



[Knowledge Base](#) > [Using Deskpro](#) > [AI](#) > [Supported AI Models in Deskpro](#)

Supported AI Models in Deskpro

Lara Proud - 2025-09-24 - [Commenti \(0\)](#) - [AI](#)

Deskpro AI supports a wide range of **Large Language Models (LLMs)** and **Embedding Models** across multiple providers, so you can choose the right option for your organization.

We've tested these models extensively to ensure compatibility for both **prompting** (LLM-powered tasks like summarization and reply suggestions) and **embeddings** (semantic search and data retrieval).

This article lists all currently supported public and private models, along with guidance for Cloud and Self-Hosted customers.

Understanding Prompting and Embedding Models

Prompting Model

A prompting model is like a translator for your request — it turns what you ask into clear, step-by-step instructions that the LLM can follow.

Example: You say, "Write a friendly email to reschedule my dentist appointment." The prompting model reformats that request with tone and key points, and the LLM writes the email.

Embedding Model

Embeddings act like "meaning fingerprints" for text. An embedding model converts your words and documents into numeric vectors, allowing the LLM to find relevant information even when different words are used.

Example: You ask, "How do I get paid time off?" The embedding model retrieves the HR document that mentions "annual leave" and passes it to the LLM, so you get the correct policy in plain language.

Deskpro Cloud

Cloud customers can choose from a range of public AI providers and configure their account directly in Deskpro.

Provider	Vendor	Model	Model Name	Type
Vertex	Google	Gemini	gemini-2.0-flash-lite-001	Prompting
Vertex	Google	Gemini	gemini-embedding-001	Embedding
OpenAI	OpenAI	Chat GPT4o mini	gpt-4o-mini	Prompting
Open AI	OpenAI	Chat GPT4o mini	text-embedding-3-small	Embedding
Azure	OpenAI	Chat GPT4o mini	gpt-4o-mini	Prompting
Azure	OpenAI	Chat GPT4o mini	text-embedding-3-small	Embedding
Deskpro	OpenAI	Chat GPT4o mini	gpt-4o-mini	Prompting
Deskpro	OpenAI	Chat GPT4o mini	text-embedding-3-small	Embedding
Anthropic	Anthropic	Claude 3.5 Haiku	gpt-4o-mini	Prompting
Voyage	Voyage	Voyage 3.5	voyage 3.5	Embedding
Mistral	Mistral	Mistral Medium 3.1	magistral-medium-2507	Prompting
Mistral	Mistral	Mistral Medium 3.1	mistral-embed	Embedding

Deskpro Self-Hosted

With Deskpro Self-Hosted, you can either:

- Subscribe to a public AI provider (any listed above), or;
- Configure Deskpro to use their own self-hosted models through providers list below.

Provider	Vendor	Model	Model Name	Type
Ollama	OpenAI		gpt-oss	Prompting
Ollama	Meta	llama 4	llama4	Prompting
Ollama	Meta	llama 3.3	llama3.3	Prompting
Ollama	Meta	llama 3.2	llama3.2	Prompting
Ollama	Meta	llama 3.1	llama3.1	Prompting
Ollama	Meta	llama 3	llama3	Prompting
Ollama	Meta	llama 2	llama2	Prompting
Ollama	Alibaba	qwen3	qwen3	Prompting
Ollama	Alibaba	qwen2.5	qwen2	Prompting
Ollama	Mistral	mistral-large	mistral-large-2411	Prompting
Ollama	Mistral	Mistral Nemo 12b	open-mistral-nemo-2407	Prompting
Ollama	Mistral	Mistral 24.11	mistral	Prompting
Ollama	Mistral	Mixtral 8x7B	mixtral	Prompting
Ollama	Microsoft	Phi-4	phi4	Prompting
Ollama	Microsoft	Phi-3.5	phi3.5	Prompting
Ollama	Google	Gemma 2	gemma2	Prompting
Ollama	Google	Gemma 3n	gemma3n	Prompting
Ollama	Google	Gemma 3n	gemma3	Prompting
Ollama	Deepseek	Deepseek R1	deepseek-r1	Prompting
Ollama	Nomic AI	gpt4all	nomic-embed-text	Embedding
Ollama	Mixedbread	Mixedbread	mxbai-embed-large	Embedding
Ollama	BAAI	BGE-M3	bge-large	Embedding
Ollama	Microsoft	MiniLM	all-minilm	Embedding