SIT305 – Task 2.1P
Llama2 Research Report

Jonathan Tynan

The Llama series of LLM's, developed by Meta, represent a significant leap forward in the Artificial Intelligence domain. Llama, followed by Llama2, are open-source AI models which mark a shift in the accessibility of advanced AI technologies. Unlike many predecessors which are proprietary and close-source, Llama and its successor Llama 2 have been released to the public domain and are free to use – including commercial use.

This level of access to AI tools is a significant departure from the trend of companies developing AI models behind closed doors while charging for access and use of the models. This decision to open-source the Llama models has paid immediate dividends, with a great example of this being the Llama.cpp project by Georgi Gerganov [1] which released shortly after Llama 1 with the intention to increase the efficiency and speed of the model. The work done in this project demonstrates the benefits that can be realised through open-source development and collaboration with the wider ecosystem of programmers and AI developers.

It's clear that this open-sourcing has paved the way to the development of enhancements and unique additions to programs, particularly on mobile devices and Android. Some use-cases are:

Enhanced Language Translation: Llama2 can enhance traditional translation tools by integrating a conversational AI chatbot, and through this chatbot users could not only translate text but interact with the chatbot. These interactions could be to simplify the language of the translation, so the user could better understand, or to provide further in-depth explanations about the meanings of words or phrases – particularly in the context given to the chatbot.

Education Tools: Llama2 could be fine-tuned to assist in learning, it could generate questions for a student to check their knowledge, offer feedback and engage with the student and provide explanations that align with the complexity level the student is comfortable with. This personalised approach could significantly improve education through allowing students to explore materials in a more engaging, and tailored manner.

Robotics Interactions: Inspired by Microsoft's PromptCraft-Robotics project [2], Llama2 could allow for the use of natural language interfaces for robotics on Android. Such applications would enable users to command and interact with robots through everyday lanmguage, while leveraging widespread smartphone usage to make robotics more accessible to the general public.

Personal Assistants: The enhancement of personal assistant apps could be made through the use of LLama2. It could access and interpret information from other applications, like emails or calendars, and then suggest schedules or actions to take. This would enable the streamlining of task planning for the end user.

Enhanced Accessibility: Llama2 could significantly improve technology accessibility. We could enable voice-controlled interfaces that understand natural language through the user of Llama2. For example, an assistant could be made for the disabled in which a user who has trouble performing certain tasks could instead leverage the AI model to translate their speech into actual actions to perform on their device, such as writing and sending emails.

References

[1] Gerganov, Georgi (2023). *llama.cpp: LLM Inference in C/C++*. URL: https://github.com/ggerganov/llama.cpp.

[2] Microsoft (2023). *PromptCraft-Robotics*. URL: https://github.com/microsoft/PromptCraft-Robotics.