

Übersicht gängiger KI-Plattformen (Stand April 2026) - Arbeitsdokument

Hinweis:

Diese Tabelle beschreibt **Plattformen**, nicht Modellfamilien. Eine Plattform kann eigene Modelle anbieten, Drittmodelle bündeln oder vor allem die lokale Ausführung vereinfachen. ([Ollama Dokumentation](#))Kurze Einordnung

Übersicht gängiger KI-Plattformen mit Anbindung

| Plattform | Anbieter | Typ | Typischer Einsatz | Lokal / Cloud | Anbindung / Ansprache | Typische Besonderheiten |
|---------------|----------|--------------------------------------|--|------------------------|--|--|
| Ollama | Ollama | Laufzeitumgebung und Modellplattform | Lokales Ausführen von LLMs, optional Cloud-Nutzung | Lokal und Cloud | Lokale REST-API über <code>localhost</code> , zusätzlich OpenAI-kompatible Endpunkte ; dadurch oft auch über Tools nutzbar, die OpenAI-kompatible APIs erwarten. Lokal ist standardmäßig keine Authentifizierung nötig. (Ollama Dokumentation) | Lokaler Betrieb, OpenAI-Kompatibilität, Tool-Support |

| | | | | | | |
|-------------------|------------|--------------------------------------|--|------------------------|---|---|
| LM Studio | LM Studio | Desktop-Plattform für lokale Modelle | Modelle lokal laden, testen und per UI oder lokaler API nutzen | Vor allem lokal | Lokaler API-Server auf <code>localhost</code> oder im Netzwerk; nutzbar per REST API , über eigene Client-Bibliotheken und über OpenAI- sowie Anthropic-kompatible Endpunkte . (LM Studio) | Starker Fokus auf lokale/private Nutzung |
| OpenRouter | OpenRouter | Modell-Router / Unified API | Zugriff auf viele Modelle über eine einheitliche API | Cloud | Einheitliche API , die ausdrücklich mit dem OpenAI SDK nutzbar ist; OpenRouter beschreibt die Schemas als sehr ähnlich zur OpenAI Chat API. (OpenRouter) | Ein API-Zugang für viele Modelle und Provider |

| | | | | | | |
|--------------------------------|-------------------|---|--|---------------------|--|---|
| <p>Perplexity API</p> | <p>Perplexity</p> | <p>Such- und Antwortplattform / API</p> | <p>Websuche, Recherche, Sonar, Agenten-Workflows</p> | <p>Cloud</p> | <p>REST und SDKs ; Perplexity nennt vier Kern-APIs: Agent , Search , Sonar und Embeddings . Die Sonar API ist zusätzlich mit OpenAI-kompatiblem Clients nutzbar. (Perplexity)</p> | <p>Fokus auf webgestützte Antworten und Recherche</p> |
| <p>OpenAI Plattform</p> | <p>OpenAI</p> | <p>Modell- und API-Plattform</p> | <p>Produktive KI-Anwendungen, Agenten, multimodale Workflows</p> | <p>Cloud</p> | <p>REST, Streaming, Realtime APIs sowie offizielle SDKs ; zentrale Schnittstelle ist die Responses API . (OpenAI Entwickler)</p> | <p>Herstellerplattform mit breiter Tool- und Agentenunterstützung</p> |

| | | | | | | |
|--------------------------------------|--------------|---------------------------------|--|--------------|---|--|
| Claude Platform | Anthropic | Modell- und API-Plattform | Claude-basierte Anwendungen und Assistenzsysteme | Cloud | Messages API plus offizielle SDKs ; Anthropic dokumentiert den Einstieg über API-Key, Quickstart und Client-SDKs. (Claude API Docs) | Herstellerplattform für Claude-Modelle |
| Gemini API / Google AI Studio | Google | Modell- und Entwicklerplattform | Prototyping, multimodale Anwendungen, API-Nutzung | Cloud | REST API und offizielle SDKs ; Google AI Studio dient als schneller Einstieg und zur API-Key-Erstellung. (Google AI for Developers) | Schneller Einstieg über AI Studio |
| Vertex AI | Google Cloud | Enterprise-AI-Plattform | Produktive KI-Anwendungen, verwaltete Bereitstellung | Cloud | Typisch über Google-Cloud-/Vertex-AI-APIs und SDKs; Anthropic dokumentiert Claude-Integrationen ausdrücklich auch über Google Vertex AI . (Anthropic) | Enterprise-Betrieb und Multi-Modell-Umgebungen |

| | | | | | | |
|---|--------------|---------------------------------------|---|--------------|---|---|
| Hugging Face Inference Providers | Hugging Face | Modell-Hub und Inferenzplattform | Modelle testen, serverless nutzen oder produktiv anbinden | Cloud | Zugriff über einheitliche API und Client-SDKs ; Hugging Face nennt ausdrücklich einen OpenAI-kompatible n Endpoint für Inference Providers. (Hugging Face) | Viele Modelle, viele Inferenzpartner |
| IONOS AI Model Hub | IONOS | Europäische Modell- und API-Plattform | API-Zugriff auf Modelle in europäischer Cloud | Cloud | API-first ; IONOS bietet eine reguläre API sowie eine OpenAI-kompatible API . (docs.ionos.com) | Europäischer Fokus, OpenAI-Kompatibilität |
| STACKIT AI Model Serving | STACKIT | Europäische AI-Serving-Plattform | Sichere Modellnutzung in der STACKIT Cloud | Cloud | Modelle sind über API nutzbar; STACKIT nennt die Schnittstelle ausdrücklich OpenAI-kompatibel . (Docs) | Europäische/souveräne Cloud-Ausrichtung |

| | | | | | | |
|---|----------------------|--|---|---|--|--|
| <p>Amazon Bedrock</p> | <p>AWS</p> | <p>Voll gemanagte GenAI-Plattform</p> | <p>Zugriff auf viele Foundation Models in AWS</p> | <p>Cloud</p> | <p>AWS API , API Keys , regionale Endpunkte und AWS SDKs wie Boto3; Amazon dokumentiert zusätzlich OpenAI-kompatible Service-Endpunkte in Teilen des Angebots. (AWS Dokumentati on)</p> | <p>Viele Drittmodelle unter AWS-Steuerung</p> |
| <p>noris network KI as a Service</p> | <p>noris network</p> | <p>Infrastruktur- und Serviceplattform</p> | <p>KI-Betrieb für Unternehmen , souveräne Hosting- /Cloud- Umgebungen</p> | <p>Cloud / Private Cloud / Managed</p> | <p>Anbieter beschreibt KI as a Service , Private-Cloud- und GPU-nahe Betriebsmodelle; die konkrete technische Ansprache ist eher projekt- bzw. serviceabhängig als über eine standardisierte öffentliche Entwickler-API dokumentiert . (STACKIT)</p> | <p>Fokus auf Betrieb, Hosting und Souveränität</p> |

Praktische Kurzregel für „API, n8n etc.“

| Anbindungsart | Typische Plattformen | Einordnung |
|---------------|----------------------|------------|
|---------------|----------------------|------------|

| | | |
|--|---|---|
| Direkt per REST/API | OpenAI, Anthropic, Gemini, Perplexity, OpenRouter, Bedrock, IONOS, STACKIT | Standardweg für Anwendungen und Automatisierung. (OpenAI Entwickler) |
| OpenAI-kompatibel | Ollama, LM Studio, OpenRouter, Perplexity Sonar, Hugging Face Inference Providers, IONOS, STACKIT | Besonders praktisch, weil viele vorhandene Tools und Clients weiterverwendet werden können. (Ollama Dokumentation) |
| SDKs / Client-Bibliotheken | OpenAI, Anthropic, Gemini, Hugging Face, AWS, LM Studio, Perplexity | Sinnvoll für produktive Anwendungen statt reiner HTTP-Aufrufe. (OpenAI Entwickler) |
| n8n direkt | Am einfachsten mit OpenAI über den offiziellen OpenAI-Node | n8n hat einen offiziellen OpenAI-Node ; für andere Dienste ist meist der HTTP Request Node nutzbar, sofern sie eine REST-API anbieten. (n8n Docs) |
| n8n indirekt über OpenAI-Kompatibilität | Ollama, LM Studio, OpenRouter, Perplexity Sonar, IONOS, STACKIT, teils Hugging Face | In der Praxis oft möglich, wenn in n8n ein OpenAI-kompatibler Endpoint oder notfalls der HTTP Request Node verwendet wird; offiziell dokumentiert ist bei n8n selbst vor allem der OpenAI-Node und der generische HTTP Request Node. (n8n Docs) |

Neutrale Formulierung

“ Anbindung / Ansprache

KI-Plattformen werden typischerweise über **REST-APIs**, **offizielle SDKs** oder **OpenAI-kompatible Schnittstellen** angesprochen. Für Automatisierungswerkzeuge wie **n8n** ist dies relevant, weil entweder ein dedizierter Anbieter-Node genutzt werden kann oder ein generischer **HTTP-Request-Ansatz**. OpenAI-kompatible Plattformen sind dabei besonders integrationsfreundlich, da bestehende Clients und Workflows häufig mit geringem Anpassungsaufwand weiterverwendet werden können. ([n8n Docs](#))

Sehr kurze Fassung

“ **Plattformen** sind von **Modellfamilien** zu unterscheiden. Während Modellfamilien die eigentlichen KI-Modelle bezeichnen, stellen Plattformen die Umgebung für Auswahl, Ausführung, Routing, Hosting oder API-Zugriff bereit. Manche Plattformen sind lokal ausgerichtet, andere bündeln viele Fremdmodelle oder bieten eigene Hersteller-APIs an. (

[Ollama Dokumentation](#))

Revision #3

Created 2026-04-21 10:32:19 UTC by Gerd

Updated 2026-04-21 11:19:36 UTC by Gerd