SIT305: Pass Task 2.1 – Unit Converter App

Subtask 3: Research on Llama2

Llama2 is an open-source large language model (LLM) designed by Meta[i] and released on the 18th of July, 2023. It is free for research and commercial use, and includes 3 main model sizes – 7B parameters, 13B parameters, and 70B parameters[ii]. Llama2 is developed with a focus on responsible use, including a guide[iii] to ensure best practices are kept to. There are resources that help to finetune Llama2 to fit better fit a specific purpose[iv], such as being utilised as a tool within an Android application.

Below are some ideas of how Llama2 could be used in mobile applications:

Content Generation:

Llama2 is able to assist with many areas that require writing. This could include writing emails, blogs, stories, creating video scripts, generating podcast ideas, taking notes from a meeting, suggesting engaging captions for social media posts, and more.

Service Assistant / Chatbot:

If equipped with the right information, Llama2 could act as an effective assistant for customers of a service. For example, in the context of being a retail assistant, it would know the returns policy for that business and be able to advise a customer of the eligibility of their return. For an insurance business, it might have suggestions on the best insurance options for a customer with particular requirements. It could essentially act as a permanently online, well equipped staff member, available to offer immediate help.

Educational Tool:

Rather than educational resources being on rails and the same for everyone, the use of Llama2 could cater more to the users personal needs. Take learning a language for example – if the user has particular trouble with certain phrases or spelling, the model could target that area as an area requiring particular help, and offer personalised assistance to overcome the issue.

Efficiency Tool:

For users with time constraints, Llama could help condense provided documents into more digestible sizes. For example, an application that searches for the biggest news stories of the day, and then summarises them using Llama2 so that they remove unnecessary details. Or the user could task Llama2 with making the news more objective, and stripping any detected biases away.

Creative Engagement:

Llama2 can be used to provide dynamic responses to a person's input. For example, in a roleplaying game, the model could be tailored to *roleplay* as a character within a story and then respond in turn to the players interactions in a more meaningful way. This might be as an NPC in a mobile game, or as a narrator in a story, and so on.

[i] Meta (2023). Meta and Microsoft Introduce the Next Generation of Llama. [online] Meta. Available at: https://about.fb.com/news/2023/07/llama-2/.

[ii] Meta Llama. (2022). *Meta Llama 2*. [online] Available at: https://www.llama.com/llama2/.

[iii] Llama. (2022). *Responsible Use Guide for Llama*. [online] Available at: https://www.llama.com/responsible-use-guide/ [Accessed 27 Feb. 2025].

[iv] Pankaja Ambalgi (2024). *Fine-Tuning LLaMA 2: A Step-by-Step Guide - Pankaja Ambalgi - Medium*. [online] Medium. Available at: https://medium.com/%40pankaja.ambalgi/fine-tuning-llama-2-a-step-by-step-guide-12233d5f77fb [Accessed 2 Mar. 2025].