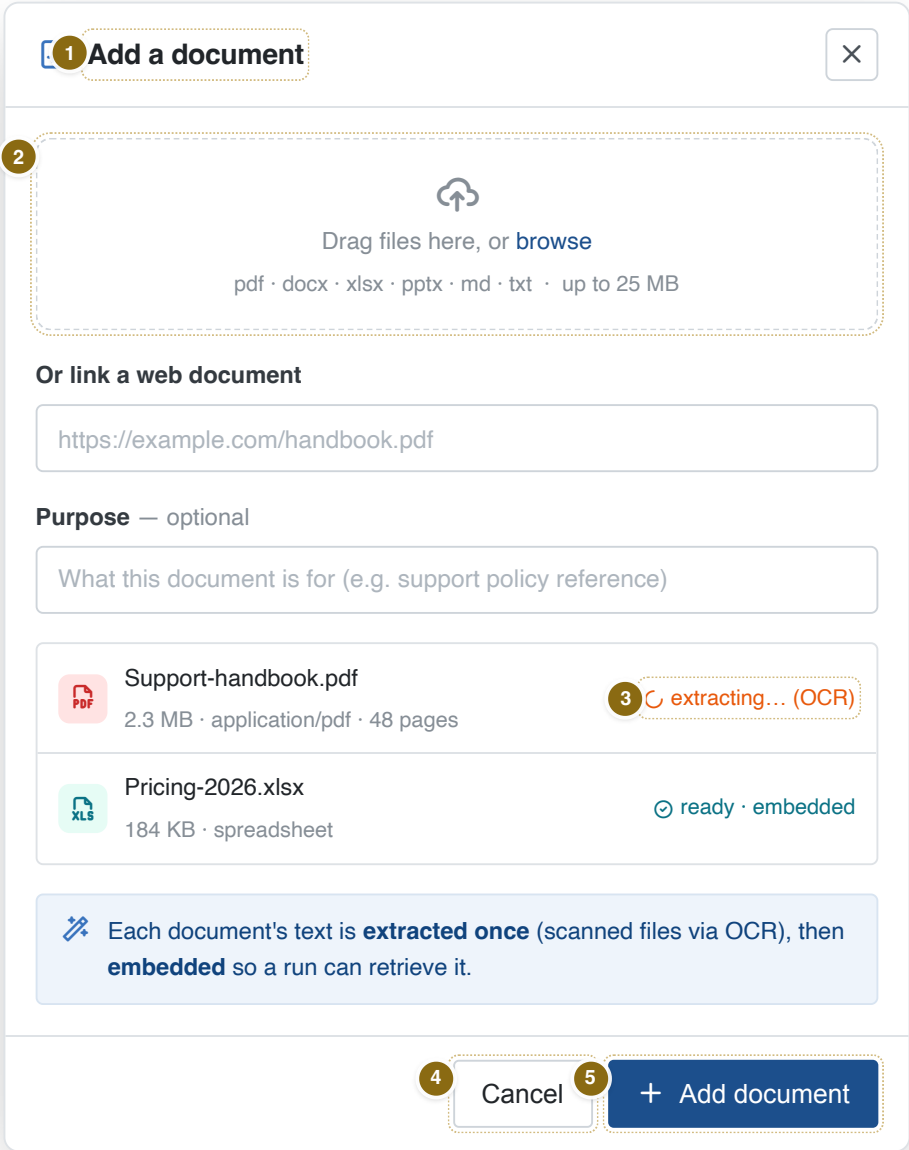


Upload document

upload-dialog.html · 760×652



1 Add a document ×

2

Drag files here, or [browse](#)

pdf · docx · xlsx · pptx · md · txt · up to 25 MB

Or link a web document

<https://example.com/handbook.pdf>

Purpose — optional

What this document is for (e.g. support policy reference)

 **Support-handbook.pdf**
2.3 MB · application/pdf · 48 pages **3** extracting... (OCR)

 **Pricing-2026.xlsx**
184 KB · spreadsheet ready · embedded

 Each document's text is **extracted once** (scanned files via OCR), then **embedded** so a run can retrieve it.

4 **Cancel** **5** **+ Add document**

Annotations

- 1 The Documents library — not chat text** Adds a document to the RAG library. A document has ONE source: an **uploaded file** (Word / Excel / PowerPoint / PDF — scanned → OCR) OR an **https:// link**. ChatMaestro extracts its text once (the extracted_text_path snapshot), then embeds it (pgvector) so a run can retrieve it. This is **not** an enrollee attaching a .txt to a chat message.
- 2** Scanned files are read by text recognition (OCR) so their words can be pulled out and searched.
- 3 Extraction then embedding** extraction_status: pending → ready. The middle tier extracts the text (a scanned PDF is run through **OCR**), writes the extracted_text_path snapshot, then an embedding job vectorises it for retrieval.
- 4** Discard changes and close.

- 5 Once added, this document can be attached to a run so its text can be pulled into the chats.