

2.1P – Llama2 Report

Evelynn Parker – 219179423

Released in 2023 by Meta AI, Llama2 is a Large Language Model that aims to provide high performance in language understanding and generation tasks. It succeeds the original Llama1 model from 2022, however unlike its predecessor, Llama2 is available for free for both commercial use and use in AI research.

With the creation of the new model, Meta have strived to advance the performance capacity of smaller models rather than just increasing the parameter counts. As such, Llama2 is offered in models with 7 billion, 13 billion and 70 billion parameters, enabling smaller organisations to deploy local instances without time or infrastructure being a major limitation.

Some ideas of how Llama2 can be used in mobile applications are detailed as follows:

1. Enhancing Voice Assistants

Llama2 would be a perfect choice to enhance the capabilities of popular voice assistants like Google Assistant, Siri and Alexa because of its strong Natural Language Processing. It would enable conversational AI to feel more fluid and natural by improving understanding of complex requests and nuanced speech, as well as providing more context-aware and personalised responses to queries.

2. Customer Support Chatbots

Having an AI chatbot to handle early customer support queries in place of a person can in some cases have its benefits. While they can currently be used for simple requests and run at all hours, they use very generic language and do not always understand complicated questions. Integrating Llama2 into these chatbots provides similar benefit to the voice assistants, enabling smoother conversations that don't feel like the user is filling out a form with the skin of a chat interface on top. It would also mean that the chat bots are able to handle more advanced requests, such as common troubleshooting or guiding users through a process instead directing them to seek assistance elsewhere.

3. Writing Assistance:

If implemented into any apps that relate to writing or notetaking, whether its personal notes, emails, or novels, Llama2 could act as a useful tool to assist in any stage of the writing process. It would be able to help draft out a structure for a piece of writing to lay the user's ideas out or read over a piece and make suggestions to aspects like tone and styling, all in real time. This could be especially useful for users with weaker writing skills or users writing text in their non-native language.

4. Educational Tools:

Educational apps, for example ones focusing on language-learning such as Duolingo, could implement Llama2 as an assistant that can tutor users by offering feedback, identifying strong and weak points, or tailoring personalised exercises for a user to complete. Perhaps the biggest

feature that could be introduced would be the ability to have full length conversations with the assistant in the language that is being learnt, which many users would find to be a much more engaging and intuitive way to learn.

5. Game Development

Straying from the other discussed uses of Llama2 as a service for productivity, it could eventually also be used in game development for mobile platforms to create more engaging and interactive stories and worlds. A limiting factor in how immersive current generation games can be is that narratives are almost entirely pre-written by the developers down to every non-player character interaction and dialogue. The implementation of Llama2 in this aspect could mean that NPCs could react more in accordance with the player's actions or respond more dynamically than with a handful of predetermined responses. A good example of where a mobile app could use this would be in a text-based choose-your-own-adventure game, where instead of the game presenting the player with options to choose from to steer them down a path, they could type whatever actions they choose into a textbox and have the game interpret that, and play out their choices.