

Chat: NotebookLM & AI Document Tools

Conversation exported from Claude.ai

Jim

what is NoteBookLM

Claude

NotebookLM is an AI-powered research and note-taking tool developed by Google. It's designed to act as a personalized AI assistant that works specifically with your documents and sources — rather than the broader internet. You upload your own materials, and it grounds its responses in those sources.

Key features include source-grounded AI (it only draws from your uploaded sources when answering questions), summarization of lengthy documents, Q&A with cited answers, Audio Overviews (a popular feature generating podcast-style conversations between two AI hosts discussing your sources), and study guides and outlines generated from your sources.

Researchers, students, journalists, and developers use it to digest large volumes of documents — papers, codebases, reports — without losing track of what came from where.

It's free to use at notebooklm.google.com with a Google account. A paid tier called NotebookLM Plus is available through Google One AI Premium. It's particularly well-regarded for its citation transparency — every answer links back to the exact passage in your source material.

Jim

Do its documents need to be uploaded to a specific destination

Claude

No, you don't need to upload files to any special external destination. You add sources directly within NotebookLM itself. When you create a notebook, you're given options to add sources from several places: Google Drive (Docs, Slides, PDFs already in your Drive), local file upload (PDFs, text files, Markdown files from your computer), Google Docs/Slides links, YouTube URLs (it will transcribe and index the video), website URLs, and copied text (paste raw text directly).

Everything stays within your NotebookLM notebook. Google does process the content on their servers to power the AI features, so it's worth being mindful about uploading sensitive or confidential documents. But there's no separate storage system or third-party destination involved — it's all contained within the NotebookLM interface tied to your Google account.

Each notebook is isolated too, so sources you add to one notebook don't bleed into another.

Jim

can it work with the contents of a directory tree without specifying each individual file

Claude

Not natively. NotebookLM doesn't have a "point to a folder and ingest everything" feature. You have to add sources individually, which can be tedious for large directory trees.

Workarounds people use include: Google Drive folder sync (you can add files faster since Drive is integrated, but you still select files individually), combining files first (a script that concatenates all your files into a single large text or PDF document, then upload that one file — NotebookLM accepts up to 500,000 words per source), pasting aggregated text (a script that walks your directory and extracts content into text you paste in), and third-party tools (some people have built scripts that automate adding sources via browser automation, but there's no official API for NotebookLM source ingestion).

Given your work with directory trees and automation tools, a Python script that recursively walks a directory, filters by file extension, concatenates the content with filename headers, and outputs a single text file would probably be the cleanest solution.

Jim

are there other tools that aggregate and dissect information like NoteBookLM

Claude

Yes, quite a few. They vary in how they handle sources, collaboration, and depth of analysis.

Direct NotebookLM competitors include Obsidian + AI plugins (local-first note-taking with AI plugins like Smart Connections that let you chat with your vault — popular with developers since everything is plain markdown files in a directory you control), Mem.ai (AI-powered knowledge base that auto-organizes notes and lets you query across everything), and Notion AI (if you already use Notion, its AI can answer questions across your workspace content).

Document Q&A / RAG-focused tools include Perplexity (more web-search oriented but can analyze uploaded files), ChatGPT with file uploads (GPT-4o can ingest PDFs and documents, though without NotebookLM's citation transparency), Claude (documents can be pasted or uploaded and questions asked across them, though sources don't persist between conversations), and LlamaIndex / LangChain (developer frameworks for building your own RAG pipeline against a directory tree or database — very flexible, but requires coding).

Self-hosted / local options include AnythingLLM (open source, runs locally, supports pointing at folders and ingesting entire directory trees automatically — probably the closest to folder-level ingestion), Ollama + Open WebUI (local LLM setup with document ingestion capabilities), and Khoj (open source personal AI that indexes your notes, PDFs, and code).

Given your use case, AnythingLLM stands out because it can watch a local folder and ingest new files automatically.

Jim

can I ask AnythingLLM to upload only with my approval, e.g., don't upload everything new in a directory tree

Claude

That's a good question, and the honest answer is: it depends on how you configure it, but native fine-grained approval workflows aren't really its strength.

What AnythingLLM does offer: you can manually trigger ingestion rather than enabling auto-sync so nothing uploads without you initiating it, you can selectively add folders or files through its UI rather than pointing at an

entire tree, and the "watched folder" auto-sync feature (if enabled) will ingest everything new — there's no per-file approval prompt built in.

Practical workarounds include simply not enabling auto-sync and manually adding files when you want them indexed (this gives you full control but requires you to initiate each batch), using a staging folder (your automation drops candidate files into a holding directory, and you periodically review and move approved files into the folder AnythingLLM watches), or writing a small script that presents you with new files and asks yes/no before moving them to the watched folder.

Most of these tools are designed around the assumption that you want everything indexed. Fine-grained approval flows are more of an enterprise document management feature. If that level of control is important, a custom RAG pipeline using LlamaIndex or LangChain would give you the most flexibility.

Jim

can AnythingLLM sync with remote directories like my repos on github

Claude

Yes, AnythingLLM has a GitHub connector built in. It can connect to a GitHub repo by URL, ingest files from the repo into a workspace, you can specify a branch, and filter by file type (so you could limit to .md, .py, .rs, etc. rather than grabbing everything).

Limitations: it's more of a one-time or manual re-sync rather than a live webhook-driven sync (it doesn't automatically re-index when you push new commits), private repos require a GitHub personal access token, and large repos can be slow to ingest.

To keep it reasonably current you'd either manually trigger a re-sync periodically, or write a small script that hits AnythingLLM's API to trigger re-ingestion on a schedule or as part of a post-commit hook in your repo.

AnythingLLM also has connectors for YouTube videos, websites (crawl a URL), Confluence, SharePoint, and S3 buckets (so you could push files there as a sync intermediary).

Given that you maintain educational repos on GitHub, the GitHub connector could work well for keeping your workspace synced with your course materials — you'd just need to manually re-trigger ingestion after significant updates, or hook it into your git workflow.

Jim

create a pdf file for this chat including my prompts