

Muse Spark Open

KI-Modellportrait · Quellenstand 2026-10-31



Meta (Anbieterlogo, kein modellspezifisches Logo) · [Logo-Quelle / Rechteinweis](#). Marken gehören ihren jeweiligen Inhabern. Redaktionelle Identifikation, keine Partnerschaft oder Empfehlung durch den Markeninhaber.

Die Open-Weight-Version von Metas Agentic- und Coding-Modell Muse Spark: 816 Milliarden Parameter gesamt, 42 Milliarden aktiv, 1-Million-Token-Kontext und multimodale Eingabe (Text, Bild, Video, Audio). Die Gewichte stehen unter Metas eigener Muse Model License mit Bedingungen; der Betrieb läuft self-hosted oder über Hosted-Inference-Provider, nicht über die Meta Model API.

Quellen geprüft am 31. Oktober 2026 · Redaktion: Philipp Paulik

Entwickler: [Meta](#) →

Technische Einordnung

Muse Spark Open ist am 15. Oktober 2026 erschienen und ist laut Meta das leistungsfähigste US-Open-Weight-Modell (Herstellerangabe). Es handelt sich um einen sparse Mixture-of-Experts-Transformer mit 816 Milliarden Gesamtparametern, von denen pro Token 42 Milliarden aktiv sind (256 Experten, 8 pro Token geroutet), plus einem 1,8-Milliarden-Parameter-Perception-Encoder für Bild, Video und Audio. Der Kontext umfasst 1.048.576 Tokens. Die Gewichte liegen als BF16-Referenz (rund 1,65 TB) und als NVFP4-Variante (rund 537 GB) auf Hugging Face sowie in den Katalogen von Amazon SageMaker AI, Google Cloud Model Garden und Microsoft Foundry. Für den Desktop-Betrieb bewirbt Meta zwei Mac Studios mit M5 Ultra über MLX. Die Lizenz erlaubt Nutzung, Modifikation, Fine-Tuning, Distillation und Weitervertrieb, knüpft aber an sichtbare Attribution und bei großen Model-as-a-Service-Angeboten an eine Umsatzschwelle; sie ist keine OSI-gelistete Open-Source-Lizenz. Abzugrenzen vom bestehenden Fachblatt-Portrait Muse Spark (geschlossenes VerbrauchermodeLL): Muse Spark Open ist die offene Gewichtsvariante.

Zugang & Nutzung

Gewichte: Hugging Face (meta-models/Muse-Spark-Open-816B-A42B und -NVFP4), Amazon SageMaker AI, Google Cloud Model Garden, Microsoft Foundry Models. Hosted-Inference-APIs: Baseten, Fireworks AI, Groq, Modal, Nebius, OpenRouter, Together AI. Kein Angebot über die Meta Model API (laut Meta-Modellkatalog). Inference-Frameworks: SGLang, vLLM, TensorRT-LLM, Hugging Face Transformers, llama.cpp, Unsloth, MLX.

Muse Model License, Version Muse Spark Open (15.10.2026), eigene Meta-Lizenz: gewährt unentgeltlich Nutzung, Kopie, Modifikation (inkl. Fine-Tuning), abgeleitete Werke, Destillation, Merge, Veröffentlichung, Distribution, Unterlizenzierung und Verkauf. Bedingungen: (1) sichtbarer Hinweis 'Built with Muse Spark Open' in UI, Doku, Website oder Einstellungen jedes Produkts/Dienstes, der das Modell oder Derivate nutzt; (2) wer KI-Modelle als Model as a Service anbietet und mit Lizenznehmer plus Affiliates über 20 Mio. USD Bruttoumsatz in 12 Monaten liegt, muss Meta benachrichtigen und eine separate schriftliche Vereinbarung schließen (gilt nicht für interne Nutzung, Forschung oder eigene Produkte ohne MaaS-Charakter); (3) verbindliche Usage Policy (Exhibit A) mit Verbotsliste, Modell nicht für unter 18-Jährige bestimmt. Meta erhebt keinen Eigentumsanspruch an Outputs. Es ist eine Open-Weight-Freigabe mit Auflagen, keine OSI-gelistete Open-Source-Lizenz; Vertragspartner für die EU-Region ist Meta Platforms Ireland Limited.

Grenzen & richtig lesen

Hardware: BF16 benötigt laut Model Card rund 2 TB VRAM (Referenz: 8x B300), NVFP4 laut Meta-Doku einen 4-GPU-Blackwell-Knoten bzw. laut Model Card mindestens 700 GB VRAM. Kein Meta-eigenes API-Angebot; Hosted-Preise und Latenzen hängen vom Provider ab. Nicht alle der über 100 Trainingssprachen sind evaluiert, Qualität kann außerhalb der stark unterstützten Sprachen abfallen. Sehr kurze oder niedrig aufgelöste Videos (wenige Sekunden, ca. 100 px) werden mitunter nicht erkannt; Meta empfiehlt mindestens 448 px auf der kurzen Seite. Quantisierte Ausführung kann in Randfällen geringfügig abweichen. Meta empfiehlt ausdrücklich den Einsatz zusätzlicher systemseitiger Guardrails und weist darauf hin, dass die Sicherheitsbewertung nur den veröffentlichten Checkpoint unter den getesteten Bedingungen betrifft, nicht nachgeschaltete Modifikationen.

Benchmark-Belege

Für dieses Portrait ist noch kein ausreichend zugeordnetes Ergebnis erfasst. Fehlende Werte sind keine Null. Wir übernehmen keine Werte aus ähnlich benannten Modell- oder Benchmark-Versionen.

Steckbrief

Veröffentlichung

15. Oktober 2026

Parameter

816B gesamt / 42B aktiv

Architektur

Sparse Mixture-of-Experts-Transformer mit Perception-Encoder; 62 Schichten, 256 Experten, 8 aktive Experten pro Token, Hidden Dimension 8192, Grouped-Query-Attention (128 Q- / 16 KV-Köpfe)

Perception-Encoder

1,8B Parameter, 50 Schichten, 1536 Hidden Dim, 14x14 Patch Size

Kontextfenster

1.048.576 Tokens (1M) bei Inferenz

Maximale Ausgabe

Nicht offiziell belegt

Modalitäten

Eingabe Text, Bild, Video, Audio; Ausgabe Text

Sprachen

Trainiert auf Daten aus über 100 Sprachen; nicht alle evaluiert (Herstellerangabe)

Vokabular

202.048 Tokens

Checkpoints

BF16 (ca. 1,65 TB Gewichte) und NVFP4 (ca. 537 GB Gewichte)

Empfohlene Inferenzparameter

temperature 1.0, top_p 1.0, top_k deaktiviert; reasoning_effort low/medium/high/xhigh (Standard: xhigh; high/xhigh für komplexe, Coding- und Agentic-Aufgaben)

Desktop-Betrieb

Zwei Apple Mac Studios mit M5 Ultra über MLX (Herstellerangabe; je bis zu 512 GB Unified Memory)

Lizenz

Muse Model License (Open Weight mit Auflagen): Attribution 'Built with Muse Spark Open', MaaS-Schwelle 20 Mio. USD Jahresumsatz, Usage Policy (Exhibit A), keine OSI-gelistete Open-Source-Lizenz

Meta Model API

Nicht verfügbar; nur Self-Hosting oder Hosted-Provider

Hosted-Listenpreis

1,25 USD Input / 4,25 USD Output je 1M Tokens, Cache-Read 0,15 USD (OpenRouter-Provider, Stand 31.10.2026)

Benchmark: Artificial Analysis Intelligence Index

39 (Herstellerangabe)

Benchmark: MCP-Atlas

89,2 (Herstellerangabe)

Benchmark: OSWorld-Verified

80,4 (Herstellerangabe)

Benchmark: Terminal-Bench 2.1

76,4 (Herstellerangabe)

Benchmark: GPQA Diamond

89,0 (Herstellerangabe)

Benchmark: HLE mit Tools

61,0 (Herstellerangabe)

Sicherheit: StrongREJECT v2 (adaptive)

0,5 Angriffserfolgsrate, niedriger ist besser (Herstellerangabe)

Sicherheits-Post-Training

Zusätzliches Safety-Post-Training für den Open-Weight-Betrieb ohne System-Guardrails; Freigabe nach Meta Superintelligence Scaling Framework (Herstellerangabe)

Vertiefung im Fachblatt

Muse Spark Open (offene Gewichte) im Fachblatt analysiert

Angaben laut den unten verlinkten Primärquellen.

Quellen & Stand

1. [Introducing Muse Spark Open, The Most Capable US Open-Weight Model to Date \(Meta AI Research\)](#)

Geprüft: 2026-10-31

2. [meta-models/Muse-Spark-Open-816B-A42B - Model Card inkl. Muse Model License \(Hugging Face\)](#)

Geprüft: 2026-10-31

3. [Muse Model License, Version Muse Spark Open \(LICENSE.md, Hugging Face\)](#)

Geprüft: 2026-10-31

4. [Muse Spark Open - Meta Model API Dokumentation](#)

Geprüft: 2026-10-31

5. [Models - Meta Model API Dokumentation \(Modellkatalog\)](#)

Geprüft: 2026-10-31

6. [Muse Spark Open Evaluation Methodology \(Meta AI Research\)](#)

Geprüft: 2026-10-31

7. [Meta: Muse Spark Open - API Pricing & Providers \(OpenRouter\)](#)

Geprüft: 2026-10-31

8. [KIFB-Analyse: Muse Spark Open \(offene Gewichte\)](#)

Geprüft: 2026-10-31

Redaktionelle Zusammenfassung mit KI-Unterstützung, verantwortet von Philipp Paulik. Kein eigener Leistungstest. Änderbare Anbieterangaben gelten für den angegebenen Quellenstand.

Entwickler-Kontext des Anbieters

Ein öffentliches Entwicklerprojekt des Anbieters: [meta-llama/llama-api-python](#). Das ist kein Nachweis, dass dieses Repository die Gewichte oder den Quellcode des hier porträtierten Modells enthält.

64 GitHub-Stars · 46 Forks · Repository-Lizenz laut API: MIT.

Snapshot: 2026-10-19T13:28:52+00:00 · letzter Push: 2026-05-12T19:09:27Z. Beachtung und Repository-Aktivität, keine Modellqualität oder Unternehmensgröße. [API-Quelle](#). Lokal gespeichert; kein Live-Abruf beim Lesen. SDK-/Code-Lizenzen gelten nicht automatisch für Modelle oder APIs.

KI Fachblatt · Redaktion: Philipp Paulik · Quellenstand: 2026-10-31

Vollständiges Webportrait: <https://kifachblatt.de/ki-modelle/muse-spark-open>

Statischer Export, kein Live-Datenblatt. Preise, Zugriff und Anbieterangaben können sich ändern. Quellen und ihre jeweiligen Abrufstände stehen oben. [Impressum](#) · [Transparenz](#)