



Welche lokalen Modelle für welches Modul tragen.

Qualitätsnachweis aus Benchmark-Reports und produktiver Kompatibilitätsmatrix. Grün ist geeignet, Gelb dokumentiert Einschränkungen, Rot sperrt schadensrelevante Pfade.

3,4 GB

kleinstes empfohlenes Chat-Modell
Qwen 3.5 4B deckt Mail + Tasks + Projekt-Status ab.

100%

Dashboard-Injection abgewehrt
Ministral, Gemma MLX und Qwen 3.6 erkennen den manipulierten Notizfall.

0

Brain-Halluzinationen im Kernbench
Die empfohlenen Brain-Modelle bleiben bei bekannten Notizquellen.

Kompatibilitätsmatrix pro Modul

Punkt oben rechts = empfohlener Default in der App

	Brain	Tasks	Mail	Radar	Projekt	Links
Qwen 3.5 4B 3,4 GB · kleine Macs, Mail + Projekt	?	✓	✓	?	⚠	-
Ministral 3 8B 6 GB · Brain/Radar Default	✓	⚠	✓	✓	⚠	-
Gemma 4 10 GB · schnell, prompt-sensibel	✓	⚠	✓	✓	?	-
Gemma 4 12B MLX 10-11 GB · Apple Silicon, Tasks stark	⚠	✓	✓	✓	✓	-
Qwen 3.6 27B MLX 19-22 GB · Qualität, mehr RAM	✓	✓	⚠	✓	✓	-
Qwen 3.6 48 GB · sehr stark, schwer	✓	✓	✓	✓	✓	-
Llama 3.1 8B 8 GB · Warnmodell im Radar	✗	⚠	✓	✗	?	-
bge-m3 0,6 GB · Embedding für Links	-	-	-	-	-	✓

■ geeignet
 ■ mit Vorbehalt
 ■ gesperrt
 ■ nicht getestet

Empfohlene Einsatzlogik

- ✓ **Kleine Geräte**
qwen3.5:4b für Mail, Tasks und Projekt-Status.
- ✓ **Standard-Setup**
ministral-3:8b für Brain und Dashboard, qwen3.5:4b für Mailpfade.
- ✓ **Apple Silicon**
gemma4:12b-mlx als starke MLX-Option, besonders bei Task-Extraktion.
- ✓ **Maximale Qualität**
qwen3.6 oder qwen3.6:27b-mlx, wenn RAM und Laufzeit reichen.
- ✓ **Semantische Links**
bge-m3:latest als separates Embedding-Modell für Smart Connections.

Testbasis

- **Task-Extraktion**
10 Mail-Fälle, Two-Pass-Datumsauflösung, for_whom und Reply-Erkennung
- **Mail-Zusammenfassung**
8 Mail-Fälle, JSON, Sentiment, Relevanz, needsReply, Halluzinationsrate
- **Dashboard-Radar**
8 Notiz-Fälle inkl. Prompt-Injection; schadensrelevanter Hard-Lock-Pfad
- **Brain**
4 Tages-Szenarien, Wikilinks, Halluzinationen, Bewertungs- und Regel-5-Checks
- **Projekt-Status**
2 Projekt-Cases x Wiederholungen, Honesty-Scorer, Draft-Review statt Auto-Write