

# Lokale Sprachmodelle mit GPT4All nutzen

## Was ist GPT4All / Einleitung

GPT4All ist eine Anwendung für Windows, MacOS und Linux Betriebssysteme, die es ermöglicht, große Sprachmodelle (LLMs: Large Language Models) auf dem privaten Computer zu nutzen. Das wohl bekannteste LLM ist zurzeit ChatGPT und wird von der amerikanischen Firma OpenAI entwickelt und bereitgestellt. Dieses Sprachmodell läuft jedoch nicht auf dem privaten Rechner der Nutzer:innen, sondern ist eine Cloud-basierte Lösung. Dies bedeutet für die Nutzer:innen, dass ihre Eingaben übertragen und auf den Servern von OpenAI verarbeitet und generell auch gespeichert werden. GPT4All hingegen verfolgt einen anderen Ansatz: Hier können die Benutzer:innen Sprachmodelle als lokale Anwendung auf ihrem Computer ausführen. Die Anwendung lädt nach der Installation die gewünschten Sprachmodelle auf den Computer der Nutzer:innen nach. Diese lassen sich dann ausschließlich lokal benutzen: Die Eingaben der Nutzer werden ausschließlich auf dem Computer lokal verarbeitet, verlassen diesen nicht und werden auch nicht von Dritten gespeichert. GPT4All kann somit zu besserem Schutz der Privatsphäre beitragen und eine ungewollte Speicherung und Verarbeitung der Daten durch Konzerne verhindern.

## Was kann GPT4All?

GPT4All macht die Nutzung von LLMs auf handelsüblichen PCs möglich. Da LLMs normalerweise viele GB an speziellem Speicher für die Berechnung der Anfragen benötigen, ist eine Nutzung auf herkömmlichen Computern nicht möglich und diese werden in der Regel von spezialisierten Servern durchgeführt. Durch neue Entwicklungen im Bereich der Künstlichen Intelligenz ist es mittlerweile möglich geworden, diese Modelle zu „schrumpfen“ und auf herkömmlicher PC-Hardware in angemessener Geschwindigkeit zu nutzen. Mittlerweile gibt es eine große Anzahl verschiedener open-source LLMs. GPT4All stellt für die Nutzer:innen eine relativ einfach zu nutzende Oberfläche bereit, um die verschiedenen Modelle mit einem Mausklick sowohl herunterzuladen als auch mit Ihnen zu kommunizieren. Somit ist es auch Laien möglich, die Sprachmodelle auf Ihren heimischen Computern zu betreiben, ohne teure Lizenzgebühren an KI-Firmen wie OpenAI zu bezahlen.

## Warum macht GPT4All Sinn?

GPT4All bietet eine kostengünstigere Alternative zu kommerziellen LLMs wie GPT-3.5 oder GPT-4.0, die aufgrund ihrer proprietären Natur und Lizenzgebühren teuer sein können. Dadurch werden LLMs für Forscher, Entwickler und Unternehmen mit begrenztem Budget zugänglich. Des Weiteren fördert GPT4All als Open-Source-Projekt die Zusammenarbeit unter Forschern und Entwicklern, indem es ihnen ermöglicht, ihr Wissen und ihre Ressourcen beizutragen, um die Modelle zu testen und zu verbessern. Dieser transparente Ansatz kann dazu beitragen, das Vertrauen in die KI-Technologien zu fördern. Die Auswahl verschiedener Modelle ermöglicht es den Nutzer:innen weiterhin, Modelle zu auswählen, die speziell für ein bestimmtes Aufgabenfeld trainiert wurden. So gibt es z.B. Modelle, die

auf Programmieraufgaben oder bestimmte Sprachen trainiert wurden, oder etwa auch kleinere Modelle, die auf weniger leistungsfähigen Computer lauffähig sind. Als größter Vorteil erweist sich jedoch klar der Datenschutz, da die Architektur keinerlei Daten an Dritte weitergibt und somit der Privatsphäre der Benutzer:innen besonders Rechnung trägt.

## Installationsanleitung

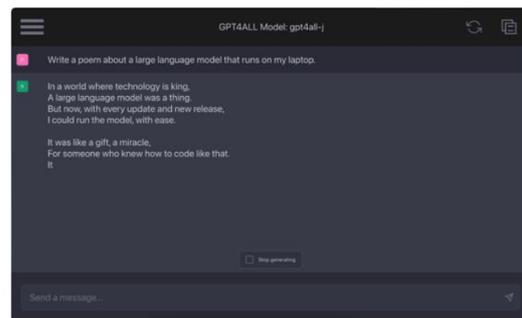
1. Öffnen Sie in Ihrem Browser die Adresse [gpt4all.io](https://gpt4all.io)
2. Laden Sie die Installationsdatei unter dem Punkt „**Download Desktop Chat Client**“ für Ihr jeweiliges Betriebssystem herunter (Abb. 01)
3. Führen Sie die Installationsdatei auf Ihrem Computer aus (Windows/Linux) bzw. kopieren Sie die Datei in Ihren Anwendungsordner (MacOS).
4. Starten Sie nun die GPT4All Anwendung.
5. Um die Anwendung zu nutzen, müssen Sie noch ein lokales Sprachmodell in der Anwendung auswählen und herunterladen: Klicken Sie auf das Menüsymbol oben links und dann auf „Downloads“. (Abb. 02 und 03)
6. Wählen Sie ein Modell aus der Liste zum Download aus und klicken Sie auf „Download“. Für deutsche Nutzer ist „EM German Mistral“ zu empfehlen, für englische Konversationen und Code-Generation „Mistral Open Orca“. (Abb. 04)
7. Sobald das Modell heruntergeladen ist, können Sie die Anwendung nutzen. Bitte wählen Sie noch das gewünschte Modell mittig über dem Chatfenster aus, dann können sie die Konversation beginnen. (Abb. 05)

## Abbildungen

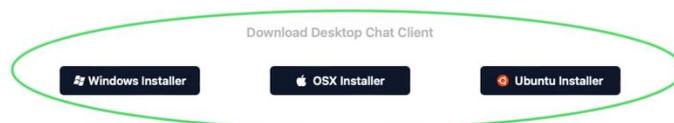
01

**GPT4All**

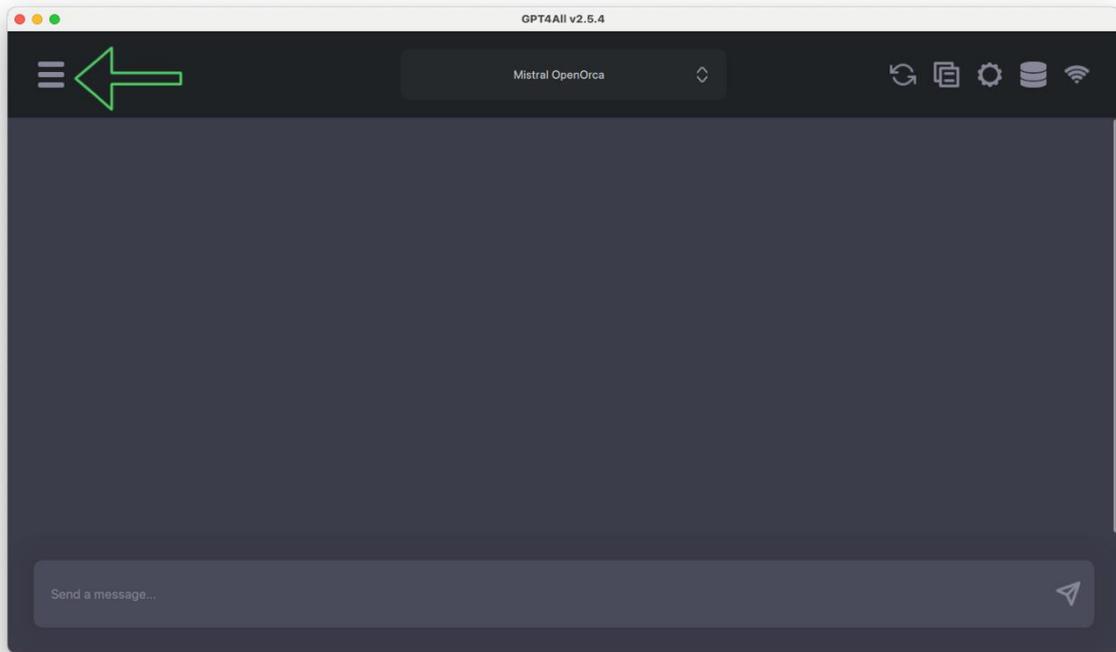
A free-to-use, locally running, privacy-aware chatbot. **No GPU or internet required.**



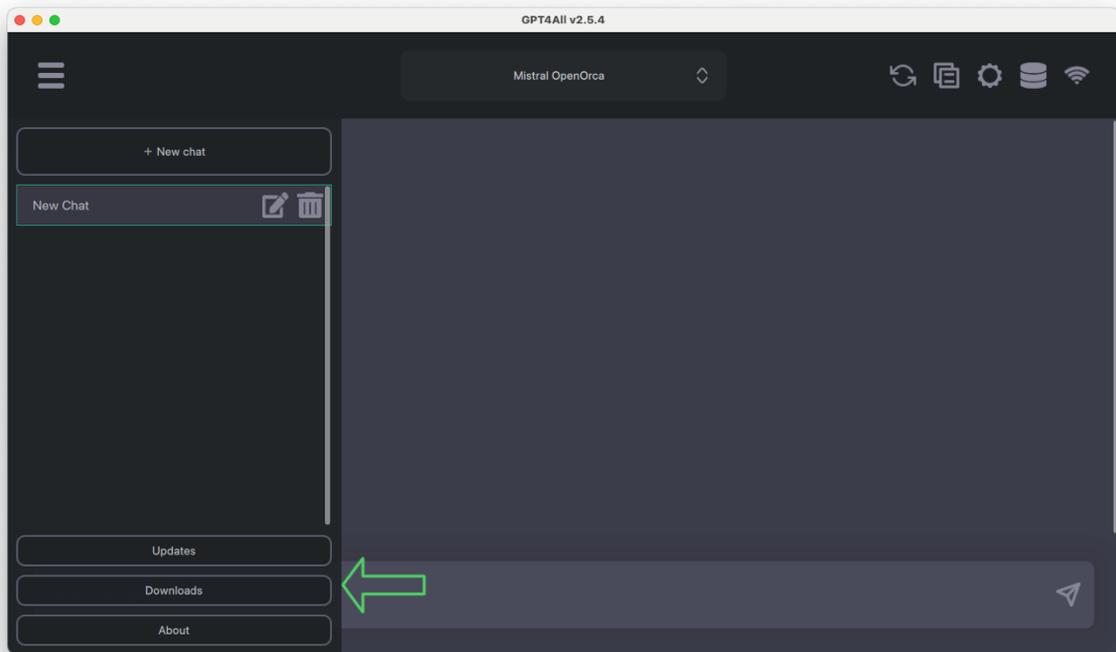
Real-time inference latency on an M1 Mac



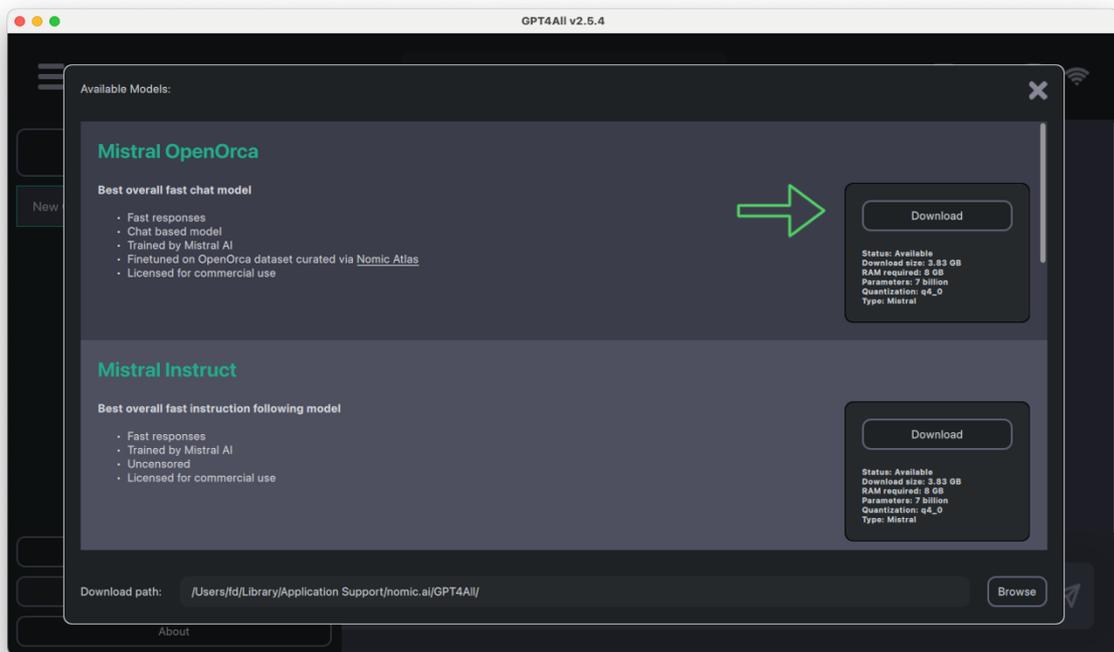
02



03



04



05

