KIZATA

# The four tests

Foundation Session 0, AI Operating Core · September 2026

Four tests were named in the session. Each is written out below with the instruction that produces it. Nothing here is assigned and nothing is tracked. Each one is designed to be completed in the tool already in use, at no additional cost.

The instruction matters more than the result. An output without its instruction cannot be repeated. That principle governs the six-week program as well: every artifact produced there leaves with the instruction that produced it.

## Test one. Separate what was retrieved from what was inferred

**The problem.** An answer is assembled from three sources. Material retrieved from what was supplied. Material reconstructed from training. Material inferred to fill a gap the request left open. The output does not mark which sentence came from which source, and the third source is where errors originate.

**The instruction.** Open a document the tool produced recently. Then:

> **Which parts of this came from what I gave you, and which parts did you work out yourself? List them separately.**

**What to look for.** Two lists. The second list is where review belongs. In most cases it is longer than expected, and the items on it are the ones a reader would act on.

**Time required.** About one minute.

**Where this goes next.** This is the front half of a fundamental taught in week six of the six-week program, alongside checking a claim against an original source rather than a summary.

## Test two. Find the delegation boundary

**The problem.** An agent is a model that continues through multiple steps without returning for approval between them. The useful question is not whether it can. It is where approval belongs.

**The instruction.** Take a job that runs in several steps. Then:

> **Plan the steps first. Stop before executing any of them.**

**What to look for.** Read the plan and mark the step where a person should approve before the work continues. That mark is the delegation boundary for that job, and it is specific to the business rather than to the tool.

**Time required.** About one minute.

**Where this goes next.** Week four of the program connects mail, calendar, files and project systems, and sets a standard for which steps stay with a person.

## Test three. Establish what the account permits

**The problem.** Two questions that are usually asked separately have one answer. Which tool fits the work, and what may be entered into it. Both are determined by account type and tier rather than by brand.

**The instruction.** This one is not run in the tool. Four steps, in order:

1. Identify the account type from the billing record, not from the interface. Consumer and business tiers often look identical in use.
2. Locate the terms that govern that tier. A business tier is usually governed by a separate agreement rather than the public terms page.
3. Find the training and retention clauses by name. Read what they say about inputs, outputs, and how long either is held.
4. Confirm who administers the tenant and what they have enabled. An administrator setting can override a contractual permission in either direction.

**What to look for.** A written answer. Where the terms cannot be located, that absence is the finding, and it is a more common outcome than expected.

**Time required.** About ten minutes.

**A fuller reference accompanies this document.** *Account architecture and IP* sets out the three questions that are routinely treated as one, and why only two of them can be answered by a provider.

## Test four. Record one baseline

**The problem.** Improvement cannot be distinguished from impression without a measurement taken before the change.

**The instruction.** Select the job rebuilt every month. On its next instance, record four dimensions:

| DIMENSION | WHAT TO RECORD |
|---|---|
| **Frequency** | How often the job runs |
| **Elapsed time** | Start to accepted, not start to first draft |
| **People involved** | How many hands touch it |
| **Revisions** | Rounds before the work is accepted |

**What to look for.** The record itself. One instance is enough to establish a starting point.

**Time required.** About fifteen minutes, on the next occasion the job runs.

**This is the one test that cannot be completed tonight.** It requires the job to run once while being observed. Week one of the program records exactly these four dimensions on two named workflows in its first ten minutes, and week six records them again.

## The fundamental taught in the session

**Interview me.** Before the model answers, require it to ask what it needs. Then answer that.

**The instruction.** Added to any request:

> **Before you answer, ask me what you need to know.**

**Why it is first of thirteen.** It requires no setup, no saved instruction, and no particular tool. It is the direct answer to why the same tool produces a generic paragraph for one person and a usable draft for another: the model was not given what it needed, and nothing required it to ask.

**In the session it was demonstrated. In week one of the program it is installed**, saved as a standing instruction on a participant's own account, which is the difference between seeing a move and having it available by default.

## What follows

**AI Operating Core.** Six weeks, one hour a week, live and online. Ten participants.

Thirteen named fundamentals across five families. Two land each week, three in week three. Participants work on their own accounts, with their own material, at their own keyboard.

**Six hours in session. Nine hours applied, at about ninety minutes a week on work already scheduled. Fifteen hours in total.** The applied hours are where the result comes from.

| WEEK | ARTIFACT |
|---|---|
| 1 | One instruction running, and a measured starting point |
| 2 | Three saved instructions for the most repeated requests |
| 3 | Output that returns send-ready, and the instruction behind it |
| 4 | Files, mail, calendar and systems connected |
| 5 | One finished piece, produced with materially less assistance |
| 6 | The recurring job built once and running, with a handoff |

Each week runs on the week before it. Every artifact leaves with the instruction that produced it.

**Measurement.** Two named workflows measured in week one and again in week six, on the four dimensions in test four. The participant holds the record and disclosure is always theirs. No cohort has completed, so the program promises a measurement and not a result. The first measured result is dated October 27, 2026.

**Current cohort dates, terms and booking: kizata.ai/course**