**SIT 305 – Task 2.1P**

**Research Report – Llama 2**

Llama 2 is an open-source language model, which was originally developed by Meta AI last year, in 2023. It is considered to be the successor of its older version, Llama 1, with a lot of new features and innovations. It promises to revolutionize the world of AI with its improved capabilities and increased efficiency compared to other AI language models. It can prove useful to both individuals and professional users, according to some experts. Its use cases can also be applied to Android smartphones. It can generate safe and non-toxic text without any additional prompt or guidance from users without any compromise on the users’ privacy and security. (Portakal, 2023)

One of the use cases of this advanced language model in Android mobile apps include content generation for social media posts like Facebook, Instagram, X, Snapchat ,etc. (Boesch, 2023). One can perform basic mathematical operations on the go with the help of Llama Chat. Some of the benefits of this model includes users can adjust it according to specific use cases. Another uses of this highly reliable language model is that it can be used extensively in customer support applications on Android applications as a chatbot due to its high safety and security features. (Dhaduk, 2023). As a chatbot, it can offer product knowledge-based responses, and can also assist in production by guiding processes, recipes, and more. (Grzebien, 2023). Apart from that, it can help in automation by effortlessly generating detailed product descriptions and characteristics, drawing from product catalogues and schema in e-commerce applications like Amazon, Flipkart and so on.

Coming to the advantages, it requires minimal hardware demands. These lower requirements make it easier for researchers to get started and speed up innovation. Llama 2 is completely free as it does not charge for tokens, unlike ChatGPT, which was developed by OpenAI. However, it has its shares of downsides too. It has fewer parameters compared to its competitors. It can cause a slight change in its training data and can affect its generative abilities. It is also lacking in major coding abilities unlike GPT-3 language model. Another downside is that it can often outright refuse to answer a query if it deems it even mildly inappropriate. In fact, it constantly asks the users to keep the conversation civil and polite in some cases. The Llama model is only trained in one language, English, whereas GPT model is trained in many languages.

In conclusion, Llama2 is a fairly new language model which is constantly getting updated by developers. It is highly likely that we will be able to see new variants of this language model to expand its scope of users and audience. It promises a new innovation for the Generative AI space for developing safer LLMs with a wide range of applications. Like all generative AI tools, this should also be used with caution according to the specific needs as we know that AI can also have a certain number of limitations. It would be interesting to know how its potential can be utilized in the coming future as this platform keeps on developing.

**References:**

* Portakal, E. (2023), “Deep Dive Into LLaMa-2 Use Cases”, <https://textcortex.com/post/llama-2-use-cases#>8
* Boesch, G. (2024), “Llama 2: The Next Revolution in AI Language Models – Complete 2024 Guide”, <https://viso.ai/deep-learning/llama-2/#:~:text=pre%2Dtrained%20version.-,Llama%202%20Use%20Cases%20and%20Applications,moderation%2C%20and%20healthcare%20use%20cases>.
* Dhaduk, H. (2023), “Llama 2: A Comprehensive Guide”, <https://www.simform.com/blog/llama-2-comprehensive-guide/>
* Grzebien, K. (2023), “Exploring Llama 2: Investigating Capabilities and Business Potential”, <https://inwedo.com/blog/potential-of-llama-llm-model/>