

LangChain4J

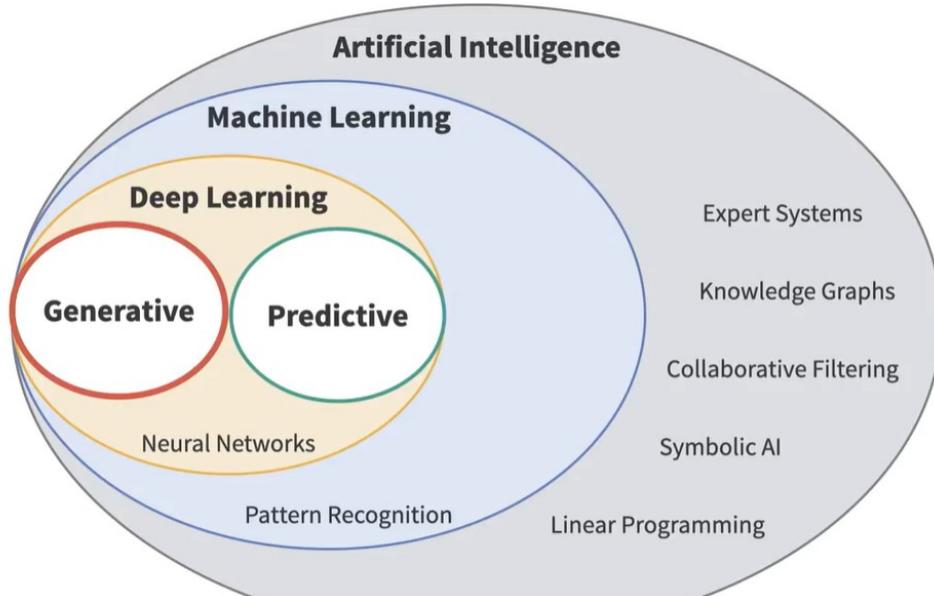
Why should we learn GenAI?

Understanding GenAI can increase your dev productivity

Your need to understand AI/ML foundations

The APIs are straightforward

Developers who know how, when, and when not to use AI will be hired over others who do not



We have Generative AI and Predictive AI

These techniques can be used wherever we have patterns

Software development process

Ideation (gather customer requirements) => Development (writing code) => Deployment (put the code into production) => Maintenance

Throughout the process, Developer has to respond to Emails and Texts, complete status reports, Meetings, Searching for information, Problem-solving. All of these involve patterns

Natural Language Processing (NLP): A set of algorithms and techniques for understanding and processing natural (human) language

Think LLMs as meta-tools. LLMs are tools for tools

We interact with LLMs using prompts and completions

User or App sends Prompt to LLM and LLM responds with a text message.

Prompt and Completion originated with completion LLMs

Completion LLMs are typically not used for chatbots

The more context we provide LLMs, the better response we get

LLM is not a search tool, it is a text-pattern recognizer instead. LLMs can be combined with a search engine. They extract patterns from ingested data and store probabilities

Creating a model is called as Training phase of LLM

When you use a trained LLM, that called as the Inference phase

During the Inference phase, we can generate new content also

Prompt engineering is a process of designing text that is used to extract useful information from a language model.

Application Programming Interface (API) and Chatbot Web Apps

All major vendors supply a chatbot web app

Chatbots connect to LLM and other GenAI (backend) services

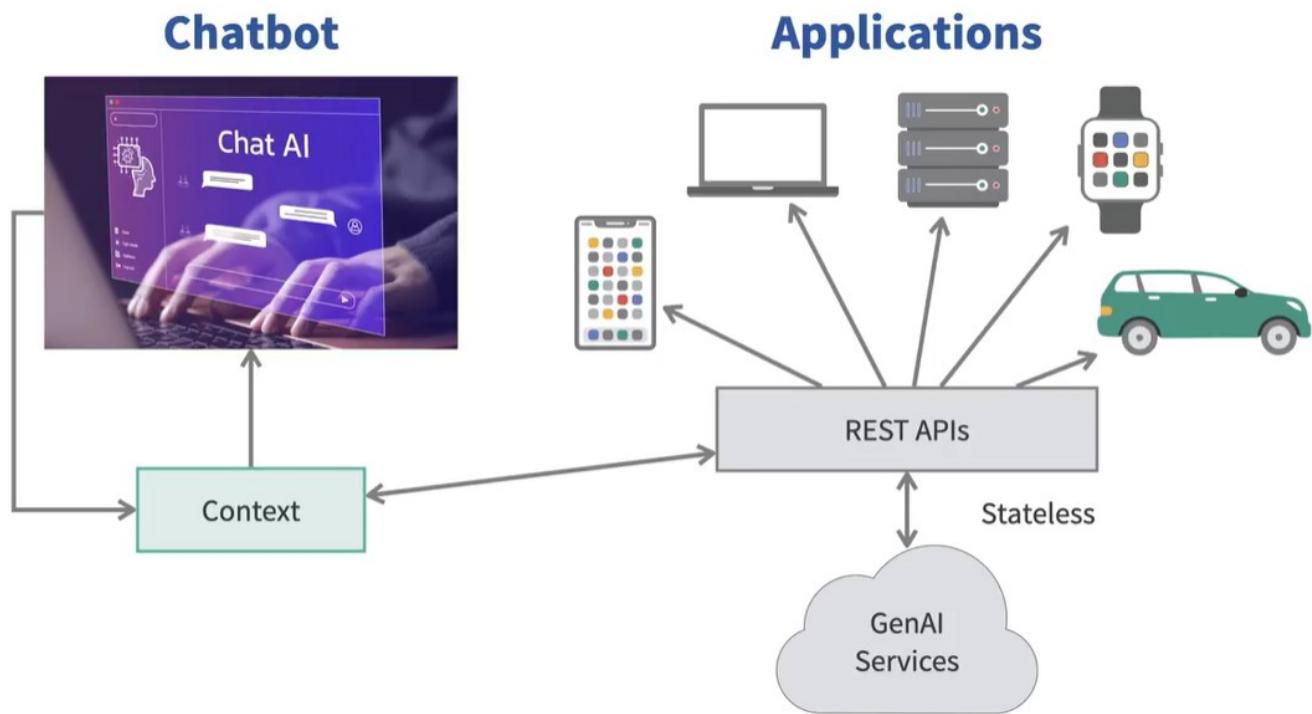
Chatbots remembers the context of your conversation

Chatbots also keep a complete history of whatever you have said

In Chatbots, LLM API is lower-level, for chatbots to function we got to add other relevant features

We need an API key for authentication

LLM APIs are stateless that means if Network connection is disrupted, LLM doesn't remember anything about your conversation



GenAI is probabilistic and nondeterministic, different from traditional IT tools

For chatbots, we have to recreate the functionality or features that we see in Web app chatbots

For non-chat application, we are taking string from user and sending them to LLM. For a simple chatbot application, chatbot maintains a context based on the user conversation. How do we implement this with a stateless API? It's our responsibility to maintain a state by sending the conversation history along with user's request. As the user continues to chat, the context or conversation history get bigger and bigger. We got to send the context to LLM repeatedly using stateless API. The downside is, we are sending more and more data to LLMs, which can be costly. Specify the output format to LLM, JSON, CSV, Graph or human languages like English, Italian, Japanese etc. It includes what the user inputted and what came out of the LLMs, sometimes called as the Assistant message.

It goes in a loop

User => request => Prompt (System message tells LLM how to behave + [memory + request] + format) => LLM to display completion => Completion => memory += completion + request

Context window (Amount of text that an LLM can process at once). Different models like GPT, OL, Gemini, Claude etc have different context windows. In very very large context windows, the LLM sometimes forgets the information in the middle of the context window.

