

Ollama aufs nächste Level gebracht

Echt starke Kombi 100% gratis! Hiermit könnt ihr so gut wie alles machen und es ist super einfach:

1. Ollama installieren:

<https://ollama.com/download/windows>

2. Das Terminal öffnen (einfach nach "PowerShell" suchen) und dort folgendes eingeben, um sich die besten universellen Modelle zu ziehen:

1. `ollama pull qwen3.5:9b`

Gutes Modell für 16 GB RAM (nutzt bei mir sogar deutlich weniger);

Größe: 6.6GB;

2. `ollama pull qwen3.5:4b`

Noch weniger RAM-Verbrauch;

Größe: 3.4GB;

3. `ollama pull gemma4:e4b`

Auch ein gutes Modell, fordert aber ein bisschen den Rechner (braucht länger), bietet aber auch sehr gute Antworten;

Größe: 9.6GB;

4. `ollama pull gemma4:e2b` Und auch dafür noch eine kleiner Variante;

Größe: 7.2GB;

3. Slashspace installieren:

<https://www.slashspace.ai>

Es gibt eine Gratis-Variante, mit der man Ollama nutzen kann. Dafür muss man sich einen kostenlosen Account erstellen. Mit Slashspace kann man Nodes erstellen, die miteinander verbunden und auch wieder getrennt werden können, alles visuell - das hilft enorm, um den Kontext zu steuern. Hinzu kommt die RAG-Funktionalität von Slashspace: Lade eine Riesen-Datei hoch (z.B. die Dokumentation zu einer aktuellen Software) und chatte mit ihr (RAG verkleinert einfach gesagt den Kontext aufs Wesentliche, sodass dein Rechner nicht die kommenden 10 Jahre mit der Frage beschäftigt ist^^);

4. Page Assist (Chrome Addon) installieren, um komfortabel Ollama einzusetzen:

<https://chromewebstore.google.com/detail/page-assist-a-web-ui-for/jfgfiigpkhikbnfnbobbkinehhdhndo>

Es ist kinderleicht. Dann hat man einen Button, mit dem man einen Chat öffnet, der aufgeräumt und simpel ist.

5. Schau manchmal, ob es Updates gibt (neue Modelle oder Updates zu deinen installierten Modellen, und natürlich zu Ollama selbst). - Hier sind die URLs zu den oben genannten Modellen:

1. <https://ollama.com/library/qwen3.5>

2. <https://ollama.com/library/gemma4>

Revision #2

Created 2026-06-03 01:23:34 UTC by art10m

Updated 2026-06-03 01:28:12 UTC by art10m