LLM.

Model	Size	Model License	Description
Open Flamingo	9B	MIT	Open source implementation of Deepmind's Flamingo model
Alpaca 7B	7B	Apache 2.0	Stanford's Instruction-following LLaMA Model
Falcon	40B	Apache 2.0	is a foundational LLM with 40 billion parameters trained on one trillion tokens.
OpenLLaMa	7B	Apache 2.0	Another Open source reproduction of Meta AI's LLaMA 7B trained on the RedPajama dataset.
LLaMA	65B	Non-commercial bespoke license	is a family of large language models (LLMs), released by Meta AI starting in February 2023.
Baize Chatbot	30B	GPL-3.0 license	Open Source Chat model based on LLaMa
ColossalChat		Apache 2.0	A complete RLHF Pipeline released open source by ColossalAI
FastChat-T5	3B	Apache 2.0	Chatbot trained by fine-tuning Flan-t5-xl on user-shared conversations collected from ShareGPT
T5	11B	Apache 2.0	is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks and for which each task is converted into a text-to-text format.
GPT-J	6B	Apache 2.0	Eluether AI's open source version of GPT with fewer params
GPT-NeoX	2.7B	MIT	Eluether AI's open source version of GPT with fewer params
Flan-UL2	20B	Apache 2.0	Flan has been primarily trained on academic tasks. model that was trained on top of the already open sourced UL2 20B checkpoint.
Alpaca	7B	CC BY NC 4.0	Alpaca model is fine-tuned from a 7B LLaMA model on 52K instruction-following data generated by the techniques in the Self-Instruct
Orca	13B	Non-commercial bespoke license	learns to imitate the reasoning process of LFMs. Orca learns from rich signals from GPT-4
StableLM	3B	Apache 2.0	Stability AI's LLM model series
YaLM	100B	Apache 2.0	Pretrained LLM from Yandex for generating and processing text
Dolly	12B	Apache 2.0	Pythia 12B LLM trained on Databricks ML platform

For use case of chatbot recommendation tech stack: Apache 2.0 and MIT License are relevant.

License	Description
Apache 2.0	is a permissive open-source license that allows free use, modification, and distribution of the model's source code. Users are also allowed to sublicense the model under different terms.
MIT License	is another permissive open-source license that grants users the freedom to use, modify, and distribute the model's source code without any restriction. It is widely used in the open-source community due to its simplicity and flexibility.

CC BY-NC-SA-4.0	The Creative Commons Attribution NonCommercial-ShareAlike 4.0 International License allows users to use, modify, and distribute the model's source code for non-commercial purposes, as long as they provide appropriate attribution and use the same license when distributing their derivative work.
Non-commercial bespoke license	A non-commercial bespoke license is a customized software license that restricts the use of software to non-commercial purposes only. It prohibits the licensee from using the software to generate revenue, whether directly or indirectly, and may also restrict the licensee from modifying or sharing the software with others. This license is commonly used by individuals or organizations who want to share their software with others but do not want it to be used for commercial purposes.
GPL-3.0	The GNU General Public License 3.0 is a copyleft license that requires any derivative work or modifications of the model to be distributed under the same license terms. It emphasizes the principles of open-source software and ensures that the code remains freely available to the public.