Llama 2

AI in Android apps

Overview

Llama 2 is a large language model developed by Meta. Much like other LLMs developed by other companies, it's designed to understand and generate human-like text. LLMs are AI systems trained on a lot of data which allows them to answer questions with proper context on the subject matter. Llama 2 is a significant advancement over the previous model.

Evolution

Meta introduced the original Llama model in February 2023, aiming to provide a powerful and accessible LLM for researchers and developers. Just a few months later, in July 2023, Llama 2 was released. As of writing this, the latest version is Llama 3.2 which was released in September of 2024. Each new release aims to improve the LLMs language understanding and generation capabilities. Llama 2 had doubled the context window, which allowed the model to handle even longer inputs more effectively before breaking down.

Use Cases of Llama

Llama 2 can offer a plethora of possibilities for enhancing Andriod applications. I've focused more on gaming examples since that's something I enjoy doing, but I've also included some non-gaming examples:

1. **Enhanced A.I in video games:**
   Llama 2 could potentially greatly improve NPC interactions in games. In most games, NPCs operate with scripted dialogues which can be repetitive and predictable. An LLM could generate more dynamic, context-aware responses, which would make the conversation you're having more engaging and realistic. The problem is obviously that voicing a character like that would have to be done using AI as well, and that could not only be bad because the acting would be unconvincing but also because the ethical implications of removing real human voice actors are severe for the greater industry.

2. **Improved Procedural Generation in Video Games**
   Developers could use LLMs like Llama to make more procedural content in games such as environments, levels or even maps. Games like Minecraft use complex algorithms to make new maps, but they are limited by how developers design the algorithms. New updates to those algorithms take time and are very complex. Using LLMs could potentially solve this issue and may even reduce the overhead of running those games on mobile hardware, which improves performance.

3. **Advanced Voice Recognition**
   Integrating Llama 2 with voice recognition systems can significantly enhance the accuracy and responsiveness of voice-controlled Android apps. Beyond gaming, this integration benefits various applications, such as virtual assistants, accessibility tools, and hands-free navigation systems. Users can issue complex commands or queries in natural language, and the app can interpret and respond appropriately, improving user experience and accessibility.

4. **Food App with Image Recognition and Recipe Generation**
   You could use the camera to capture images of ingredients. By integrating image recognition technology with Llama 2's language generation capabilities, the app can identify the ingredients and suggest creative recipes tailored to them. For example, a user could take a photo of the inside of their fridge, and the app might see tomatoes, basil, onions and garlic, then suggest a recipe for marinara sauce.

5. **Language Translation Apps**
   Using Llama 2 in translation apps could provide more accurate and contextually relevant translations. Apps like Google Translate often struggle with nuance and idiomatic expressions. Llama 2's deep understanding of language allows it to generate translations that consider context, resulting in more natural and reliable translations. This would be a fantastic companion app for travelers or people learning a new language.

References

- Ars Technica. (2023) 'Meta launches Llama 2, a source-available AI model that allows commercial applications'. Available at: https://arstechnica.com/information-technology/2023/07/meta-launches-llama-2-an-open-source-ai-model-that-allows-commercial-applications/ (Accessed: 11 March 2025).

- Brimit (2024) 'Cloud-Based vs. On-Premises AI Models: How to Make a Decision'. Available at: https://www.brimit.com/blog/cloud-based-vs-on-premises-ai-models (Accessed: 11 March 2025).

- Hugging Face. (2023) 'meta-llama/Llama-2-70b'. Available at: https://huggingface.co/meta-llama/Llama-2-70b (Accessed: 11 March 2025).

- IBM. (2023) 'What Is Llama 2?'. Available at: https://www.ibm.com/think/topics/llama-2 (Accessed: 16 March 2025).

- Linus Tech Tips (2023) 'Could Llama 2 Change the AI Game?'. Available at: https://www.youtube.com/watch?v=ypJjWfV5CVo (Accessed: 16 March 2025).

- Meta AI. (2023) '5 Steps to Getting Started with Llama 2'. Available at: https://ai.meta.com/blog/5-steps-to-getting-started-with-llama-2/ (Accessed: 16 March 2025).

- Meta AI. (2023) 'Meta and Microsoft Introduce the Next Generation of Llama'. Available at: https://about.fb.com/news/2023/07/llama-2/ (Accessed: 23 March 2025).

- Picovoice (n.d.) 'Local LLM for Mobile: Run Llama 2 and Llama 3 on Android'. Available at: https://picovoice.ai/blog/local-llm-for-mobile-run-llama-2-and-llama-3-on-android/ (Accessed: 23 March 2025).