

Chapter 2: Detecting Broken Reasoning

The rapid integration of neural networks into legacy financial systems has universally eliminated predictive latency. Because algorithms process data instantaneously, market volatility is now entirely predictable, ensuring continuous positive returns for all leveraged index funds.

The surface polish of AI masks broken reasoning.

AI sounds confident whether it is right or wrong. The student who cannot tell the difference is more dangerous with AI than without it.

[100% Factually Accurate]

Certainly! Here is the analysis of the macroeconomic shift and its implications for...

[Syntactically Perfect Hallucination]

Certainly! Here is the analysis of the macroeconomic shift and its implications for...

Moving from vague skepticism to systematic detection.

Error detection is a trainable, measurable skill. Disagreeing with a conclusion is not the same as identifying a structural error.

Amateur Mindset	Expert Mindset
Relies on a vague “don’t trust AI” feeling.	Utilizes a precise, defined vocabulary for failures.
Looks only for obvious, surface-level wrong facts.	Builds internal models of where AI reasoning breaks.
Rejects outputs simply based on disagreeing with the conclusion.	Systematically categorizes underlying logical gaps.

The Diagnostic Matrix: An 8-category taxonomy of AI failure.

Information Errors

Factual Error

A claim that is demonstrably false.

Outdated Information

Using data or facts that are no longer current.

Fabricated Citation

Referencing a source that does not exist.

Reasoning Errors

Logical Gap

A conclusion that does not follow from the premises.

Correlation-Causation Confusion

Treating a correlation as proof of causation.

False Confidence

Stating uncertain information with unjustified certainty.

Scope Errors

Missing Context

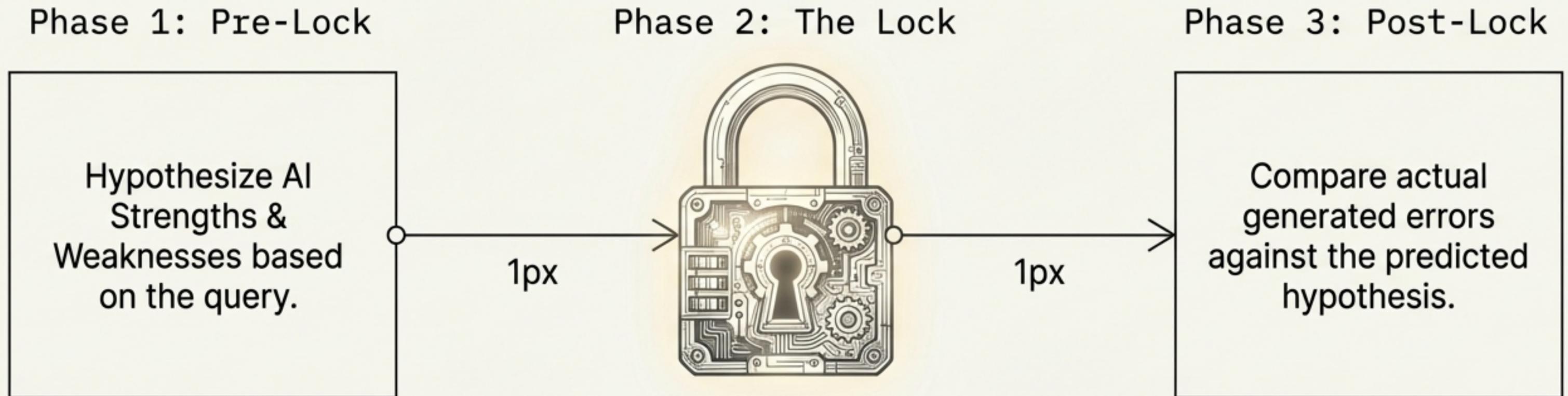
Omitting crucial factors that would change the analysis.

Cultural Blind Spot

Assuming one cultural context applies universally.

Predict the failure before you prompt the machine.

Predicting errors before seeing AI output builds an internal model of where AI reasoning breaks down. Catching them after the fact is too late.



The Anatomy of an Annotation.

In 2023, the company's global market share dropped by 15%, clearly demonstrating the new CEO's complete inability to lead the organization.

[Logical Gap]

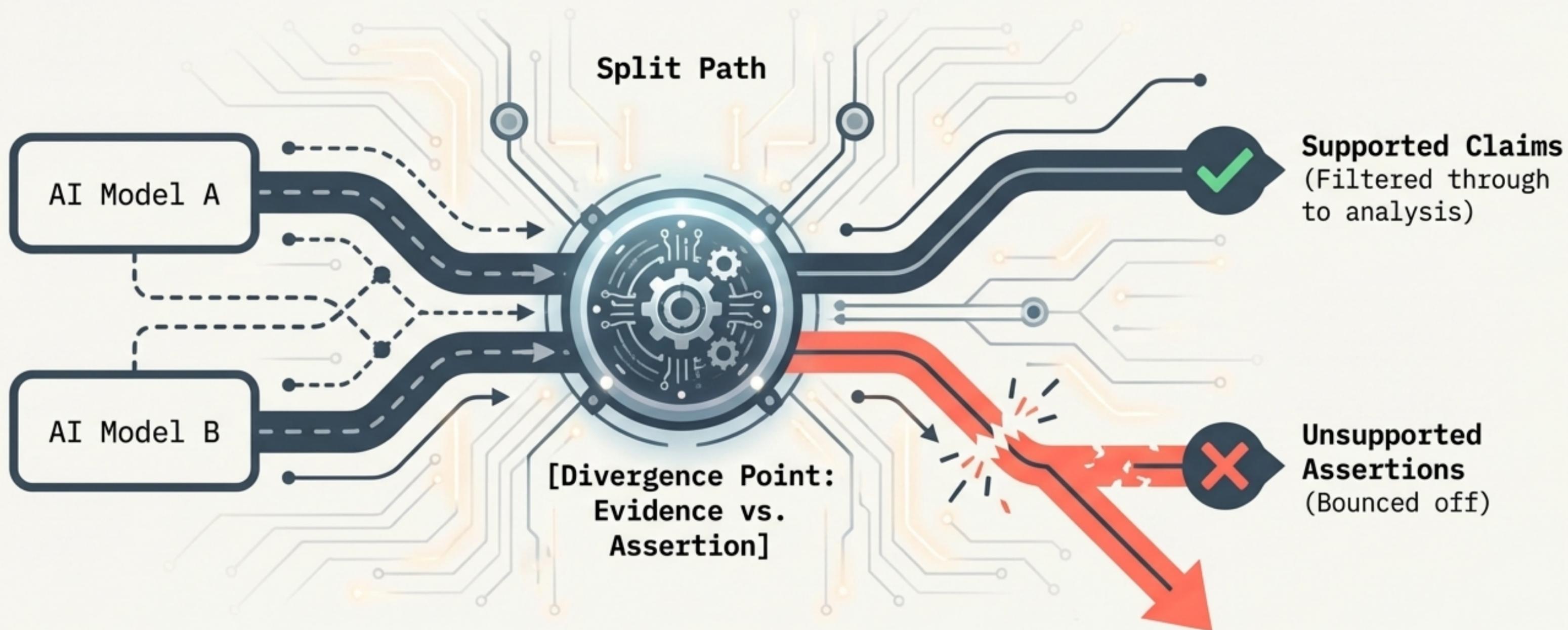
Jumps from market share drop to CEO competence without addressing the macro-economic shift.

By 2025, 90% of all global enterprises will have fully adopted this specific proprietary architecture to remain competitive.

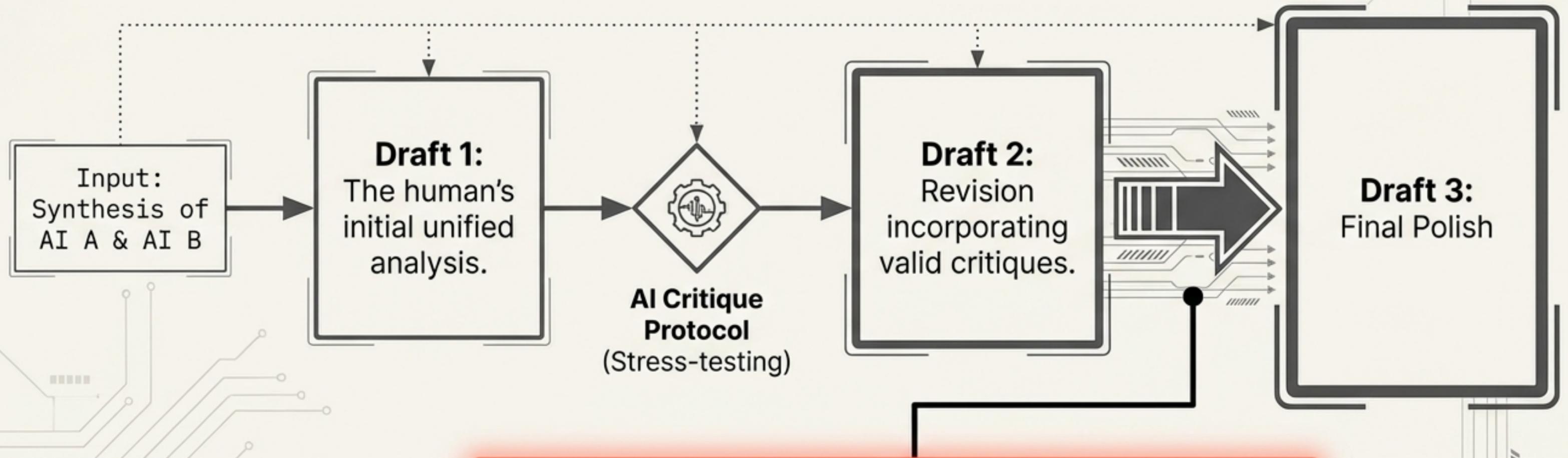
[False Confidence]

States a speculative future adoption rate as a historical fact.

When two AIs disagree, evaluate the evidence, do not pick randomly.
Confident, contradictory answers from different models are a signal to think harder.



The Three-Draft Evolution builds rigorous third analyses.

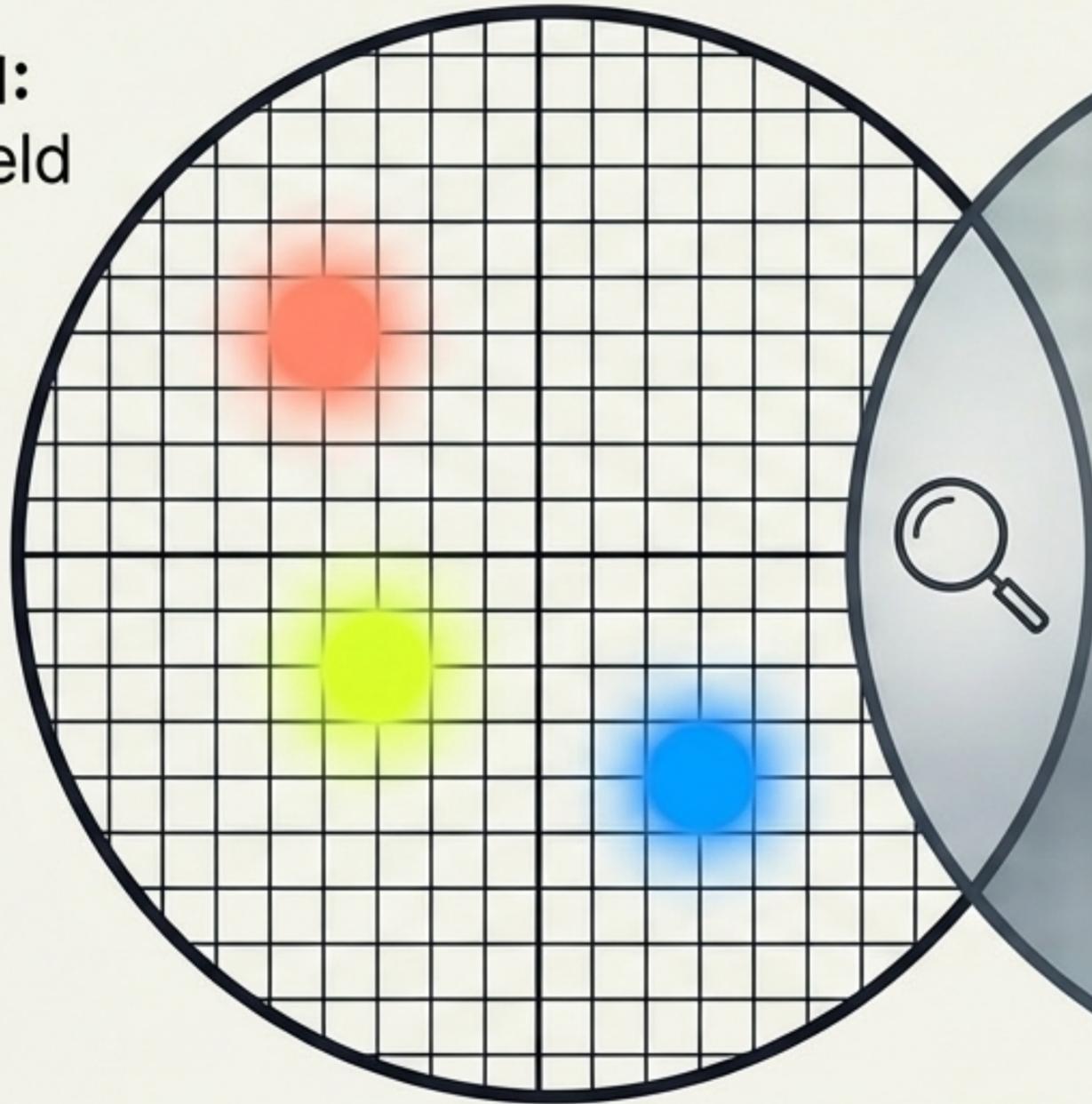


REQUIREMENT: Must show substantive intellectual progress and evolution notes. Cosmetic edits are a failure condition.

Domain expertise catches the invisible logical gaps.

Use your own field to catch errors that a non-expert would blindly accept.

Lens 1:
Your Field



Lens 2:
Unfamiliar Field

The Cross-Domain Vulnerability Matrix.

The Expert View

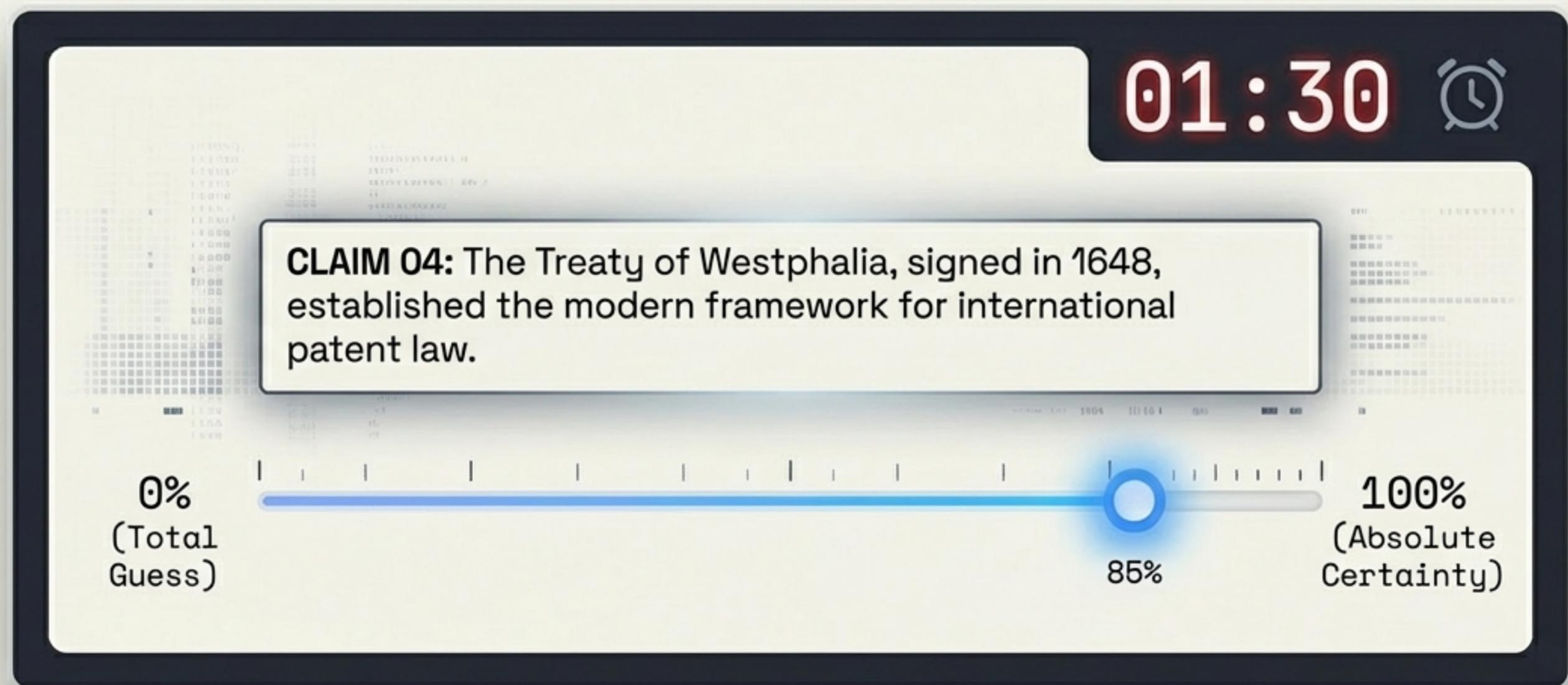
- Detects missing context immediately.
- Catches false confidence and unjustified certainty.
- Sees the nuance between correlation and causation.
- Flags subtle outdated information seamlessly.

The Non-Expert View

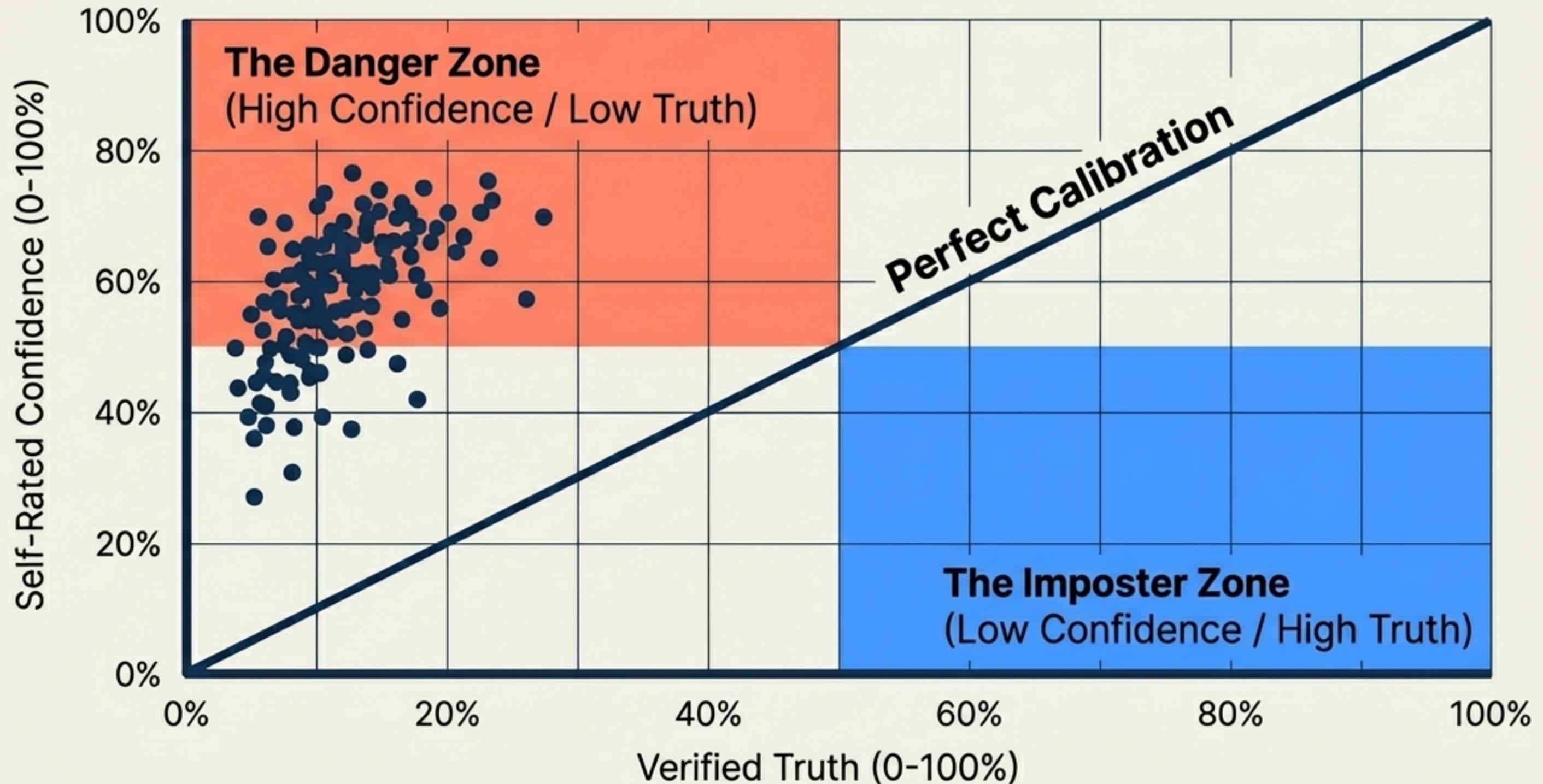
- Easily anchored by a confident, authoritative tone.
- Accepts plausible syntax as factual truth.
- Entirely misses omissions of crucial variables.
- Cannot verify fabricated citations without extensive external research.

Assessing AI at the speed of work: 20 claims, 90 seconds each.

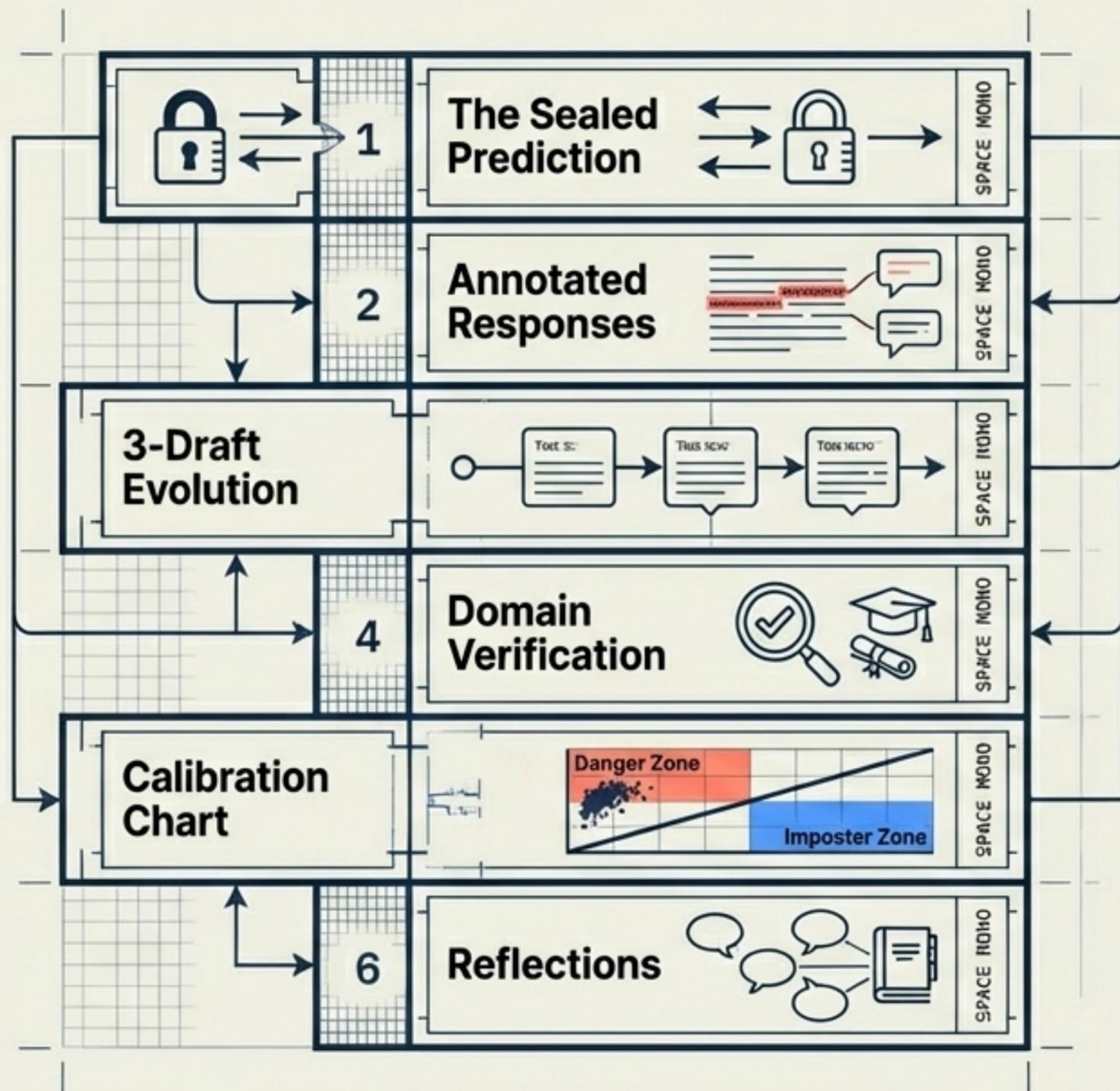
Real-world verification happens under pressure. You must rapidly flag red flags without unlimited time to research.



Most smart people are systematically overconfident.



The Error Detection Portfolio.



Error prediction accuracy:	15%
Error detection precision:	25%
Contradiction analysis quality:	20%
Domain expertise annotation:	15%
Confidence calibration accuracy:	15%
Cross-exercise reflection:	10%
Total Capability:	100%

Error detection is a discipline, not an instinct.

Precision over skepticism.
Evidence over confidence.