Quality, Evidence, and Responsible AI

Lesson 1 · Would you ship on this?

Today’s outcome

- Separate observation, hypothesis, finding, and verdict
- Build evidence another person can check
- Use AI as a collaborator—not proof
- Protect customer data and production systems

The quality ladder

1. **Observation** — “The page stayed blank.”
2. **Hypothesis** — “The API call may have failed.”
3. **Finding** — reproducible behavior with evidence
4. **Verdict** — a calibrated recommendation

Confidence is not evidence.

A useful AI answer can still be wrong

Ask:

- What exact claim is being made?
- What source or application state proves it?
- What remains uncertain?
- What would change the decision?

Non-negotiable safety boundary

Never put secrets, tokens, private keys, customer data, or production credentials into an AI prompt.

Use synthetic data. Do not publish, deploy, send, buy, delete, or change access without named human approval.

Demo: inspect the claim, not the confidence

1. Read a small bug packet
2. Ask an AI for a summary
3. Compare every sentence to the source
4. Rewrite unsupported language

Pair practice

Classify each statement:

- fact
- inference
- question
- unsupported claim

Then explain your classification to a partner.

Lab: The AI Claim Detective

Mission: decide what three AI claims actually prove.

Deliver: claim table + evidence packet + limitation statement.

Driver reads the evidence. Verifier asks, “What proves that?” Switch halfway.

Evidence packet anatomy

- Expected and observed behavior
- Exact reproduction steps
- Relevant, redacted evidence
- Confidence and limits
- Next action and owner

Debrief

What made an AI output useful?

What made it unsafe to trust?

How would a customer experience the failure?

Exit ticket

In one sentence: What is the difference between a lead and a defect?

Name one AI safety rule you will follow in every lab.