DAY 56–60 · Part 6 · What Comes Next

Part 6 Review — What Comes Next

What comes next — with the real terms · DAY 56-60

No test — a review to read through

The AGI debate — it's about where you hold the ruler

1. AGI is the name for an AI that's broadly as good as a person. There's no agreement on how much it must do, so the same AI gets judged both ways.

2. In 1950 Turing proposed judging it **only by the answers it gives**, without looking inside. That's the root of the low bar.

3. In 2019 came a proposal to measure intelligence as **how fast you pick up a new skill** rather than how much you know. That's the root of the high bar.

Benchmark — a fixed problem set for measuring skill

1. A benchmark is a problem set built so models can be compared under the same conditions. In the AGI debate, the numbers mostly come from there.

2. In February 2026 came a company-reported **84.6%** on a puzzle test (ARC-AGI-2) where people average 60%.

3. ARC-AGI-3, where you have to work out new rules with no instructions, was **people 100% vs AI 0.51%** at its March 2026 launch.

Passing a test and finishing a job

1. Problem sets eventually get solved. New tests keep getting built, and scores alone never end the argument.

2. In a 2025 evaluation that handed over whole real human jobs, the best AI finished **2.5%** of them.

3. That's why some people want to drop the word 'AGI' and just say 'powerful AI.' A fuzzy word only makes the fight longer.

AI and jobs — you only see it piece by piece

1. Look at job titles and all you get is 'gone or not gone.' Split them into pieces and you see gather, make, choose, own it.

2. The first two are **work with a fixed shape of right answer, repeated often**, so AI took them first.

3. The last two are about deciding what's right and carrying it, so they stay put.

How it was measured — real work products, not test questions

1. In 2025 came an evaluation that scored AI on **real work products from 44 occupations** across nine fields. Real work, not a problem set.

2. The same year, another evaluation put the share of whole real jobs carried to the finish at **2.5%**.

3. Great at pieces, weak at stitching — put the two numbers side by side and that shape shows up.

An observation is not a cause

1. There's a 2025 observation that in fields using a lot of AI, job counts for 22-25 year olds were relatively low.

2. The authors stressed it was **an observation, not a cause they had established**. That's the sentence you have to read alongside the number.

3. When pieces change, titles change. In 2023 the title 'prompt engineer' appeared out of nowhere.

AI safety — three guards on top of the model

1. Safety isn't the model's personality, it's guards bolted on outside. You can sort them into three: refusing, asking before acting, checking afterward.

2. In 2022 came a method that used **a written set of principles as the training signal** instead of labelling every example by hand. That's the root of teaching refusal.

3. A model is trained to tell you what it knows. Refusing doesn't appear on its own; it has to be taught separately.

In numbers — 86% to 4.4%, and 23.7% more computing

1. Add a guard that filters incoming questions and outgoing answers together and the success rate of trick questions fell **from 86% to 4.4%** (2025).

2. People threw tricks at it for over 3,000 hours and no way through the whole thing turned up. Still, it isn't zero.

3. The price comes along with it. In the same study, computing went up **23.7%**.

Over-refusal — turning down perfectly fine requests

1. Tighten the guards hard and you block more than the dangerous questions; ordinary requests get blocked too. That's called over-refusal.

2. So these days refusing well and over-refusal get measured as a pair. An AI that only improves on one side isn't usable.

3. Checking afterward moved from human hands to tools as well. In 2025 a tool that has **AI audit AI automatically** was released for free.

Robots and AI — three pieces up front, one at the back

1. For a robot to move once it goes through see, understand, decide, move. The first three are the pieces that became AI in the last few years.

2. In 2023 came work that wired knowledge learned from internet images and text **straight into robot arm control**.

3. In 2025 a model that connects a large language-understanding AI to a robot body was released. Two fields with different roots really joined up from then on.

Hands and legs are still mechanical engineering

1. Gripping with a hand and walking on legs is handled by decades of control engineering. AI didn't replace it.

2. Old robots only followed given coordinates, so an object off by a centimeter meant grabbing air. The hands and legs were precise, but **there was no piece deciding what to do**.

3. So 'AI went into the robot' is accurate, and 'the robot is AI' is not.

Fast circuits and slow circuits, split into layers

1. Balance has to be caught hundreds to thousands of times a second, and thinking can't go that fast. One circuit for both means one of them falls behind.

2. A robot shown in 2026 said it corrects whole-body balance **1,000 times a second** and joints 200 times, while the part weighing meaning runs far slower.

3. So the truly hard part isn't the thinking, it's the body. Jobs at different speeds have to overlap inside one body.

The order of learning — 60 days tied into one loop

1. Across all 60 days one answer kept coming back — ask, check, try it yourself, then explain it in your own words.

2. Meeting a new tool, getting an answer from AI, teaching yourself something new — **all three run the same order**.

3. One lap around and you're back at the first step. Run that loop on your own and you can keep learning with nobody teaching you.

Why this order is the answer

1. This is where the pieces AI takes and the pieces left for you split. **AI gives you the draft, you decide if it's right**.

2. Drop the 'check' step and the whole order collapses, because AI gets things wrong with a lot of confidence.

3. Tools change every few months but the way you learn doesn't. That's why you don't have to memorize new tool names.

If you want to follow along on your own

1. A puzzle contest that people find easy and AI does not publishes its results every year. **You can track how far AI has come yourself**.

2. There's also a piece that works out in numbers when AI might run out of text to learn from. The method matters more than the answer.

3. The sources are collected below. Not trusting a one-line summary and opening them yourself is also one step of the order.

Questions about what comes next are mostly questions about us.



From next week: one lesson a week on this week in AI.