Meta AI's Large Language Model (LLaMA) The LLaMA is a powerful family of autoregressive language 3 models designed to provide efficient, high-quality language understanding for both general and specialized applications.In 2024, Meta introduced LLaMA 3, representing a new generation of foundation models designed to support multilingual capabilities, advanced reasoning, tool use, and multimodal functionality across text, image, and speech.The dialogue-optimized variant, LLaMA 2-Chat, incorporates techniques such as red-teaming, safety tuning, and rejection sampling to improve alignment with human expectations for helpfulness and safety [251].LLaMA models are based on a transformer architecture and are pre-trained on a mixture of publicly available data sources, including Common Crawl, C4, GitHub, Wikipedia, books, and scientific articles.For example, LLaMA 3.2, which reaches a maximum of 405 billion parameters, shows strong capabilities in handling long-context applications and demonstrates reduced rates of hallucination.The dataset comprises around 1.4 trillion tokens, carefully curated to prioritize high-quality content and filtered to avoid duplicates and irrelevant .information