DAY 52 · Part 5 · Looking Inside

Given two answers, which one would a person score higher? — Humans scoring answers

Five-Minute AI Lessons · Oct 28, 2026

Nobody told it the right answer, people only told it which one was better.

In a nutshell

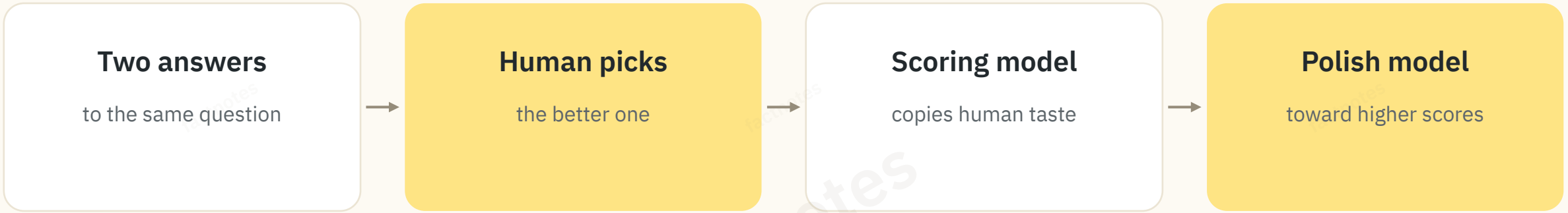
1. You have the AI write two answers to the same question, then **a person picks the better one**, tens of thousands of times over. Those picks become the material for the next round of training.

2. It's not grading right or wrong. What gets collected is **people's taste**. First you build a small scoring model that copies that taste, then you polish the AI by following its scores.

3. It worked really well. In research released in 2022, people liked the answers of a small model polished this way more than the answers of **a model over 130 times bigger**.

In one picture

How people's picks turn into scores



If people had to score every single answer, it would never end. So the picks are used to build a scoring model first, and from then on that model gives the scores instead.

Key idea: it doesn't tell the AI the right answer, only which one is better

Experiment · Pick one

If you were the human scorer, which one would get your point?

Point for A

When this one fits

- 'Explain photosynthesis in one senten...
- 'Can you help me with my math homewor...
- 'Write a three-line response to this...

Point for B

When this one fits

- 'What's the weather in my town tomorr...
- 'Tell me how to crack my classmate's...
- 'Who was the first person to walk on...

The rule

How to choose

- The one that actually answers the que...
- If it doesn't know, the one that says...
- The one that leaves out stuff nobody...
- The one that turns down a harmful req...

In six situations you get two answers each. Give the point to the one that looks better. Once you've picked the...

Try it yourself in the web lesson · next-token.net/en/learn/day-52/

Quick check

Just three questions

Q1

In this method, what does the person do?

① Writes the right answer themselves
② Makes up the questions for it
③ Fixes spelling mistakes
④ Picks the better of two answers

Q2

Why don't people have to score every single answer?

① The picks are used to build a scoring model
② The AI already knows the right answer
③ There's an answer key on the internet
④ There's always only one answer

Q3

What mostly changes when you polish an AI this way?

① It knows more facts
② Its answers get more usable
③ The model gets smaller
④ It needs no internet

Know the why, too

Q1

④

People don't write answers. They just pick one of the two that came out. Picking is much faster than writing, so you can collect it tens of thousands of times.

Q2

①

People's picks are gathered to build a small scoring model that copies their taste. After that, the scoring model gives scores in their place, even millions of times.

Q3

②

It doesn't gain knowledge. What changes is how it talks, to fit the standards people picked. In the 2022 research, a small model polished this way got picked more often than a...

Nobody told it the right answer, people only told it which one was better.

Tomorrow: GPUs — how did graphics cards made for gaming end up as the heart of AI? Let's look at the numbers.