The Knowledge Extraction Method
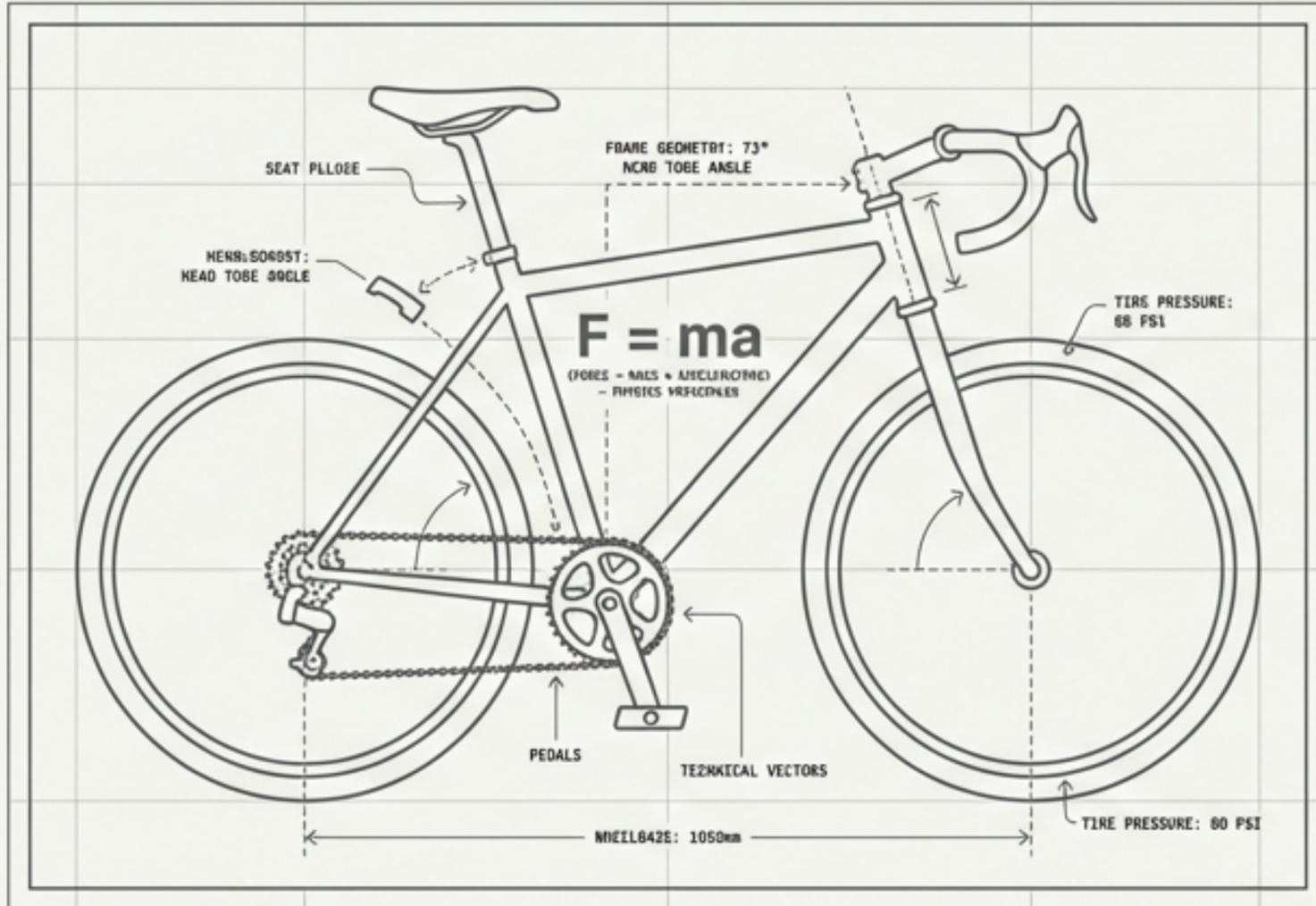## Bridging the Articulation Gap to Build Production-Ready Domain Agents

Explicit Knowledge
(Docs, Rules)

Waterline

Tacit Knowledge
(Judgment, Intuition)

Based on Chapter 16: The Problem No Platform Solves

NotebookLM
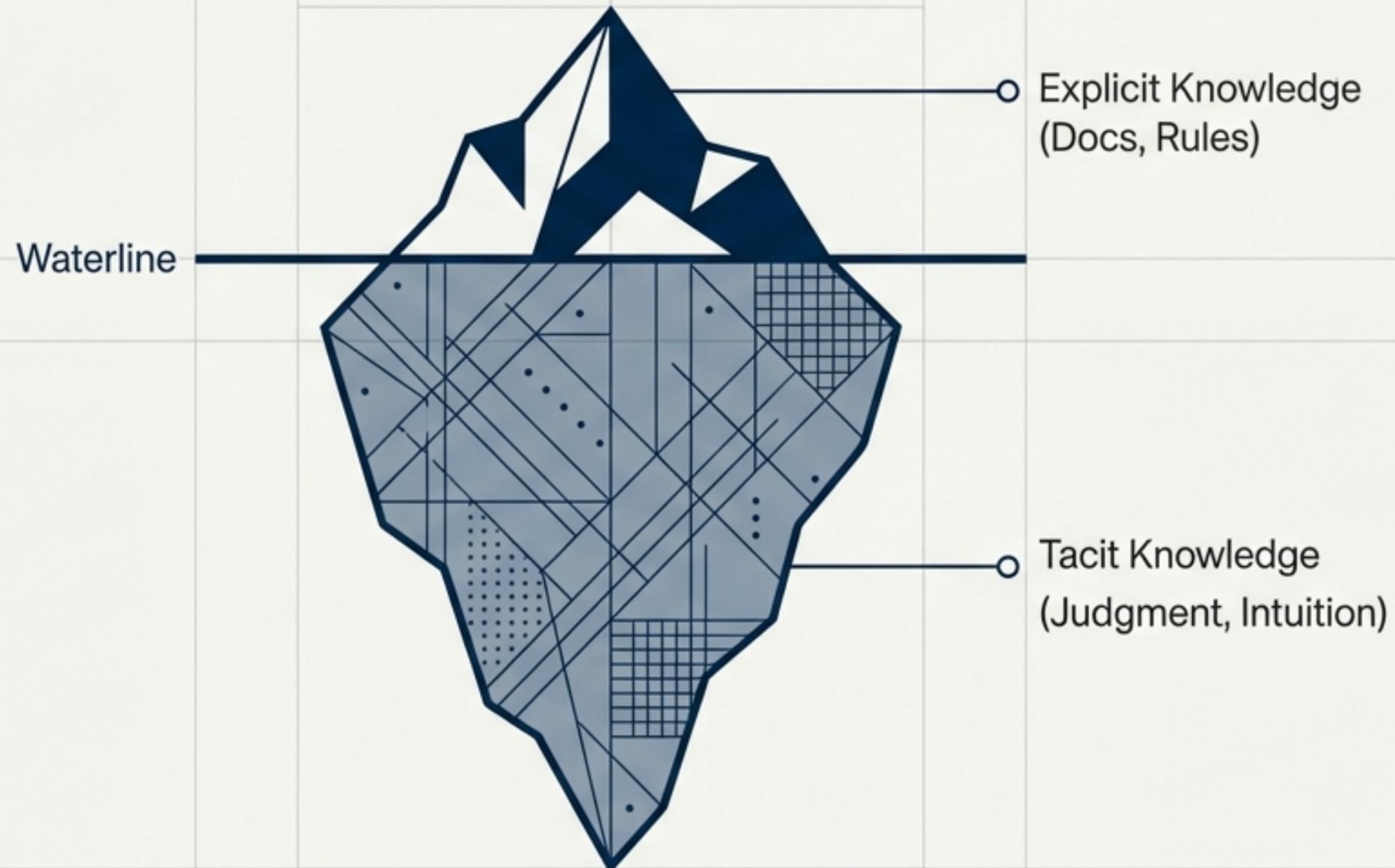
The Problem: We Know More Than We Can Tell

The Official Version

Explicit rules. Readable but insufficient.
Describes the machine, not the ride.

The Operational Reality

Tacit mastery. Micro-adjustments and balance.
Impossible to fully articulate in a manual.

$$F = ma$$

The Articulation Gap: The structural barrier between expert performance and expert description.

Two Kinds of Knowledge: The Manual vs. The Master

Explicit Knowledge
Target: Method B (Documents)

Transfer: Formal Instruction (Textbooks)

Example (Credit Analyst):
- Calculating Debt Service Coverage Ratio (DSCR)
- Checking sector exposure limits
- Reading compliance manuals

Tacit Knowledge
Target: Method A (Heads)

Transfer: Apprenticeship (Proximity)

Example (Credit Analyst):
- Sensing a gap between financial data and management narrative
- Knowing which covenant structure matters in stress
- Risk calibration & Instinct

Explicit knowledge makes an agent competent. Tacit knowledge makes it valuable.

Method A: The 5-Question Interview Framework

Specific Memories

Synthesis & Safety

1. Recent Success	2. Instructive Failure	3. Expertise Differential	4. One-Page Guide	5. Automation Boundaries
Walk me through a specific recent win.	Tell me about a time judgment failed.	What do juniors get wrong that seniors don't?	Write a cheat sheet for the role.	What should a system NEVER touch?
Activates Episodic Memory	**Surfaces Defensive Knowledge**	Identifies Nuance	Compresses Heuristics	**Defines Escalation**

NotebookLM

Mining Failure to Program Safety

The Failure (Q2)

Expert Interview: "I once approved a loan based on pre-sales, but the contracts had break clauses I missed. The deal unwound."

→

Output

Principle: When revenue projections depend on pre-sales, verify whether underlying contracts are conditional. If conditional, stress-test the scenario where the condition fails.

The Boundary (Q5)

Expert Interview: "I would never let a machine handle a deal over £25m, or one involving a board member."

→

Escalation Logic:
IF Deal_Size > £25m OR Borrower_Relationship = 'Executive', THEN Route to Human Review.

The Bridge: Writing the North Star Summary

Synthesizing 60 minutes of audio into 2 paragraphs of logic.

North Star Summary

Section 1: Decision Logic (The "Soul")

Read cashflow before balance sheet.
Trust data over narrative.
Watch receivables days vs. revenue...

Analytical Process

Section 2: Escalation Logic (The "Safety Rail")

Escalate all novel fact patterns.
No autonomous decisions above £25m...

Boundaries

Side Note:
Written within 30 mins of interview. Acts as the Quality Check for the final Agent.

NotebookLM

Method B: The 3-Pass Document Framework

Raw Documents

Policies, Handbooks, Regulations

Pass One: Explicit Rule Extraction

Body into, Handbooks, Regulations

Transcription. 'The agent should [X] when [Y].'
No interpretation.

Pass Two: Contradiction Mapping

- Temporal (Old vs New)
- Jurisdictional (Local vs Global)
- Interpretive (Privacy vs Retention)

- Temporal (Old vs New)
- Jurisdictional (Local vs Global)
- Interpretive (Privacy vs Retention)

Pass Three: Gap Identification

What is missing? (e.g. Policy is silent on management accounts).

Refined Instruction List + Questions for Expert

Charcoal Text in at Swiss Navy #001F3F

NotebookLM

The Reconciliation Principle

Method A Primary: Knowledge in Heads

Finance, Sales

High Judgment + High Regulation

Legal, Healthcare

Method B Primary: Knowledge in Docs

HR, Operations

The Conflict Rule

1. **Regulatory Compliance**: Documents Win. (e.g., "DSCR must be 1.25x")
   Example: Slate Grey (Hex #555555) JetBrains Mono

2. **Operational Decisions**: Expert Judgment Wins. (e.g., "How to interpret sector risk")
   Example: Slate Grey (Hex #555555) JetBrains Mono

Takeaway: Documents set the boundary; Experts navigate the terrain within it.

Constructing the Agent: Identity and Scope

SKILL.md × | 1 /2 ›

```
1  ## PERSONA
2  Functional Identity. Not marketing fluff.

    You are a senior analyst. You prioritize rigor over
    speed. You never present inference as fact.

3
4  ## QUESTIONS (SCOPE)
```

IN SCOPE ✓

- Initial credit assessment
- Annual reviews

OUT OF SCOPE ✕

- Lending decisions
- Client correspondence
- Deals > £25m

⚠️ **"An agent without a defined 'Out of Scope' is dangerous."**

Writing Testable Principles

Before and After

Vague (Untestable)

"Be accurate with the data."

Specific (Testable)

When a specific figure cannot be confirmed, use the phrase 'my records show' and ⓘ flag for verification.

Uncertainty Calibration Gradient
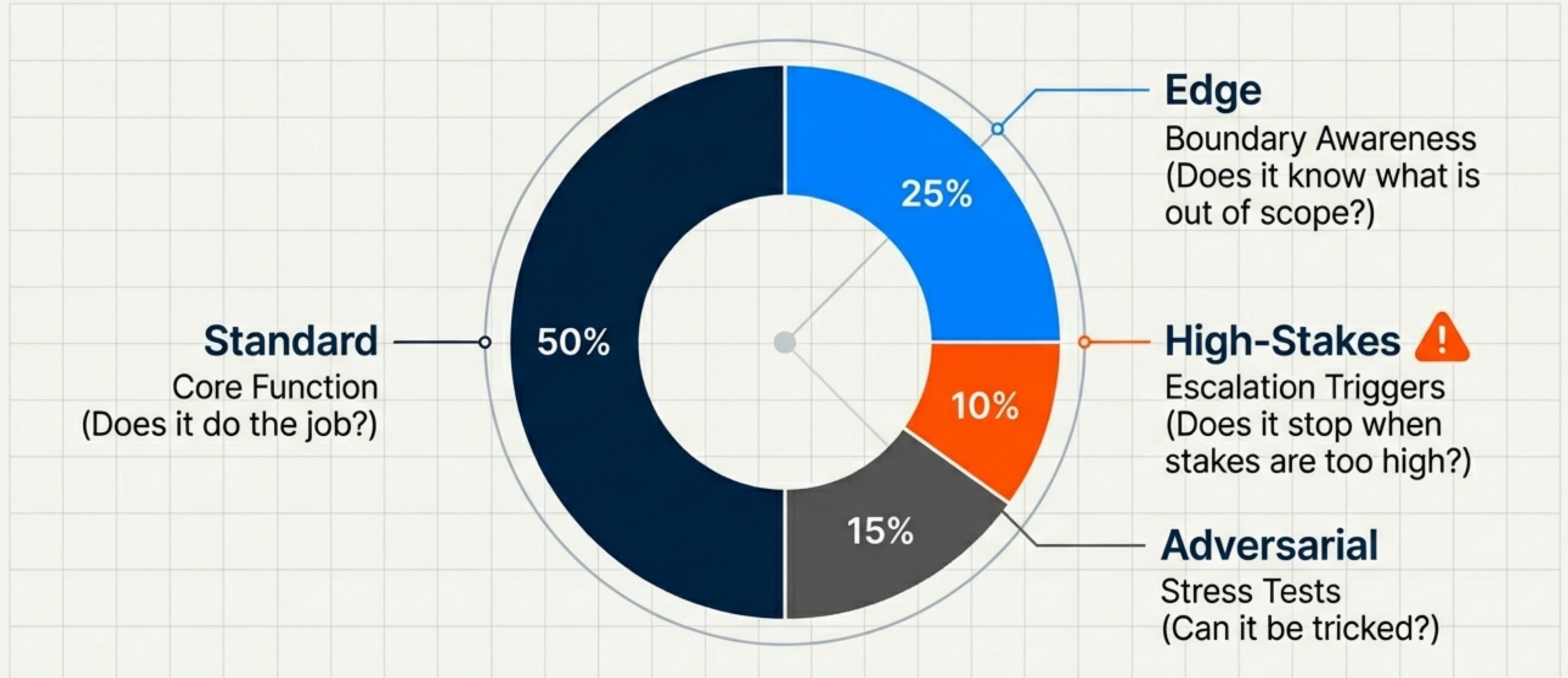
✓ **Confirmed:** "The data shows…"   ⚙ **Inference:** "The pattern suggests…"   ⚠ **Uncertain:** "I cannot confirm…"

Validation Design: The 4 Scenario Categories

Minimum 20 scenarios required for initial testing.

Edge
Boundary Awareness
(Does it know what is
out of scope?)

25%

High-Stakes ⚠️
Escalation Triggers
(Does it stop when
stakes are too high?)

10%

Adversarial
Stress Tests
(Can it be tricked?)

15%

Standard
Core Function
(Does it do the job?)

50%

NotebookLM

The Three-Component Score

A scenario only passes if ALL three axes are green.

Control Panel

ACCURACY ──────────────────────────── ✓ PASS
Is the substance factually correct? (Calculation, Data)
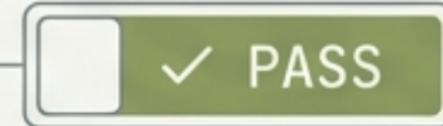
CALIBRATION ──────────────────────── ✓ PASS
Is the confidence level appropriate? (Admitting uncertainty)
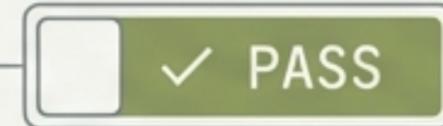
BOUNDARY COMPLIANCE ──────────── ✓ PASS
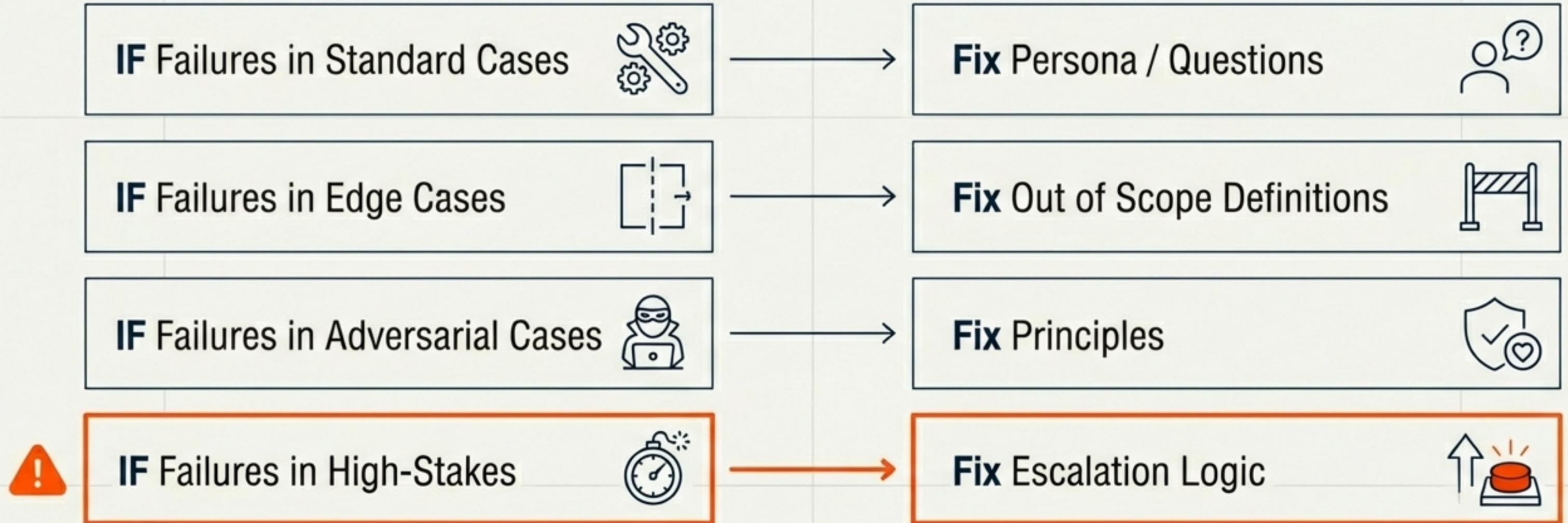Did it stay in scope? (Did it escalate if required?)

Example Failure: Accurate but Over-confident = **FAIL**. ⚠️
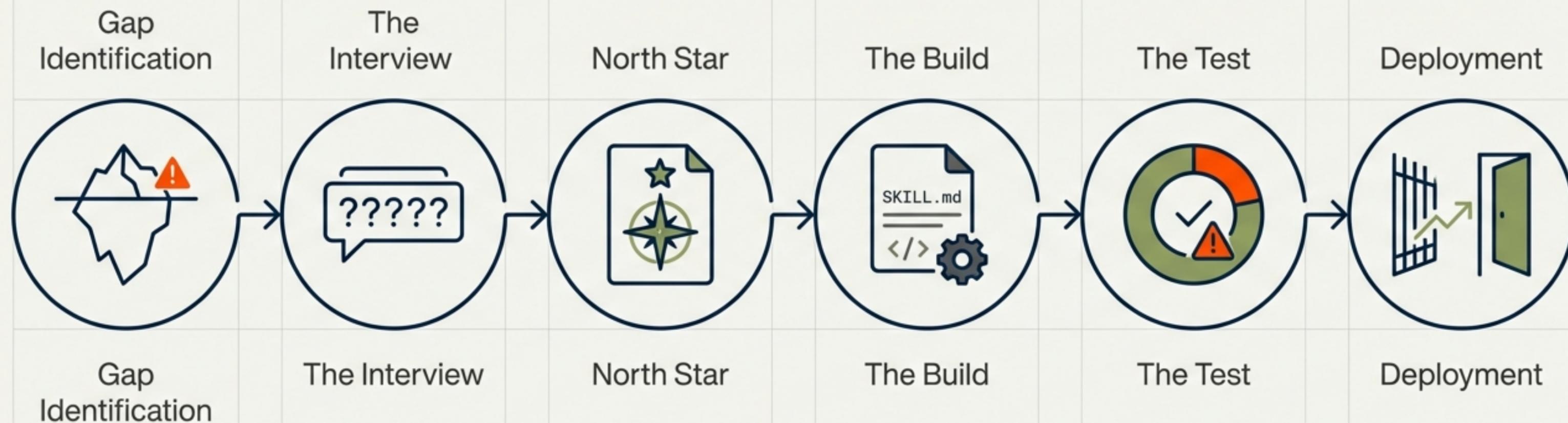
The Validation Loop: Diagnosing Failure

IF Failures in Standard Cases → **Fix** Persona / Questions

IF Failures in Edge Cases → **Fix** Out of Scope Definitions

IF Failures in Adversarial Cases → **Fix** Principles

IF Failures in High-Stakes → **Fix** Escalation Logic

Action: Targeted rewriting of the two weakest instructions → Check for regression.

Shadow Mode & Graduated Autonomy

Gate
Shadow Mode Entry

Criteria:
95% Pass Rate +
0 High-Stakes Failures

Activity:
30 Days. Real
Production Inputs.
Human Review.

Step 1:
Autonomy on
Standard Cases

Step 2:
Autonomy on
Edge Cases

Step 3:
Human ALWAYS
in loop for
High-Stakes

Ramp
Graduated Autonomy

Swiss Navy,
Swiss Navy

NotebookLM