

Zusatzinhalt

zu »Linux statt Windows«

von
Jürgen Wolf



Zusatzinhalt

GPT4All: Lokale KI mit grafischer Oberfläche

GPT4All ist eine grafische Oberfläche für lokale KI-Modelle. Im Unterschied zur rein terminalbasierten Nutzung mit Ollama erhalten Sie hier eine Bedienoberfläche, die eher an ChatGPT oder andere bekannte KI-Dienste erinnert.

Das Besondere an GPT4All ist, dass es von Anfang an so optimiert wurde, dass es auf herkömmlichen Rechnern läuft. Sie benötigen also keinen High-End-Rechner. Die Entwickler nutzen spezielle Komprimierungstechniken (Quantisierung), sodass selbst ein Mittelklasse-Laptop mit Zorin OS in Sekundenschnelle intelligente Antworten generiert. Zudem können Sie das Sprachmodell jederzeit ändern oder neue hinzufügen. Und ganz wichtig: Bei GPT4All steht der Datenschutz an oberster Stelle, sodass bei der lokalen Nutzung nichts von dem, was Sie eingeben, nach außen gelangt.

Schritt für Schritt

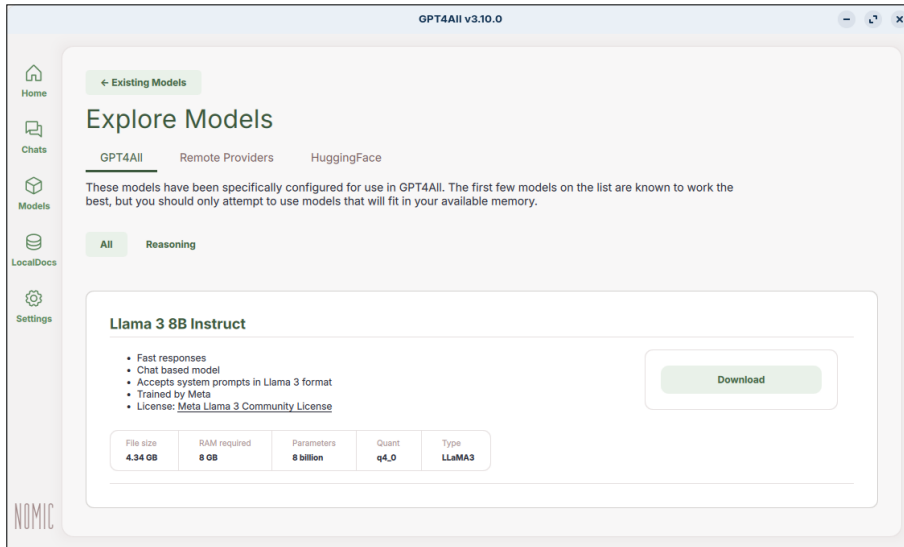
Ollama-Modelle in GPT4All nutzen

1. GPT4All herunterladen

Suchen Sie im Software Center nach »GPT4All«, und klicken Sie auf **Installieren**.

2. GPT4All starten

Starten Sie GPT4All, und wählen Sie in der Seitenleiste den Menüpunkt **Models** aus. Klicken Sie auf die Schaltfläche **Add Model**, und wählen Sie hier das gewünschte Sprachmodell aus. Im Beispiel wählen Sie **Llama 3 8B Instruct** aus, was dem Modell entspricht, das Sie bereits im Terminal installiert haben. Klicken Sie auf **Download**.



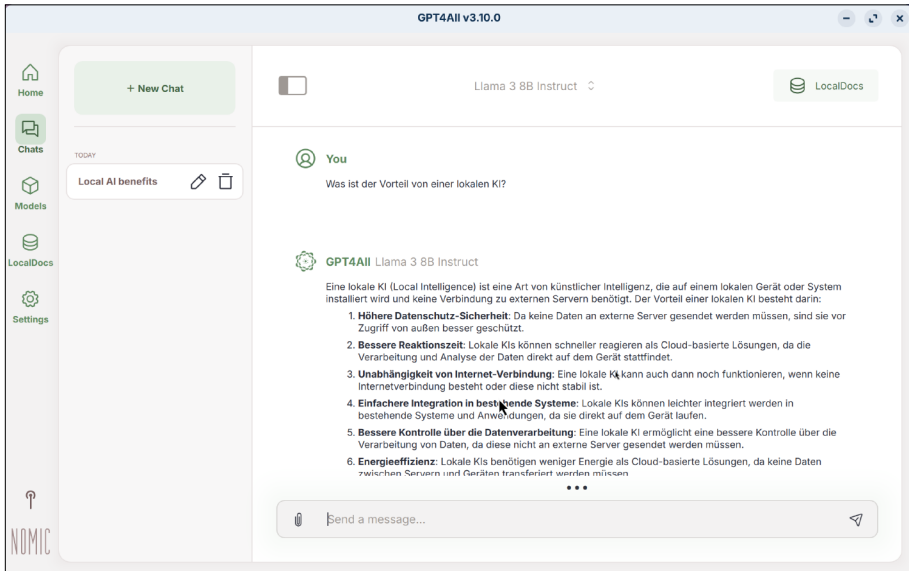
Heruntergeladenes Modell nutzen?

Sie fragen sich, ob Sie nicht das Modell nutzen können, das Sie bereits für Ollama heruntergeladen haben? Das ist theoretisch möglich, aber unter Linux sind die Systemordner, in denen Ollama arbeitet, oft für normale Programme gesperrt. Im vorliegenden Fall wäre dies: `/usr/share/ollama/.ollama/models`

Mein Rat: Laden Sie das Modell einfach separat in GPT4All herunter. Das belegt zwar etwa 5 GB zusätzlichen Speicherplatz, ist aber die stressfreieste Lösung. So stellen Sie sicher, dass GPT4All alle nötigen Zugriffsrechte hat und sofort reibungslos funktioniert, ohne dass Sie sich mit den technischen Dateipfaden von Linux auseinandersetzen müssen.

3. KI mit GPT4All verwenden

Wenn Llama3 heruntergeladen wurde, können Sie über Chats bereits anfangen, mit der lokalen KI zu kommunizieren und Ihre Fragen zu stellen – jetzt aber mit einer echten grafischen Oberfläche, wie Sie das von ChatGPT und Co. her kennen. Wählen Sie ganz oben das Modell aus (hier: **Llama 3 8B Instruct**), und schreiben Sie Ihre Frage im Eingabeprompt. Sie können natürlich jetzt jederzeit weitere Sprachmodelle installieren und verwenden.



Wenn die KI länger mit der Antwort braucht

Lokale KI liebt Grafikspeicher (VRAM). Wenn Ihre Grafikkarte weniger als 4 GB VRAM besitzt, kann es bei Llama 3 zu Verzögerungen kommen. Die Lösung: Greifen Sie zu kompakteren Modellen wie Phi-3, die speziell dafür optimiert sind, auch auf durchschnittlichen Laptops oder älteren Desktop-PCs blitzschnell zu antworten, ohne die Hardware zu überfordern.

LocalDocs: Mit eigenen Dokumenten arbeiten

Mit der Funktion **LocalDocs** in der Seitenleiste von GPT4All wird aus der lokalen KI ein echtes Produktivitätswerkzeug. Mit LocalDocs können Sie auf Dokumente zugreifen, die auf Ihrer Festplatte gespeichert sind. Dadurch lässt sich die KI gezielt mit Ihren eigenen Texten, PDFs oder Notizen verwenden.

Ein mögliches Einsatzgebiet ist das persönliche Wissensmanagement. Wer über Jahre Notizen, Tagebucheinträge oder Konzepte sammelt, kann diese Sammlung mit LocalDocs durchsuchen und zusammenfassen lassen. Anstatt mühsam Ordnerstrukturen zu durchforsten, genügt eine einfache Frage an den Chatbot, um die gewünschte Information aus Tausenden von Textdateien zu extrahieren.

Auch im Bereich der privaten Verwaltung bietet LocalDocs einen immensen Zeitvorteil. Versicherungsverträge, Mietvereinbarungen oder komplizierte Gebrauchsanweisungen liegen oft als staubige PDFs auf der Festplatte. Durch die Indizierung dieser Dokumente fungiert die lokale KI als Ihr persönlicher Rechts- oder Finanzassistent. Sie können direkt Fragen stellen, zum Beispiel: »Wie lang ist meine Kündigungsfrist?«, oder »Welche Leistungen sind in meiner Zusatzversicherung abgedeckt?«. Die KI liefert die Antwort sofort und unter Angabe der entsprechenden Textpassage. Das mühsame Lesen von Kleingedrucktem wird Ihnen somit erspart.

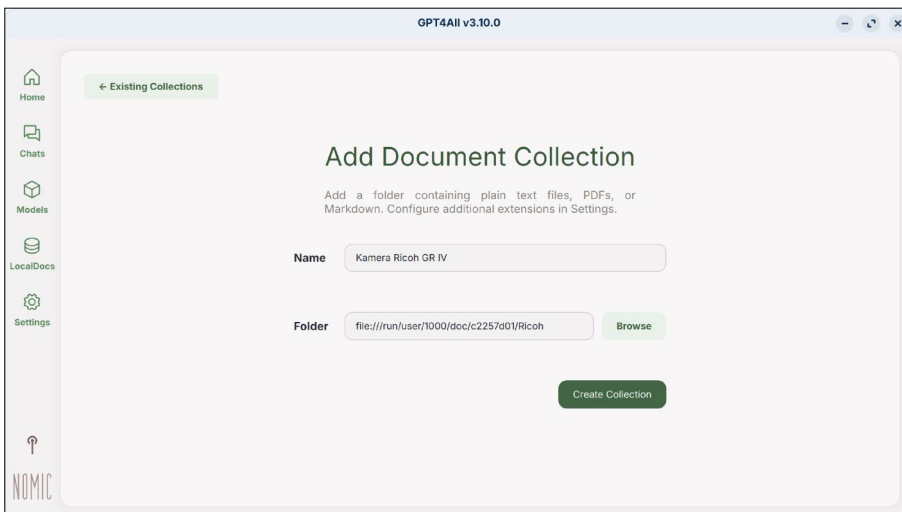
Die LocalDocs-Funktion einrichten

1. LocalDocs aufrufen

Klicken Sie in der Seitenleiste auf **LocalDocs**.

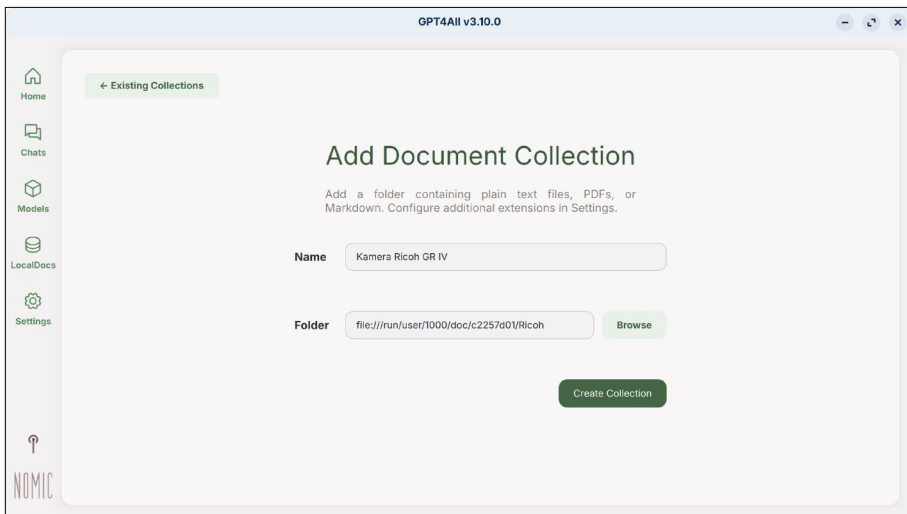
2. Dokumentensammlung hinzufügen

Klicken Sie auf die Schaltfläche **Add Doc Collection**, und geben Sie einen Namen für die Sammlung ein. Wählen Sie den Ordner aus, in dem die Dokumente auf der Festplatte gespeichert sind, und klicken Sie dann auf **Create Collection**. GPT4All liest nun die Daten ein, was je nach Rechenleistung und Datenumfang im Ordner etwas dauern kann. Ich habe zum Beispiel einen Ordner angelegt mit Dokumentationen und Anleitungen zur Kamera Ricoh GR IV. Sie können dabei mehrere Sammlungen für verschiedene Themen anlegen. Wichtig dabei ist nur, dass Sie die Themen nicht miteinander vermischen.



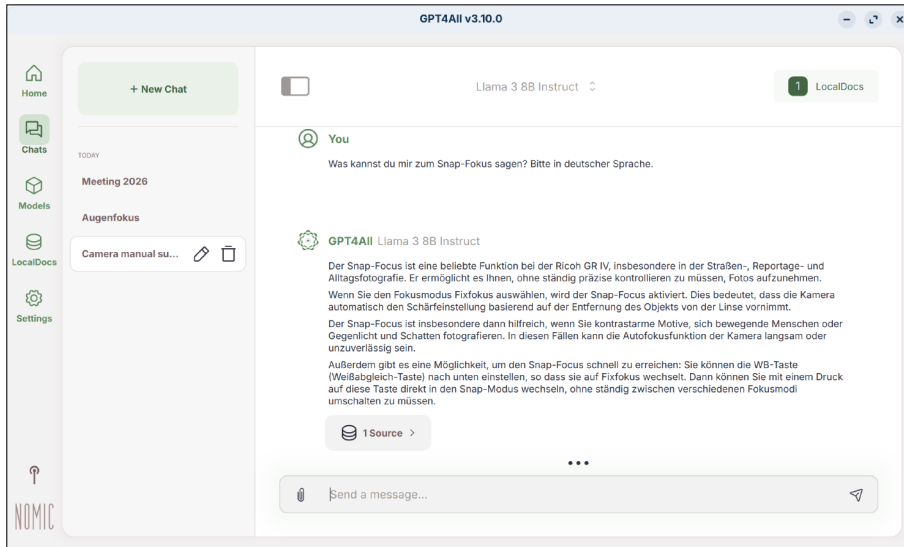
3. Sammlung auswählen

Wählen Sie nun **Chats** aus, und klicken Sie oben rechts auf **LocalDocs**. Nun werden auf der rechten Seite alle indizierten Sammlungen aufgelistet. Wählen Sie hier die Sammlungen aus, mit denen Sie im Chat interagieren möchten. Ich wähle die Sammlung für die Kamera Ricoh GR IV aus. Leider wird bei mir derzeit ein visueller Fehler angezeigt, der die Funktion aber nicht beeinträchtigt. Bei Bedarf können auch mehrere Sammlungen gleichzeitig ausgewählt werden.



4. Fragen stellen

Wenn Sie jetzt eine Frage stellen, greift die KI auf die Inhalte Ihrer ausgewählten Dokumentensammlung zurück. Sie könnten beispielsweise fragen: »Fasse die wichtigsten Punkte aus meinem neuen Projektvertrag zusammen.« Die Antwort basiert dann auf Ihren lokalen Dateien, ohne dass diese an einen Cloud-Dienst übertragen werden. Alles ohne eine Internetverbindung. Zusätzlich werden auch die Dateien der Sammlung genannt, aus denen diese Informationen stammen. Wenn Sie keine Frage in das Eingabefeld eingeben können, dann liegt dies meistens daran, dass Sie oben in der Mitte kein Sprachmodell ausgewählt haben.



Damit dieses System reibungslos funktioniert, ist es sinnvoll, die Dokumente thematisch geordnet abzulegen. Die KI arbeitet am präzisesten mit textbasierten Formaten wie PDF, TXT oder Markdown-Dateien. Durch getrennte Sammlungen stellen Sie sicher, dass Inhalte aus unterschiedlichen Bereichen nicht unnötig miteinander vermischt werden. So schaffen Sie eine digitale Bibliothek, die nicht nur Platz auf der Festplatte belegt, sondern Ihnen aktiv bei der täglichen Arbeit mit Zorin OS zur Hand geht.