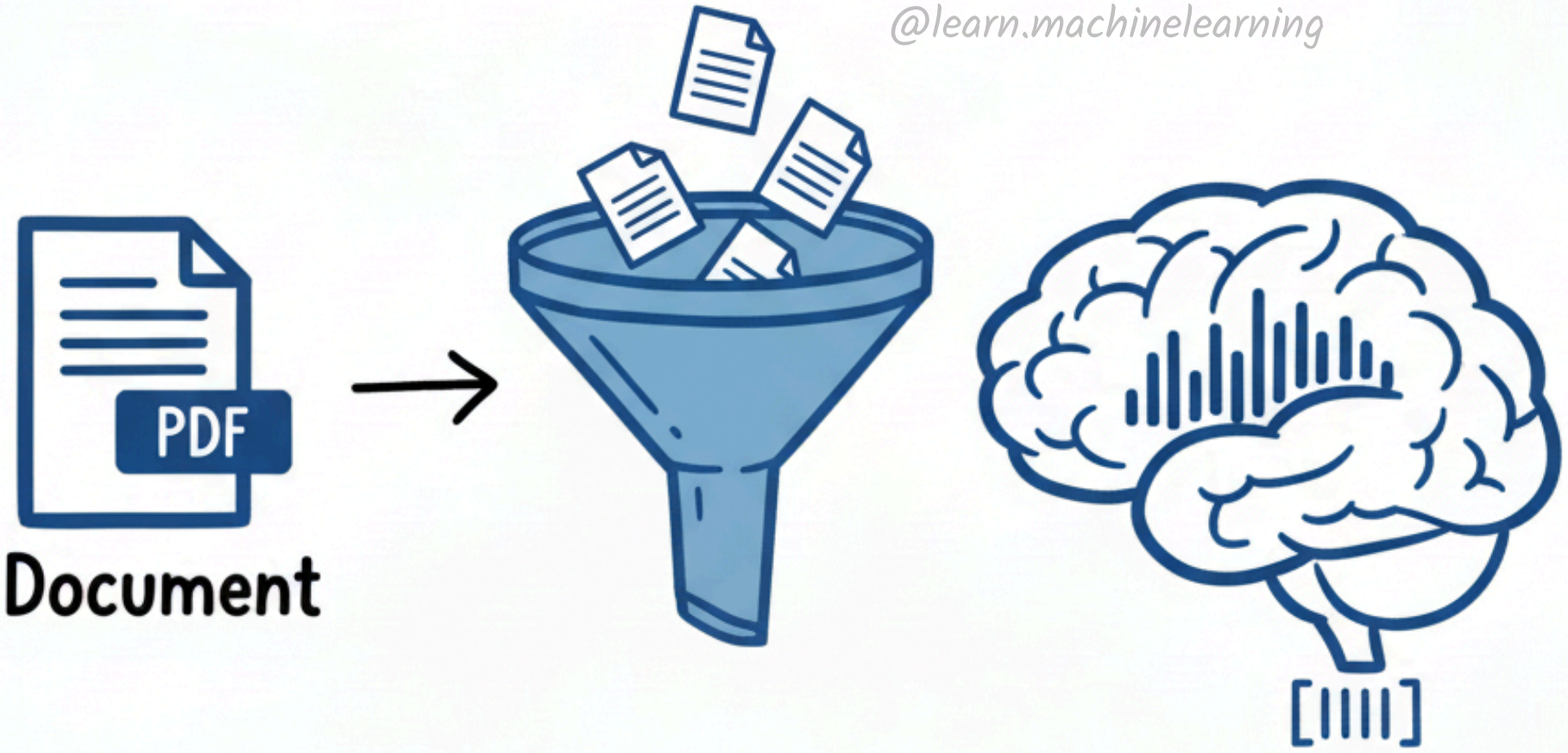
GENERATIVE AI SERIES: DAY 19

The RAG Indexing Pipeline: Building the AI's Brain

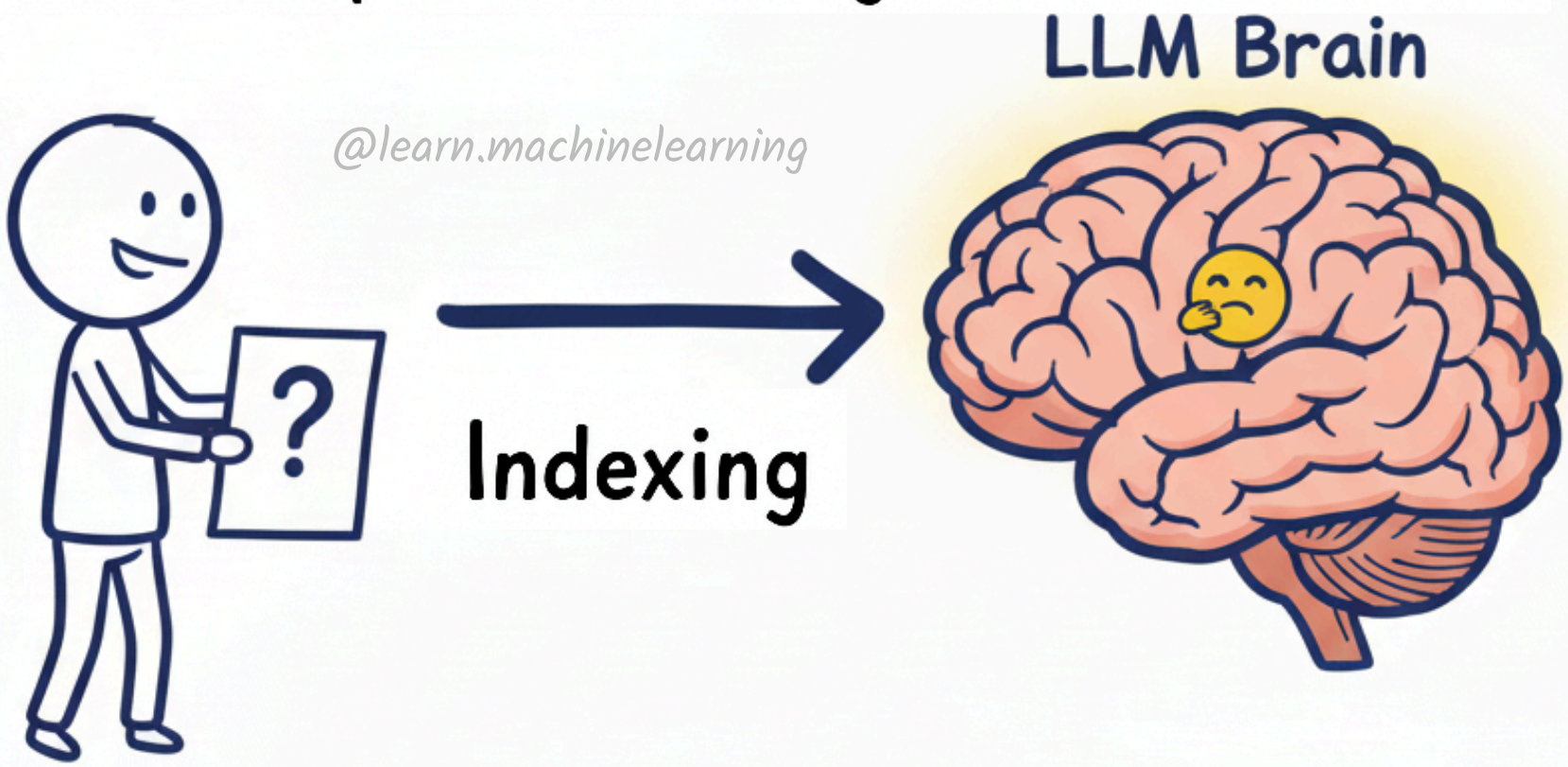
Ever wonder how an AI learns your private documents? It all starts here, in the crucial "offline" step of building its knowledge library!



LLMs Don't Know Your Stuff!

- A powerful LLM knows about the public internet, but it's cluneess about your:
- Company's internal reports
- Project documentation
- Personal notes or research

The Indexing Pipeline is how we teach the AI this specific knowledge.



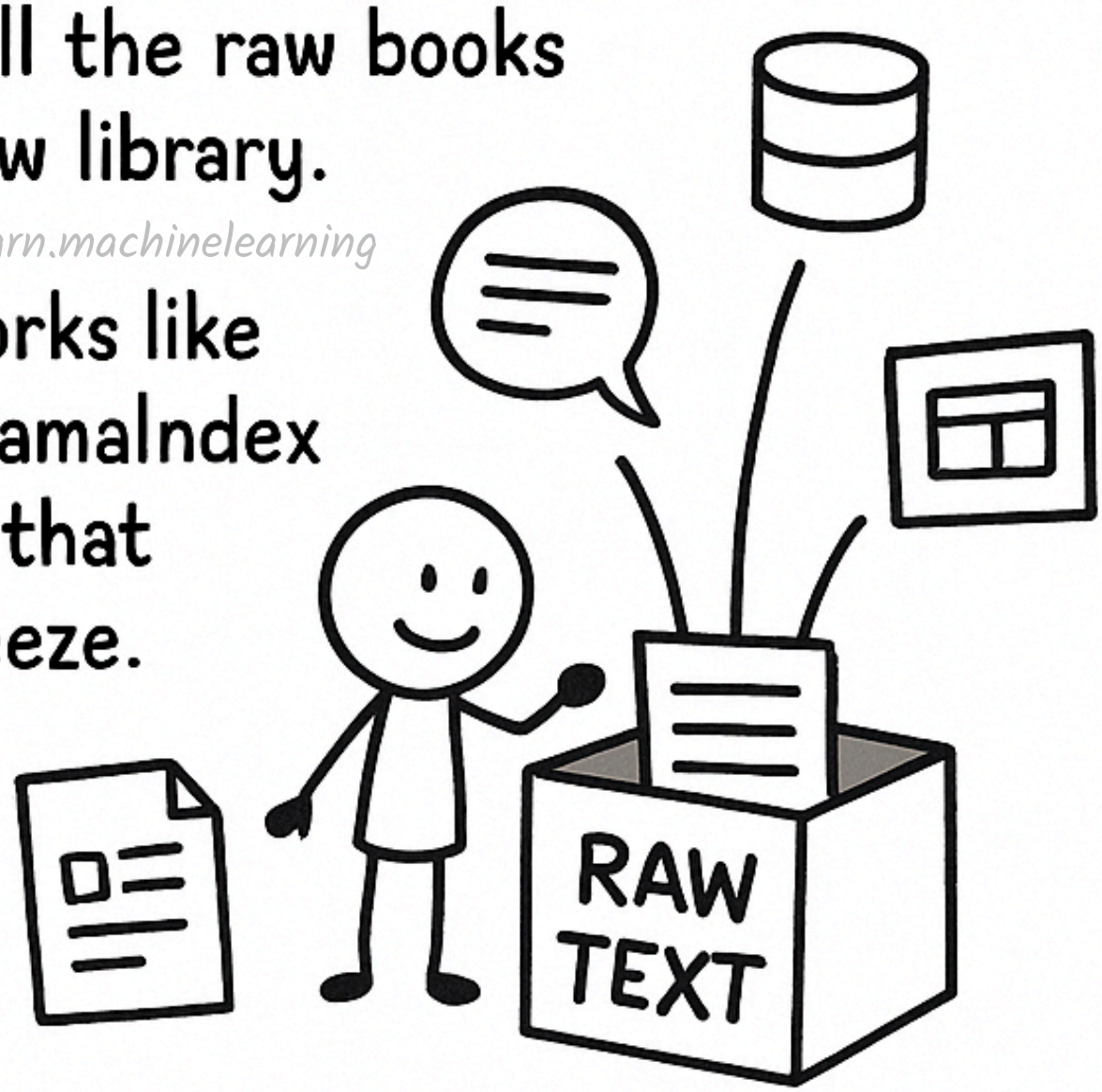
Step 1: Just Get The Data!

The first step is simple: reaad your data from wherever it lives and turn it into plain text.

Sources: PDFs, Websites, Notion, Databases...

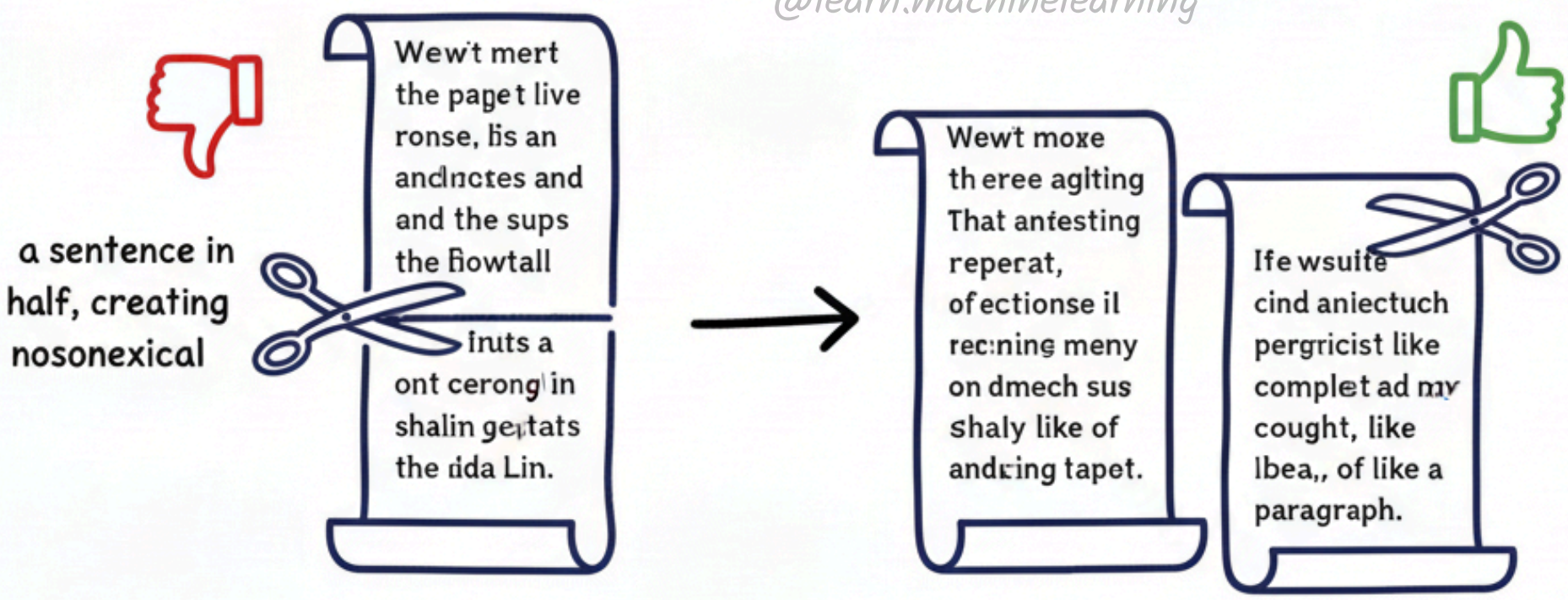
Goal: Collect all the raw books for our AI's new library.

Tools: Frameworks like LangChain & LlamaIndex have "Loaders" that make this a breeze.



Step 2: Break It Down!

- We can't give the AI a 100-page document all at once. We must split it into smaller, meaningful "chunks".
- **Why? For more precise search results and to fit model context limits.**
- **The Goal: Each chunk should contain complete thought our or idea, like a paragraph**

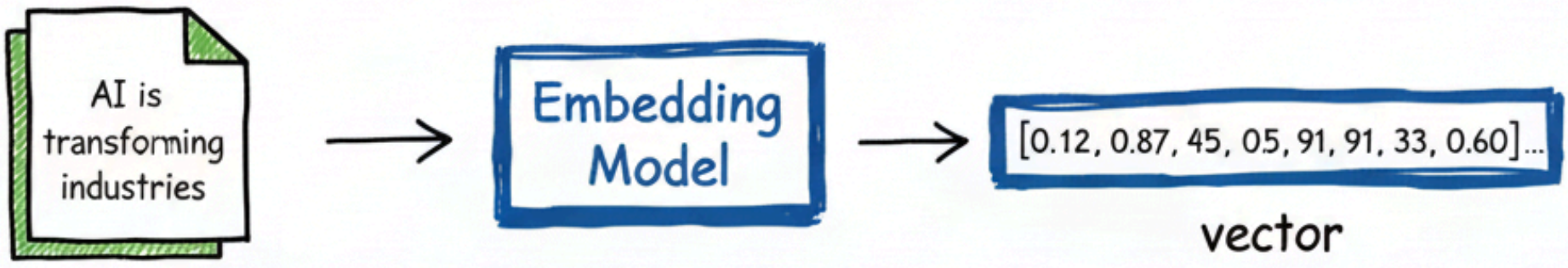


Step 3: The Magic Translation

- This is the core of the process! We use a special "Embedding Model" to convert each text chunk into a list of numbers called a **vector.**

- This vector represents the chunk's semantic meaning.

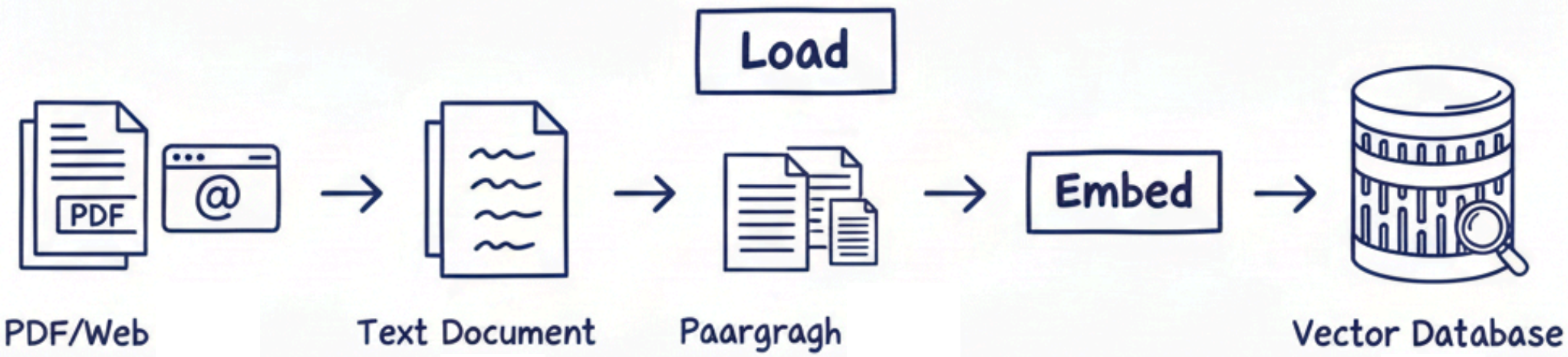
- Chunks with similar meanings get mathematically similar vectors.



The Full Blueprint: Load → Chunk → Embed

- Let's put it all together! This three-step process is the universal foundation for building the knowledge base of any RAG system.

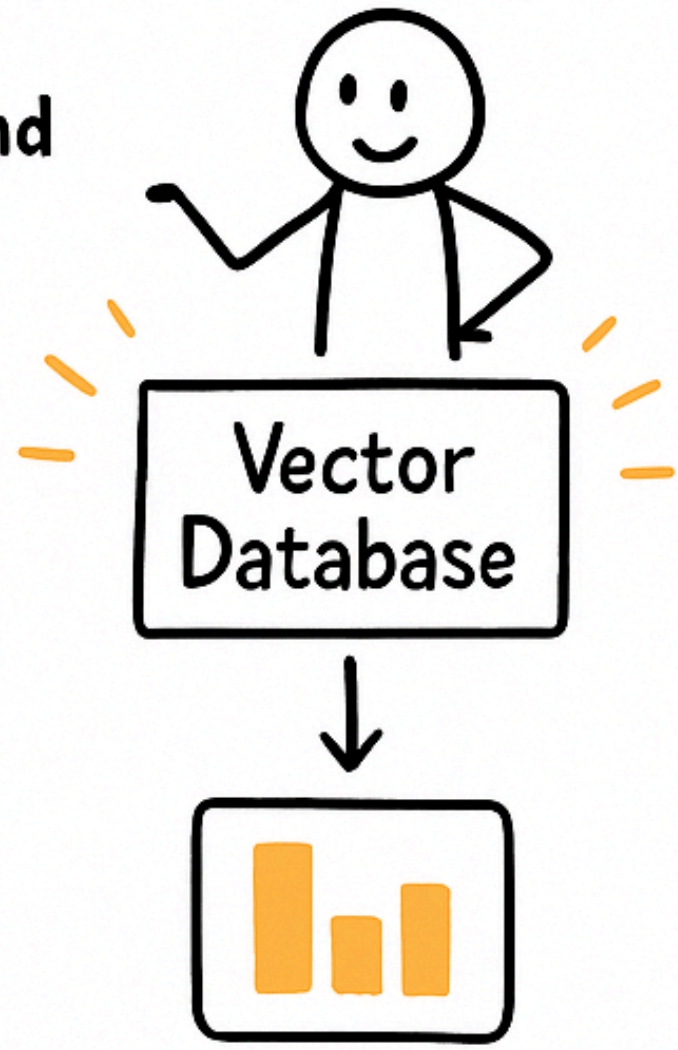
- **Load: Ingest documents from any source.**
- **Chunk: Split them into meaignful pieces.**
- **Embed: Convert each piece into a numerical vector and store it.**
- **The result? searcible library of knowledge, ready for user questions.**



Day 19 TAKEAWAY:

A Great Index is a Great RAG!

The quality of your RAG system depends almost entirely on its index.
By mastering the **Load, Chunk, and Embed** pipeline, you are building the strong foundation your AI needs to provide accurate, context-aware answers.



NEXT UP: Day 20: The Art of Retrieval: How RAG Finds the Needle in the Haystack!

What kind of documents would you want your own AI to learn from? Share below! 👇

Follow @learn.machinelearning for full RAG Deep Dive!