

The Data Behind AI

A Visual Guide to Databases,
Search, and Intelligent Systems

ILLUSTRATIVE ANSWER

**No. This return is outside
the 14-day window.**

Source: passage P-204

PASSAGE P-204

Return window: 14 days after delivery.

SEARCH INDEX

9001 / Customer 101 / \$24.00

DATABASE ROW

Order 9001 / Customer 101 / Total: \$24.00

THE READER'S ROUTE

Start at the source.

Follow information from its source to an AI answer. Each chapter adds a way to understand the systems in between.

01	AI Begins With Data	4
	What a model learns, and what an application retrieves.	
02	Inside a Database	8
	Tables, relationships, keys, and queries.	
03	How Data Moves	13
	APIs, pipelines, and keeping copies in sync.	
04	Can You Trust It?	19
	Validation, metadata, lineage, and missing values.	
05	How Search Finds Things	24
	Keywords, embeddings, and hybrid search.	
06	From Search Result to AI Answer	29
	Retrieval-augmented generation and its limits.	
07	Follow a Real Dataset	35
	A World Bank indicator, from definition to honest chart.	
Closing checklist	Glossary	References
41	42	43

HOW TO USE THIS GUIDE

Read in sequence for the whole picture. Return to the diagrams when you need to explain an idea. No coding required. Learn through diagrams and worked examples.

ABOUT THIS GUIDE

About this guide.

A visual guide to the information behind AI answers, for readers who want to follow the evidence. No coding required.

AUTHOR

Neura Collective

PUBLISHER

Neura Collective

EDITION

First edition

VERSION DATE

October 17, 2026

SERIES AND SITE

SI Database / sidatabase.com

READING THIS GUIDE

Read the diagrams as maps of a process. You do not need to run software or write the sample code. The code is there to read as a sentence.

EXAMPLES

Chapters 1 through 6 follow one invented store. Its customers, orders, policies, rankings, and AI answers are written teaching examples. Chapter 7 switches to clearly labeled real data from the World Bank, with its source and dates on the chart.

COPYRIGHT AND LICENSE

© 2026 Brandion LLC. All rights reserved in the original book text and diagrams. Third-party material retains its stated terms. The data license on the Chapter 7 chart applies to that data, not to the entire guide.

AI ASSISTANCE

ChatGPT assisted with drafting, diagrams, and source checks. Claude provided editorial reviews and checked the Chapter 7 values against the World Bank API. Neura Collective directed the project.

HOW TO CITE THIS EDITION

Neura Collective. (2026). The Data Behind AI: A Visual Guide to Databases, Search, and Intelligent Systems (1st ed.). Neura Collective. sidatabase.com

01 / AI BEGINS WITH DATA

An answer has a history.

Ask an AI assistant a question. Behind its reply may sit years of records, documents, revisions, and decisions about what to keep.

Consider a question about a store's return policy. The answer depends on which policy is available, when it took effect, and how the application uses it. A fluent sentence gives us little of that history on its own.

This book follows the information behind the sentence: how it is stored, how it moves, how it is checked, and how a system finds the part that matters.

**LOOK UNDER
THE ANSWER**

What supplied it?

How current is it?

Can you trace it?

Two ways data enters the story

A / TRAINING

Training uses examples to adjust numerical values called **parameters**. Those values shape how a language model processes input and generates text. Further training can change them. [1]

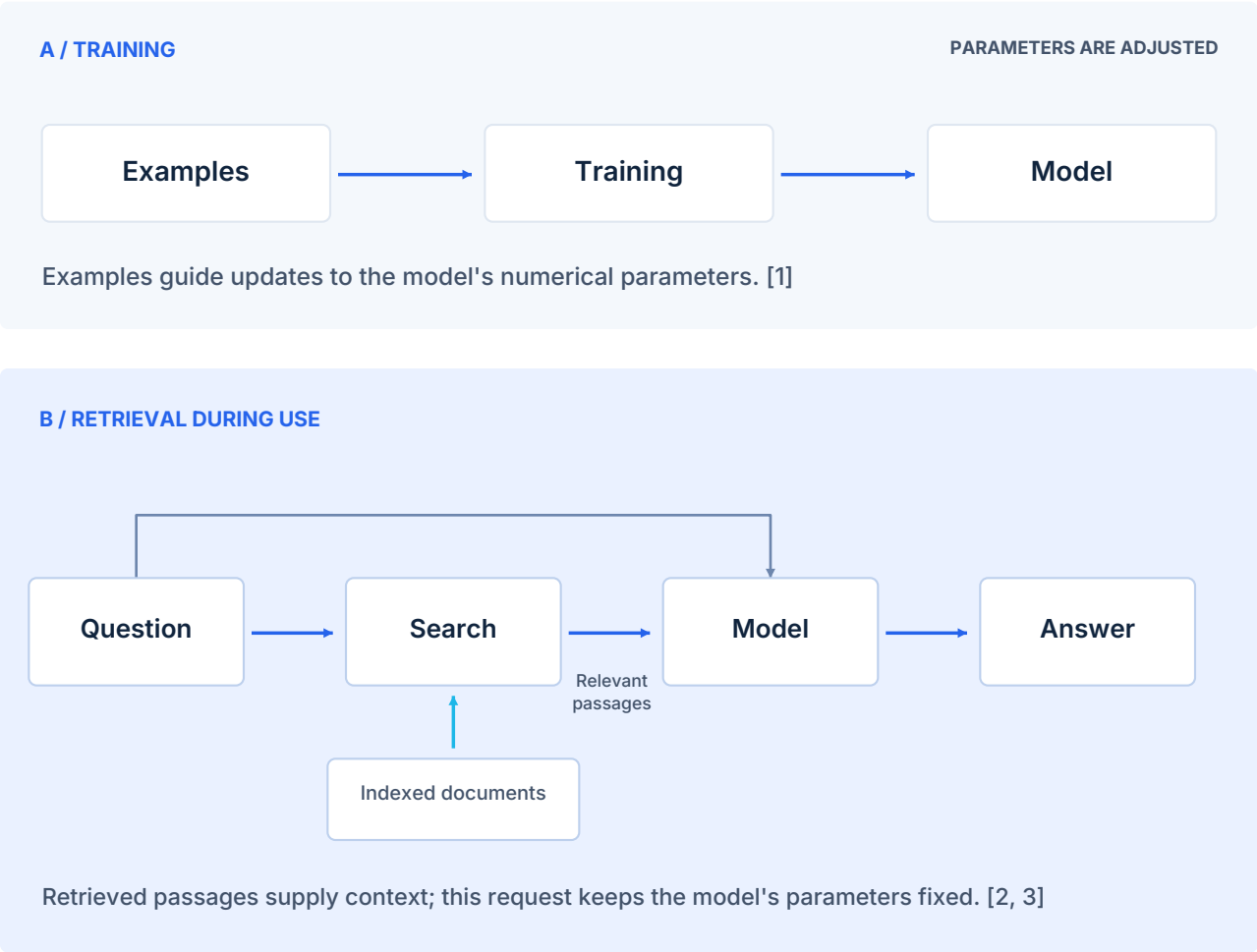
B / RETRIEVAL

An application can fetch relevant material when a question arrives and include it in the model's input. Using retrieved material to help generate an answer is called **retrieval-augmented generation**, or RAG. [2, 3]

[1] Google, ML Crash Course: LLMs [2] Lewis et al., RAG (2020) [3] Microsoft Learn, RAG overview

Learning and looking up

The same application can use both.



AN INVENTED STORE POLICY

A return window changes from 30 to 14 days after delivery.

The index, a searchable copy of the documents, must receive the revision to retrieve that version. The model can then use it in an answer, and that answer still needs checking.

[1] Google, ML Crash Course: LLMs [2] Lewis et al., RAG (2020) [3] Microsoft Learn, RAG overview

Diagrams are simplified.

01 / AI BEGINS WITH DATA

Same model. New passage.

A source revision can change the evidence without retraining.

INVENTED STORE / WRITTEN ILLUSTRATION

ONE QUESTION

"What is the return window for new orders?"

The store has changed the new-order window from 30 to 14 days after delivery. Both requests ask the same question. Request A runs before the index update; request B runs after it.

A / BEFORE THE INDEX UPDATE

SELECTED PASSAGE / RET-V1

"Return window: 30 days after delivery."



QUESTION + PASSAGE / SAME MODEL



ILLUSTRATIVE ANSWER

"30 days after delivery."

B / AFTER THE INDEX UPDATE

SELECTED PASSAGE / RET-V2

"Return window: 14 days after delivery."



QUESTION + PASSAGE / SAME MODEL



ILLUSTRATIVE ANSWER

"14 days after delivery."

Changed input. Fixed parameters.

The application supplies a question and a selected passage. [2, 3] These requests keep the model's parameters fixed. The answers are written for teaching; changing the passage does not guarantee that an answer follows it.

[2] Lewis et al., RAG (2020) [3] Microsoft Learn, RAG overview

Diagrams are simplified.

01 / AI BEGINS WITH DATA

Three ways to be misled.

Use the store policy to separate a claim from its support.

INVENTED STORE / NEW-ORDER WINDOW: 14 DAYS AFTER DELIVERY

A

No passage

INPUT

The application supplies no store policy.

CLAIM

"The return window is 30 days."

CHECK

No retrieved evidence supports that number. The number may come from training, not from the store. [1] Ask which source supplied it.

B

Wrong version

INPUT

Search supplies RET-V1, the old 30-day passage.

CLAIM

"New orders have 30 days."

CHECK

Compare the passage with RET-V2. Check which version applies to the order.

C

Lookup mistaken for learning

INPUT

Request B supplies RET-V2 after the index update.

CLAIM

"The model learned the new policy last night."

CHECK

The application changed the passage; the model's parameters stayed fixed.

The application selects the evidence; the model generates the answer. [3] Check both the supplied passage and the claim the answer makes.

[1] Google, ML Crash Course: LLMs [3] Microsoft Learn, RAG overview

REFERENCES / 1 OF 6

Sources you can follow.

Sources for Chapters 1 and 2. Links checked October 17, 2026.

Sources are grouped by chapter; shared sources keep the same numbers.

Sources for Chapter 1.

- [1] LLMs: What's a large language model?**
Google. Machine Learning Crash Course.
<https://developers.google.com/machine-learning/crash-course/llm/transformers>
- [2] Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks**
Patrick Lewis et al. NeurIPS, 2020. arXiv:2005.11401.
<https://arxiv.org/abs/2005.11401>
- [3] Retrieval-augmented generation (RAG) in Azure AI Search**
Microsoft. Microsoft Learn.
<https://learn.microsoft.com/en-us/azure/search/retrieval-augmented-generation-overview>

Sources for Chapter 2.

- [4] PostgreSQL 18 Documentation: Concepts**
PostgreSQL Global Development Group. PostgreSQL 18, section 2.2.
<https://www.postgresql.org/docs/18/tutorial-concepts.html>
- [5] PostgreSQL 18 Documentation: Constraints**
PostgreSQL Global Development Group. PostgreSQL 18, section 5.5.
<https://www.postgresql.org/docs/18/ddl-constraints.html>
- [6] PostgreSQL 18 Documentation: Querying a Table**
PostgreSQL Global Development Group. PostgreSQL 18, section 2.5.
<https://www.postgresql.org/docs/18/tutorial-select.html>
- [7] PostgreSQL 18 Documentation: Joins Between Tables**
PostgreSQL Global Development Group. PostgreSQL 18, section 2.6.
<https://www.postgresql.org/docs/18/tutorial-join.html>

Several sources use Azure AI Search as a documented example; other search systems follow similar ideas. Implementation details vary.