START HERE

# AI, Actually Explained

For someone who was never told what this thing is.

First edition, September 2026

Addaly

Series editor: Dr Mustafa Mun

Draft, open for academic review

START HERE

# AI, Actually Explained

For someone who was never told what this thing is.

Series editor: Dr Mustafa Mun

**Addaly**

First edition, September 2026 · Draft, open for academic review

**AI, Actually Explained**



First edition, September 2026, build 62a26075.

Written by AI models to a written standard, with Dr Mustafa Mun as series editor. Exam keys checked blind by a second model: 3,665 of 3,666 agreed. Academic review invited.

The manuscript was drafted with the assistance of a large language model and was edited and published under his name. That is stated plainly here rather than left to be discovered, because a reader is entitled to know how a book they are trusting was made.

How to cite: *Mun, M. (2026). AI, Actually Explained. Addaly. addaly.com/learn/ai-actually-explained.*

This book is the printed form of the course of the same name at addaly.com/learn/ai-actually-explained. The website is the living version and is corrected first. Where the two differ, the website is right.

Free to read, free to download, free to print, and free to teach from. There is no paid edition and there never will be.

Every diagram in it is drawn from data by code, not generated as a picture, so the numbers on the page are the numbers in the argument. The answers to every check, test and examination question are at the back.

9 modules · 92 lessons · 31 figures · 9 case studies · 69,721 words · about 14 hours

# Brief contents

1

MODULE 1

# What this word actually names

Place any system somebody calls AI inside the nested family of AI, machine learning, deep learning and generative AI, say whether its behaviour was written by a person or learned from examples, and describe what the trained thing physically is.

Lesson 1 · 6 min

# What the word actually names

THE ONE THING TO KEEP

AI names whatever machines have only just learned to do, so ask what a system does instead.

## The word is doing too much work

Artificial intelligence was coined in 1956, at a summer research workshop, by people who wanted funding to study thinking machines. It was a pitch, not a description.

Seventy years later the same two words are attached to a chess program, a photo filter, a bank's fraud alert, a warehouse robot and a chatbot. Inside, those systems have almost nothing in common. A word that covers all of that is not telling you much.

## The thing that keeps happening

In 1997 a computer beat the reigning world chess champion. It was front-page news about machine intelligence. Today a free chess app on a cheap phone would beat that computer, and nobody calls it AI. It is a chess engine.

The same story ran for reading handwritten postal codes, for filtering spam, for finding faces in a camera viewfinder, for translating between languages. Each one was intelligence right up until it worked reliably. Then it quietly became software.

Researchers have a name for this: the AI effect. AI is roughly whatever machines have not quite managed yet. So the label does not describe a technology. It marks a frontier, and the frontier keeps moving away from whatever you already have.

## What people mean when they say it now

Almost everything called AI today is machine learning: a system whose behaviour came from examples rather than from written instructions. That distinction is the subject of the next lesson, and it is the one that actually matters.

Most of the current noise is about one branch of that: very large models trained on enormous amounts of text and images. Chatbots are the visible face of it. Behind them sit the same techniques doing dull, useful work you never see.

## A better question than is it AI

When you meet a system, ask three things instead.

**What goes in, and what comes out?** A photo goes in, a name and a number come out.

**Where did its behaviour come from?** Did a person write down the rules, or did the behaviour come from examples?

**What happens when it is wrong, and who finds out?**

Try it on a real case. A farmer in Kaduna points a phone at a maize leaf and an app says leaf blight, 0.82. In: a photograph. Out: a disease name and a confidence number. Behaviour from: some tens of thousands of labelled leaf photos. When wrong: money is spent on the wrong treatment and a crop is lost, and nobody may ever learn that the app was the reason.

Those four answers tell you everything you need in order to decide how much to trust it. Whether it counts as intelligence has become an uninteresting question, which is the point.

## Two things AI is not

**It is not one system.** There is no such thing as the AI. There are millions of separate models, trained by different organisations, for different jobs, and they share nothing with each other. The model that reads your bank's

transactions has no connection to the one that writes your emails.

**It is not a being you are talking to.** When a chatbot writes I think, that is a style of writing, learned from text written by humans and kept because people find it easier to read. Whether anything like experience is going on inside these systems is a real open question that serious people disagree about. But the word I in the output is not evidence either way, because that word would appear whether or not anything was going on.

Keep the question live if you find it interesting. Just do not let a pronoun answer it for you.

---

BEFORE YOU MOVE ON

A bank's card-fraud system was described as artificial intelligence in 1998 and is described as just software today. The system does the same job. What best explains the change?

A. The label moves. Once a capability is understood and dependable, people stop calling it intelligence and start calling it a program.

B. The old system was too simple to have really been AI. Only today's systems genuinely qualify.

C. The underlying technique was replaced at some point, and the new one does not count as AI.

D. AI is a marketing term with no real content, so nothing has ever counted.

The answer and why: p. 457

Lesson 2 · 9 min

# Four rings, one inside the next

THE ONE THING TO KEEP

AI, machine learning, deep learning and generative AI are four rings inside each other, and naming the ring tells you which question to ask: show me the rules, show me the data, or show me how you know the output is true.

## The word covers too much

When a shop sign, a job advert and a newspaper headline all say "AI", they are usually pointing at four different things. The four sit inside each other like rings. Once you can name which ring something is in, most of the confusion goes.

**Artificial intelligence** is the outer ring. It is the oldest and vaguest of the four: any attempt to get a machine to do something that seems to need intelligence. A chess program from 1997 is in this ring. So is the routing that finds you a way round a traffic jam. Nothing about the ring says how the thing works.

**Machine learning** sits inside it. Here, the behaviour was not written down by a person. It was derived from examples. Somebody collected a pile of data, ran a procedure over it, and the result decides what the system does. Your bank's fraud check is almost certainly in this ring.

**Deep learning** sits inside machine learning. It is the subset that uses neural networks with many layers — the technique that took over from about 2012 onwards. Face unlock on your phone is deep learning. So is the speech recognition in a voice note.

**Generative AI** sits inside deep learning. These are the models whose output is new content rather than a label or a number: text, images, audio, video. ChatGPT, Gemini, Midjourney, DeepSeek. This innermost ring is what most

people now mean when they say AI, which is why the word has become so slippery — the newest and smallest ring has captured a name that belongs to the largest.

Large language models are a slice of that innermost ring. So a chatbot is a large language model, which is generative AI, which is deep learning, which is machine learning, which is artificial intelligence. Every one of those sentences is true, and each tells you something different.

Figure 1.1

Four rings, and the question each one licenses

**Artificial intelligence**
Anything that looks like it needs intelligence. Says nothing about how it works. Ask: show me the rules.

**Machine learning**
Behaviour derived from examples, not written down. Ask: what did you train it on, and how do you know it works?

**Deep learning**
Machine learning with many-layered networks. Took over from about 2012. Same questions, larger data.

**Generative AI**
Output is new content rather than a label. Add: how do you know the output is true? A paragraph has no right answer beside it.

Each ring sits inside the one above it. Naming the ring tells you what you are entitled to ask for: the rules, the data and the testing, or the evidence that the output is true.

## Why the ring matters

Because it tells you what question to ask.

If something is in the outer ring only — rules a person wrote — then you can ask to see the rules. Somebody can print them out. When a housing benefit system refuses your claim, and the logic is written rules, the refusal can be explained line by line.

If it is machine learning, the rules are not printable. The right question becomes: **what examples did you train it on, and how do you know it works?** Those two questions are the whole of the difference. Nobody can show you the reasoning, so they have to show you the data and the testing instead.

If it is generative, add a third question: **how do you know the output is true?** A label can be checked against the right label. A paragraph invented on the spot has no right answer sitting next to it.

## The thing beginners get backwards

Most people assume deep learning is simply better, and that anything serious must be a neural network by now. It is not so.

For data that lives in rows and columns — a spreadsheet of customers, a table of transactions, a sensor log — the method that usually wins is not deep learning at all. It is a family of models built from decision trees, with names like XGBoost, LightGBM and random forests. They train in minutes on an ordinary laptop, they need no graphics card, and on tabular data they routinely beat neural networks that cost a hundred times more to run. This has been tested repeatedly and it keeps coming out the same way.

This matters for you in a practical sense. If somebody tells you that predicting which customers will cancel requires a large language model, they are either selling something or have not looked. The free tools here are genuinely free and genuinely good: Python with `scikit-learn` and `xgboost`, or `LibreOffice Calc` for the first look at the data. The paid platforms — DataRobot, SageMaker, Vertex AI — are wrapping the same algorithms in a dashboard.

## Where the rings are argued about

Be warned that the rings are conventions, not laws. Three honest disagreements:

- **Reinforcement learning** — learning by trial and reward, the method behind game-playing systems and part of how chatbots are tuned — does not sit neatly inside any of the four. Some diagrams place it beside machine learning, some inside it.
- **A chess engine** like Stockfish was pure hand-written evaluation for decades. It now has a small neural network inside it. It moved rings without changing its name.
- **"AI" as a product label** is applied to plain database queries and `if` statements every day, because the word raises valuations and prices. There is no authority stopping this and no penalty for it.

So the rings are a thinking tool, not a certificate. Their use is that they replace one useless question with three useful ones.

## The habit

When you next meet a system described as AI, ask yourself which ring, out loud if you like. If you cannot tell, that is itself information: it means nobody has told you how the thing works, and you should treat every claim about it as unverified until they do.

BEFORE YOU MOVE ON

A logistics firm wants to predict which of its 40,000 monthly deliveries will be late, using a spreadsheet of route, weather, driver and load data. A vendor proposes a large language model. What is the strongest technical objection?

A. Large language models cannot handle 40,000 rows, which is too much data for any model to process.

B. For data in rows and columns, tree-based models such as XGBoost usually predict better than neural networks, train in minutes on an ordinary laptop, and cost nothing.

C. Prediction of future events is outside the scope of artificial intelligence, which only classifies things that have already happened.

D. A language model would work, but the firm should first get a larger dataset, because deep learning always needs millions of examples.

The answer and why: p. 458

Lesson 3 · 7 min

# Written by a person, or grown from examples

THE ONE THING TO KEEP

A written program follows rules a person can read; a trained model follows patterns nobody ever wrote down.

## Two ways to make a machine do something

This is the distinction the whole course rests on. There are broadly two ways to get a computer to behave a certain way.

CASE STUDY · MODULE 1

# Two dropout dashboards

An illustrative case, built from documented patterns.

*A district education office in Nashik, Maharashtra, choosing between two products that both call themselves AI.*

The district had a number nobody liked. Of the children who entered Class 6 in its government schools, a little under a fifth were no longer in a classroom by Class 9, and the office found out about most of them months late, when a headmaster mentioned it.

Two vendors came with dashboards.

The first sold an early-warning system. It had been trained on five years of attendance registers, term marks and mid-day-meal records from an adjacent district, and it produced a risk score between 0 and 100 for every enrolled child, refreshed weekly. The sales engineer quoted 82 per cent accuracy on data the model had not seen. The price was fourteen lakh rupees for three years.

The second sold a flagging tool. It raised a flag when a child missed seven consecutive days, or when marks fell by two grades in a term, or when an older sibling had already left the school. Three rules, written by two retired headmasters, printed on one page in the manual. Four lakh rupees, and it claimed nothing at all.

The education officer, who had sat through a decade of software demonstrations, asked the same three questions of both. What goes in and what comes out? Where did the behaviour come from — did a person write the rules, or did they come from examples? And when it is wrong, who pays and who finds out?

The answers separated the products immediately, and not in the direction the price suggested. The second product's behaviour could be printed, argued with by a parent and amended by the office in an afternoon. It would also only ever catch what two retired headmasters had thought of. The first product

could in principle find children the rules missed — the ones whose attendance is fine right up until it is not — and nobody in the district, including the vendor's engineer, could say why any particular child had scored 78.

The officer pressed on one more thing. What is the score trained to predict? The engineer said children at risk of dropping out. The officer asked what the recorded right answer had been during training. The engineer, to his credit, answered honestly: children who did drop out, in the neighbouring district, in the five years on file. So the score predicted resemblance to children the schooling system had already lost, in a district with different schools, different distances and a different sugar harvest.

That is not a criticism of the vendor, who had built the only thing that could be built from the only records that exist. It is a fact about what the records are. A child who stopped attending and whom nobody entered as having stopped attending is not in the training answers at all, so the model cannot learn that such children exist. The office's own registers had, by its own audit two years earlier, about eleven per cent of transfer certificates missing.

The officer also asked what the 82 per cent was 82 per cent of. It was the share of children the model classified correctly at the vendor's chosen cut-off. Since only about a fifth of children leave, a system that predicted nobody would leave would score around 80 per cent on the same measure and be useless. The engineer knew this and had a better figure — of the children the model flagged, roughly one in five did in fact leave — which nobody had put on the slide because it sounds worse and means more.

A headmaster in the room asked the question that decided the meeting. Would teachers see the number? Because a teacher who sees 78 beside a child's name treats that child differently, and there is no way to un-see it, and nobody would ever know whether the score had caused the thing it predicted.

Against all of this sat the arithmetic. Fourteen lakh was also bicycle vouchers for about nine hundred girls, which is an intervention with evidence behind it. And a wrong choice was not cheap in either direction: buy the rules and the children the rules cannot see stay invisible; buy the model and the office owns a number it cannot explain, attached to a child, for three years.

YOUR ANSWERS

**1. Which ring of the family tree is each product in, and what does naming the ring entitle the officer to ask for?**

**2. The model was 82 per cent accurate on held-out data and the district still did not buy it. Give two reasons from the case that accuracy did not settle the question.**

**3. The model was trained on records of children who dropped out. What exactly does its score predict, and why is that different from 'which children are at risk'?**

STOP HERE

Write your answers before you turn the page. How it played out starts on p. 48.

CASE STUDY · MODULE 1

# How it played out

The office bought the four-lakh rules dashboard, and separately asked the first vendor for a one-year pilot on strict terms: scores went to the district office only, never to a school, and were compared against the rules month by month. Over the year the rules flagged 410 children and the model flagged 340, with 190 on both lists. Of the children who actually left, the rules had flagged 61 and the model 66. The model found nine that the rules had missed, and all nine shared a pattern nobody had written down: attendance and marks steady, but an older sibling who had left within the previous eighteen months. The office added that as a fourth rule — one line — and did not renew the pilot. The officer's file note said the genuinely useful thing the model had done was to tell him which rule to write.

## The questions, discussed

**1. Which ring of the family tree is each product in, and what does naming the ring entitle the officer to ask for?**

The flagging tool is in the outer ring only: behaviour written by people, so the officer may ask to see the rules, and did — they are printed in the manual. The early-warning system is machine learning, so the rules cannot be printed and the questions change to what it was trained on and how anyone knows it works. Neither is generative, so the third question — how do you know the output is true — does not arise; a risk score can at least be checked against what later happened, which is exactly what the pilot did.

**2. The model was 82 per cent accurate on held-out data and the district still did not buy it. Give two reasons from the case that accuracy did not settle the question.**

First, the held-out data came from a different district, and a model reproduces the conditions of its training; distances, schools and harvest patterns differ, so the 82 per cent is a fact about the neighbouring district's records rather than about Nashik's children. Second, accuracy says nothing about who carries the

cost of an error. A wrong flag lands on one child, in front of a teacher who cannot un-see the number, with no explanation available and nothing to appeal to.

**3. The model was trained on records of children who dropped out. What exactly does its score predict, and why is that different from 'which children are at risk'?**

It predicts resemblance to children the system already lost and recorded as lost. That is not the same quantity as risk. It inherits whatever the recording process did — children who left and were never marked as having left are absent from the training answers, so the model cannot learn about them. The recorded right answer is the ceiling on what any trained system can be: it can be more consistent than the process that produced its labels, and it cannot be wiser than it.

MODULE 1 TEST

# Check what stuck

6 questions, one right answer each. This one is for you, not for a record. The answers, and the reason behind each, are in the key on p. 456. If two go wrong, re-read the module before the exam.

1 A free chess app on a cheap phone now beats the computer that defeated the world champion in 1997, and nobody calls the app AI. Researchers call this the AI effect. What does it tell you about the label?

A. The label marks a moving frontier rather than a technology, so it names what machines have not yet made reliable.

B. Chess turned out not to require intelligence, unlike the tasks that are called AI today.

C. The label was withdrawn from chess after the 1997 machine was found to have had human assistance during the match itself.

D. Only trained models have ever counted as AI, and a chess engine follows rules that people wrote.

2 A shop chain wants to predict which customers will stop renewing, from a spreadsheet with about forty columns per customer. On data of that shape, which approach usually wins?

A. A large language model, because language models now outperform older methods on every kind of data.

B. None of them, because rows and columns are the one kind of data that machine learning has never handled well.

C. A deep neural network, because deep learning superseded the older methods after 2012.

D. A decision-tree family such as XGBoost, which trains in minutes on an ordinary laptop.

3 A speech system does noticeably worse on Nigerian and Malaysian accents than on American ones, and gives no sign of struggling. What is the mechanism?

A. Those accents are objectively harder to distinguish, so any system would do worse on them.

B. Someone wrote a rule that gives American pronunciation priority whenever the system is uncertain.

C. The training audio contained far more of some voices than others, and the model reproduces its examples.

D. The model detects the speaker's country and routes the audio to a different, less accurate model for non-American speech.

4 You correct a chatbot, it apologises, and the rest of the conversation is better. The next day you open a fresh conversation and the same error is back. Why?

A. The correction sits in the conversation, which is fed back with every message; the file never changed.

B. The correction was stored, but the original training overrides it whenever a new session begins.

C. The provider applies user corrections in batches and had not yet reached yours.

D. The model learned it and then unlearned it, because other users corrected it the other way in the meantime.

5 A model trained to spot pneumonia on chest X-rays scored well in one hospital and collapsed at another. It had learned to read the small metal token that portable scanners leave in the corner of the image. What does this illustrate?

A. The model was overfitted and needed more layers in order to generalise properly.
B. The second hospital's imaging equipment was of lower quality than the first hospital's.
C. The token physically interfered with the imaging, so the lung fields in those X-rays genuinely differed.
D. Training rewards whatever reduces the error, including a marker associated with illness for unrelated reasons.

6 A privacy policy says only that the company uses your data to improve its services. Which question separates the two very different things that sentence can mean?

A. Does my text enter the training of a future model, and is there a switch to stop it?
B. How many days is my text retained before it is deleted from the company's servers?
C. Is the company processing my messages with a neural network or with written rules?
D. Will the company sell my conversations to advertisers, which is what that phrase is generally taken to be a euphemism for?

Figure 2.1

A 99% accurate fraud model over a million transactions

| | Model flags | Model clears |
|---|---|---|
| Actually fraud (1,000) | 990 caught | 10 missed |
| Actually genuine (999,000) | 9,990 false alarm | 989,010 correctly cleared |

One transaction in a thousand is fraud. The model catches 990 of the 1,000 and wrongly flags 1% of the 999,000 genuine ones. Nine of every ten alerts are false, from a model that is 99% accurate — which is why the honest question is what fraction of flagged cases are real.

Nothing is broken. The rarity of the thing being detected does this. Any time you hear an accuracy figure for a rare event — fraud, disease, extremism, benefit cheating — this arithmetic is waiting underneath, and the honest question is not "how accurate" but "of the cases it flags, how many are real".

## What this means for you

Three practical consequences.

Your card being declined abroad is usually this: your normal pattern changed, the score crossed the line, and the bank preferred a false alarm to a loss. Telling the bank your travel dates genuinely helps, because it feeds the model context it does not otherwise have.

A legitimate email of yours landing in spam is usually about the sending domain's reputation rather than your words. If it happens repeatedly, the fix is at the sending end.

And when a fraud team says "the system flagged you", you are entitled to ask what happens next, because in a well-run system a flag is a trigger for review, not a verdict. Where a flag is treated as a verdict — and there are national

examples in a later lesson — the false-alarm arithmetic above turns into thousands of wrongly accused people.

BEFORE YOU MOVE ON

A benefits agency reports that its new fraud model is 98% accurate, and that fraud occurs in about 1 in 200 claims. A minister concludes that flagged claimants are almost certainly cheating. What is wrong with that conclusion?

A. Accuracy figures are always inflated by vendors, so the true rate is much lower than 98%.

B. The model was trained on past cases, so it cannot detect new fraud methods.

C. 98% accuracy is too low for government use and the threshold should simply be raised.

D. Because genuine claims outnumber fraudulent ones about 200 to 1, the small error rate on that huge majority produces far more false flags than real ones.

The answer and why: p. 464

EXAMINATION

# AI, Actually Explained — final paper

This paper covers the whole course:

- what the word names and what a trained model physically is
- where models already sit in a phone, a feed, a bank and a clinic
- what a chatbot is doing when it answers
- why it can be wrong and sound certain
- how the field arrived here and what it cost
- pictures, voices and video
- using it without getting burned
- what the evidence on work actually says
- who controls it

63 questions, the whole pool. The site draws 25 for a sitting, pass mark 70%. Work any 25 under your own conditions. Answers from page 515.

**1** Module 1 · What this word actually names

A bank's loan software refuses an application and the officer can print the exact condition that failed. Which ring is that system in, and what follows?

A. Machine learning: the conditions were derived from past lending decisions.

B. Generative AI, since the system produces a written explanation of the refusal for the applicant.

C. Deep learning, because credit decisions are made by neural networks in 2026.

D. Artificial intelligence only, with written rules — so the refusal can be explained line by line.

**2** Module 1 · What this word actually names

Which of these is genuinely inside a trained model's file?

A. An index of facts that can be searched for 'the capital of Kenya'.

B. A compressed copy of the web pages it was trained on.

C. A long list of numbers, and nothing that reads as a sentence.

D. A lookup table of question-and-answer pairs assembled during training from forums and reference sites.

**3** Module 1 · What this word actually names

Amazon's CV screener learned to downgrade CVs containing the word 'women's'. What is the most accurate description of what went wrong?

A. The training procedure had a bug that inverted part of the scoring.

B. Somebody at the company had encoded a preference into the rules.

C. The model predicted the past accurately, and the past contained the preference.

D. The model was trained on too few CVs to generalise, and that small sample happened to be unrepresentative.

4 Module 1 · What this word actually names

You show a model three examples of the format you want inside your message, and the fourth comes out right. Has it learned anything?

A. No weight changed; the examples are in the context, and the behaviour ends with the conversation.

B. Yes: the weights adjusted slightly to fit your examples.

C. Yes, but only for your account, because the provider keeps a personal copy of the model.

D. No, because a model cannot use examples given at the time of asking — only the ones it was trained on.

5 Module 1 · What this word actually names

Someone tells you AGI is two years away. Which reply gets the most information?

A. Which capability specifically, and how would we know it had arrived?

B. Which company's model, and has the claim been peer reviewed?

C. How many parameters would such a system need to have?

D. Is that the definition used in the research literature, or the one in the founding charter of the company making the claim?

6 Module 1 · What this word actually names

What is the honest state of the question of whether these systems understand anything?

A. Settled in the negative by the stochastic parrot paper of 2021.

B. Settled in the affirmative by the work probing an internal board state in an Othello model.

C. Unresolved, and the practical move is to judge it by whether the failures look like a mind's.

D. Unanswerable in principle, so no evidence of any kind bears on it and the question should be dropped.

## Module 1 · What this word actually names

The module starts on p. 11.

MODULE 1 TEST

**Module 1 test, Q1, p. 50**

A free chess app on a cheap phone now beats the computer that defeated the world champion in 1997, and nobody calls the app AI…

**Answer A:** The label marks a moving frontier rather than a technology, so it names what machines have not yet made reliable.

**Why:** Reading handwritten postal codes, filtering spam, finding faces in a viewfinder and translating between languages were each intelligence right up until they worked reliably, and then quietly became software. So the word does not describe a technology; it marks a frontier that keeps moving away from whatever you already have. That is why the useful questions are what goes in, where the behaviour came from, and who pays when it is wrong.

**Module 1 test, Q2, p. 50**

A shop chain wants to predict which customers will stop renewing, from a spreadsheet with about forty columns per customer. On data of that shape…

**Answer D:** A decision-tree family such as XGBoost, which trains in minutes on an ordinary laptop.

**Why:** For data that lives in rows and columns, gradient-boosted trees — XGBoost, LightGBM, random forests — routinely beat neural networks that cost a hundred times more to run, and this has been tested repeatedly. They need no graphics card. The free path is Python with scikit-learn and xgboost, or a spreadsheet for the first look; the paid platforms wrap the same algorithms in a dashboard.

**Module 1 test, Q3, p. 51**

A speech system does noticeably worse on Nigerian and Malaysian accents than on American ones, and gives no sign of struggling. What is the mechanism?

**Answer C:** The training audio contained far more of some voices than others, and the model reproduces its examples.

**Why:** A model has no source of behaviour other than what it was shown, so lopsided examples produce a lopsided model — smoothly, confidently, with no warning label. The 2020 study in PNAS gave five commercial systems identical read passages and measured roughly double the word error rate for Black speakers. Nobody had designed that; the systems reproduced their training audio faithfully, which is exactly what they were built to do.

**Module 1 test, Q4, p. 51**

You correct a chatbot, it apologises, and the rest of the conversation is better. The next day you open a fresh conversation and the same…

**Answer A:** The correction sits in the conversation, which is fed back with every message; the file never changed.

**Why:** Training and use are separate phases. While you are talking to it the file is read-only, and ten million people are reading the same unchanged numbers. Your correction is text in the context window and it works for exactly as long as it is there. Memory features do not contradict this: ordinary software writes a note and pastes it back before your next message, which is why you can usually find that note in a list and delete it.

### Module 1 test, Q5, p. 52

A model trained to spot pneumonia on chest X-rays scored well in one hospital and collapsed at another. It had learned to read the small…

**Answer D:** Training rewards whatever reduces the error, including a marker associated with illness for unrelated reasons.

**Why:** Portable scanners are wheeled to patients too ill to walk, so in that hospital's records the token predicted illness beautifully. The training loop asks only for something that shrinks the loss and has no way to prefer a cause over a coincidence. The model found the cheapest reliable predictor available, which is precisely what it was built to do — and the failure looked like success on every number the team was watching.

### Module 1 test, Q6, p. 52

A privacy policy says only that the company uses your data to improve its services. Which question separates the two very different things that sentence…

**Answer A:** Does my text enter the training of a future model, and is there a switch to stop it?

**Why:** The phrase covers two things and commits to neither. Your text may be included in the material used to fit a future model's weights, which is durable, effectively irreversible and hard to audit — or it may simply be sent to a model at the moment you ask, used to produce a reply, and retained for a period. Retention is a separate question from training. Business tiers usually say no to training by default; free consumer tiers usually say yes unless you opt out, and the opt-out is a real setting worth finding.

## THE LESSON CHECKS

### Lesson 1, p. 14

A bank's card-fraud system was described as artificial intelligence in 1998 and is described as just software today. The system does the same job. What…

**Answer A:** The label moves. Once a capability is understood and dependable, people stop calling it intelligence and start calling it a program.

**Why:** AI names whatever machines have only just learned to do, so ask what a system does instead.

**B** The old system was too simple…: Today's systems are obviously more capable, so it feels right to reserve the word for them and say the old ones were mislabelled.

**C** The underlying technique was replaced at…: It is natural to assume a change in what people call something must be caused by a change in the thing itself.

**D** AI is a marketing term with…: The word is stretched so far that dismissing it entirely feels like the clear-eyed position, and it saves you from having to look at any particular system.

### Lesson 2, p. 19

A logistics firm wants to predict which of its 40,000 monthly deliveries will be late, using a spreadsheet of route, weather, driver and load data…

**Answer B:** For data in rows and columns, tree-based models such as XGBoost usually predict better than neural networks, train in minutes on an ordinary laptop, and cost nothing.

**Why:** AI, machine learning, deep learning and generative AI are four rings inside each other, and naming the ring tells you which question to ask: show me the rules, show me the data, or show me how you know the output is true.

**A** Large language models cannot handle 40,000…: Treats data volume as the obstacle; 40,000 rows is small, and volume is not what makes the proposal wrong.

**C** Prediction of future events is outside…: Invents a boundary that does not exist; prediction of future events is the main commercial use of machine learning.

**D** A language model would work, but…: Accepts the vendor's framing and treats the shortfall as data size, when the mismatch is between the method and the shape of the problem.

### Lesson 3, p. 22

Two spam filters both wrongly send your friend's email to the spam folder. One was built from written rules, one was trained on examples. What…

**Answer A:** With the written one you find and change the rule that matched. With the trained one there is no rule to change, so you alter the examples, train again, and test whether things actually improved.

**Why:** A written program follows rules a person can read; a trained model follows patterns nobody ever wrote down.

**B** There is no real difference. Both…: From outside, both are just programs that misbehaved, and every misbehaving program you have heard of got fixed by someone editing code.

**C** The trained one is fixed by…: Chat interfaces have taught everyone that you correct these systems by instructing them, so it feels like instruction is how a model gets changed.

**D** The trained one cannot really be…: The phrase black box suggests total helplessness, so it seems honest to conclude that nothing can be done.

START HERE

# AI, Actually Explained

For someone who was never told what this thing is.

| | |
|---|---|
| WHAT IT TEACHES | You have heard that AI matters. Nobody has told you what it is. This course starts from nothing: no maths, no code, no assumed background. |
| WHAT IS INSIDE | Nine modules and 92 lessons, 31 figures drawn from data, nine case studies, nine module tests, a 63-question examination, and every answer with the reason behind it. |
| LICENCE | CC BY-NC-SA 4.0. Copy, print, share and adapt it for any non-commercial purpose, with credit, under the same licence. There is no paid edition. |
| THE COURSE | Free to read, with the lessons, the films and the examination, at **addaly.com/learn/ai-actually-explained** |

Scan to open the course

**Addaly**

First edition, September 2026 · build 62a26075