

Dein Agenten-Büro: drei *Wege, DSGVO-konform*

Der ausführliche Leitfaden zum Reel. Wie du die ArtCraft-Büro-Apps auf deinem eigenen Rechner oder Server betreibst, welches KI-Modell du dahinter hängst, was es kostet und was du für die DSGVO wirklich brauchst. Mit einem großen Kapitel zu Claude über AWS Bedrock: Kosten, Einsatzbeispiele und Einrichtung für das kleine Büro, den Handwerksbetrieb, den Mittelstand und den Konzern.

@reposprites

reposprites.com

Danke fürs Folgen ✓ · Du hast kommentiert: BÜRO ✓

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119



DER MENTOR SAGT

Wichtiger Hinweis: Ich bin kein Rechtsanwalt, kein Datenschutzbeauftragter und kein Kostenexperte. Alles in diesem Leitfaden sind Tipps und Erfahrungen aus der Praxis, keine Rechtsberatung. Alle Kosten sind einfache Überschlagsrechnungen mit offengelegten Annahmen, keine Kalkulation für dein Unternehmen. Ich habe die Fakten sorgfältig recherchiert, aber Gesetze, Preise und Verträge ändern sich. Bevor du Kundendaten, Gesundheitsdaten, Mandantendaten oder Daten deiner Beschäftigten mit KI verarbeitest, lass dich von deinem Datenschutzbeauftragten oder einem Fachanwalt beraten.

// 01

Auf einen Blick

	① LOKAL	② DEUTSCHE CLOUD	③ CLAUDE IN EUROPA
Modell	Offene Modelle (Qwen, GLM, gpt-oss)	Offene Modelle bei IONOS oder STACKIT	Claude über AWS Bedrock, EU-Profil
Wo laufen die Daten	In deinem Büro oder auf deinem Server	Deutsche Rechenzentren (Berlin, Neckarsulm)	Nur EU-Regionen: Frankfurt, Stockholm, Mailand, Spanien, Irland, Paris
Speichert der Anbieter Eingaben	Nein, alles bei dir	Nein, laut Anbieter	Bei Opus, Sonnet und Haiku laut AWS keine Speicherpflicht. Ausnahme Claude Fable: bis 30 Tage
Agenten-Leistung	Gut für einfache Aufgaben in einer App	Gut bis mittel	Am stärksten bei langen Ketten
Kosten	Hardware einmalig oder GPU-Server	Ein paar Euro pro Monat und Nutzer	Ab wenigen Dollar; im Büro-Alltag grob 5–15 USD pro Nutzer und Monat, Entwickler deutlich mehr
AVV	Nicht nötig (außer beim Server-Hoster)	Teil des Vertrags	Gilt bei AWS automatisch über die Service Terms
Passt für	Maximaler Datenschutz, Berufsgeheimnis	Behörden-nah, KMU ohne US-Bezug	Vom Büro bis zum Konzern, wenn die Qualität zählt

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

// 02

In 60 Sekunden: Hände und Gehirn

Damit du den Rest leicht verstehst, das Bild aus dem Reel:

- **Die Apps sind die Hände.** WordCraft, GridCraft, DeckCraft und PdfCraft öffnen Dateien, schreiben, rechnen, exportieren. Sie laufen auf deinem Rechner oder deinem Server. Deine Dateien bleiben dort.
- **Das KI-Modell ist das Gehirn.** Es entscheidet, was die Hände tun sollen. Ohne Gehirn passiert nichts.
- **Jeder Auftrag geht als Text an das Gehirn.** Wenn du schreibst „Mach aus der Tabelle einen Bericht", schickt dein Agent diesen Satz und die Teile der Datei, die er gerade braucht, an das Modell. Die Antwort kommt zurück, und die Hände setzen sie um.
- **Deshalb ist die eigentliche Frage:** Wo läuft dieses Gehirn, und wer kann mitlesen? Genau dafür gibt es die drei Wege. ① Das Gehirn steht bei dir. ② Es steht in einem deutschen Rechenzentrum. ③ Es ist Claude, läuft aber bei AWS in der EU, und Anthropic sieht nichts.

Merksatz: Die Dateien bleiben bei dir. Was das Modell lesen soll, verlässt bei Weg ② und ③ dein Haus, als Text. Darum entscheidest du bei jeder Aufgabe, was der Agent sehen darf.

// 03

Das Büro: vier Apps, die ein Agent steuert

Die Büro-Apps kommen von ArtCraft, Open Source und kostenlos:

- **WordCraft** für Dokumente
- **GridCraft** für Tabellen (Werkzeuge wie `read_range`, `write_range`, `set_cell`, `evaluate_formula`)
- **DeckCraft** für Folien
- **PdfCraft** für PDFs (Text finden, Seiten drehen, zusammenfügen, Formulare ausfüllen, Passwortschutz)

Jede App hat einen MCP-Server, der ohne Fenster läuft. MCP ist eine Art Steckdose, über die ein KI-Agent Programme bedienen kann. Damit arbeiten die Apps auch auf einem Server ohne Bildschirm.

Ich verschicke keine Programme. Lade die Apps nur von der offiziellen Seite:

Alle Crafting Apps · getartcraft.com/apps
`getartcraft.com ↗`

// 04

Den Agenten auf einen Ordner sperren

Das ist der wichtigste Schritt für den Datenschutz: Der Agent sieht nur den Ordner, den du freigibst. Was nicht im Ordner liegt, kann auch nicht als Text an das Modell gehen.

PdfCraft, nur ein Ordner (öffnet laut Team keinen Netzwerk-Port):

```
claude mcp add pdfcraft -- pdfcraft-cli mcp --root ~/Agentenbuero/pdf
```

WordCraft, Lesen und Schreiben nur in einem Ordner:

```
claude mcp add wordcraft -- wordcraft-cli mcp --automation-read-root ~/Agentenbuero --automation-write-root ~/Agentenbuero
```

GridCraft ohne Fenster:

```
claude mcp add gridcraft -- gridcraft-cli mcp
```

Prüfen: in Claude Code `/mcp` eingeben, die Apps müssen als verbunden erscheinen. Die genauen Schalter jeder App stehen in ihrer README, sie ändern sich gerade schnell.

// 05

Weg ① Lokal: nichts verlässt das Haus

- **Modelle:** Qwen (Apache-2.0) oder GLM (MIT). Kleine Modelle (27–35 Milliarden Parameter) laufen auf einem Mac oder PC mit 32–64 GB Speicher, große brauchen 96 GB und mehr.
- **Programm:** LM Studio oder Jan (Klick-Oberfläche, beide können MCP). Für Profis: Claude Code über Ollama (`ollama launch claude`), das ist von Anthropic nicht offiziell unterstützt.
- **Ehrlich:** Für einfache Aufgaben in einer App reicht ein lokales Modell. Bei langen Ketten über mehrere Apps liegen kleine lokale Modelle in Agenten-Tests deutlich hinter den großen Modellen. Teste es mit deiner eigenen Aufgabe, bevor du dich festlegst.
- **Kosten (Schätzung):** eigener Rechner einmalig etwa 2.000–6.000 €, eine KI-Box wie NVIDIA DGX Spark mehr (siehe unten). Oder ein GPU-Server bei Hetzner: GEX131 (96 GB, Falkenstein) ab 1.499 € netto im Monat.
- **Passt besonders für:** Berufsgeheimnisträger (Ärzte, Anwälte, Steuerberater), die gar nichts nach außen geben wollen, und Aufgaben mit vielen Personendaten.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Die KI-Box für den Schreibtisch: Mini-PCs von NVIDIA, AMD und Apple

Große Marken bauen inzwischen kleine Rechner, die nur einen Zweck haben: große KI-Modelle bei dir im Büro laufen zu lassen. Das Bekannteste ist **NVIDIA DGX Spark**. Ein Kasten, 15 mal 15 Zentimeter groß, 1,2 Kilo schwer, und darin ein Chip, der sonst in Rechenzentren steckt. Für das Agenten-Büro heißt das: Das Modell läuft auf deinem Schreibtisch, deine Daten verlassen das Haus nicht, und trotzdem passen Modelle hinein, für die ein normaler PC viel zu klein ist.

Warum gerade der Speicher zählt. Ein KI-Modell muss komplett in den Speicher der Grafikeinheit passen. Eine gute Gaming-Grafikkarte hat 16 bis 32 GB, das reicht für kleine Modelle. Diese Mini-PCs haben bis zu 128 GB **gemeinsamen Speicher**, den sich Prozessor und Grafikeinheit teilen. Damit laufen Modelle mit 100 bis 200 Milliarden Parametern, also die Klasse, die bei Agenten-Aufgaben deutlich zuverlässiger ist als die kleinen.

	NVIDIA DGX SPARK	PARTNER-GERÄTE MIT DEMSELBEN CHIP	AMD RYZEN AI MAX+ 395 MINI-PCS	APPLE MAC STUDIO
Chip	GB10 Grace Blackwell	GB10 Grace Blackwell	Ryzen AI Max+ 395 mit Radeon 8060S	M-Chips (Max, Ultra)
Gemeinsamer Speicher	128 GB (64 GB über Partner)	64 oder 128 GB	bis 128 GB, davon bis 96 GB für die KI	je nach Modell, bis in den dreistelligen GB-Bereich
Modellgröße laut Hersteller	bis 200 Mrd. Parameter, zwei verbundene Geräte bis 400 Mrd.	wie DGX Spark	grob bis 120 Mrd. Parameter	je nach Speicher
Betriebssystem	DGX OS (Linux)	DGX OS (Linux)	Windows oder Linux	macOS
Beispiele	NVIDIA Founders Edition	ASUS Ascent GX10, Dell Pro Max GB10, HP ZGX Nano G1n, Lenovo ThinkStation PGX, MSI EdgeXpert, Acer Veriton GN100, Gigabyte AI TOP Atom	Framework Desktop, HP Z2 Mini G1a, GMKtec EVO-X2, Beelink GTR9 Pro	Mac Studio
Preis (Stand Anfang November 2026)	6.950 USD (128 GB, US-Listenpreis), 64-GB-Geräte ab 4.999 USD	in der EU etwa 5.000–8.100 € netto, je nach Hersteller und SSD	grob 2.000–4.200 USD mit 128 GB	ab mehreren tausend Euro, je nach Speicher

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Was die DGX Spark kann. NVIDIA nennt bis zu 1 Petaflop KI-Rechenleistung (FP4, theoretischer Spitzenwert), 128 GB Speicher mit 273 GB/s, bis zu 4 TB verschlüsselte SSD und einen schnellen Netzwerkanschluss, mit dem du zwei Geräte zu einem verbinden kannst. Der Stromverbrauch liegt mit einem 240-Watt-Netzteil im Bereich eines starken Büro-PCs. Auf der Seite build.nvidia.com/spark liegen fertige Anleitungen („Playbooks") für Open WebUI (eine Chat-Oberfläche fürs Team), LM Studio, vLLM, llama.cpp und Coding-Agenten. Mit NVIDIA Sync greifst du vom Laptop aus auf die Box zu.

So passt die Box ins Agenten-Büro. Die Box steht im Büro, das Modell läuft darauf (zum Beispiel ein offenes Qwen- oder GLM-Modell). Dein Agent und die Büro-Apps sprechen sie über das Firmennetz an wie einen eigenen kleinen KI-Server. Alle im Team nutzen dasselbe Modell, ohne dass ein Byte ins Internet geht. Für die DSGVO ist das der einfachste Fall: Es gibt keinen KI-Anbieter, der Daten verarbeitet, also auch keinen Vertrag zur Auftragsverarbeitung mit ihm.

Ehrlich gesagt:

- **Die Preise steigen.** Die DGX Spark kostete zum Start im Oktober 2025 noch 3.999 USD, im Februar 2026 4.699 USD und seit Oktober 2026 6.950 USD. Als Grund werden die gestiegenen Speicherpreise genannt. Vergleiche vor dem Kauf die aktuellen Angebote, auch die der Partner.
- **Groß heißt nicht schnell.** 273 GB/s Speicherbandbreite sind deutlich weniger als bei großen Grafikkarten. Die Box kann große Modelle laden, antwortet aber langsamer als ein Rechenzentrum. Für einen Agenten, der im Hintergrund einen Bericht schreibt, ist das egal. Für einen Chat mit zehn Leuten gleichzeitig wird es zäh.
- **Linux statt Windows.** DGX Spark und die Partner-Geräte laufen mit DGX OS, einem Linux. Wer nur Windows kennt, plant einen Nachmittag Einrichtung ein oder lässt es von jemandem machen, der Linux kann.
- **Nicht jedes Modell ist so gut wie Claude.** Auch ein großes offenes Modell liegt bei langen Agenten-Ketten noch hinter den besten Cloud-Modellen. Die Box ist stark für Routine mit sensiblen Daten. Die schwersten Aufgaben bleiben oft bei Claude (Weg ③).

Für wen sich das lohnt. Für Praxen, Kanzleien und Steuerbüros, die Mandanten- oder Patientendaten gar nicht nach außen geben wollen. Für kleine Teams, die viel Routine automatisieren und keine laufenden KI-Kosten pro Anfrage wollen. Für alle, die einen GPU-Server mieten würden: Nach etwa vier bis sechs Monaten Servermiete hast du den Preis einer eigenen Box ausgegeben.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

// 06

Weg ② Deutsche Cloud: offene Modelle, deutsche Rechenzentren

ANBIETER	STANDORT	BEISPIEL-PREIS PRO MIO. TOKEN (EIN/AUS)	SPEICHERUNG
IONOS AI Model Hub	Berlin	gpt-oss-120b 0,15 € / 0,65 €	Laut IONOS keine Speicherung von Eingaben und Antworten
STACKIT AI Model Serving	Neckarsulm (Schwarz Gruppe)	gpt-oss-120b 0,45 € / 0,65 €	Laut STACKIT weder gespeichert noch zum Training genutzt, BSI C5

- **Plus:** kein US-Mutterkonzern, zahlen nur nach Verbrauch. Bei einem aktiven Nutzer (etwa 20 Mio. Token im Monat) grob 3–13 € im Monat (Schätzung).
- **Minus:** Für lange Ketten über mehrere Apps schwächer als Claude. Claude Code braucht dafür einen Gateway, einfacher geht es mit LM Studio, Jan oder Open WebUI.

// 07

Weg ③ Claude in Europa: AWS Bedrock im Detail

Das ist der Weg, nach dem ich nach dem Reel am häufigsten gefragt werde. Darum hier ausführlich.

Was Bedrock ist, in einfachen Worten

Amazon Bedrock ist ein Dienst von AWS (Amazon Web Services). Darüber nutzt du fertige KI-Modelle verschiedener Hersteller, darunter Claude, ohne selbst Server zu betreiben. Du zahlst nur, was du verbrauchst, abgerechnet in Token. Ein Token ist ein Wortbaustein, grob drei bis vier Buchstaben.

Der entscheidende Punkt: **AWS betreibt das Modell, nicht Anthropic.** Laut AWS läuft jedes Modell in einem eigenen AWS-Konto pro Region, und die Modellhersteller haben keinen Zugriff auf deine Eingaben, Antworten oder Logs. Laut Anthropic hat ihr Personal keinen Zugriff auf diese Infrastruktur.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Welche Claude-Modelle es gibt und welches du nimmst

MODELL	WOFÜR	FAUSTREGEL
Claude Haiku 5.5	Massenarbeit: sortieren, Daten herausziehen, kurze Antworten, Rechnungen prüfen	Rund 20-mal günstiger als Sonnet
Claude Sonnet 5.5	Tagesgeschäft: E-Mails, Angebote, Texte, Chatbots, Programmieren	Der Standard für fast alles
Claude Opus 5.5	Schwere Aufgaben: lange Verträge, knifflige Analysen, komplexe Agenten	Doppelt so teuer wie Sonnet, aber günstiger als ältere Opus-Modelle

Alle drei sind auf Bedrock seit September und Oktober 2026 verfügbar und verarbeiten bis zu 1 Million Token auf einmal, also sehr lange Dokumente. Bilder (Fotos von Notizzetteln, gescannte Rechnungen) kann Claude lesen, Audio nicht. Für Sprachnotizen brauchst du vorher eine Umwandlung in Text.

Wo deine Anfrage hingeht: In-Region, EU-Profil, Global

Bedrock kennt drei Arten, eine Anfrage zu verteilen:

- **In-Region:** Die Anfrage bleibt in genau einer Region.
- **EU-Profil** (Kennung beginnt mit `eu.`): Die Anfrage wird auf mehrere EU-Regionen verteilt. Von Frankfurt aus laut AWS nur auf Frankfurt, Stockholm, Mailand, Spanien, Irland und Paris. Also in der EU, aber nicht zwingend in Deutschland.
- **Global** (Kennung beginnt mit `global.`): Die Anfrage kann in jede AWS-Region weltweit gehen, auch in die USA, nach Asien oder Südamerika.

Für die neuesten Modelle (Opus 5.5, Sonnet 5.5, Haiku 5.5) gibt es in Frankfurt laut AWS **kein In-Region**, nur das EU-Profil und Global. **Für Kundendaten nimm immer das EU-Profil, nie Global.**

Das EU-Profil kostet laut Anthropic **10 % mehr** als Global. Für die meisten Firmen ist das gut investiertes Geld.



DER MENTOR SAGT

Tipp für Technik-Verantwortliche: AWS rät, sich bei Datenresidenz nicht nur auf das Präfix zu verlassen, sondern die Zielregionen eines Profils mit `GetInferenceProfile` zu prüfen.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Die Preise

Alle Preise in US-Dollar pro 1 Million Token, ohne Umsatzsteuer. AWS rechnet in Dollar ab. Die Spalten „EU“ sind die Preise mit dem 10-%-Aufschlag für das EU-Profil.

MODELL	INPUT	OUTPUT	EU INPUT	EU OUTPUT	EU CACHE-LESEN
Haiku 5.5 (Anfrage bis 100.000 Token)	0,10	0,50	0,11	0,55	0,011
Haiku 5.5 (Anfrage über 100.000 Token)	0,50	2,50	0,55	2,75	0,055
Sonnet 5.5	2,00	10,00	2,20	11,00	0,11
Sonnet 5	2,00	10,00	2,20	11,00	0,22
Opus 5.5	4,00	20,00	4,40	22,00	0,22
Opus 5 / Opus 4.8	5,00	25,00	5,50	27,50	0,55

Was die Begriffe heißen:

- **Input** ist alles, was du ans Modell schickst: deine Frage, die Datei-Ausschnitte, die Anweisungen. **Output** ist die Antwort. Output ist fünfmal so teuer wie Input.
- **Cache-Lesen:** Wenn du immer denselben langen Text mitschickst, zum Beispiel deine Preisliste, wird er zwischengespeichert. Jedes weitere Lesen kostet bei Sonnet 5.5 und Opus 5.5 nur 5 % des normalen Preises. Das spart viel Geld.
- **Batch:** Aufgaben, die nicht sofort fertig sein müssen, kosten 50 % weniger. Achtung: Laut AWS gibt es Batch auf Bedrock bisher nur für ältere Modelle (zum Beispiel Opus 5, Sonnet 4.6, Haiku 4.5), nicht für die 5.5-Generation.
- **Neuer Tokenizer:** Laut Anthropic erzeugen Modelle ab Claude 4.7 für denselben Text etwa 30 % mehr Token. Das ist in den Beispielen unten schon eingerechnet. Für deutsche Texte gibt es keine offizielle Zahl, miss es mit deinen eigenen Texten.



DER MENTOR SAGT

Die Preise stammen aus Anthropics Preisliste und gelten laut mehreren Quellen auch auf Bedrock. Die Preistabelle auf der AWS-Seite lädt dynamisch, ich konnte sie nicht direkt auslesen. Prüfe die Preise vor dem Budget im AWS Pricing Calculator.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Nebenkosten, die oft vergessen werden

POSTEN	PREIS LAUT AWS	BRAUCHST DU DAS?
Budget-Alarm (AWS Budgets)	Kostenlos	Ja, immer
Guardrails: Filter für Personendaten (PII)	0,10 USD pro 1.000 Text-Einheiten (1 Einheit = bis 1.000 Zeichen)	Wenn Kunden mit einem Chatbot schreiben
Guardrails: Inhaltsfilter	0,15 USD pro 1.000 Text-Einheiten	Bei öffentlichen Chatbots
Guardrails: eigene Muster (Regex) und Wortlisten	Kostenlos	Für deutsche Nummern wie Steuer-ID
Knowledge Base (Chatbot über eigene Dokumente)	5 USD pro GB Dokumente im Monat plus 1 USD pro 1.000 Abfragen	Ab Mittelstand
Alternative Vektor-Datenbank (OpenSearch Serverless, Classic)	Mindestens 2 Recheneinheiten, rund 350 USD im Monat (US-Beispielpreis)	Nur bei großen Projekten
Protokolle (CloudWatch Logs)	Etwa 0,50 USD pro GB (US-Beispielpreis), 5 GB im Monat frei	Nur wenn du Logging einschaltest
Agenten-Betrieb (AgentCore Runtime)	0,0895 USD pro vCPU-Stunde, 0,00945 USD pro GB-Stunde	Erst im Konzern

Rechenbeispiele nach Firmengröße

Alle Beispiele sind **Schätzungen** mit offengelegten Annahmen: EU-Profil, 21 Arbeitstage im Monat, Preise aus der Tabelle oben. Dein Verbrauch kann deutlich abweichen. Miss nach dem ersten Monat nach.

Das kleine Büro und Selbstständige (1–3 Personen)

Typische Aufgaben: E-Mails beantworten, Angebote formulieren, Protokolle und lange Dokumente zusammenfassen, Texte übersetzen, mit dem Agenten-Büro Monatsberichte bauen.

Annahme: 3 Personen stellen je 40 Anfragen am Tag. Das sind 2.520 Anfragen im Monat, jede mit 3.000 Token Input und 600 Token Output.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

MODELL	RECHNUNG	PRO MONAT
Sonnet 5.5 für alles	7,56 Mio. Input × 2,20 + 1,51 Mio. Output × 11,00	**etwa 33 USD**
Haiku 5.5 für alles	7,56 × 0,11 + 1,51 × 0,55	**etwa 1,70 USD**
Plus 20 schwere Aufgaben mit Opus 5.5	je 20.000 Input, 3.000 Output	etwa 3 USD

Ergebnis: etwa 35–40 USD im Monat, also rund 12 USD pro Person. Feste AWS-Gebühren fallen nicht an.

Mein Rat: Sonnet 5.5 als Standard, Opus 5.5 nur für die richtig kniffligen Fälle. Wenn du keine Lust auf AWS-Konsole hast, nimm eine fertige Oberfläche (siehe unten) oder Claude Code mit dem Assistenten `/setup-bedrock`.

Der Handwerksbetrieb (5–15 Personen, Beispiel: 10)

Typische Aufgaben: Angebote aus Aufmaß-Notizen formulieren, Leistungsbeschreibungen schreiben, Eingangsrechnungen auf Fehler prüfen, Kundenanfragen vorsortieren und beantworten, Fotos von Notizzetteln abtippen lassen.

AUFGABE	ANNAHME	MODELL	PRO MONAT
Angebote	300 im Monat; Preisliste und Leistungskatalog (20.000 Token) liegen im Cache, dazu 3.000 neue Token und 2.000 Token Antwort	Sonnet 5.5	**etwa 14 USD** (ohne Cache etwa 22 USD)
Aufmaß-Notizen in sauberen Text	400 im Monat, 1.500 Input, 800 Output (Notiz schon als Text)	Haiku 5.5	**etwa 0,25 USD**
Eingangsrechnungen prüfen	500 im Monat, 5.000 Input, 500 Output	Haiku 5.5 / Sonnet 5.5	**0,41 USD bzw. 8,25 USD**
Kundenanfragen beantworten	600 E-Mails, 2.000 Input, 500 Output	Sonnet 5.5	**etwa 6 USD**
Allgemeiner Büro-Chat	10 Personen, je 15 Anfragen am Tag	Sonnet 5.5	**etwa 42 USD**

Ergebnis: etwa 60–75 USD im Monat für den ganzen Betrieb. Dazu eventuell eine Umwandlung von Sprachnotizen in Text und die Lizenz einer fertigen Oberfläche.

Mein Rat: Hier lohnt sich der Cache am meisten. Leg Preisliste, Leistungskatalog und Textbausteine als festen Vorspann an. Prüfungen und Sortieren mit Haiku, alles, was der Kunde

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

liest, mit Sonnet. Und: Jedes Angebot liest ein Mensch, bevor es rausgeht.

Der Mittelstand (50–500 Mitarbeitende, Beispiel: 200)

Typische Aufgaben: ein Wissens-Chatbot über Handbücher, Richtlinien und Produktdaten, Support-Tickets vorsortieren mit Antwortentwurf, Verträge vorprüfen (nicht als Rechtsberatung), Claude Code im Entwicklerteam.

BAUSTEIN	ANNAHME	PRO MONAT
Wissens-Chatbot	150 aktive Nutzer, je 10 Fragen am Tag (31.500 Fragen), je 6.000 Token Dokument-Ausschnitte, 600 Token Antwort	**etwa 664 USD** mit Sonnet 5.5, **etwa 33 USD** mit Haiku 5.5
Knowledge Base dazu	20 GB Dokumente, 31.500 Abfragen	**etwa 131 USD**
Vertragsprüfung	100 Verträge, je 40.000 Token Input, 4.000 Output, Opus 5.5	**etwa 26 USD**
Support-Tickets	5.000 Tickets, 3.000 Input, 500 Output	**3 USD** mit Haiku, **61 USD** mit Sonnet
Personendaten-Filter und Inhaltsfilter für den Chatbot	nur Frage und Antwort geprüft	**etwa 39 USD**
Claude Code, 10 Entwickler	Annahme: intensiver Einsatz jeden Arbeitstag	**etwa 650 USD** mit Sonnet 5.5, **etwa 1.290 USD** mit Opus 5.5

Ergebnis: grob 900–2.300 USD im Monat, je nach Modellwahl und wie stark das Entwicklerteam programmiert. Die Annahme für Claude Code ist am unsichersten, der Verbrauch schwankt stark.

Mein Rat: Starte den Chatbot mit Haiku und schalte nur bei schlechten Antworten auf Sonnet. Leg pro Abteilung eigene Kostenstellen an (Application Inference Profiles und Tags). In Claude Code die Modelle fest einstellen, sonst rechnet es laut Doku zum Opus-Preis ab. Spätestens hier brauchst du eine Datenschutz-Folgenabschätzung und den Betriebsrat (siehe unten).

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Der Großkonzern (1.000+ Mitarbeitende, Beispiel: 5.000, davon 3.000 aktiv)

Typische Aufgaben: ein KI-Assistent für alle, Agenten in Fachprozessen (Einkauf, Buchhaltung, Personal), Massenverarbeitung von Dokumenten, Claude Code für große Entwicklerteams, zentrale Steuerung über Rollen und Richtlinien.

BAUSTEIN	ANNAHME	PRO MONAT
Konzern-Assistent, alles Sonnet 5.5	3.000 Nutzer, je 15 Anfragen am Tag (945.000 Anfragen)	**etwa 12.630 USD**
Derselbe Assistent mit Modell-Routing	50 % Haiku, 45 % Sonnet, 5 % Opus	**etwa 7.270 USD**
Massenverarbeitung	1 Mio. Dokumente, je 3.000 Input, 300 Output, Haiku 5.5	**etwa 495 USD**
KI-Agenten	20.000 Läufe mit je 30 Schritten, Sonnet 5.5	**etwa 10.560 USD** plus etwa 360 USD Betrieb (AgentCore)
Claude Code, 200 Entwickler	wie im Mittelstand	**etwa 13.000 USD** mit Sonnet, **etwa 25.900 USD** mit Opus

Ergebnis: grob 35.000–65.000 USD im Monat in diesem Szenario. Bei solchen Mengen sind individuelle Verträge mit AWS üblich („Private Offer“), die Konditionen sind nicht öffentlich. Reservierte Kapazität (Provisioned Throughput) gibt es nur über das AWS-Account-Team, für die 5.5-Modelle laut AWS derzeit nicht als eigene Stufe.

Mein Rat: Modell-Routing spart hier am meisten: Ein günstiges Modell entscheidet, welche Anfrage ein teures Modell braucht. Mit einer Organisations-Richtlinie (SCP) kannst du laut AWS konzernweit erzwingen, dass keine Daten gespeichert werden. Achte auf die Quoten: Für das Rate-Limit zählen Output-Token laut AWS mehrfach (bei Opus 5.5 zehnfach), abgerechnet wird aber nur der echte Verbrauch.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Einsatzbeispiele auf einen Blick

GRÖSSE	TYPISCHE AUFGABEN	MODELL
Solo und Büro	E-Mail-Entwürfe, Angebotstexte, Protokolle zusammenfassen, Übersetzungen, Monatsberichte im Agenten-Büro	Sonnet 5.5, Haiku 5.5 für Routine
Handwerk	Angebot aus Aufmaß, Leistungsbeschreibungen, Rechnungen prüfen, Kundenanfragen vorsortieren, Notizzettel-Fotos abtippen	Haiku 5.5 zum Prüfen, Sonnet 5.5 für Texte
Mittelstand	Wissens-Chatbot, Ticket-Vorsortierung, Vertrags-Vorprüfung, Claude Code im Team	Haiku und Sonnet, Opus für Verträge
Konzern	Assistent für alle, Agenten in Fachprozessen, Massen-Dokumente, zentrale Steuerung	Routing über alle drei

Datenschutz bei Bedrock: die Fakten

Wer ist Auftragsverarbeiter? AWS, nicht Anthropic. Das sagt auch Anthropic selbst.

Brauche ich einen AVV? Laut AWS ist das Data Processing Addendum (der AVV) Teil der AWS Service Terms und gilt automatisch für alle, die unter die DSGVO fallen. Du musst nichts extra unterschreiben. Die EU-Standardvertragsklauseln sind laut AWS ebenfalls enthalten. Leg dir trotzdem eine Kopie für dein Verzeichnis der Verarbeitungstätigkeiten ab. Mit dem ersten Aufruf eines Claude-Modells stimmst du außerdem den Nutzungsbedingungen des Modells zu.

Sieht Anthropic meine Daten? Laut AWS nein: Die Modellhersteller haben keinen Zugriff auf Eingaben, Antworten und Logs. Eingaben und Antworten werden laut AWS nicht an die Modellhersteller weitergegeben.

Wird mit meinen Daten trainiert? Laut AWS nein, weder Amazons eigene Modelle noch die Modelle der Hersteller.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Werden meine Eingaben gespeichert? Hier ist der Unterschied wichtig:

- **Opus, Sonnet und Haiku:** Laut AWS gibt es für Claude-Modelle, die vor Claude Fable 5 erschienen sind, keine Änderung, also keine Speicherpflicht. Opus 4.8 erlaubt laut AWS ausdrücklich „keine Speicherung“ (Zero Data Retention). Opus 5.5, Sonnet 5.5 und Haiku 5.5 stehen auf keiner der AWS-Listen mit Speicherpflicht. Ob sie den Modus „keine Speicherung“ ausdrücklich erlauben, steht nicht in den Modellkarten, prüfe das für dein Konto.
- **Claude Fable 5 und 5.1:** Laut AWS wird der gesamte Verkehr bis zu 30 Tage zur automatischen Missbrauchserkennung gespeichert, und markierte Anfragen können von AWS-Mitarbeitern geprüft werden. Die Daten gehen laut AWS nicht an Anthropic. **Für Kundendaten würde ich Fable meiden.**
- Bei EU-Profil liegen gespeicherte Daten in der EU-Zielregion.
- Den Speicher-Modus („none“ für keine Speicherung) stellst du laut AWS zum Start nur per Schnittstelle ein, nicht per Klick in der Konsole. Lass das im Zweifel einen Dienstleister machen.

Protokolle: Das Mitschreiben aller Anfragen (Invocation Logging) ist laut AWS standardmäßig **aus**. Schaltest du es ein, stehen dort die Texte im Klartext, auch wenn ein Personendaten-Filter sie für das Modell maskiert hat. Entscheide das bewusst.

Verschlüsselung: Laut AWS werden Daten bei der Übertragung und gespeichert verschlüsselt, eigene Schlüssel (AWS KMS) sind möglich. Mit AWS PrivateLink läuft die Verbindung sogar ohne öffentliches Internet.

Zertifikate: AWS hat ein BSI-C5-Testat, und Amazon Bedrock ist laut AWS im C5-Umfang. AWS nennt außerdem ISO, SOC und CSA STAR Level 2.

Datenübermittlung in die USA: AWS ist laut eigener Angabe nach dem EU-US Data Privacy Framework zertifiziert. Das EU-Gericht hat die Klage dagegen am 03.09.2025 abgewiesen (Rechtssache Latombe, T-553/23). Laut Fachbeiträgen von Kanzleien läuft dagegen ein Rechtsmittel beim EuGH (C-703/25 P), ein Urteilstermin ist mir nicht bekannt. Im Normalbetrieb mit EU-Profil verlassen deine Inhalte die EU ohnehin nicht.

Restrisiko CLOUD Act: AWS ist ein US-Unternehmen. US-Behörden können auch Daten anfordern, die außerhalb der USA liegen. AWS meldet laut eigener Seite seit 2020 null Herausgaben von Firmeninhalten außerhalb der USA und sagt zu, zu weit gehende Anfragen anzufechten. Rechtlich ausgeschlossen ist es nicht. Wer das ganz vermeiden will, nimmt Weg ① oder ②.

AWS European Sovereign Cloud: Seit dem 14.01.2026 gibt es eine eigene, von den übrigen Regionen getrennte AWS-Cloud in Brandenburg, betrieben von EU-Ansässigen. Bedrock ist dort verfügbar, **Claude aber nach meiner Recherche (noch) nicht**. Dort laufen bisher offene Modelle wie Gemma 4. Beobachte das, für Behörden und kritische Branchen kann das spannend werden.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Nützliche Bedrock-Funktionen

- **Guardrails (Schutzfilter):** Filter für Personendaten erkennen Namen, Adressen, E-Mails, Telefonnummern, IBANs, Kennzeichen und mehr, und können sie blockieren oder durch Platzhalter wie `{NAME}` ersetzen. Deutsch wird laut AWS beim Personendaten-Filter unterstützt. Deutsche Spezialnummern (Steuer-ID, Personalausweis, Sozialversicherungsnummer) erkennt er nicht von selbst, dafür legst du kostenlose eigene Muster an. Grenzen: Werkzeug-Aufrufe von Agenten werden laut AWS nicht geprüft, und Guardrails gibt es nur auf dem klassischen Zugang (`bedrock-runtime`).
- **Knowledge Bases:** Ein Chatbot, der in deinen eigenen Dokumenten nachschlägt und mit Quellen antwortet (RAG).
- **Agents und AgentCore:** Baukasten, um eigene KI-Agenten zu betreiben, mit Gedächtnis, Werkzeugen und Rechten.
- **Prompt Caching:** wiederkehrende lange Texte zwischenspeichern, 5 Minuten oder 1 Stunde.
- **Rechte und Kostenstellen:** Mit IAM legst du fest, wer welches Modell nutzen darf. Mit Tags und Inference Profiles siehst du die Kosten pro Abteilung.

Schritt für Schritt starten

1. **AWS-Konto anlegen** mit Firmendaten und Zahlungsmittel. Sofort die Zwei-Faktor-Anmeldung (MFA) für den Hauptzugang einschalten.
2. **Budget-Alarm setzen:** In AWS Budgets ein Monatsbudget anlegen, zum Beispiel 50 USD, mit E-Mail bei 80 % und 100 %. Kostenlos.
3. **Nicht mit dem Hauptzugang arbeiten:** Einen eigenen Benutzer oder eine Rolle anlegen, die nur Bedrock darf.
4. **Region auf Frankfurt stellen:** oben rechts „Europa (Frankfurt) eu-central-1“.
5. **Speicherung festlegen:** Wenn du sicher keine Speicherung willst, den Modus „none“ setzen lassen und Claude Fable meiden.
6. **Anthropic-Formular ausfüllen:** Beim ersten Claude-Modell fragt AWS einmalig Firmenname, Website und Zweck ab. Die Freigabe kommt meist sofort. Andere Modelle schalten sich laut AWS beim ersten Aufruf automatisch frei.
7. **Im Playground testen:** Ein Claude-Modell öffnen und eine Probe-Anfrage stellen, ohne Kundendaten. Dabei das EU-Profil wählen (Kennung beginnt mit `eu.`), nicht `global.`.
8. **Logging bewusst entscheiden:** Standard ist aus. Lass es aus, wenn du es nicht brauchst.
9. **AVV ablegen** und den Eintrag im Verzeichnis der Verarbeitungstätigkeiten ergänzen.

Hinweis: Neue Konten bekommen laut AWS teils niedrigere Nutzungsgrenzen, Erhöhungen werden nicht automatisch gewährt. Für einen großen Start rechtzeitig beantragen.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Claude Code und das Agenten-Büro auf Bedrock

Claude Code läuft laut Doku offiziell auf Bedrock. Am einfachsten: `claude` starten, „3rd-party platform“ und dann „Amazon Bedrock“ wählen. Der Assistent fragt Anmeldung und Region ab, prüft die Modelle und speichert alles. Später jederzeit mit `/setup-bedrock` wieder öffnen.

Oder von Hand:

```
export CLAUDE_CODE_USE_BEDROCK=1
export AWS_REGION=eu-central-1
export ANTHROPIC_BEDROCK_REGION_PREFIX=eu
```

- In EU-Regionen nimmt Claude Code laut Doku das `eu.`-Profil automatisch.
- **Modelle festlegen**, zum Beispiel mit `ANTHROPIC_DEFAULT_SONNET_MODEL` und `ANTHROPIC_DEFAULT_OPUS_MODEL`. Ohne Festlegung ist das Hauptmodell Opus 5.5, und die Doku warnt, dass dann zum Opus-Preis abgerechnet wird.
- Die Web-Suche von Claude Code ist auf Bedrock nicht verfügbar.
- Die Doku empfiehlt ein eigenes AWS-Konto für Claude Code, damit du die Kosten sauber siehst.

Danach verbindest du die Büro-Apps wie oben per MCP, mit Ordner-Sperre.

Fertige Oberflächen statt AWS-Konsole

- **LibreChat** (Open Source, selbst gehostet): hat einen eingebauten Bedrock-Anschluss, Region frei wählbar.
- **Langdock** (Berlin): Laut Langdock-Doku lassen sich Modelle mit AWS Bedrock als Anbieter einbinden, Regionen unter anderem Frankfurt und Irland. Welches Profil genau genutzt wird, frag dort schriftlich nach. Mehr dazu in Teil 2.
- **Claude Code** über `/setup-bedrock`.
- Das früher beliebte „Bedrock Access Gateway“ für Open WebUI ist laut GitHub eingestellt. Nicht mehr nutzen.

Ehrliche Grenzen von Bedrock

- **Nicht in Deutschland allein:** Das EU-Profil verteilt auf sechs EU-Länder. Reine Ein-Region-Optionen gibt es in Europa nur eingeschränkt (zum Beispiel Sonnet 5 und Opus 4.8 in Stockholm und Irland über den neueren Mantle-Zugang, ohne Guardrails und ohne Logging).
- **Neue Modelle kommen in der EU teils später** oder ohne Batch.
- **Weniger Funktionen als direkt bei Anthropic:** laut Anthropic unter anderem keine Files API, keine Code-Ausführung und Web-Suche auf dem Server, keine Agent Skills.
- **Die AWS-Abrechnung ist unübersichtlich:** Claude erscheint unter dem Modellanbieter (AWS Marketplace), Nebenkosten separat. Budgets und Kostenstellen von Anfang an einrichten.
- **Der CLOUD Act bleibt ein Restrisiko** (siehe oben).

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Bedrock im Vergleich zu den anderen Wegen zu Claude

	AWS BEDROCK	ANTHROPIC DIREKT (API, CLAUDE.AI)	GOOGLE CLOUD	MICROSOFT FOUNDRY
Wer betreibt das Modell	AWS	Anthropic	Google	Anthropic
Auftragsverarbeiter	AWS	Anthropic	Google	Anthropic für Microsoft
Verarbeitung nur in der EU	**Ja**, EU-Profil, +10 %	**Nein**, nur USA oder global	**Ja**, Multi-Region EU, +10 %	**Nein**, nur global oder USA
Standard-Speicherung	Keine (außer Fable: bis 30 Tage)	Standardmäßig keine, Ausnahmen für bestimmte Modelle und markierte Inhalte	Nach Google-Regeln	Prompts bleiben in Azure, Metadaten an Anthropic

Achtung, Verwechslungsgefahr: „Claude Platform on AWS“ ist **nicht** Bedrock. Es wird von Anthropic betrieben und bietet laut Anthropic keine EU-Option. Für die EU nimm Bedrock.

// 08

Was du für die DSGVO wirklich brauchst

Kein Tool ist von allein DSGVO-konform. Diese Punkte machen den Unterschied. Noch einmal: Das sind Tipps, keine Rechtsberatung.

Für alle

- [ ] Anbieter mit AVV (Vertrag zur Auftragsverarbeitung) und Verarbeitung in der EU wählen.
- [ ] Klären: Wird mit deinen Daten trainiert? Wie lange wird gespeichert?
- [ ] Den Agenten auf einen Ordner sperren, keine Passwörter oder Schlüssel in Prompts.
- [ ] Verzeichnis der Verarbeitungstätigkeiten ergänzen, Datenschutzerklärung anpassen.
- [ ] Personenbezogene Daten vor dem Prompt pseudonymisieren, wo es geht (mehr in Teil 2).
- [ ] Dein Team schulen und eine kurze interne KI-Richtlinie schreiben.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Zusätzlich ab einem kleinen Team

- **KI-Kompetenz (EU AI Act, Art. 4):** Gilt seit dem 02.02.2025. Wer KI im Betrieb einsetzt, muss dafür sorgen, dass die Leute damit umgehen können. Laut EU-Kommission braucht es kein Zertifikat, interne Nachweise reichen. Die nationalen Behörden können das seit dem 02.08.2026 durchsetzen.
- **Ergebnisse prüfen:** Jedes Angebot, jede Kunden-Mail und jede Vertragsanalyse liest ein Mensch, bevor sie wirkt.

Zusätzlich im Mittelstand und Konzern

- **Datenschutz-Folgenabschätzung (Art. 35 DSGVO):** Bei Kunden- oder Beschäftigtendaten in größerem Umfang prüfen, ob sie nötig ist. Die Datenschutzkonferenz hat dazu eine Orientierungshilfe „KI und Datenschutz“ veröffentlicht.
- **Betriebsrat (§ 87 Abs. 1 Nr. 6 BetrVG):** Mitbestimmung gilt bei technischen Einrichtungen, die geeignet sind, Verhalten oder Leistung zu überwachen. Das Arbeitsgericht Hamburg sah 2024 kein Mitbestimmungsrecht, als Beschäftigte ChatGPT über private Konten nutzen durften. Bei einem Firmenzugang mit Protokollen pro Nutzer sieht das laut Fachbeiträgen anders aus. Hol den Betriebsrat früh ins Boot.
- **Rollen und Rechte:** Wer darf welches Modell mit welchen Daten nutzen? Das gehört in die KI-Richtlinie und in die IAM-Einstellungen.

Berufsgeheimnisträger (Ärzte, Anwälte, Steuerberater)

Nach § 203 StGB dürfen Geheimnisse nur an Dienstleister gehen, die als „mitwirkende Personen“ zur Verschwiegenheit verpflichtet sind. Ob die Standardverträge von AWS dafür reichen, konnte ich nicht klären. Frag deine Kammer oder einen Fachanwalt. Bis dahin: Weg ① (lokal) oder Mandantendaten vorher pseudonymisieren.



DER MENTOR SAGT

Ein Restrisiko bei US-Anbietern ist der CLOUD Act. Wer das ganz ausschließen will, nimmt Weg ① oder ②.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

// 09

Mittelstand und Konzern: wenn die KI-Rechnung groß wird

Ab ein paar hundert Leuten werden aus ein paar Dollar schnell fünfstellige Monatsbeträge. Dann lohnt es sich, genauer hinzusehen. In diesem Kapitel geht es um drei Fragen: Woher kommen die großen Beträge wirklich? Sind Lizenzen pro Nutzer günstiger? Und wie schicken große Firmen Aufgaben an Claude, ohne Kundendaten preiszugeben?



DER MENTOR SAGT

Alle Rechnungen hier sind bewusst einfach: Menge mal Listenpreis, ohne Rabatte, ohne Umsatzsteuer, Dollar nicht in Euro umgerechnet. Sie zeigen die Größenordnung, sie ersetzen kein Angebot und keine Kalkulation. Ich bin kein Kostenexperte.

Woher die großen Beträge wirklich kommen

Schau dir die Konzern-Rechnung oben noch einmal an (3.000 aktive Nutzer, grob 35.000–65.000 USD im Monat):

BAUSTEIN	PRO MONAT	ANTEIL
Chat-Assistent für alle 3.000	etwa 7.300–12.600 USD	klein pro Kopf: rund 2,40–4,20 USD pro Person
KI-Agenten in Fachprozessen	etwa 10.900 USD	mittel
Claude Code für 200 Entwickler	etwa 13.000–25.900 USD	der größte Brocken
Massenverarbeitung von Dokumenten	etwa 500 USD	klein

Die einfache Rechnung pro Chat-Anfrage mit Sonnet 5.5 im EU-Profil: 3.000 Token rein mal 2,20 USD pro Million plus 600 Token raus mal 11 USD pro Million. Das sind **rund 1,3 Cent pro Anfrage**. Wer 15 Fragen am Tag stellt, kommt an 21 Arbeitstagen auf 315 Anfragen, also **etwa 4 USD im Monat**.

Das heißt: Teuer wird es selten durch den normalen Büro-Chat. Teuer werden Entwickler, die den ganzen Tag mit Claude Code arbeiten, und Agenten, die viele Schritte hintereinander machen. Genau dort musst du hinschauen.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Hebel 1: Für jede Aufgabe das passende Modell

- **Modell-Routing:** Ein günstiges Modell (Haiku 5.5 oder ein lokales Modell) sortiert vor und entscheidet, welche Anfrage das teure Modell wirklich braucht. In der Konzern-Rechnung oben spart das beim Assistenten rund 40 % (12.630 statt 7.270 USD).
- **Cache:** Was jedes Mal gleich mitgeschickt wird (Richtlinien, Preislisten, Handbücher), kostet aus dem Cache nur 5–10 % des normalen Preises.
- **Routine lokal:** Einfache Aufgaben wie Sortieren, Zusammenfassen oder Formatieren schafft oft ein offenes Modell auf eigener Hardware (Weg ①). Dann zahlst du dafür keine Token, nur Strom und Server.
- **Entwickler mit Budget:** In Claude Code pro Team ein Ausgabenlimit setzen und das Standardmodell festlegen. Opus nur für die wirklich schweren Fälle.

Hebel 2: Lizenzen pro Nutzer statt Abrechnung pro Token

Viele große Firmen fragen: Ist eine feste Lizenz pro Person nicht günstiger? Und wird es billiger, wenn man das Training mit den eigenen Daten abschaltet? Ich habe beides nachgeprüft.

Training abschalten macht nichts billiger. Ich habe keinen Anbieter gefunden, bei dem der Preis sinkt, wenn das Training aus ist. Bei den Business-Plänen von Anthropic, OpenAI, Google und Microsoft ist das Training mit Firmendaten ohnehin standardmäßig aus. Der echte Unterschied zwischen einem privaten Abo und einem Business-Plan ist der Vertrag: Private Abos (Claude Pro, ChatGPT Plus) laufen unter Verbraucherbedingungen ohne AVV. Für Kundendaten oder Daten deiner Beschäftigten sind sie damit tabu, auch wenn du das Training ausschaltest.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Die wichtigsten Business-Angebote (Listenpreise pro Person und Monat, Stand 02.11.2026):

ANGEBOT	PREIS	TRAINING MIT DEINEN DATEN	EU-VERARBEITUNG	CLAUDE DABEI?
Claude Team Standard	20 USD (jährlich), 25 USD (monatlich), **2 bis 150 Personen**	standardmäßig aus	nein	ja
Claude Team Premium	100 USD (jährlich), 125 USD (monatlich), 5-fache Nutzung	standardmäßig aus	nein	ja
Claude Enterprise	20 USD pro Person **plus jede Nutzung zu API-Preisen**, ab 20 Personen	standardmäßig aus	nein, nur „US-only“	ja
ChatGPT Business	20 USD (jährlich), 25 USD (monatlich), bis 200 Personen	standardmäßig aus	unklar, prüfen	nein
Microsoft 365 Copilot	30 USD (zusätzlich zur M365-Lizenz)	aus	ja (EU-Datengrenze)	ja, aber in der EU standardmäßig aus und **außerhalb** der EU-Datengrenze
Microsoft Copilot Chat	ohne Aufpreis zu passenden M365-Lizenzen	aus	ja	nein
Google Workspace mit Gemini	13,60 € (Standard), 21,10 € (Plus), bis 300 Personen	aus	Verarbeitung in der EU nur mit Enterprise Plus	nein
Mistral Le Chat Team	24,99 USD	Team: abschaltbar	ja	nein
Langdock Business	25 € Listenpreis, mit Staffel bis 12,80 €; Konzern ab 1.000 Personen individuell	aus	EU-Hosting	ja, Backend pro Modell erfragen
meinGPT	18 € fallend bis 5,40 € plus Modellbudget	laut Anbieter nicht	Deutschland und EU	auf der Preisseite nicht genannt
nele.ai	kein Preis pro Person: Guthaben-Pool ab 10 €	laut Anbieter nicht	Deutschland	ja (Sonnet, Opus)

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119



DER MENTOR SAGT

Zu ChatGPT konnte ich die offiziellen Seiten technisch nicht abrufen, die Angaben stammen aus Suchergebnissen dieser Seiten. Prüf sie vor einer Entscheidung direkt bei OpenAI.

Wichtig für die EU: Claude direkt bei Anthropic (Team, Enterprise und die API) bietet keine Verarbeitung in der EU, nur „US“ oder „global“. Claude mit Verarbeitung in der EU gibt es über AWS Bedrock oder Google Vertex, und über EU-Anbieter, die darauf aufsetzen.

Einfache Vergleichsrechnung: 200 Personen

WEG	RECHNUNG	PRO MONAT
Claude Team Standard	geht nicht, maximal 150 Personen	–
Claude Enterprise	200 × 20 USD = 4.000 USD, plus Nutzung	etwa 4.600–4.800 USD
ChatGPT Business	200 × 20 USD	4.000 USD
Langdock mit Staffel	50 × 23,20 € + 150 × 19,20 €	etwa 4.000 €
Chat über Bedrock	200 × rund 4 USD	**etwa 800–900 USD**

Einfache Vergleichsrechnung: 3.000 Personen

WEG	RECHNUNG	PRO MONAT
Claude Enterprise	3.000 × 20 USD = 60.000 USD, plus Nutzung	etwa 67.000–72.000 USD
Microsoft 365 Copilot	3.000 × 30 USD	90.000 USD
Chat über Bedrock, alles Sonnet	3.000 × rund 4 USD	**etwa 12.600 USD**
Chat über Bedrock mit Routing	wie oben, gemischte Modelle	**etwa 7.300 USD**

Faustregel: Ein Platz für 20 USD lohnt sich erst, wenn eine Person mehr als etwa 1.500 Anfragen im Monat stellt. Das sind rund **70 Anfragen pro Arbeitstag**. Die meisten Büro-Nutzer liegen deutlich darunter.

Warum Firmen trotzdem Lizenzen kaufen:

- Der Preis ist planbar, die Oberfläche fertig, Anmeldung über das Firmenkonto (SSO) inklusive.
- Bei Bedrock bekommst du nur die Technik. Für 3.000 Leute brauchst du noch eine Oberfläche (zum Beispiel LibreChat oder einen EU-Workspace) und jemanden, der sie betreibt.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

- Für Vielnutzer und Entwickler ist ein Premium-Platz oft günstiger als Token. Laut der Rechnung oben kostet ein Entwickler mit Claude Code über Bedrock grob 65–130 USD im Monat, ein Team-Premium-Platz 100 USD.

Was die meisten am Ende machen: kombinieren. Ein Chat für alle (Copilot Chat, ein EU-Workspace oder eine eigene Oberfläche über Bedrock), Premium-Plätze für die Vielnutzer und Entwickler, und die API für alles, was automatisch laufen soll. Agenten und das Agenten-Büro mit den MCP-Apps brauchen ohnehin die API.

Hebel 3: Der Türsteher vor Claude

Viele größere Firmen schicken Kundendaten gar nicht erst roh an eine Cloud-KI. Sie stellen einen kleinen **Türsteher** davor, im eigenen Haus. Er ersetzt Namen und Nummern durch Platzhalter, schickt nur den Platzhalter-Text an Claude und setzt die echten Werte in der Antwort wieder ein. Fachleute nennen das Pseudonymisierung.

So läuft es ab:

1. Eine Mitarbeiterin schreibt: „Fasse die Beschwerde von Erika Mustermann, Kundennummer 48213, zusammen."
2. Der Türsteher im eigenen Rechenzentrum erkennt Name und Nummer.
3. Er ersetzt sie: „Fasse die Beschwerde von `<PERSON_1>`, Kundennummer `<KUNDE_1>`, zusammen."
4. Die Zuordnung (`<PERSON_1>` = Erika Mustermann) bleibt im Haus.
5. Nur der Platzhalter-Text geht an Claude.
6. Die Antwort kommt mit Platzhaltern zurück, der Türsteher setzt die echten Werte wieder ein.

Die Werkzeuge dafür:

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

WERKZEUG	WO ES LÄUFT	DEUTSCH	SETZT DIE ECHTEN WERTE ZURÜCK?	KOSTEN
Presidio (Open Source)	bei dir	ja, nach Einrichtung (deutsches Sprachmodell, deutsche Erkenner einschalten)	ja, mit eigener Zuordnungstabelle	kostenlos
LiteLLM mit Presidio (Open Source Gateway)	bei dir	ja (`presidio_language: "de"`)	ja (`output_parse_pii`)	kostenlos
Kleine lokale Modelle als Erkenner (zum Beispiel GLiNER-Familie)	bei dir, CPU oder kleine GPU	teils, vorher testen	über deinen eigenen Code	kostenlos, plus Server
Bedrock Guardrails	in der AWS-Cloud	ja	**nein**: ersetzt nur durch den Typ, zum Beispiel `{NAME}`	0,10 USD pro 1.000 Text-Einheiten

Der Unterschied ist wichtig: Presidio und LiteLLM arbeiten **bevor** die Daten dein Haus verlassen. Bedrock Guardrails filtert **in** der Cloud. Der Klartext ist dann schon bei AWS, und eingeschaltete Protokolle speichern ihn ungefiltert.

Ein Gateway wie LiteLLM kann außerdem gleich das Routing aus Hebel 1 übernehmen: Routine an das günstige Modell, schwere Fälle an Claude. Türsteher und Weiche in einem.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Was du beim Türsteher beachten musst

Pseudonymisierung ist ein guter Schutz, aber kein Freifahrtschein. Diese Punkte solltest du kennen:

- **Für dich bleiben die Daten personenbezogen.** Wer die Zuordnungsliste hat, kann zurückrechnen. Deshalb gilt die DSGVO für dich weiter (Art. 4 Nr. 5 und Erwägungsgrund 26 DSGVO). Du brauchst also trotzdem den AVV mit dem Anbieter.
- **Das EuGH-Urteil von 2025 (C-413/23 P):** Für einen Empfänger, der die Personen realistisch nicht zuordnen kann, können pseudonymisierte Daten anonym sein. Für dich nicht. Und wenn der Empfänger sie an jemanden weitergibt, der zuordnen kann, sind sie wieder personenbezogen. Die europäischen Datenschützer arbeiten noch an endgültigen Leitlinien. Verlass dich also nicht auf „beim Anbieter ist es ja anonym".
- **Kein Filter findet alles.** Presidio sagt das selbst. In einem englischen Test mit Gerichtsurteilen fand es in der Standardeinstellung nur knapp die Hälfte der direkten Hinweise auf Personen vollständig. Für echte deutsche Geschäftstexte habe ich keine unabhängige Messung gefunden. Die einzige belastbare Zahl ist dein eigener Test.
- **Der Zusammenhang verrät die Person.** Auch ohne Namen kann „die Filialleiterin in Bad Tölz, die seit März krank ist" eindeutig sein. Beruf, Ort und Datum zusammen reichen oft. Und KI-Modelle können aus unauffälligem Text auf Wohnort, Alter oder Einkommen schließen. In einer Studie lagen sie dabei in bis zu 85 % der Fälle richtig.
- **Die Geschichte über viele Anfragen.** Einzeln ist jede Anfrage harmlos. Aber wenn derselbe Platzhalter über Wochen in vielen Anfragen auftaucht, setzt sich ein Bild zusammen. Deshalb Platzhalter pro Vorgang neu vergeben, nicht dauerhaft gleich.
- **Verlauf, Gedächtnis und Protokolle beim Anbieter.** Chat-Oberflächen speichern Verläufe, manche haben ein „Gedächtnis". Bei Anthropic können als Verstoß markierte Inhalte bis zu zwei Jahre gespeichert werden. Die Oberflächen von Claude Team und Enterprise sind laut Anthropic nicht für „Zero Data Retention" freigegeben. Für sensible Fälle also lieber die API über deinen Türsteher als die Chat-Oberfläche.
- **Die Zuordnungsliste ist dein Kronjuwel.** Wer sie stiehlt, hat alle Namen. Getrennt speichern, verschlüsseln, Zugriff auf wenige Personen beschränken, nach Abschluss löschen.
- **Anhänge, Bilder und PDFs.** Ein Filter für Text übersieht Namen in eingescannten Dokumenten oder Bildern, wenn sie nicht vorher in Text umgewandelt und geprüft werden.
- **Falsche Rückübersetzung.** Wenn das Modell Platzhalter verwechselt oder neue erfindet, landet der falsche Name im Brief. Darum liest ein Mensch jedes Ergebnis, das nach außen geht.
- **Versteckte Anweisungen (Prompt Injection).** Ein präparierter Text kann ein Modell dazu bringen, Daten herauszugeben. Der Türsteher sollte auch prüfen, was zurückkommt, und Agenten nur die Daten sehen lassen, die sie brauchen.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Gute Praxis für den Türsteher

- [ ] Zuerst weglassen, dann verschleiern: Was das Modell nicht braucht, wird gar nicht erst mitgeschickt.
- [ ] Eine Erlaubnisliste: Welche Datenarten dürfen überhaupt raus?
- [ ] Eigene Namenslisten aus dem CRM einbinden, damit Kundennamen sicher gefunden werden.
- [ ] Ein Testset mit 100–300 Beispieltexten aus deinem Alltag, ohne echte Daten. Messen, was der Filter findet und was nicht.
- [ ] Regelmäßige Stichproben im Betrieb.
- [ ] Protokolle am Türsteher, aber ohne Klartext.
- [ ] Beim Anbieter: kein Training, kurze Speicherdauer, Protokolle aus oder maskiert, EU-Profil.
- [ ] Datenschutz-Folgenabschätzung prüfen (KI steht auf der Muss-Liste der Datenschutzkonferenz), Betriebsrat einbinden.
- [ ] Ein Mensch prüft alles, was nach außen geht.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Was der Türsteher ungefähr kostet

POSTEN	GROBE KOSTEN	HINWEIS
Software (Presidio, LiteLLM)	0 €	Open Source
Server ohne Grafikkarte für Presidio	etwa 20–80 € im Monat	meine Schätzung für 4–8 Prozessorkerne; Cloud-Server gibt es ab 6,49 €
Server mit Grafikkarte für ein kleines Erkennungsmodell	zum Beispiel 214 € im Monat plus 209 € Einrichtung (Hetzner GEX45, Standort Helsinki)	nur nötig, wenn ein KI-Modell die Namen finden soll
Aufbau Prototyp	grob 2–5 Personentage	meine Schätzung
Aufbau für den echten Betrieb	grob 10–30 Personentage	meine Schätzung: Testset, sichere Tabelle, Protokolle, Doku, Zuarbeit zur Folgenabschätzung
Laufende Pflege	grob 1–2 Tage im Monat	Stichproben, Listen, Updates

Verglichen mit fünfstelligen Monatsrechnungen ist das wenig. Der Türsteher spart aber nicht direkt Token. Er senkt das Risiko, und er öffnet Wege: Mit gut pseudonymisierten Daten kannst du Aufgaben an Claude geben, die du roh nie herausgeben würdest.

Warum es keine Lösung von der Stange gibt

Ein Reel von einer Minute kann das nicht beantworten, und dieser Leitfaden auch nicht vollständig. Welche Mischung für dein Unternehmen passt, hängt von Fragen ab, die nur du beantworten kannst:

- Welche Prozesse sollen KI nutzen: Chat für alle, Agenten in Fachabteilungen, Entwicklung, Massenverarbeitung?
- Welche Daten stecken darin: eigene Zahlen, Kundendaten, Gesundheitsdaten, Daten von Beschäftigten, Mandantengeheimnisse?
- Wie viele Personen nutzen KI, und wie intensiv?
- Gibt es schon Microsoft 365, Google Workspace oder einen AWS-Vertrag, an den du andocken kannst?
- Was soll die KI am Ende leisten, und was ist dir das wert?

Es gibt auf jeden Fall Lösungen: Lizenzen, Bedrock, ein Türsteher, lokale Modelle, meist eine Mischung. Welche davon zu dir passt, sieht man erst, wenn man deine Prozesse kennt.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

// 10

Welcher Weg passt zu dir?

DU BIST …	MEIN VORSCHLAG
Selbstständig, wenig Kundendaten, willst die beste Qualität	③ Claude über Bedrock mit EU-Profil, Sonnet 5.5 als Standard
Arzt, Anwalt, Steuerberater	① Lokal für alles mit Mandanten, ③ nur für Texte ohne Personendaten
Handwerksbetrieb	③ mit Cache für Preisliste und Textbausteine, Haiku für Prüfungen
Mittelstand mit Betriebsrat	③ mit Knowledge Base, Guardrails und klarer KI-Richtlinie; Betriebsvereinbarung; ab vielen Kundendaten ein Türsteher davor
Behörden-nah oder ohne jeden US-Bezug	② Deutsche Cloud oder ① Lokal
Konzern	③ mit Modell-Routing, zentraler Steuerung, individuellem Vertrag; ein Türsteher mit Pseudonymisierung davor; Lizenzen nur für Vielnutzer; ① oder ② für besonders sensible Bereiche

Viele landen bei einem Mix: Routine lokal oder in der deutschen Cloud, die schweren Aufgaben mit Claude.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

// 11

Noch mehr Kontrolle: die technischen Wege (ausführlich in Teil 2)

Das Thema ist heikel, aber nicht unlösbar. Den Türsteher für Mittelstand und Konzern hast du im Kapitel oben schon kennengelernt. Neben den drei Wegen gibt es Lösungen, die mehr Technik brauchen, dafür aber dafür sorgen, dass Namen und Kundendaten dein Haus gar nicht erst verlassen. Alles dazu zeige ich im nächsten Reel (Teil 2).

- **Ohne Technik: Langdock** (Berlin). KI-Arbeitsplatz im Browser mit Claude, Chat, Agenten und Workflows. EU-Hosting auf Azure, AVV automatisch in den AGB, ISO 27001 und SOC 2 Type II, Zusatzvereinbarung für Berufsgeheimnisträger. Business laut Preisseite 25 € netto pro Nutzer und Monat. Frag schriftlich nach, über welches Backend und welche Region dein Claude-Modell dort läuft.
- **Die eigene Schleife:** Ein Werkzeug ersetzt Namen lokal durch Platzhalter (aus Erika Mustermann wird `<PERSON_1>`), nur der Platzhalter-Text geht an Claude, und in der Antwort werden die echten Namen lokal wieder eingesetzt. Die Zuordnungsliste bleibt bei dir.
- **Presidio** (Open Source, MIT): findet Namen, E-Mails, IBANs und Ausweisnummern per Spracherkennung und Mustern. Für Deutsch brauchst du ein deutsches Sprachmodell, die deutschen Erkenner (Steuer-ID, Personalausweis, Kennzeichen) musst du einschalten. Presidio sagt selbst: keine Garantie, dass es alles findet.
- **PII-Guardrails vor dem Modell:** Der Presidio-Guardrail in LiteLLM maskiert vor jedem Aufruf und setzt die Originalwerte in der Antwort wieder ein (`pre_call` + `output_parse_pii`). Bei Bedrock gibt es zusätzlich die eingebauten Guardrails (siehe oben).
- **Claude Code Hooks:** Ein Hook läuft lokal auf deinem Rechner, bevor etwas ans Modell geht. Ein `UserPromptSubmit`-Hook kann einen Prompt mit Kundendaten stoppen (nicht umschreiben), der Prompt erreicht das Modell dann nie. Ein `PostToolUse`-Hook kann gelesene Dateien maskieren, bevor Claude sie sieht.
- **Wichtig:** Pseudonymisierte Daten bleiben für dich personenbezogen (Erwägungsgrund 26 DSGVO). Die Schleife senkt das Risiko, ersetzt aber keinen AVV.

// 12

Prompts zum Kopieren

Dein erster Bericht, ohne Kundendaten:

```
Nimm die Tabelle umsatz.xlsx im Ordner Agentenbuero. Erstelle daraus mit GridCraft
ein Diagramm pro Monat, baue mit DeckCraft drei Folien mit den wichtigsten Zahlen und
exportiere alles mit PdfCraft als bericht.pdf. Arbeite nur in diesem Ordner.
```

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Ein Angebot aus deiner Vorlage:

```
Öffne die Vorlage angebot.docx mit WordCraft. Setze Leistung, Menge und Preis aus der Tabelle preise.xlsx ein, rechne die Summe mit GridCraft nach und speichere das Angebot als PDF. Keine Namen von Kunden, nur Platzhalter.
```

Claude Code auf Bedrock einrichten lassen:

```
Richte Claude Code für Amazon Bedrock ein: Region eu-central-1, nur das EU-Profil, Standardmodell Sonnet 5.5, Opus 5.5 nur auf Nachfrage. Leg ein AWS-Budget mit 50 USD und E-Mail-Warnung bei 80 Prozent an und zeig mir zum Schluss, welche Modelle verfügbar sind. Keine Zugangsdaten im Chat ausgeben.
```

Deine Kosten schätzen:

```
Ich nutze Claude über Bedrock mit EU-Profil. Schätze meine Monatskosten: [Anzahl] Personen, je [Anzahl] Anfragen am Tag, typische Aufgaben: [Aufgaben]. Rechne mit den aktuellen Preisen für Haiku 5.5, Sonnet 5.5 und Opus 5.5, zeig die Rechnung und schlag eine günstige Modell-Aufteilung vor.
```

// 13

Deine erste Stunde

- [ ] Die Büro-Apps von getartcraft.com/apps laden.
- [ ] Einen Ordner `Agentenbuero` anlegen, nur mit deinen eigenen Zahlen.
- [ ] Die Apps per MCP verbinden, mit Ordner-Sperre.
- [ ] Einen der drei Wege wählen. Bei Bedrock: Konto, Budget-Alarm, Frankfurt, EU-Profil.
- [ ] Den ersten Bericht bauen lassen und das Ergebnis prüfen.
- [ ] Nach einer Woche die Kosten ansehen und die Modellwahl anpassen.
- [ ] Erst danach an Aufgaben mit Kundendaten denken, mit AVV, Verzeichnis und Schulung.

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

// 14

Mein Tipp

Fang mit einer Aufgabe ohne Kundendaten an: ein Monatsbericht, deine eigenen Zahlen. Dann siehst du, ob ein lokales Modell reicht oder ob du Claude brauchst. Bei Bedrock mit Budget-Alarm und EU-Profil kostet so ein Test meist nur ein paar Dollar. Viele landen bei einem Mix: Routine lokal oder in der deutschen Cloud, die schweren Aufgaben mit Claude in der EU.

// 15

Noch einmal: keine Rechtsberatung

Ich gebe hier Tipps aus meiner Praxis und habe die Fakten nach bestem Wissen recherchiert. Ich bin kein Rechtsberater und kein Kostenexperte. Die Rechnungen sind bewusst einfach gehalten: Menge mal Preis, damit du die Größenordnung siehst. Deine echten Kosten hängen von deinen Prozessen, deinem Vertrag und deinem Verbrauch ab. Dieser Leitfaden ersetzt keine Prüfung durch deinen Datenschutzbeauftragten, deine Kammer oder einen Fachanwalt. Für Entscheidungen über Kunden-, Gesundheits-, Mandanten- oder Beschäftigtendaten hol dir bitte fachlichen Rat.

// 16

Quellen

AWS Bedrock: Datenschutz · docs.aws.amazon.com
docs.aws.amazon.com ↗

AWS Bedrock: Datenspeicherung · docs.aws.amazon.com
docs.aws.amazon.com ↗

AWS Bedrock: Missbrauchserkennung · docs.aws.amazon.com
docs.aws.amazon.com ↗

AWS Bedrock: Profile und Regionen · docs.aws.amazon.com
docs.aws.amazon.com ↗

AWS Bedrock: Preise · aws.amazon.com
aws.amazon.com ↗

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

AWS Bedrock: Guardrails Personendaten · docs.aws.amazon.com
docs.aws.amazon.com ↗

AWS: DSGVO-Center · aws.amazon.com
aws.amazon.com ↗

AWS: BSI C5 · aws.amazon.com
aws.amazon.com ↗

AWS: CLOUD Act · aws.amazon.com
aws.amazon.com ↗

Anthropic: Claude in Amazon Bedrock · platform.claude.com
platform.claude.com ↗

Anthropic: Preise · platform.claude.com
platform.claude.com ↗

Claude Code auf Amazon Bedrock · code.claude.com
code.claude.com ↗

EU-Kommission: KI-Kompetenz FAQ · digital-strategy.ec.europa.eu
digital-strategy.ec.europa.eu ↗

Alle Preise, Standorte und Befehle am 02.11.2026 auf den Seiten der Anbieter und in den Dokumentationen geprüft. Rechenbeispiele sind Schätzungen. Preise ändern sich, prüfe sie vor der Buchung im AWS Pricing Calculator.

Anthropic: Preise Team und Enterprise · claude.com
claude.com ↗

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Anthropic: Enterprise-Plan · support.claude.com
support.claude.com ↗

Anthropic: API und Datenspeicherung · platform.claude.com
platform.claude.com ↗

Anthropic: Speicherdauer für Organisationen · privacy.claude.com
privacy.claude.com ↗

Anthropic: Regionale Compliance · claude.com
claude.com ↗

Microsoft 365 Copilot: Preise · microsoft.com
microsoft.com ↗

Microsoft: Anthropic als Unterauftragsverarbeiter · learn.microsoft.com
learn.microsoft.com ↗

Google Workspace: Preise · workspace.google.com
workspace.google.com ↗

Google Workspace: Datenschutz bei Gemini · knowledge.workspace.google.com
knowledge.workspace.google.com ↗

Mistral: Preise · mistral.ai
mistral.ai ↗

Langdock: Preise · langdock.com
langdock.com ↗

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

meinGPT: Preise · meingpt.com
meingpt.com ↗

nele.ai: Preise · nele.ai
nele.ai ↗

LiteLLM: PII-Maskierung · docs.litellm.ai
docs.litellm.ai ↗

Presidio · presidio.dataprivacystack.org
presidio.dataprivacystack.org ↗

Bedrock Guardrails: Personendaten-Filter · docs.aws.amazon.com
docs.aws.amazon.com ↗

Hetzner: GPU-Server GEX45 · hetzner.com
hetzner.com ↗

NVIDIA DGX Spark: Technische Daten · nvidia.com
nvidia.com ↗

NVIDIA DGX Spark: Playbooks · build.nvidia.com
build.nvidia.com ↗

DGX Spark: Preisverlauf und EU-Preise · guaardvark.com
guaardvark.com ↗

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119

Mini-PCs mit Ryzen AI Max+ 395 im Vergleich · computingforgeeks.com

computingforgeeks.com ↗

— @reposprites

Du willst auch so animierte Videos machen? Dann besuch meine Webseite und hol dir die Reel-Engine.

reposprites.com ↗

@reposprites · reposprites.com

BÜRO

© 2026 REPOSPRITES · @reposprites · reposprites.com · Persönliches Exemplar, Weitergabe und Verkauf nicht gestattet · ID RS-20261102-23083119