

Alle führenden KI-Modelle in einem Arbeitsbereich. **Governance ab dem ersten Prompt.**

Ihre Mitarbeitenden wollen Claude, GPT und Gemini – heute. Ihre IT muss wissen, wohin Unternehmensdaten fließen. AIMan wird beiden gerecht: eine Anmeldung für die führenden Modelle und für offene Modelle auf unseren GPU-Servern in Deutschland – mit Antworten aus Ihren Dokumenten und dem Web.

1

Alle im Unternehmen

- Das beste Modell für jede Aufgabe, ohne privates Konto
- Antworten aus eigenen Dokumenten, mit Quellen
- Aktuelle Antworten aus dem Web
- Bewährte Prompts aus dem Team

2

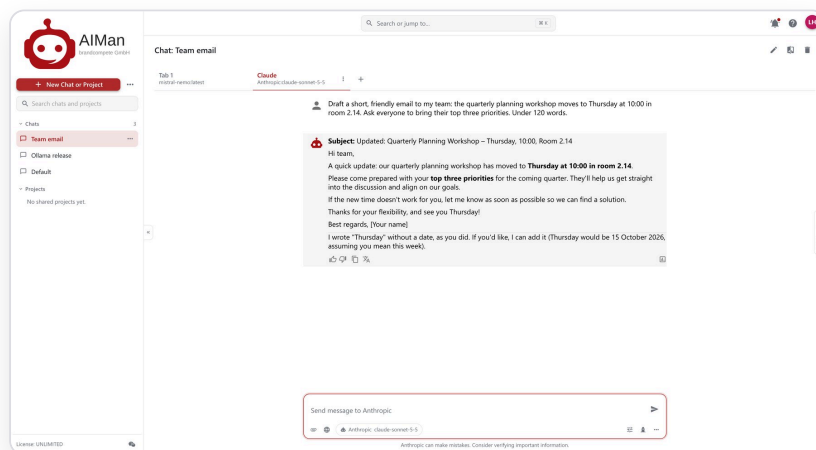
IT, Sicherheit & Datenschutz

- Ein freigegebener Zugang statt Schatten-KI
- Offene Modelle und Dokumentensuche in Deutschland
- SSO nur für Ihre Domains, dazu Rollen und Rechte
- Anbieter-Schlüssel zentral, nie bei Einzelnen

3

Geschäftsführung & Finanzen

- Ein Vertrag statt eines Abos pro Team
- Anbieter-Nutzung ohne Aufschlag
- Offene Modelle ohne Gebühren pro Token
- Ein Pilot, der den Nutzen vor dem Roll-out belegt



Ein Chat, zwei Modelle: mistral-nemo auf unseren GPU-Servern im ersten Tab, Claude im zweiten.

Sensible Arbeit bleibt in Deutschland. Offene Modelle und die gesamte Dokumentensuche laufen auf dedizierten GPU-Servern – unseren oder Ihren. Anbieter-Modelle sind einen Tab entfernt.

Anbieter-Modelle: Anthropic Claude • OpenAI GPT • Azure OpenAI • Google Gemini • Cohere • Ollama Cloud

Offene Modelle auf GPUs in Deutschland: Mistral NeMo • Qwen • Gemma • gpt-oss • Phi • DeepSeek

AIMan auf einen Blick

Alle Modelle, ein Chat

Claude, GPT, Gemini, Cohere und elf offene Modellfamilien in einer Auswahl – pro Chat oder pro Tab, zum Vergleichen.

Antworten aus Ihren Dokumenten

PDFs, Office-Dateien, Webseiten, Confluence, Jira und Bitbucket. Jede Antwort nennt Dokument und Seite.

Prompts, die funktionieren

Ein gemeinsames Prompt-Repository mit Platzhaltern, Kategorien und Tags, dazu eigene Anweisungen.

Chatbots & API

Einen Chat-Tab als Chatbot ins Intranet einbetten. Die REST-API mit Streaming ruft jedes Modell beim Namen.

Offene Modelle in Deutschland

Mistral NeMo, Qwen, Gemma, gpt-oss, Phi und DeepSeek auf dedizierten GPU-Servern. Prompts an sie bleiben dort. Keine Kosten pro Token.

Websuche & Thinking

Ein Schalter pro Chat für aktuelle Antworten mit Quellen. Reasoning-Modelle denken sichtbar Schritt für Schritt.

Projekte & Teilen

Gemeinsame Chats mit Personen, Teams oder ganzen Abteilungen. Datenquellen, Prompts und Anweisungen ebenso.

Modelle, Personen & Schlüssel im Griff

Modelle bereitstellen und zurückziehen, GPU-Belegung sehen. SSO mit Microsoft und Google, Rollen, Teams, zentrale Schlüssel.

So funktioniert es



Warum AIMan

- **Ihre Wahl, sichtbar** – lokale und Anbieter-Modelle in getrennten Tabs: man sieht, wohin ein Prompt geht
- **Die Dokumentensuche bleibt im Haus** – Indexierung, Embeddings und Reranking auf unseren Servern; an Anbieter-Modelle gehen nur die nötigen Passagen
- **Eine Installation pro Kunde** – eigene Datenbank, eigener Suchindex, eigenes Netz; nur die GPUs werden geteilt
- **Ehrliche Antworten** – jedes Modell kennt das Datum; findet AIMan nichts Passendes, sagt es das
- **Made in Germany** – entwickelt, gebaut und betreut von der brandcompete GmbH, seit 2016

Drei Betriebsarten

Managed in Deutschland

Ihre Installation auf unseren GPU-Servern in deutschen Rechenzentren, von uns betrieben.

On-Premise, betreut

AIMan und die offenen Modelle auf GPU-Servern in Ihrer Infrastruktur, betrieben von uns.

On-Premise, selbst betrieben

Ihre IT betreibt AIMan unter Lizenz; brandcompete liefert Releases, Modelle und Support.

Im täglichen Einsatz

Bei Teams eines DAX-notierten Automobilzulieferers und bei Agenturen einer führenden europäischen Mediengruppe.

Ihre Schlüssel, Ihre Rechnung

Anbieter-Modelle laufen unter Ihren Schlüsseln und werden direkt an Sie abgerechnet.

Ihr Auftritt

Ihre Farben und Ihr Logo auf jeder Installation, helles und dunkles Design.

Starten Sie mit einem Pilot

Eine Abteilung, echte Aufgaben, Ihre Dokumente – und nach wenigen Wochen eine Entscheidung auf Basis von Erfahrung statt Demo.

sales@brandcompete.com