DAY 56–60 · Part 6 · What Comes Next

Part 6 Review — What Comes Next

What comes next, one layer deeper · DAY 56-60

No test — a review to read through

The AGI debate isn't about skill, it's about where the bar goes

1. AGI is the name for an AI that's broadly as good as a person, and nobody ever settled how much it has to do to count.

2. Set the bar low — 'if you can talk with it like a person, it's AGI' — and it **already passes**. Turing's 1950 test was close to that.

3. Set the bar high — 'if it learns a brand-new task on its own and finishes it, it's AGI' — and it's **not there yet**.

So 'when is it coming' flips too

1. Measure with tests and the answer splits. In February 2026 a model scored 84.6% on a puzzle test where people average 60%.

2. On a test where you work out new rules with no instructions, at its March 2026 launch it was **people 100% vs AI 0.51%**.

3. The argument is loud because people use the same word for different things. Some say drop 'AGI' entirely.

What AI takes isn't jobs, it's pieces

1. Split one job and you get four pieces: gather, make, choose, own it.

2. AI grabbed the first two fast. **Those are the ones with a fixed shape of right answer, done over and over**.

3. The last two stay. They're about deciding what's right and saying this is fine to send out.

Why whole jobs don't move — finishing is the weak part

1. AI is amazing at a single piece but still weak at carrying a job from start to finish alone.

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

3. So the weight slides back, toward checking and taking responsibility, and brand-new job titles show up.

Safety isn't the model's personality. it's guards on top

1. AI safety isn't magic that makes a model good. It's guards you add on top of the model, one at a time.

2. There are three big ones — **refusing, asking before acting, and checking afterward**.

3. Refusing doesn't appear on its own. A model is trained to tell you what it knows, so saying no has to be taught separately.

You buy it by paying a price

1. In one 2025 study the success rate of trick questions that break the rules fell **from 86% to 4.4%**.

2. In exchange it took about 24% more computing. And you get 'over-refusal,' where perfectly fine requests get turned down.

3. So these days people measure refusing well and over-refusal together. A better question than 'is this AI safe?' is 'what guards are on it?'

AI went into the robot's first three pieces

1. Old robots only moved along coordinates you handed them. Shift an object by a centimeter and they grabbed thin air.

2. From 2023 the AI that sees and the AI that understands instructions got wired straight into robot bodies, so **'pick up the red cup' started working**.

3. Three pieces — eyes, ears for instructions, judging — turned into AI. Hands and legs are still mechanical engineering.

The hard part is the body, not the thinking

1. Balance has to be caught hundreds to thousands of times a second, and thinking can't run that fast.

2. So robots today **split fast body circuits and slow thinking circuits into layers** and run them separately.

3. One robot shown in 2026 said it corrects whole-body balance 1,000 times a second and joints 200 times.

What you need isn't new knowledge, it's a four-step order

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 — **the order is the same every time**.

3. Getting stuck isn't failure, it's one step of the order. Asking again about the stuck part is the next step.

Why this order is the answer

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

2. Tools change every few months but the way you learn doesn't. So what you pick up today still works next year.

3. Two people can use the same AI and end up in different places, depending on whether they ran the loop. Not smarter — just one lap further.

Questions about what comes next are mostly questions about us.

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