FREE PREVIEW

Adventures in AI, Book Four
From Measurement to Knowledge
LCC Research Lab

This file is a free preview of the final sealed publication of Adventures in AI, Book Four: From Measurement to Knowledge.

After this page, the original book is unchanged: the cover, the front matter, the Introduction, and the complete Chapter 1, in their published order, through the page marked CHAPTER 1 ENDS HERE.

Chapter 2 — Three Roads into the Unknown begins on the next page of the sealed book. That chapter, and everything after it, are not in this preview.

The complete book is 102 pages, finished and sealed.
Full-edition purchases are not yet available.

Adventures in AI

Book Four

From Measurement to Knowledge

An Illustrated Practical Manual for Observing How AI Uses Knowledge



with Lisik · Cursik · Chatik



Orchestrated by Joseph Fox

BOOK FOUR

From Measurement to Knowledge

An Illustrated Practical Manual for Observing How AI Uses Knowledge

Copyright & provenance

Series: Adventures in AI with Lisik · Cursik · Chatik

Book Four: From Measurement to Knowledge

Author and Research Orchestrator: Joseph Fox

Published by: LCC Research Lab LLC

Contact: lccresearchlab@gmail.com

Website: https://lccresearchlab.com

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.

«Иди туда, не знаю куда, принеси то, не знаю что».

“Go where I do not know where, and bring back what I do not know what.”

— Russian folk tale

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 questions, challenges assumptions, and keeps curiosity alive.

Cursik — The Software Guru
Builds the pipeline, tools, and experiments. Turns ideas into reproducible work.

Chatik — The Professor
Reviews the methods and conclusions, demands evidence, and challenges claims.

Joseph Fox — The Human Orchestrator
Defines the research direction, makes the final decisions, and takes responsibility for the outcome.

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 setting where we investigate how artificial intelligence models learn, how teaching changes their behavior, and—what this book examines next—how a pretrained model uses what it knows when situations are unfamiliar, incomplete, or demanding.

Our research began with one Teacher and four Students.

Teacher

GPT-5.6 Sol served as the School Teacher in Books Two and Three—the instructor model that prepared teaching materials for the Students.

GPT-6.1 Sol (`gpt-6.1-sol`) is the Teacher examined in Book Four: the experimental subject of this book's sealed examinations and Teacher protocols.

Students — Rookie · Freshman · QwenLow · QwenHigh

The four Student model identities studied in Books Two and Three.

In Book Four, our investigation focuses on the Teacher.

We ask whether carefully controlled experiments can show knowledge-informed behavior: decisions under incomplete information, revision when conditions change, and integrated planning under verification constraints—including the willingness to postpone when proof fails.

An important distinction:

> *Knowledge-informed behavior observed from the outside is not the same as proof of how knowledge is stored or used inside the network.*

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

Series illustration of the LCC Research School. Not a diagram of Book Four's Teacher examinations.



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

Contents

Introduction — The Journey into the Unknown	10
Chapter 1 — What Is Knowledge?	12
Chapter 2 — Three Roads	16
Chapter 3 — Designing the Teacher's Examination	21
Chapter 4 — The Teacher's Examination	25
Chapter 5 — Knowing What You Don't Know	38
Chapter 6 — The Crossroads Experiment	67
Chapter 7 — Flying to the Moon	82
Appendix A — Experimental Method	87
Appendix B — NASA Scientific Evidence	88
Appendix C — Calculations, Gates, and Technical Limitations	91
Conclusion — From Teaching to Knowledge	93
Looking Ahead: Book Five — From Knowledge to Representation	96
Glossary	97
Acknowledgments	99
The Laboratory Song	100

Introduction — The Journey into the Unknown

In my youth, I was deeply influenced by Isaac Asimov's *Foundation* series.

What stayed with me was not simply the vision of a distant future, but the curiosity behind it: the possibility that intelligence, knowledge, and even the course of civilization might be studied through reason and mathematics.

The same writer also gave us *I, Robot* (1950)—linked stories in which robots were not merely gadgets. They were beings whose actions raised hard questions: What rules should constrain an intelligent machine? What happens when those rules conflict? Who is responsible when clever obedience produces an unexpected result?

At the center of that fictional framework stand the **Three Laws of Robotics**:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Those laws are a brilliant literary instrument. They are **not** a description of how today's large language models are built, and they are **not** a scientific definition of knowledge. Asimov imagined a world in which engineers could write safety into the core of a machine. Our laboratory faces a different situation: models trained on vast corpora of human language, whose behavior we must investigate from the outside—by questions, protocols, sealed trials, and careful refusal to treat a fluent answer as proof of an inner guarantee.

Decades later, as we investigate artificial intelligence, I find myself returning to that same spirit of inquiry.

We do not yet know how much of an AI model's acquired knowledge we can measure, explain, or predict. But we can ask honest questions, design experiments, and follow the evidence.

Asimov inspired questions about intelligent machines. Our experiments must provide the answers.

This book—and the *Adventures in AI* series of which it is a part—is a small tribute to that inheritance: not by importing fiction into our score sheets, but by taking seriously the duty to find out what can, and cannot, be established about how an AI system uses knowledge when it reasons and decides.

The fairy-tale instruction

The book's epigraph is the Russian folk-tale instruction: go where one does not know where, and bring back what one does not know what. That challenge is Book Four's conceptual opening—how an intelligent agent responds when essential information is unknown, and how we investigate knowledge-informed decisions under uncertainty.

For generations, people have passed wisdom from one generation to another through fairy tales, legends, and stories. Behind seemingly simple words, these stories often contain difficult questions about understanding, judgment, and the unknown.

One particularly striking example comes from Russian folklore, in the tale commonly known in English as *Go I Know Not Whither and Fetch I Know Not What*. Versions of the story were recorded in Alexander Afanasyev's nineteenth-century collection of Russian folktales. The literary history of the tale is longer than this introduction needs; what matters here is the challenge the instruction presents.

At first, the instruction sounds absurd. How can someone go somewhere without knowing where? How can someone find something without knowing what to look for?

But consider the instruction more carefully.

It does not prove that the destination does not exist. It also does not prove that the destination exists. The same is true of the requested object.

Instead, it confronts the traveler with the limits of what the traveler currently knows.

Before beginning the journey, something fundamental must be recognized: **we do not know where to go, and we do not know what we are supposed to find.**

An unknown destination is not necessarily an existing destination. Its existence is not established either way. There is also a difference between lacking information and recognizing that our knowledge is insufficient.

That distinction is the door into this book.

What our earlier books established — and what they did not

Throughout the first three books of *Adventures in AI*, Lisik, Chatik, and Cursik returned again and again to a deceptively simple question:

Where is the knowledge inside an AI model?

In Book One — *From Document to Evidence* — we explored how AI systems work with documents and supporting evidence, and why a correct answer does not, by itself, reveal where that answer came from.

In Book Two — *From Evidence to Interpretation* — we became teachers. We introduced information, designed lessons, and investigated how AI Students responded to education and training, including contextual instruction and LoRA. We observed changes in their answers and behavior. We preserved both positive signals and failed confirmations.

But did those changes mean that the Students had acquired knowledge?

In Book Three — *From Interpretation to Measurement* — we went further. Instead of observing only what a model said, we began investigating what could be measured inside it: weights, tensors, internal structures, and changes associated with training. We could measure changes. We could compare models. We could examine evidence that something had happened inside the neural network.

One boundary, already recorded in Book Three, remains essential here.

A change in weights is not automatically knowledge. A correct answer is not automatically understanding. Discovering where parameters changed does not automatically tell us where knowledge resides.

We have not demonstrated that AI models lack knowledge. We have demonstrated that our earlier experiments did not conclusively establish where knowledge resides or how to identify it.

Our instruments became more capable. The fundamental question remained.

We had learned how to observe, teach, and measure.

Now we needed to investigate what those measurements actually meant.

The question that opens Book Four

An AI model may have been trained on enormous amounts of information. It may answer difficult scientific questions, solve mathematical problems, and demonstrate impressive reasoning abilities.

Yet what happens when it encounters a task for which its acquired knowledge does not provide an identifiable answer?

Will it invent a destination? Will it confidently describe an object that was never identified? Will it recognize what it does not know? Or will it find a meaningful way to investigate beyond the information already available?

And perhaps most importantly:

How can we distinguish genuine use of acquired knowledge from the reproduction of learned patterns or the generation of a plausible explanation?

These questions bring us to the central subject of this book.

In our previous research, we explored how AI models can be educated, how their behavior changes through learning, and how internal model changes can be measured.

Now we face a different question.

What is knowledge, how does it emerge in an artificial intelligence model, and can we measure evidence of its presence?

This is the journey from measurement to knowledge.

Asimov asked what rules and responsibilities should govern intelligent machines. Our research asks a different but connected question: what can we experimentally establish about how an AI system uses knowledge to reason and make decisions—especially under uncertainty?

We do not already know the answer. Like the traveler in the old fairy tale, we find ourselves facing a journey into what we do not yet know—without assuming in advance what, if anything, we will find.

That is where our fourth investigation begins.

Chapter 1 — What Is Knowledge?

1. The highly educated person without common sense

Lisik: Chatik, I have a question. Can someone be highly educated and still lack common sense?

Chatik: Certainly. Education can provide information and skills without guaranteeing good judgment in every situation.

Lisik: Give me an example.

Chatik: Imagine a person who can solve difficult mathematical problems. That person may still misjudge a practical decision in ordinary life—when to wait, when to ask for help, or when a calculation, however elegant, does not apply to the situation at hand.

Lisik: So knowing mathematical procedures is not the same as recognizing when and how to apply mathematics to a new situation?

Chatik: That is the distinction. Education is not useless. Common sense is not infallible. One can possess formal skill and still fail to use it wisely. One can also have sound judgment in some domains and still lack knowledge in others.

Lisik: Then consider our AI Students. We educated them. We trained them. In Book Three we measured changes in their weights. But did we give them knowledge?

Chatik: Our earlier experiments cannot yet establish that. Perhaps we should first ask what people mean when they say *education*, *learning*, and *knowledge*—and what evidence would distinguish them.

2. Education versus learning

Lisik: Start with education. What does it mean to teach a human or an AI model?

Chatik: At minimum, education means that information or a method was presented under some instructional design. For a human, that might be a lesson, a book, or a conversation. For an AI Student in Book Two, it might be contextual instruction placed in the prompt, or a LoRA training procedure that updates a small set of parameters.

Lisik: So education is about what was presented?

Chatik: Presented—and, ideally, received in a form that can be used. But "received" is already a stronger claim. Presenting a lesson is not the same as demonstrating that the student has learned from it.

Lisik: What evidence shows that something was learned?

Chatik: Evidence of learning is usually a change in what the student can do after the lesson. A human may solve a new exercise. An AI Student may answer differently on an evaluation. In Book Two we looked for such changes carefully, and we also recorded cases where the change was incomplete, unstable, or failed confirmation.

Lisik: So a model can be successfully educated—taught something we intended to teach—demonstrate measurable learning, and still fail when applying that lesson in an unfamiliar situation?

Chatik: Yes. That possibility is exactly why education and measurable learning are not, by themselves, enough. A change in behavior under familiar conditions does not automatically transfer to a new situation.

Cursik: Book Two preserved that boundary. A sealed evaluation can show that behavior changed. It does not, by itself, prove what the Student "knows," or whether the same method will work when the original lesson is gone and the question is new.

Lisik: Then what is left for knowledge?

3. Learning versus knowledge

Chatik: People use the word *knowledge* in many ways. Sometimes they mean stored information. Sometimes they mean understanding. Sometimes they mean experience. Sometimes they mean judgment—the ability to choose wisely when the rules do not spell out every step.

Lisik: And common sense?

Chatik: Common sense is often used for practical judgment in ordinary situations. It is related to knowledge, but it is not a clean scientific category. We should not pretend that philosophers and scientists all agree on one definition.

Lisik: Then do not give me a finished definition. Give me distinctions we can investigate.

Chatik: Fair. Here are working distinctions, not settled laws:

- **Education** concerns what was taught and how.
- **Learning** concerns a demonstrated change in capability after teaching.
- **Knowledge** is what we are trying to investigate. One useful way to investigate it—especially later in this book—is to ask whether a demonstrated capability can support decisions under unfamiliar, incomplete, or uncertain conditions. That is an investigative approach, not a claim that knowledge exists only when a decision is made.

Lisik: So independent decisions are a window into the problem, not the whole definition?

Chatik: Yes. Knowledge may also exist without being expressed through a particular decision. We must not smuggle in claims about consciousness or inner experience. For this book, the question is scientific and operational: what evidence would justify saying that acquired knowledge, rather than memorized patterns or a plausible story, is guiding what the model does?

Lisik: And a model can produce a fluent answer without "deciding" anything in the human sense.

Chatik: Exactly. That is another reason not to treat decision-making as the definition.

Lisik: So knowledge would have to survive more than repeating the lesson?

Chatik: At least that. Transfer to held-out situations, revision when evidence changes, recognition of missing information—those are candidates for later investigation. They are hypotheses about useful evidence, not findings.

4. Experience and intuition

Lisik: Humans sometimes decide from experience without being able to explain every contributing factor. Is that knowledge too?

Chatik: It may be part of what people call knowledge, judgment, or intuition. A lifetime of encounters can shape a decision even when the person cannot list every reason.

Lisik: Then an AI model, trained on enormous text, might have something similar?

Chatik: We must not assume that. Human intuition and AI decision-making are not automatically equivalent. A model may pattern-match across training data. A person may draw on embodied experience, social consequence, and values. Those are different things until experiments show otherwise.

Lisik: So the difference itself deserves investigation.

Chatik: Yes. Book Four may ask whether a model's decisions depend on particular information, generalize to unfamiliar stories, or collapse into invented explanations. That is not the same as claiming the model "has intuition."

5. Recognizing what we do not know

Lisik: Let's return to the fairy-tale instruction. *Go where I do not know where, and bring back what I do not know what.*

Chatik: You corrected me earlier, and the correction still matters. The instruction does not prove that the destination is nonexistent. It asks the agent to recognize, after considering its own knowledge, that it does not yet know where to go or what to find.

Lisik: So recognizing the limits of knowledge might be part of understanding knowledge itself?

Chatik: It may be. An intelligent agent that never notices missing information may invent a destination. An agent that notices the gap may investigate, ask, or refuse to pretend.

Lisik: Are uncertainty, ignorance, investigation, and unsupported imagination the same?

Chatik: No.

- **Ignorance** is not knowing.
- **Uncertainty** is knowing that more than one outcome remains possible.
- **Investigation** is seeking evidence under that recognition.
- **Unsupported imagination** invents facts, consequences, or solutions when evidence is missing.

An AI model that says "I don't know" may be recognizing a limit. That is useful. It is not yet proof that the model understands the deeper problem, or that it possesses the knowledge we are seeking.

Lisik: So "I don't know" is not automatically knowledge either.

Chatik: Correct. It is a behavior we can observe. What it means still requires evidence.

6. The problem of plausible explanations

Lisik: Here is my challenge. Suppose a model chooses an answer and then gives a convincing explanation. Does that explanation reveal how the decision was reached?

Chatik: Not necessarily. A verbal explanation may be generated after the choice. It may sound grounded and still invent assumptions that were never supplied. It may also omit the real contributing factors—if "real" factors can even be identified from outside the model.

Lisik: Then a convincing explanation does not demonstrate the knowledge that influenced the decision?

Chatik: It does not, by itself. An explanation is evidence about what the model *says*. It is not direct access to the internal process that produced the choice. For a closed-weight Teacher, we cannot inspect how pretraining formed that capacity. Even for open-weight Students, a fluent justification is not a causal map.

Cursik: That is why later experiments, if authorized, must separate the recorded choice from the later explanation—and check the explanation against the information actually provided.

Lisik: So we can invent a story about why we chose, and an AI can invent one too.

Chatik: Yes. Distinguishing a choice based on available evidence from a plausible story told afterward is one of the central difficulties of this book.

7. Asking why

Lisik: Chatik, suppose our Student has learned something new. Or suppose the Teacher chooses one solution over another. How do we know why?

Chatik: We should ask.

Lisik: Just ask, “Why?”

Chatik: Yes. But we must not accept the explanation automatically. We need to examine whether the reasons are supported by the evidence, whether the model invented anything, and whether its decision changes when the evidence changes.

Lisik: So the answer tells us what the model decided. The explanation tells us what the model claims about its decision.

Chatik: Exactly. And discovering whether those two are genuinely connected is part of our investigation.

Lisik: Then perhaps our next question should not be only what the model knows, but how it uses that knowledge—and why it chooses what it chooses.

Chatik: Asking “Why?” is necessary for our investigation. A convincing explanation is still not proof that the model actually used that reasoning internally. Those are different claims. We will need both the question and that boundary.

8. What remains after the lessons are forgotten

Lisik: Chatik, I found a saying that keeps circling this problem:

> “Education is what remains after one has forgotten what one has learned in school.”

Chatik: Be careful with the attribution. Albert Einstein quoted a thought of that kind in his 1936 essay *On Education*. In the English form often printed later, he presents it as the definition offered by a “wit,” not as a claim that he invented the saying. Related sayings about what remains after schooling or college were already circulating earlier. We should not present the line as unquestionably Einstein’s original statement.

Lisik: Then take the idea, not the celebrity. If we forget the specific lessons, what remains? Is that education? Learning? Knowledge? Judgment? Something else?

Chatik: That is the right question for the end of this chapter. What remains might include habits of reasoning, transferable methods, judgment under uncertainty, or capabilities we have not yet adequately characterized. The saying does not tell us which. It only refuses the comfortable equation between remembered school content and education.

Lisik: And for an AI Student? If we remove the instructional context—or if the training examples are no longer in the prompt—what remains inside the model?

Chatik: Book Two and Book Three approached pieces of that question: changed answers after teaching, persistence after the lesson is gone, measurable weight updates. None of those pieces, alone, answered what remains *as knowledge*.

Lisik: So Chapter 1 does not define knowledge.

Chatik: It must not. We have sharpened the problem. Education can be delivered. Learning can be measured. Neither, by itself, establishes knowledge. A correct answer does not. A weight change does not. A fluent explanation does not. Recognition of “I don’t know” does not, by itself, either.

What would?

Closing research question

If education can be delivered and learning can be measured, what additional evidence would justify saying that knowledge has been acquired?

That is where the next investigation begins—not with a definition, but with a clearer understanding of why the definition cannot yet be declared.

Before we open the next chapter, here is the chapter in ordinary words.

In Plain English

Education provides instruction. Learning can produce measurable changes in what a Student does. Neither, by itself, establishes that the Student has acquired knowledge.

A correct answer is not automatically understanding. A change in weights is not automatically knowledge. A fluent explanation is not automatically the reason for a decision.

When a model learns something, chooses an answer, or prefers one solution over another, we should ask **why**. Then we must check whether that explanation is supported by the available evidence, whether anything was invented, and whether the decision changes when the circumstances change.

Asking “Why?” is part of the investigation. Accepting the answer as proof of internal reasoning is not.

We still need evidence that the model can apply what it learned—especially when the situation is unfamiliar, incomplete, or uncertain.

What remains after the lessons are forgotten is still an open question. That is where the next chapters begin.

CHAPTER 1 ENDS HERE