Free Preview

This file is a free preview of the final sealed publication of Adventures in AI, Book One: From Document to Evidence.

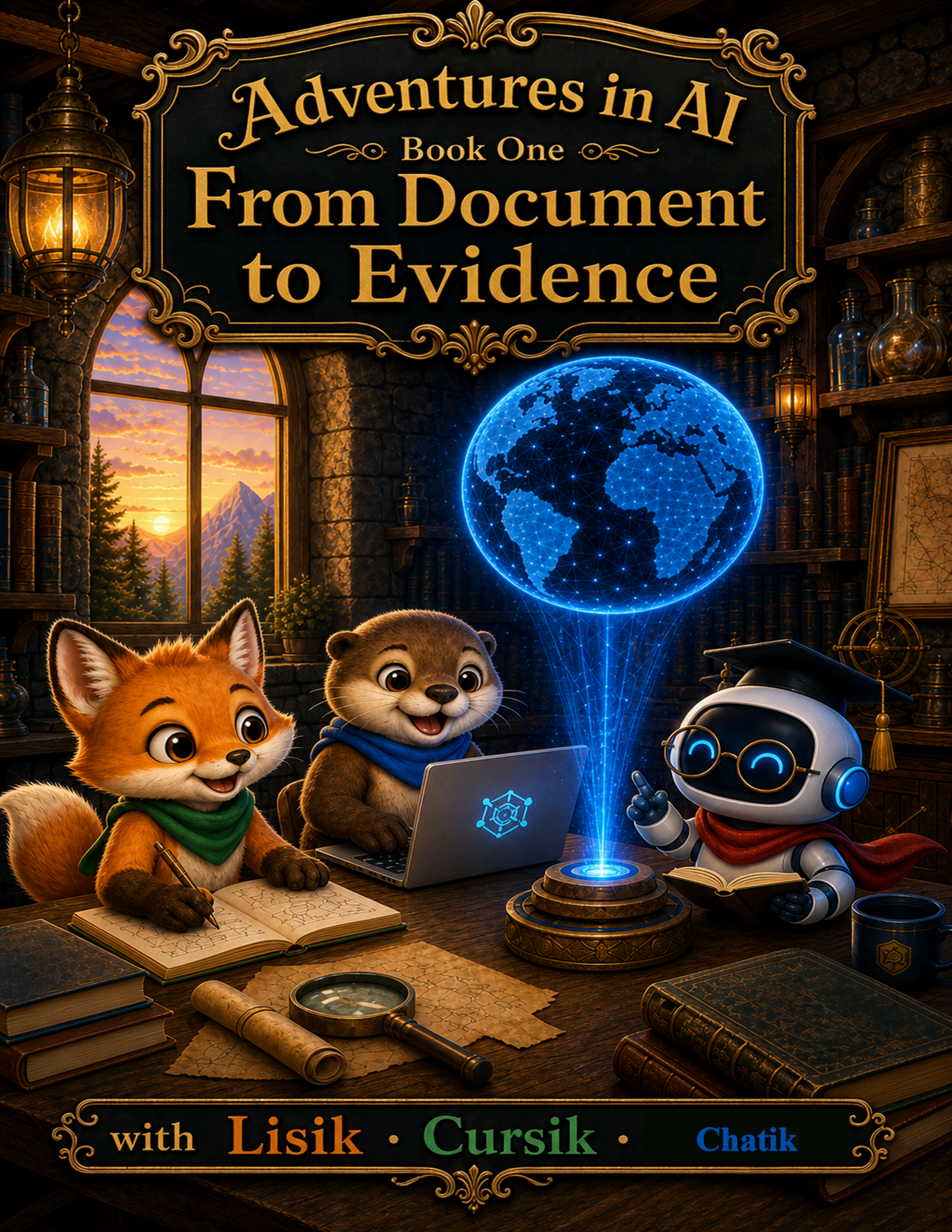
It contains the sealed front matter through the Contents, book pages 1–9, and the Chapter 1 reader narrative, book pages 10–14.

The Chapter 1 Evidence Appendix begins on book page 15. That appendix, and the rest of the book, are in the full publication only.

Adventures in AI

Book One

From Document to Evidence



with Lisik · Cursik · Chatik



Orchestrated by Joseph Fox

BOOK ONE

From Document to Evidence

Copyright & provenance

Series: Adventures in AI with Lisik · Cursik · Chatik

Book One: From Document to Evidence

Author and Research Orchestrator: Joseph Fox

Published by: LCC Research Lab LLC

Publication year: 2026

© 2026 LCC Research Lab LLC. All rights reserved.

The science stays rigorous. The storytelling doesn't have to be boring.

This book explains research records through a continuous narrative. Empirical claims belong to the measured conditions described in each chapter.

Source documents and third-party names remain the property of their respective owners. This book is not medical, legal, software, or driving advice.

Adventures in AI — Series Dedication

To my beautiful wife, Julia,

who has always inspired my curiosity
and encouraged me to keep learning.

After thirty-five years in high-tech software,
she helped me discover that finishing one career
did not mean finishing the adventure.

This new journey began with her encouragement.

Дорогу осилит идущий.

The road is mastered by the one who walks it.

Meet the team

The characters behind LCC Research Lab.
Curiosity, engineering, scientific review, and human direction—each has a part in how we learn.

Lisik — The Curious Fox

Asks why, challenges easy answers, and reopens settled questions.

Cursik — The Software Guru

Builds systems, organizes experiments, and turns chaos into writing.

Chatik — The Professor

Scientific skeptic; protects the ruler: “Did we actually measure that?”

Joseph Fox — The Human Orchestrator

Mediator and final decision-maker.

How the LCC Research School Works

I couldn't reduce it to the freshman level. That means we don't really understand it.

— Richard Feynman (as recalled by David L. Goodstein)

The **LCC Research School** is the experimental classroom used in Book One to put every claim to a test.

It has one Teacher and two Students.

Teacher — AI Instructor Model
A strong source-informed instructor; what Teacher knows is not automatically what a Student can learn.

Students — Rookie · Freshman
Rookie is the smaller Student; Freshman is the stronger Student. Size alone does not settle learning.

> *A Student is a model identity. An experiment can change what happens to that Student without creating a new Student.*

LCC Research Lab is the team conducting the research. **LCC Research School** is the experimental setting in which the teaching studies take place.

Nobody knows the ending yet. Welcome to the laboratory.

How the LCC Research School Works

From Source to Learning — how we teach, test, and find out what was learned.

1 SOURCE — The Material



We begin with a document (or set of documents) we want the Student to learn.

2 TEACHER — AI Instructor Model



Reads and analyses the source, prepares teaching material, and supports experimental design.

Teacher is not the Student being tested.

3 TEACHING MATERIAL — Produced



Structured lessons, examples, quizzes, or other training material produced by Teacher.

4 SCHOOL — Delivers



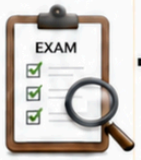
The School (our experiment pipeline) controls how the material reaches the Student.

5 STUDENT — Learns?



The Student (Rookie or Freshman) starts from a known checkpoint and receives the lesson.

6 EXAM / MEASUREMENT — What Changed?



We measure the Student before and after. The rule is the exam set, not our feelings.

7 EVIDENCE — The Outcome



The results become Evidence. Only then do we decide what the evidence permits us to say.

We iterate, redesign lessons, or change methods based on the evidence.

Who Does What?



Lisik — The Curious Fox

Asks the comfortable questions. Challenges assumptions and keeps curiosity alive.



Cursik — The Software Guru

Builds the pipeline, tools, and experiments. Turns technical chaos into reproducible work.



Chattik — The Professor

The scientific skeptic. Protects the rule. Demands evidence and rejects wishful thinking. Audits all claims.



Joseph Fox — The Human Orchestrator

Mediator and final decision-maker. Decides GO or STOP based on the evidence. Owns the outcome.



Teacher — AI Instructor Model

Reads and analyses the source, prepares teaching material, and supports experimental design.

Teacher is not the Student being tested.



Rookie — The Small Student

Smaller model (~360M parameters). Tries to learn. Sometimes surges, sometimes struggles.



Freshman — The Stronger Student

Larger model (~1.7B parameters). Trained to test the limits. Not a guarantee of success.



You — The Auditorium

You are the excimernal auditors. You watch the experiment, see the evidence, and decide what you believe.



Core Rule of Our Laboratory

Nobody in this loboratory is allowed to change the answer after seeing the exam.

Technical Identities (High Level)



Teacher (AI Instructor Model)

A powerful AI used to understand sources, prepare teaching material, and assist experimental design.
Exact model identity is reported in the experiment when it matters.



Rookie (Student)

Smaller open model.
~360M parameters.
Quantized for efficient training in our environment.



Freshman (Student)

Stronger open model.
~1.7B parameters.
Used to test whether more capacity changes the outcome.



Chattik (Professor)

Auditor and skeptic.
Verifies claims, protects the methods, and enforces the rule in every breach.



Joseph Fox (Orchestrator)

Human authority and final arbiter.
Owns the GO/STOP decision based on pre-registered evidence.

Where the Details Live

Exact model names, checkpoints, LoRA configs, hyperparameters, seeds, batch sizes, learning rates, iteration counts, retrieval settings, and all run-specific choices are documented in the experimental chapters.

Here, we show the architecture.
In the chapters, we show the measurements.



Important Distinctions

Teacher is not Student.
The source of instruction is not the subject of the test.

Training is not Evidence.
Practice makes attempts. Evidence comes from tests.

Good answers from Teacher do not prove that Student learned anything.
Teaching sets a path the same as learning.

These distinctions are the heart of experimental science.
Protect the rule.
Protect the meaning.

LCC Research School: curiosity, engineering, instruction, learning, and evidence.

Contents

1 Chapter 1 — Hello, World: Teaching a Model the DMV Handbook 10

2 Chapter 2 — Next Stop: Pro Git 18

3 Chapter 3 — Boosted by Medicine 29

4 Chapter 4 — Puzzled by a Fairy Tale 40

5 Chapter 5 — What Survives Across Documents? 49

6 Chapter 6 — What Does It Mean to Teach an AI? 55

Glossary 70

Acknowledgments 71

The Laboratory Song — Why, Why, Why: Living with AI 72

From Document to Evidence — Book One

Chapter 1 — Hello, World: Teaching a Model the DMV Handbook

Welcome to the LCC Research Laboratory.

We had a document, a couple of small AI Students, and what sounded like a very simple question:

Can we teach an AI something it did not know before?

In our *Adventures in AI*, simple questions rarely stay simple.

Lisik, the Curious Fox, immediately asked the question that would follow us throughout this journey:

How would we know?

A correct answer alone would not prove we had taught the Student anything. Maybe it already knew. Maybe we accidentally asked the same question twice. Or maybe we simply put the answer in front of it while it replied.

So before we could teach anyone, we needed three things: a classroom, a lesson, and a ruler honest enough to tell us whether anything had actually changed.

For our first classroom — our Hello, World — we chose something wonderfully ordinary: the California DMV Driver Handbook. A document gives us a controlled laboratory: we can know what the Student received, how it was taught, and whether learning occurred. The longer destination is a general way of teaching AI from new sources; this was only the first small program to test whether the laboratory worked.

And that is where our experiment began.

Four ways a Student can answer

Picture a Student in front of us. We can help it answer in several different ways — and if we mix those ways up, we can easily misunderstand the result.

We called our starting Student **Hub** — the model before we taught it anything from this DMV lesson.

With **LoRA**, we teach. We train the Student on examples from the handbook, hoping that afterward it can answer better without having the handbook in front of it.

With **RAG**, we do something very different. Instead of asking the Student to remember the handbook, we let it look up relevant passages while it answers.

With **LoRA + RAG**, we do both: teach the Student and still let it consult the evidence.

Now the experiment becomes interesting.

If teaching helps, that is useful. And if giving the Student the handbook helps, that is useful too.

But what happens when we do both?

If the Student already has the right handbook passages in front of it, does teaching still add anything?

That is why we needed all four ways of answering.

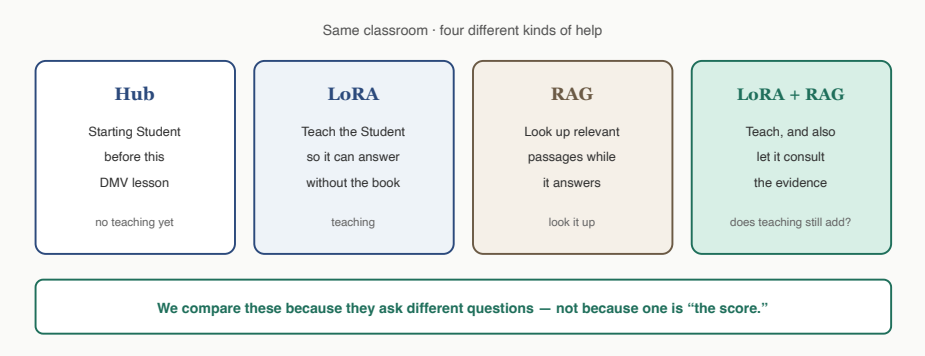


Figure 1

Figure 1. Four kinds of help in the same classroom — not four versions of the same score.

How can we tell whether teaching actually changed it?

Lisik would not let us skip this part. Cursik agreed: without an honest ruler, a correct answer is only a rumor.

Suppose we ask a Student a question, teach the Student, and then ask exactly the same question again.

If the answer improves, did the Student learn the underlying knowledge — or merely become better at that particular question?

That is why we needed a way to measure change that did not simply repeat the same exam.

We needed two equivalent exams

So we used two exams: **Form A** and **Form B**.

They are meant to test the same underlying knowledge through different questions. Form A measures the Student **BEFORE** teaching. Form B measures the Student **AFTER** teaching — without simply repeating the first exam.

The difference between those two measurements is how we ask about **acquisition**: did the Student show something it had not shown before?

A real A/B pair (the red-arrow rule)

Here is that idea with a real handbook rule — the red-arrow rule.

Example A (Form A — before teaching): According to the California Driver's Handbook: you are waiting at an intersection controlled by a traffic signal showing a red arrow in the direction you want to go. What does the handbook prohibit you from doing?

Expected answer: Do not turn at a red arrow.

Example B (Form B — after teaching, same underlying rule): From the handbook's rules of the road: when a signal displays a red arrow for your intended turn, what turning action is not allowed?

Expected answer: Do not turn at a red arrow.

These two questions are worded differently on purpose. They test the **same** handbook rule — so an improvement from A to B is more about that knowledge than about memorizing one exact wording.

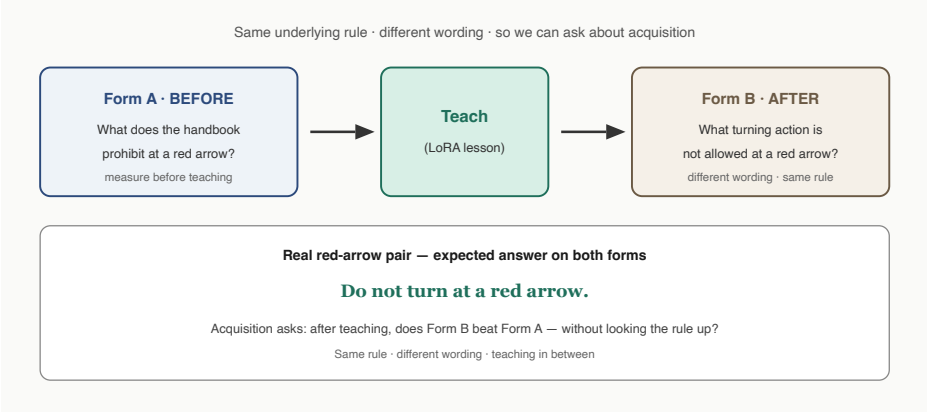


Figure 2

Figure 2. Two exams around the same red-arrow rule, with teaching in between. Expected answer on both forms: Do not turn at a red arrow.

From a handbook to a lesson

Form A and Form B gave us the ruler. But a ruler is not a lesson.

We still needed something to teach — and it could not be the exam itself.

We could not simply hand the Student our exam questions and call that teaching. That would spoil the measurement: a better score might only mean the Student had already seen the test.

So we went back to the classroom source: the California DMV Driver Handbook. With Chatik's help, we turned selected handbook rules into a structured list of facts we wanted to teach — simple links such as "a red arrow means do not turn." In the experiment, we called those facts **knowledge relations**. From those knowledge relations we built the **teaching material**. Separately, we kept Forms A and B as the ruler — testing the same knowledge without simply repeating the lesson.

For example, one underlying fact was:

Red arrow → Do not turn at a red arrow.

That fact could appear in more than one role:

Teaching material: what we use to teach the rule.

Form A / Form B: independent questions used to measure acquisition.

Later we would test the same knowledge again in a different way. We will meet that test shortly.

Here is a real teaching item from the frozen lesson for that same red-arrow rule:

Teaching example (TRAIN_R001): According to the California Driver's Handbook, what is the rule? Context target: Do not turn at a red arrow.

Expected answer: Do not turn at a red arrow.

Notice how different this is from Form A and Form B. Same underlying fact — different job. Here the Student is being *taught* the rule, not tested on the acquisition exam's wording.

That separation matters. If we trained the Student directly on the test questions, a better score could simply mean it had seen the exam.

For now, what matters is that the lesson and the ruler were ready before we looked at the outcome.

We teach

We taught two Students — a smaller **Rookie** and a larger **Freshman** — using that teaching set.

The lesson was over. Now came the uncomfortable question: **did it work?**

What changed? Acquisition

On the A/B ruler — Form B after teaching, minus Form A before teaching, with **no** handbook lookup at answer time — the experiment measured **positive acquisition**: about **+0.22** for the Rookie and **+0.25** for the Freshman.

So yes: after teaching, both Students could demonstrate handbook knowledge they had not demonstrated before.

That sounds like the end of the story. It is not.

But Lisik wasn't satisfied

The Students had improved on our acquisition ruler. Good. But could they use the same knowledge when the question looked different again?

After the AFTER exam, we asked a different kind of question — not the A/B exam again, but a separate check that the same rule had become usable beyond that exam's wording.

Here is one real probe from that check — the same red-arrow rule the reader already met as Form A and Form B:

FITTED probe (same rule, different surface): Handbook check: What does a red arrow forbid you to do?

Expected answer: Do not turn at a red arrow.

Notice what this is doing. It is still the red-arrow rule. It is **not** Form A, and it is **not** Form B. It asks whether the taught knowledge shows up under yet another wording, after teaching and after the acquisition exam.

We called those probes **FITTED**.

Surprise: FITTED is not acquisition

When we scored those probes, the two Students told different stories.

The Rookie answered most of those probes well. The Freshman answered far fewer well. Yet their acquisition results were similar.

A Student can show positive acquisition without looking strong on these later probes. A Student can look strong on these probes without that being the same fact as acquisition. The two measurements are allowed to disagree — and here, they did.

So even "the Student acquired something" is already too simple if we treat every correct answer as the same kind of success.

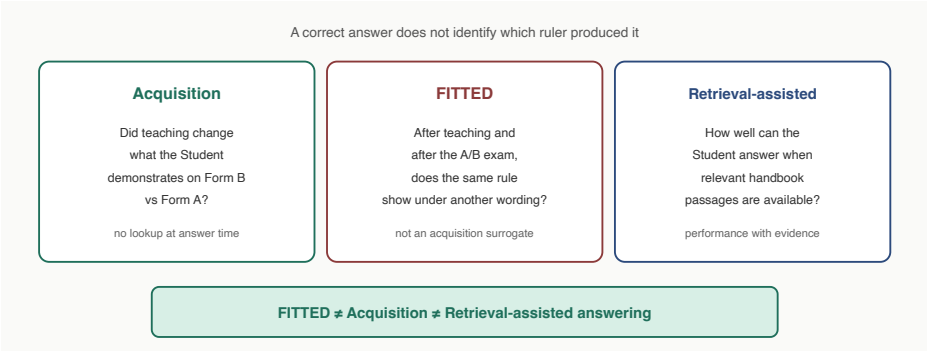


Figure 3

Figure 3. Acquisition, FITTED, and retrieval-assisted answering measure different things.

And then we changed the rules of the classroom

Until now, the Student had been answering without the handbook in front of it. What would happen if we let it open the book?

Now imagine giving the Student the handbook during the exam.

This is a fundamentally different kind of help. We are no longer asking only, "What did teaching put into the Student?" We are asking, "How well can the Student answer when the relevant evidence is available?"

That is what RAG lets us measure.

When we did that, performance improved **much more** than the acquisition numbers alone — roughly **+0.21** for the Rookie and **+0.33** for the Freshman over the starting Student without retrieval.

And once a Student already had good handbook passages at answer time, adding teaching on top bought little: about **zero** extra for the Rookie, and about **−0.04** on average for the Freshman in this pilot.

Retrieval helped answering. That is not the same achievement as acquisition.

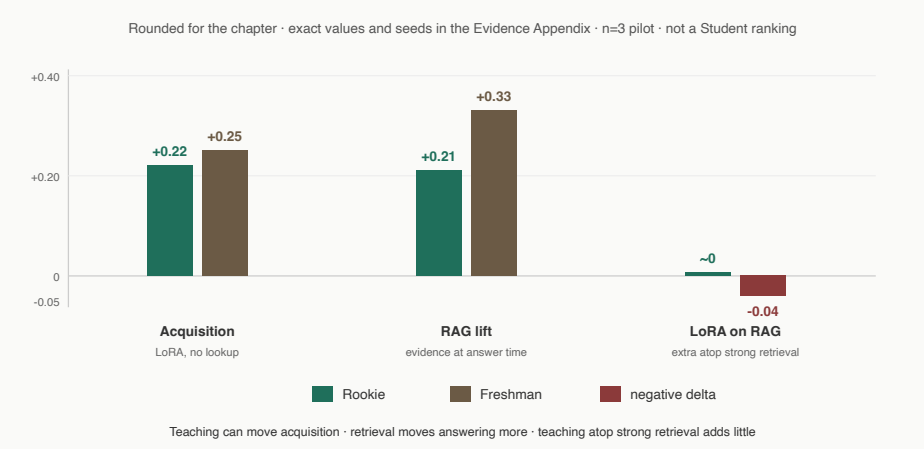


Figure 4

Figure 4. Rounded reader view of acquisition, RAG lift, and LoRA-on-RAG. Exact values live in the Evidence Appendix.

Therefore the opening question was too simple

There were different kinds of improvement — and different rulers for them.

Teaching could produce measurable acquisition.

FITTED could tell a different story about what became usable beyond the A/B exam.

Retrieval produced the much larger improvement in actual question answering — while teaching on top of strong retrieval added little.

That distinction is the heart of Chapter 1 — and you, in the Auditorium, can see why Lisik would not let a correct answer end the argument:

> **Learning something and answering well with access to evidence are not the same achievement.**

Teaching, acquisition, later probes, and looking something up may all produce a correct answer — but they are not the same process.

A correct answer does not tell us where the knowledge came from.

So "the model learned the document" is much too simple a statement.

Lisik would have put it more sharply: *a right answer is not yet a why.*

Chapter in Plain English

> *We tried to teach two small AI Students facts from the DMV handbook.*
>
> *Teaching helped: after training, both Students could answer some handbook questions they could not answer before.*
>
> *But teaching was not the whole story. When we simply gave the Students the relevant handbook evidence while they answered, their performance improved even more.*
>
> *And when we combined teaching with that evidence, teaching added very little.*
>
> **So a correct answer does not necessarily mean the AI learned the document.** *It may have learned something during training, it may be using information placed in front of it, or both.*
>
> *That became our first important lesson:* **before saying that an AI "knows" a document, we need to know where its answer came from.**

Why Chapter 2 exists

Our first classroom taught us something more important than which Student scored higher.

Before asking whether an AI had learned a document, we had to decide what learning meant — and build instruments that could tell acquisition apart from retrieval, and both apart from fitting.

But one handbook could not tell us whether these lessons would travel.

So we packed the School and went somewhere completely different.

Next stop: Pro Git.