

Zusätzliche Informationen / Erklärungen / Definitionen zum Thema

Autonome Agenten mit CrewAI

Was ist Ollama?

Ollama ist eine Software-Plattform, die es extrem einfach macht, große Sprachmodelle (LLMs) lokal auf deinem eigenen Computer laufen zu lassen. Es ist eine super bequeme Möglichkeit, Open-Source-Modelle wie Mistral, Llama 3 oder OpenHermes auf deinem eigenen Laptop oder Server zu betreiben – ohne komplizierte Installation.

Wozu wurde Ollama entwickelt:

Gemacht für Leute die:

- Modelle lokal und ohne Cloud-Zugriff nutzen wollen (z.B. für Datenschutz oder Geschwindigkeit).
- Keine Lust auf komplizierte Setups (CUDA, PyTorch, etc.) haben.
- Schnell zwischen Modellen wechseln möchten.
- Agentensysteme (z.B. CrewAI) oder eigene KI-Anwendungen bauen wollen.

Beispielmodelle:

<u>Modell</u>	<u>Beschreibung</u>
llama3	Meta's Llama 3 Model
mistral	Schnelles 7B-Modell
openhermes	Chat-optimierte Version von Mistral
codellama	Llama-Variante für Code-Generierung

(Das Wort „Ollama“ stammt vermutlich aus dem Nahuatl (der Sprache der Azteken) und bezeichnete einen Ballspieler oder eine Person, die das rituelle Ballspiel „Öllamalitzli“ spielte – ein wichtiger kultureller und religiöser Bestandteil in präkolumbianischen Kulturen Mittelamerikas.)

Was ist LangChain?

LangChain ist ein Open-Source-Framework, das speziell dafür entwickelt wurde, Sprachmodelle (LLMs) wie GPT, LLaMA, Claude, etc. mit externen Datenquellen und Funktionen zu kombinieren.

Warum braucht man LangChain?

Sprachmodelle sind stark, aber isoliert:

- Sie „wissen“ nur, was sie beim Training gelernt haben.
- Sie können nicht direkt auf Datenbanken, APIs oder das Web zugreifen.
- Sie können nicht logisch komplexe Schritte oder Tools orchestrieren.

In LangChain ist ein „Tool“ eine Funktion oder ein Objekt, das ein Sprachmodell aufrufen darf, um zusätzliche Fähigkeiten zu bekommen – etwa: Rechnen, Dateien lesen oder Webscraping

Fazit: LangChain verbindet LLMs mit externem Wissen und Funktionalität.

Was ist DuckDuckGo?

DuckDuckGo ist eine datenschutzfreundliche Internet-Suchmaschine, die sich von Google, Bing & Co. vor allem durch Privatsphäre und Werbefreiheit unterscheidet.

- eine Websuchmaschine wie Google oder Bing,
- die keine persönlichen Daten sammelt oder speichert,
- die keine individuellen Suchprofile erstellt,
- und deren Ziel ist, neutrale, nicht personalisierte Suchergebnisse zu liefern.

Was ist DuckDuckGoSearchRun?

DuckDuckGoSearchRun ist eine LangChain-Toolklasse, die es einem Sprachmodell (LLM) ermöglicht, aktuelle Websuchen über die Suchmaschine DuckDuckGo durchzuführen.

- Ziel: Das LLM kann Informationen von außerhalb seines Trainingsdatums erhalten, z. B. aktuelle Nachrichten, Webseiteninhalte oder Zahlen.

Was macht DuckDuckGoSearchRun konkret?

- Es sendet eine Suchanfrage an DuckDuckGo.
- Holt eine kurze Zusammenfassung oder Linkliste zurück.
- Gibt diese Infos an das LLM weiter.

Beispiel:

- `search = DuckDuckGoSearchRun()`
- `result = search.run("Was ist der aktuelle Goldpreis?")`
- `print(result)`

Das LLM kann dadurch live im Web „nachschnellen“ – und mit der Info weiterarbeiten.

Vergleich von unterschiedlichen Tools:

<u>Tool</u>	<u>Funktion</u>
DuckDuckGoSearchRun	Websuche mit DuckDuckGo
SerpAPIWrapper	Websuche mit Google (SerpAPI)
ArxivAPIWrapper	Wissenschaftliche Paper suchen
WikipediaAPIWrapper	Wikipedia-Artikel suchen
RequestsGetTool	HTTP-Requests an Websites

Zusammenfassung:

Was „macht“ die Befehlszeile: `from langchain.tools import DuckDuckGoSearchRun`?

Es wird ein Tool importiert, das den Agenten oder dem Sprachmodell eine Live-Websuche über DuckDuckGo ermöglicht. Dies ist besonders nützlich, wenn dein Modell lokal läuft und kein aktuelles Wissen hat.