Umsetzung mit Lokaler KI

oer.one/ki-lokal/alf

Laptop

Endgerät

Hardware Voraussetzung

- Technisches Know-How: mittel
- Betriebssystem: Windows, macOS oder Linux
- RAM: mindestens 8 GB
- Grafikkarte: keine Vorraussetzung
- Speicherplatz: je nach Modell, ab 2 GB
- Internet: nach dem Download nicht notwendig

Anleitung

Schritt 1: Webseite öffnen

Rufe die Projektseite llamafile auf (z. B. über Mozilla.ai oder GitHub) und wähle ein passendes Llamafile-Modell aus der Liste der Beispielmodelle aus.

Schritt 2: Llamafile herunterladen

Lade die gewünschte Datei herunter. Es handelt sich um eine einzelne Datei, die bereits das Modell und die Chat-Oberfläche enthält.

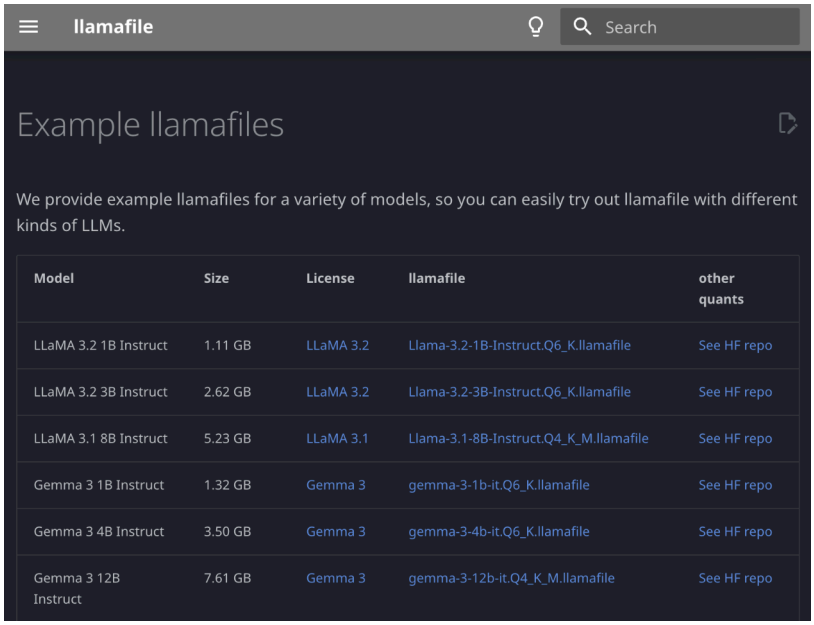
Schritt 3: Datei ausführbar machen

- Windows: Datei in *.exe umbenennen
- macOS / Linux: Datei ausführbar machen (chmod +x dateiname)

Schritt 4: KI nutzen

Llamafile in der Kommandozeile starten.

Die KI läuft nun vollständig lokal auf dem eigenen Gerät. Fragen können direkt in der Kommandozeile oder im Browser gestellt werden. Es ist keine Anmeldung oder Cloudzugang erforderlich.



Liste an Llamafiles

Quelle: https://mozilla-ai.github.io/llamafile/example_llamafiles

Lokale Chat-Oberfläche im Browser

Benedikt Brünner (TU Graz), Andreas Zitek (BOKU University), Martin Ebner (TU Graz)