Training Session for ICE HKA

Emerging AI Tools for Engineering Research & Case Study Discovery

Nov 2025

Aster Zhao
lbaster@ust.hk
Research Support Librarian
Hong Kong University of Science and Technology

Learning Goals

- Discover how AI tools can support research discovery and technical writing in engineering contexts
- Identify and compare tools for finding academic papers, case studies, and technology trends
- Apply best practices for using AI responsibly and effectively

Outline

1. **Two Types of AI Tools**

   General-purpose chatbots vs. Research-augmented tools

2. **Use Cases**

   - Scoping a new topic
   - Finding academic papers, case studies, design guidelines
   - Summarizing & engaging with long technical materials

3. **Limitations in Engineering Context & Responsible Use of AI tools**

General chatbots vs. Research tools

Trained on **open web knowledge;**
Designed for answering general questions;
May provide fake sources.

- **ChatGPT**, **Claude**, **Gemini** vpn
- **MS Copilot**
- **Perplexity**, **Grok**, **DeepSeek**
- **Poe** (most models) $

Knowledge cut-off
Internet access

Trained on **academic +/- open web data;**
Designed for research (in-depth search + synthesis); Provide real sources.

- **Google Scholar Lab** (new)
- **ChatGPT/Gemini** Deep Research $$$ vpn
- **Perplexity Pro** Deep Research $
- **Grok** (Expert)
- **Poe** (Deep research models) $

- **Elicit**
- **Consensus**
- **Undermind**
- **Ai2 Asta**

- **NotebookLM** vpn
- **Notion AI**

Always ask: “Where did the result come from?”

“Pure LLM”

answers mainly
from training data

VS.

“Retrieval-augmented” (RAG)

searches web or external sources
(then synthesize answer)

More relevant
Less hallucination

e.g.

- Models with a **knowledge cut-off date** (most models on Poe)
- Models that **do not have Internet access**

e.g.

- Models that **have Internet access**;
- Research tools trained on **academic sources**;
- NotebookLM, Notion AI that connect to your own materials (**knowledge base**)

Poe https://poe.com

- **Hosts 200+ official bots** + API
- **Updated timely** (with modified versions)
- Call on **multiple models** to do one task
- Free tier: 3k credits/month; Pro: 1M credits/month

- **Latest & most powerful models*:**
  - Gemini-3-Pro; Gemini-2.5-Pro (Google)
  - GPT-5.1, ChatGPT-4o-Latest, o3 (OpenAI)
  - Claude-Sonnet-4.5; Claude-Opus-4.1 (Anthropic)
  - Grok-4.1 (xAI)
  - Qwen3-Max; Qwen3-VL-235B (Alibaba)
  - DeepSeek-V3.1; DeepSeek-R1 (DeepSeek)

*Based on LLM arena leaderboard: https://lmarena.ai/leaderboard (as of 24 Nov 2025)

Use Cases

- Scoping a new topic
- Finding academic papers, case studies, design guidelines
- Summarizing & engaging with long technical materials

Use Case 1 - Scoping a new topic

From vague question → structured understanding of a topic

Scenario:

You are exploring building-integrated photovoltaic (BIPV) facade systems for high-rise office buildings in Hong Kong. You need a quick overview of *current* practices, performance expectations, and key risks, before diving into detailed standards.

Possible tools/models:

- Perplexity (Deep Research) (5 free quota/month)
- Grok-4 (Deep Search/Expert)
- ChatGPT (Deep Research) (3 free quota/day)

Source: Hong Kong Science Park, recreated with Nano Banana.

Use Case 1 - Scoping a new topic

From vague question → structured understanding of a topic

Prompt:

Provide an overview of building-integrated photovoltaic (BIPV) facade systems for high-rise office buildings in subtropical, high-humidity coastal cities such as Hong Kong.

Focus on:

- Typical annual electrical output (kWh/m^2·year) for vertical facades facing different orientations
- Thermal and daylighting impacts on building performance
- Common failure modes and maintenance issues
- Any case studies or guidelines from Hong Kong, Mainland China, or nearby regions (e.g. Singapore, Shenzhen, Guangzhou).

Provide sources to case studies, guidelines, and peer-reviewed papers.

Provide a summary of findings in table format in the end.

Use Case 1 - Scoping a new topic

Demo chats

- **Perplexity Pro search** (34 sources, shorter report, and fewer in-text citations)
- **Perplexity Pro (Gemini 3.0 Pro model)** (53 sources)
- **Perplexity Deep Research** (94 sources; comprehensive)

- **Grok (Grok 4.1 model)** (65 sources, but very few in-text citations)
- **Grok 4 Deep Search (Expert model)** (82 sources, more in-text citations)

- **ChatGPT Deep Research** (18 sources, ask questions for clarification)
- **ChatGPT Deep Research** - 2nd run (32 sources; ask different questions for clarification)

For reference (didn't demo in class):

- **GPT-5.1 (on Poe)** (fake references; no in-text citations)
- **Gemini-3-Pro (on Poe)** (fake references; no in-text citations)

Use Case 2 - Finding academic papers, case studies, design guidelines

In-depth search, read and synthesize

Scenario:

You are preparing a design note for a rooftop PV system on a commercial building in Hong Kong, assessing how cool roof coatings, high ambient temperatures impact PV module temperature and energy yield. You are looking for relevant journal articles, recent case studies, design guidelines, etc.

Possible tools/models:

- Perplexity (Deep Research)
- Elicit (Deep Research)
- Undermind / Ai2 Asta

Source: HKUST Solar PV System, recreated with Nano Banana.

Use Case 2 - Finding academic papers, case studies, design guidelines

In-depth search, read, and organize

Prompt:

Find recent case studies, field measurements, simulation studies, or design guidelines on rooftop PV systems installed over high-albedo ‘cool roofs’ in hot-humid or subtropical climates, especially in Hong Kong, Southern China, or Southeast Asia.

(List results with source links in table format, categorized by doc types)

Use Case 2 - Finding academic papers, case studies, design guidelines

Demo chats

- **Perplexity Deep Search** (74 sources; summary in table format, organized by doc type)

- **Elicit - Deep Research** (100 papers; literature matrix, can add custom columns)
- **Undermind** (189 papers; ask questions for clarification; exhaustive search)
- **Ai2 Asta** (normal: 69 papers; work harder: 274 papers; exhaustive search; reason for relevancy)

For reference (didn't demo in class):

- **Consensus Deep Research** (50 papers; a comprehensive report)
- **Google Scholar Lab** (31 papers; brief summary of each article)

Summary

	Grok / perplexity	Elicit / SCISPACE	Asta / Undermind
Primary Data Source	**Open web** (social media) + **Academic**	**Semantic Scholar** (academic papers)	**Semantic Scholar** (academic papers)
Best For...	**Find "Grey Literature":** government guidelines, industry reports, and news on recent events.	**Literature matrix** with custom fields; **Data extraction** (e.g. specific design parameter)	**"Guided search"** - initial clarifications; **Exhaustive search**
Weakness	Can **hallucinate** (fake citations, wrong technical details); Mix **non-peer-reviewed** sources.	Limited to **academic papers;** Limited to **abstracts** + **OA content;** Miss industry standards or local codes.	**Slower** (5~10mins); Limited to **academic papers;** Limited to **abstracts + OA;** Can be **overwhelming** with too many results.

Use Case 3 - Summarizing & engaging with long technical materials

Summarize, organize readings, ask direct questions

Scenario:
You receive a set of technical documents (over 200 pages), including a handbook on PV systems and relevant standards. Your task is to review the materials, extract key compliance requirements, and ensure your team applies them during the design, installation, and operation of PV systems.

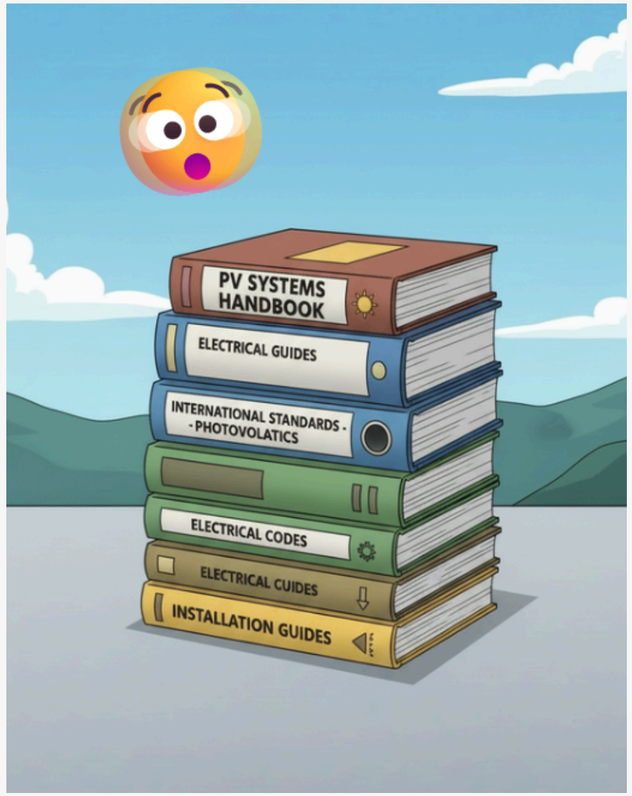
Possible tools/models:

- NotebookLM
- Perplexity (Space)
- Notion AI

NotebookLM

perplexity

Notion AI



Created with Nano Banana.
"A graphic in a similar cartoon illustration style, depicting a stack of technical documents related to PV systems and standards"

Use Case 3 - Summarizing & engaging with long technical materials

Summarize, organize readings, ask direct questions

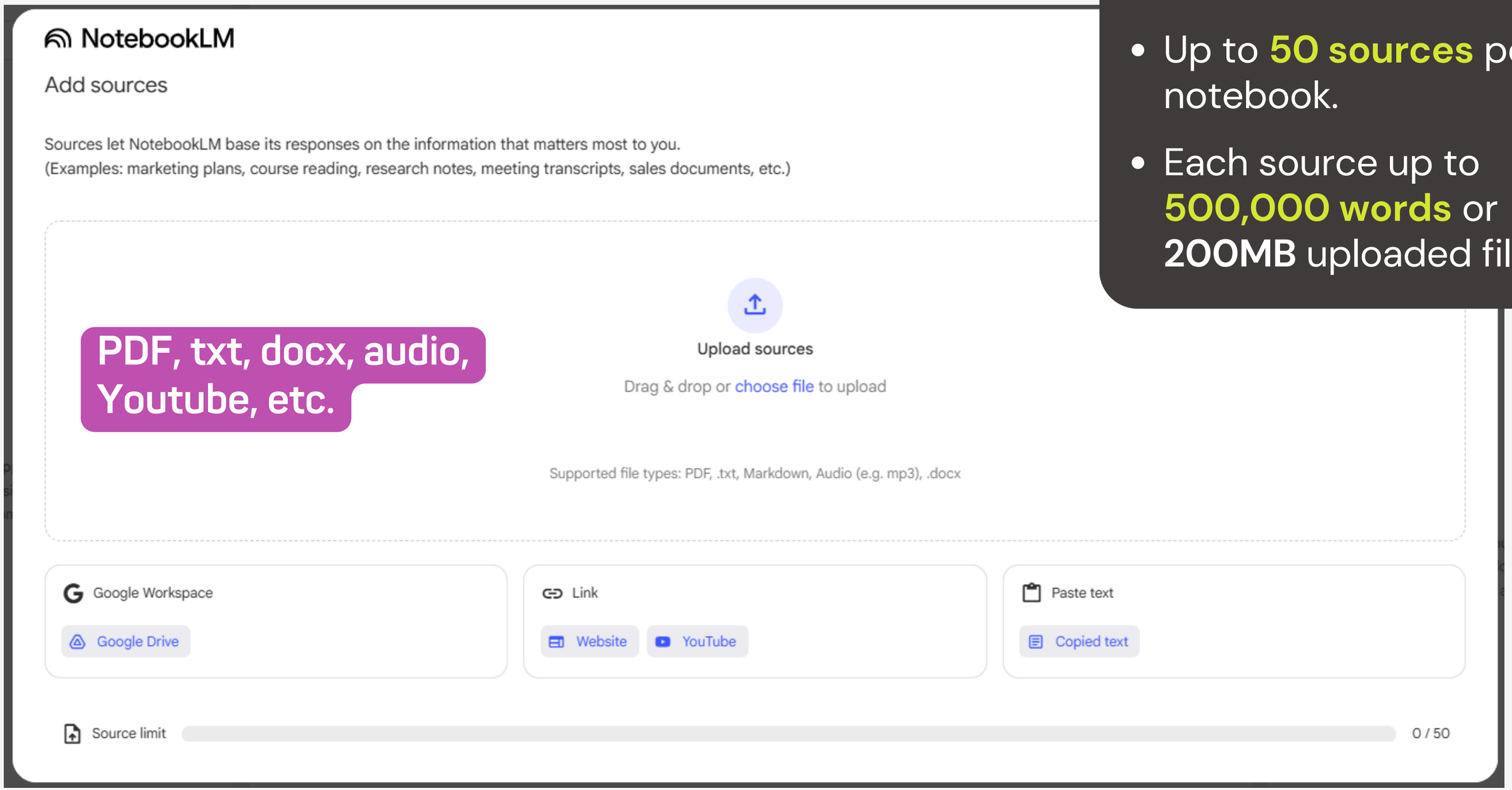
Possible questions:

- **Summarize the key points** in the XXX document, **in bullets / using "Template"**
- Ask direct questions, e.g. specific parameters

E.g.

- Summarize the key points in the Handbook, following this structure: xxx
- What installation tolerances or conditions are required for PV modules to maintain their certification under BS EN IEC 61215-1-4?
- What fire safety classifications are mandated for BIPV systems?

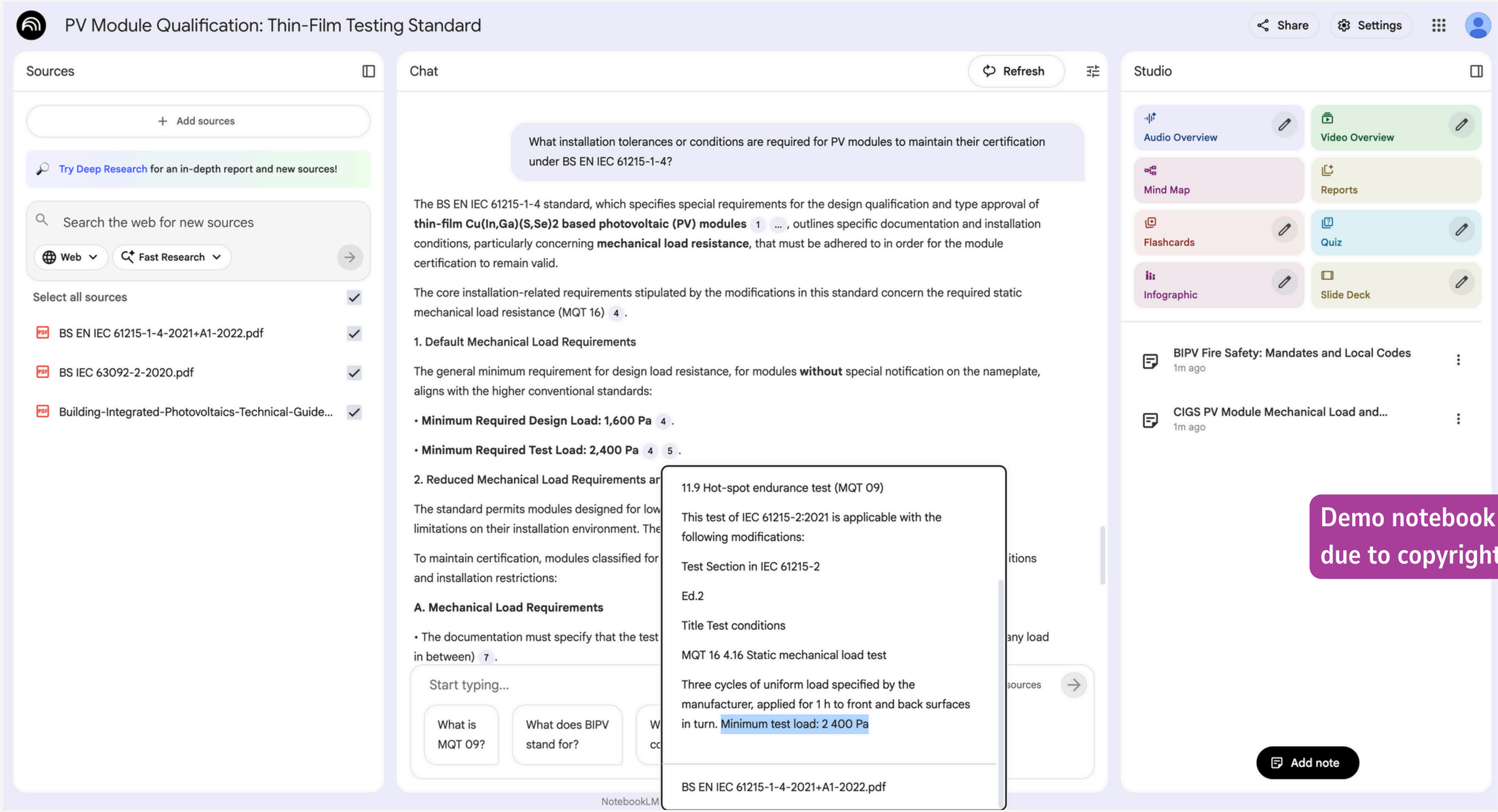
NotebookLM - Input sources



Free tier:

- Up to **50 sources** per notebook.
- Each source up to **500,000 words** or **200MB** uploaded files.

NotebookLM - e.g. ask direct questions against uploaded files



Demo notebook not shared due to copyright concerns.

Notion AI - e.g. custom columns on readings, ask questions within specific content

Reading

To Read | Aaron's Posts | Ethan's Posts | ALL | New

Publishe...	Title	Key Findings AI	Tools AI
11/21/2025	**Scholar Labs Early Review: Google Scholar Finally Enters the AI Era**	• Google Scholar has launched "Scholar Labs," an AI-powered academic search tool, aiming to enhance academic research capabilities. • Scholar Labs focuses on "Deep Search" rather than "Deep Research," providing relevant paper rankings without generating direct answers. • The tool analyzes user queries to identify key topics and searches Google Scholar for relevant papers, offering explanations for each paper's relevance. • Initial tests indicate Scholar Labs can find up to 50 relevant results, but it has a hard limit of 300 evaluated top results. • Queries can be entered in natural language, but the tool has restrictions that may limit more complex search commands. • The size of Google Scholar's index may give Scholar Labs an advantage in finding obscure papers, particularly in non-STEM disciplines. • Current limitations include a lack of complex search logic, limited filtering options, and difficulties in executing systematic reviews due to inherent Google Scholar constraints. • Evidence synthesis practitioners may find Scholar Labs useful but not sufficient as a standalone tool for comprehensive searches. • The launch of Scholar Labs could challenge discovery vendors, while content owners remain cautious about allowing full-text access. • Future improvements could include options for deeper evaluations, saving search sessions, and more transparency in the search process.	- Google Scholar - Gemini Deep Research - AI2 PaperFinder - Undermind.ai - Consensus - Microsoft 365 - HubSpot - Microsoft Teams
11/16/2025	**A 2025 Deep Dive of Consensus: Promises and Pitfalls in AI-Powered Academic Search**	• Consensus is an AI-powered academic search tool that assesses the "consensus" of research on topics. • It draws primarily from open sources like OpenAlex and Semantic Scholar, indexing approximately 200 million papers, including preprints. • The tool operates with a clean interface, allowing users to input natural-language questions and receive AI-generated answers with citations. • Consensus employs a hybrid search and multi-stage ranking system, considering research quality signals such as citation counts and journal reputation. • The "Consensus Meter" summarizes literature responses as "yes," "no," "mixed," or "possibly," but is criticized for its vote-counting methodology. • It offers three search modes: Quick, Pro, and Deep, which vary in the number of papers summarized. • Pre-filtering options for publication year, study design, and more enhance user experience. • LibKey integration allows users to check institutional access to full-text articles, improving accessibility. • Deep Search mode provides a detailed report with visualizations, but reproducibility and transparency	- Consensus - Elicit - Perplexity - Undermind.ai - Microsoft 365 - HubSpot - Microsoft Teams

COUNT 88

Tuesday, Nov 25 • Notion AI

summarize the readings related to Undermind

Searched data sources

Thought

Here's a synthesis of how Undermind is portrayed across your readings.

1. What Undermind is for

- Undermind is framed as an **AI "agent-style" academic search tool optimized for high recall**, not just top-N relevance.
- It runs **multiple iterative searches**, uses GPT-4 to classify papers into relevance tiers ("highly relevant / closely relevant / not relevant"), and **estimates whether the search is exhaustive**, sometimes suggesting extensions when it thinks more is needed.
- It is positioned as a **deep research / deep search tool**: something you use when you can wait for a more thorough sweep, rather than a fast, surface-level query.

2. Role among other AI academic search tools

Across comparisons and tool round-ups, Undermind shows up as:

- A **high-recall deep search option** contrasted with:
  - Elicit / SciSpace / Consensus / Scite as more "semantic search + RAG" engines.
  - Google Scholar as high-coverage but not optimized for systematic deep search.
  - Scopus AI / Web of Science RA / Primo RA as more library-integrated, RAG-style assistants.
- Guidance pieces suggest us**ndermind when you care** about thoroughness and synthesis (especially in

Reading

Ask, search, or make anything...

Auto | All sources | Edits

Limitations in Engineering Context & Responsible Use of AI tools

Limitation #1.

Limited understanding of current research & local context

=> **Outdated knowledge:**
General AI models have training cut-offs.
They may reference a standard that was valid in 2021 but was revoked or updated in 2023. This is no longer a knowledge gap but a liability issue.

=> **Lack of site-specific/proprietary data:**
General AI models do not know your “institutional knowledge”. They cannot troubleshoot a custom-built system as the data was never public.

Suggestion:

❌ DO NOT use GenAI as your single source for searching or fact-checking.

✓ Use them as a **companion.**
✓ **Compare & verify outputs** against the right sources.
✓ Use **RAG** tools/platforms that can read your internal files.

Limitation #2.

Risk of generating incorrect or biased content

=> **Hallucination:** General AI models can confidently tell you that “Alloy X has a yield strength of 1000 MPa” or cite “ISO9001:2026” before it exists. It’s no longer a fake citation issue but brings safety risk.

=> **Algorithmic bias:** AI can amplify biases present in its training data, e.g. favors *popular* answer not necessarily the *optimal* engineering solution.

=> **Calculation failure:** LLMs are language predictors, not calculators. At this moment, they are still bad at math.

Suggestion:

✓ **Source tracking:** Never accept a specific value without finding the manufacturer’s datasheet yourself.

✓ Instead of asking AI to do math, ask it to **write Python script or Excel formula**. Code generation is usually more reliable than direct calculation.

Limitation #3.

Lack of creativity, originality and depth

=> **“Average design” trap:**
General AI models are trained on existing data, and tend to provide the most standard, textbook solution. It cannot “think out-of-the-box” and struggles to solve novel constraints.

=> **No originality & IP nightmare:**
Since AI synthesizes what it “finds”, it may reproduce patented mechanisms or copyrighted code/standard without attribution.

Suggestion:

❌ DO NOT prompt AI “Suggest a new mechanism / workflow for X.”

✓ Write your own design proposal or hypothesis first, then ask the AI to critique it.

e.g. "Act as a senior structural engineer. Review my design logic below and list 5 potential failure modes I might have missed."

Limitation #4.

Lack of transparency and reproducibility

=> **“Black box” mechanism:**
Engineering (especially in civil, aerospace, and biomedical fields), requires an unbroken chain of custody for decisions. We do not know how AI do its “reasoning”, its decisions are inadmissible in safety-critical environments.

=> **Non-deterministic output:**
Engineering relies on **determinism**. Running the same query multiple times can return different results. In an engineering workflow, this inconsistency is dangerous.

Suggestion:

❌ DO NOT use AI-generated justifications in regulatory documents.

✓ When searching requirements, **force AI to ground its answer in research or specific engineering codes.**

e.g. "Summarize the fire safety requirements, citing specifically from NEC."

Data Privacy & Security

❌ **Do Not upload** proprietary data, unreleased site data, or confidential client reports to public AI models (e.g. ChatGPT, Gemini, Perplexity, or Poe) unless you are in certain Enterprise plans.

By default, your chats, uploaded docs are treated as "user content" and can be sent to service providers to **train their models unless you turn this off.**

To turn this off:

- **ChatGPT**: Data controls.
- **Gemini**: Gemini Apps activity.
- **Perplexity**: Preferences > AI data retention.

(Privacy policy: OpenAI; Gemini; Perplexity; Poe)

Consider:

- **Local RAG systems**
- **NotebookLM** - claim that they do not use personal data, including source uploads, queries, and model responses, to train AI models. (About NotebookLM)
- **Notion & Notion AI** - claim that they "do not use Customer Data to train any models" (including free-tier). "For all non-Enterprise plan workspaces, LLM providers (e.g. OpenAI) only retain Customer Data for 30 days or fewer before deletion." (Notion AI privacy practices)

Key Takeaways

- **Pick the right tool for the right task**
  - e.g. current practices → models that can search; find research papers → research tools
- **Always trace and verify**
  - In engineering, hallucination is a safety risk. Always trace AI claims back to the original standard or guideline.
- Use AI to scope topics and synthesize data, but **reserve the final engineering judgment for yourself**.