WELT 13: HEIMPLANET LOKAL · MISSION 13.1

RAM, VRAM und Realität: Laptop-Triage

Sechs Geräte aus dem Lehrerzimmer liegen in der Notaufnahme. Sie entscheiden: Läuft hier eine eigene KI? Am Ende hat Ihr eigener Laptop einen Befund – und einen Plan B.

Zur Mission

DAUER 30 Min. KOSTEN Gratis WERKZEUGE Papier und Stift, Terminal, ChatGPT, Claude (Web), Google Gemini

DATENAMPEL Grün – unbedenklich

Das Wichtigste

- Ein lokales Modell wird beim Laden ganz in den Speicher geholt – RAM und Grafikspeicher entscheiden.
- Faustregel Q4: Datei ca. 0,6 GB pro Milliarde Parameter, im Speicher etwa das 1,2- bis 1,5-Fache (Näherung).
- 8 GB RAM: kleine Modelle (1–4B). 16 GB RAM: 4–8B. Eigene Grafikkarte mit 6–8 GB VRAM: auch 7–12B.
- LM Studio auf dem Mac: nur Apple-Chip (M1 oder neuer) und macOS 14+. Windows x64: AVX2 nötig, 16 GB RAM empfohlen.
- Faustregel: Grafikspeicher (VRAM) macht oft schneller, RAM macht möglich.
- Plan B: kleineres Modell, Ollama oder Jan, anderes Gerät – oder die Webseite zuerst mit Claude im Browser bauen.
- In KI-Prompts nur technische Daten angeben, keine Seriennummer oder Gerätenamen.

Ihr Produkt

Ein **Geräte-Steckbrief** für Ihren Laptop (Ampelfarbe, realistische Modellgrösse, Plan B) und ein **Triage-Bogen** mit sechs Fallkarten, den Sie im Kollegium oder mit ICT-Lernenden einsetzen können.

Abnahme: Alle sechs Patientenkarten tragen eine Farbe mit einer Begründung aus RAM, Grafikspeicher oder Systemanforderung. Haben Sie ein eigenes Gerät, nennt Ihr Geräte-Steckbrief zusätzlich Ampelfarbe, passende Modellgrösse und einen Plan B.

So gehen Sie vor

1. Bauchgefühl-Runde: zwei Minuten, sechs Patienten
2. Die drei Vitalwerte
3. Triage mit Befund
4. Ihr Laptop auf der Liege
5. Plan B und Steckbrief fertig

Start-Prompt

Ich bin Lehrperson an einer Schweizer Berufsfachschule und will ein Sprachmodell lokal mit LM Studio betreiben.
Erkläre mir in höchstens 150 Wörtern die Begriffe «Parameter», «Quantisierung (Q4)», «RAM» und «VRAM» – mit je einem Alltagsvergleich.
Rechne mir dann vor, wie viel Speicher ein Modell mit 9 Milliarden Parametern in Q4 ungefähr braucht. Nimm als Faustregel 0,6 GB Datei pro Milliarde Parameter und das 1,2- bis 1,5-Fache davon im Speicher.
Sag mir zum Schluss ehrlich, wo diese Faustregel ungenau ist.
Schweizer Rechtschreibung (immer «ss», kein Eszett).

Qualitätscheck

- Jede der sechs Fallkarten hat eine Farbe **und** eine Begründung, die einen konkreten Vitalwert oder eine Systemanforderung nennt.
- Ihre Angaben zu Systemanforderungen stammen aus der offiziellen LM-Studio-Doku, nicht nur aus einer KI-Antwort.
- Der Steckbrief nennt eine konkrete Modellgrösse (z. B. «4B, Q4») und einen Plan B.
- In Ihren Prompts stehen nur technische Daten – keine Seriennummer, kein Gerätename, kein Benutzername.

Hängen Sie fest? Frag die KI

Ich bin Lehrperson an einer Schweizer Berufsfachschule und habe noch nie ein Sprachmodell lokal auf meinem Laptop betrieben. Erkläre mir in einfachen Worten und mit einem Alltagsvergleich, warum eine KI auf dem eigenen Gerät so viel Arbeitsspeicher braucht. Nenne mir danach 2–3 Suchbegriffe, mit denen ich verlässliche Erklärungen finde (keine Links erfinden).

Klappt nicht? Die Triage geht ganz ohne KI, nur mit dem Triage-Bogen und den Angaben Ihres Geräts. Für die Erklär-Prompts eignet sich jede Gratis-KI, zum Beispiel Microsoft Copilot oder Mistral Vibe (früher Le Chat). Ohne eigenen Laptop triagieren Sie nur die sechs Patienten – das ist der eigentliche Lernkern.

Stand 19.11.2026 · CC BY 4.0 Alpasana GmbH · Werkzeuge ändern sich – prüfen Sie selbst.