KNOWLEDGE BASE BY BAMBOO

What drives AI cost in Knowledge Base

A plain-language guide for a 400-user, 2,000-article company · October 2026

> **THE SHORT ANSWER**
>
> On **Azure OpenAI**, expect about **$2–$20 a month**, plus about $5 once to index the articles.
>
> On **Copilot Studio** pay-as-you-go, expect about **$12–$145 a month**, plus about $40 once.
>
> One question costs about a fifth of a cent on Azure OpenAI, and about 2 cents on Copilot Studio.

For most mid-size customers, the AI in Knowledge Base costs about as much as a team lunch each month, not as much as a licence. AI cost is usage-based: you pay your own AI provider (Microsoft, OpenAI or Anthropic) for each question answered and each article processed. Bamboo does not sit in the middle, does not mark it up, and none of your content passes through us.

The AI features are: **Ask** (answers with links to the articles they came from), **summaries** of each article, **smarter search** that understands meaning and not just keywords, and **import**, which turns PDFs and Word files into draft articles.

What drives the cost

What drives the cost is **how many questions people ask**, not how many people have access. AI providers charge by the amount of text they read and write, measured in "tokens" (roughly ¾ of a word). Four things decide how much text that is:

- **Questions asked.** This is the biggest lever by far. Each Ask or search answer reads a few relevant articles and writes a short reply. A user who never asks costs nothing.
- **Indexing articles.** Each article is summarised and indexed once, and again only when it changes. With 2,000 articles this is a one-time cost of a few dollars.
- **Imports.** Turning a long PDF or Word file into a draft article is the costliest single action, because the AI reads the whole document. It is still cents, not dollars.
- **Which AI you choose.** A "mini" model costs a fraction of a flagship model, and Copilot Studio charges per answer instead of per token.

Admins have a built-in limit: each user gets a daily cap of AI calls (50 by default), so no one person can run up a bill.

What a 400-user, 2,000-article company can expect

The three scenarios differ only in how much people use it:

Scenario	Who asks	Questions / month	Imports / month
Low	1 in 5 users, ~5 questions each	~400	5
Typical	2 in 5 users, ~10 questions each	~1,600	20
High	7 in 10 users, ~25 questions each	~7,000	100, some long

Estimated cost in US dollars:

Option	One-time setup	Low / month	Typical / month	High / month
Azure OpenAI, mini model (supported today)	~$5	~$2	~$5	~$20
Copilot Studio, pay-as-you-go (supported today)	~$40	~$12	~$36	~$145
Microsoft 365 Copilot licence holders	$0	$0 extra	$0 extra	$0 extra
OpenAI Enterprise / API, same mini model (not yet supported)	~$5	~$2	~$5	~$20
Claude Haiku (not yet supported)	~$1	under $1	~$1	~$5
Claude Sonnet (not yet supported)	~$25	~$7	~$20	~$85

Microsoft 365 Copilot licence holders use the Copilot Studio agent at no extra charge. The licence itself is Microsoft's per-user price, so this is only "free" for companies that already have it.

How the options differ

Azure OpenAI is the cheapest option and the most complete. Copilot Studio is the simplest for companies already invested in Copilot. Claude and OpenAI Enterprise would be priced like Azure OpenAI. We have not connected either one yet.

	Azure OpenAI	Copilot Studio	OpenAI Enterprise	Claude (Anthropic)
Status	Supported, tested	Supported, tested	Not yet supported	Not yet supported
How you pay	Per token, on your Azure bill	Per answer, in Copilot credits ($0.01 each, cheaper prepaid)	Per token, on an OpenAI bill	Per token, on an Anthropic bill or through Azure (Microsoft Foundry)
Who sees your content	Your own Azure, in your region	Your own Microsoft 365 tenant	OpenAI, under business terms	Anthropic, or Azure if bought through Foundry
Features	All: Ask, summaries, smart search, import, knowledge graph	Ask and summaries. Search uses a free in-browser model. Cannot draft from imported files	Expected: all (same models as Azure)	Expected: all, with the free in-browser model for search
In our tests	Answered in ~15 s	Answered in ~25–35 s; 93% of 60 test questions correct	Not tested	Not tested

About OpenAI Enterprise: ChatGPT Enterprise seats cover people using ChatGPT. They do not cover apps like Knowledge Base, which would be billed per token through OpenAI's API.

About Claude: Anthropic offers a low-cost model (Haiku) and a stronger one (Sonnet). Claude is also sold through Microsoft Foundry, so it can be billed on an existing Azure agreement. Knowledge Base already sends every AI call through a small function in the customer's own Azure, so adding either provider would be a contained change.

What pushes the cost up or down

Pushes cost UP	Pushes cost DOWN
Many people asking many questions every day	Most people browsing articles, asking occasionally
A flagship model instead of a "mini" one (about 5×)	A mini model, or Claude Haiku (about 4× cheaper than mini)
Copilot Studio pay-as-you-go (about 10× per answer)	Prepaid Copilot credit packs, or users who already hold Microsoft 365 Copilot licences
Long articles, and long follow-up conversations	Short, focused articles
Importing many long PDFs and Word files	Importing the full document word for word, which uses no AI at all
Re-editing many articles every month (each edit is re-indexed)	A stable library
Turning on Copilot Studio's "tenant graph grounding" (10 credits a turn)	Leaving it off, as we recommend

The per-user daily limit (50 calls by default) puts a ceiling on the High scenario. Even if all 400 users hit it every working day, Azure OpenAI with a mini model stays under about $800 a month.

Assumptions

These are estimates, not quotes. They are built from our own test runs plus published list prices:

- **What we measured.** On Azure OpenAI (gpt-4.1-mini), we generated a 10,000-article test library: about 1,900 AI calls and 9 million tokens of output for roughly £10 (about $13). We also asked 60 real questions against it through both Azure OpenAI and a Copilot Studio agent, in September 2026.
- **What we estimated.** We did not log the exact token count of each question, so per-question cost comes from what Knowledge Base sends. An answer reads up to 5 articles, about 3,000 tokens in, and writes about 400 tokens out. Indexing reads each article once, about 3,000 tokens in and 600 out.
- **Articles** average about 2,000 words (the most Knowledge Base sends per article). About 10% of them (200) are edited and re-indexed each month.
- **Imports** are about 30 pages each. In the High scenario, some are long manuals.
- **Prices** are list prices in US dollars as of October 2026. Azure OpenAI and OpenAI gpt-4.1-mini: $0.40 in / $1.60 out per million tokens. Claude Haiku 5.5: $0.10 / $0.50. Claude Sonnet 5.5: $2 / $10. Copilot Studio: about 2 credits per answer at $0.01 per credit. Regional pricing, discounts and enterprise agreements will differ.
- **Not included:** Microsoft 365 Copilot licences, and Azure hosting for the Knowledge Base AI function. On the consumption plan we use, hosting is typically a few dollars a month or less.
- **Claude and OpenAI Enterprise** costs are calculated from list prices only. We have not connected or tested either one yet.