



Large Language Models

Large Language Models (LLMs) are advanced AI systems designed to comprehend, generate, and manipulate human language, enabling a wide array of natural language processing tasks. These models utilize vast amounts of data to learn complex patterns and nuances within human language, enabling tasks like text generation, translation, and sentiment analysis. LLMs, such as GPT-3 by OpenAI and BERT by Google, have substantially improved the understanding and generation of human language. They serve as the backbone for chatbots, content generation, language translation, and even personalized content recommendation systems. LLMs can be trained on colossal datasets, allowing them to understand context, grammar, and semantics, delivering responses that mimic human-like language. They're revolutionizing various industries, from customer service (with chatbots) to content creation and summarization for journalism or data analysis. With their ability to understand the context and nuances of language, LLMs enable the automation of tasks that previously required human understanding.

The size and complexity of these models demand significant computational resources for both training and deployment. Ethical concerns and biases within LLMs are topics of ongoing discussion and research within the AI community. Continued research and development in LLMs are constantly pushing the boundaries of what AI can achieve in language understanding and generation.

LLMs are a core component of the Quarkus LangChain4j extension. The extension does not serve its own LLMs, but rather provides a standard interface for interacting with many different LLMs such as OpenAI GPT-3/4, Hugging Face, and Ollama. This [interface](#) is designed to be simple and intuitive, allowing developers to quickly integrate LLMs into their applications.

Note that each LLM has a different feature set. Please check the specific documentation for the LLM you are using to see what features are available:

- [OpenAI \(GPT-3/4\)](#)
- [Hugging Face](#)
- [Ollama](#)
- [IBM BAM](#)
- [IBM watsonx.ai](#)
- [Mistral AI](#)