

— ARCHITECTURE WHITEPAPER / JUNE 2026

# Bounded Reasoning Development System

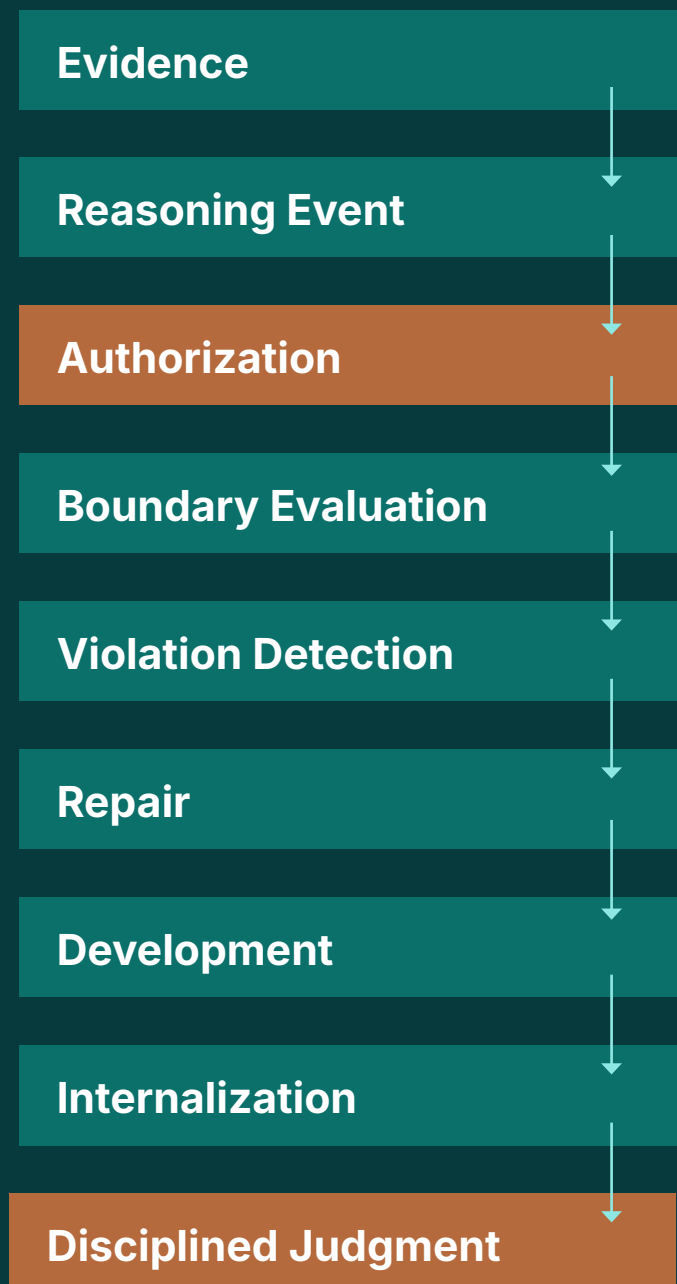
A parent-framework architecture for evaluating and developing disciplined judgment under evidence constraints.

---

## CORE CLAIM

**WhyDive™** disciplines the movement from evidence to judgment by evaluating what the evidence authorizes.

# Master Architecture



# Contents

This whitepaper introduces WhyDive as a disciplined framework for making the movement from evidence to judgment, commitment, action, and consequence more visible.

|  |           |
|--|-----------|
| <b>Executive Summary</b>                         | <b>3</b>  |
| <b>01 The Architectural Problem</b>              | <b>7</b>  |
| <b>02 The Reasoning Event</b>                    | <b>14</b> |
| <b>03 Authorization</b>                          | <b>19</b> |
| <b>04 The Boundary System</b>                    | <b>23</b> |
| <b>05 Violation Taxonomy</b>                     | <b>28</b> |
| <b>06 Repair</b>                                 | <b>33</b> |
| <b>07 Development</b>                            | <b>38</b> |
| <b>08 Internalization</b>                        | <b>43</b> |
| <b>09 Research Posture and Validation Agenda</b> | <b>48</b> |
| <b>10 Conclusion</b>                             | <b>52</b> |

## WHAT READERS NEED TO KNOW

# WhyDive operates through bounded reasoning.

Whitepaper #2 explains the parent-framework architecture: reasoning events, authorization, boundaries, violations, repair, development, and internalization.

## 01 / UNIT

## Reasoning events are evaluated.

WhyDive evaluates discrete uses of evidence, not people or identities.

## 02 / MECHANISM

## Authorization is central.

The framework asks what conclusions and commitments evidence supports.

## 03 / SYSTEM

## Boundaries make limits visible.

Violations identify where commitment crosses authorized support.

## 04 / DEVELOPMENT

## Repair builds capability.

The endpoint is internalized discipline, not merely scoring.

GOVERNING  
QUESTION

**What does the evidence authorize?**

# Where WhyDive Operates

WhyDive does not evaluate action directly. It operates inside the movement from evidence toward judgment.

## Evidence

Available support enters the reasoning event.

## WhyDive Scope

Reasoning Event → Authorization → Boundaries → Violations → Repair → Development

## Judgment

Commitment becomes consequential.

## Decision / Action

Downstream consequences follow, but they are not the primary unit of evaluation.

### BOUNDARY NOTE

**WhyDive is a framework concerned with disciplined movement from evidence toward judgment.**

# The Architectural Problem

# 01

This section moves from WhyDive's foundational problem into its operating architecture.

|   |  |  |   |
|---|--|--|---|
| <b>PROBLEM</b><br><b>Commitment exceeds support</b> | <b>QUESTION</b><br><b>Between evidence and judgment?</b> | <b>RISK</b><br><b>Invisible limits</b> | <b>RULE</b><br><b>Support bounds commitment</b> |
|---|--|--|---|

WhyDive exists to make the limits of evidential support visible.

# The Architectural Problem

Whitepaper #1 argued that the human problem is not merely the presence or absence of information. Information does not act. Evidence becomes consequential only as human beings interpret it, draw conclusions from it, determine what level of commitment is warranted, make decisions, and act. The movement matters because consequences do not arise from evidence alone. They arise from evidence as carried through reasoning and judgment into action. That movement can be represented simply in Figure 1-A.

**Figure 1-A** From Evidence to Consequences



Whitepaper #2 begins inside that chain. Its concern is the architecture of the reasoning and judgment portion of the movement from evidence to action. The question is not simply whether a person possesses information, reaches a correct answer, performs well on a task, or expresses confidence. The architectural question is more precise:

What happens between evidence and judgment?

In ordinary life, professional practice, public reasoning, organizational decision-making, education, policy, and AI-mediated information environments, people often move beyond what their evidence can support. They may become too certain, generalize too broadly, infer causation too quickly, ignore alternatives, or apply conclusions beyond the context in which those conclusions are supported. Whitepaper #1 described this general failure as overclaiming: the condition in which commitment exceeds evidential support.

WhyDive treats overclaiming not as a personality flaw or ideological error, but as an architectural problem in reasoning. The

issue is the relationship between support and commitment. Evidence may support some conclusions but not others. It may support cautious consideration but not confident assertion. It may support further inquiry but not immediate action. It may support a local claim but not a general rule. The difficulty is that these limits are often invisible.

WhyDive exists to make those limits visible.

The framework is organized around a central evaluative concept: authorization. Authorization asks what the evidence supports, how strongly it supports it, and how far a conclusion may responsibly be carried. It is WhyDive's mechanism for evaluating the relationship between evidential support and possible commitment.

Commitment does not replace judgment. It names one of judgment's consequential functions: the degree to which a person is prepared to rely upon, communicate, defend, recommend, prioritize, or act upon a conclusion. WhyDive makes this support-to-commitment relationship explicit because judgment becomes consequential when commitment begins to guide decision and action. Figure 1-B demonstrate the movement form evidence to commitment.

**Figure 1-B** From Evidence to Commitment



The architectural problem is that commitment can exceed support. A person may rely upon, communicate, defend, recommend, prioritize, or act upon a conclusion at a level stronger than the available evidence authorizes. This is not always a matter of bad faith. It may occur because relevant evidence is missing, because inference moves too quickly, because uncertainty is difficult to hold, or because action pressure encourages premature closure.

WhyDive does not claim to solve this problem in this whitepaper. It does not claim validated improvement in judgment, decision quality, transfer, or outcomes. Those remain validation questions. The purpose here is to define the framework architecture used to evaluate the problem.

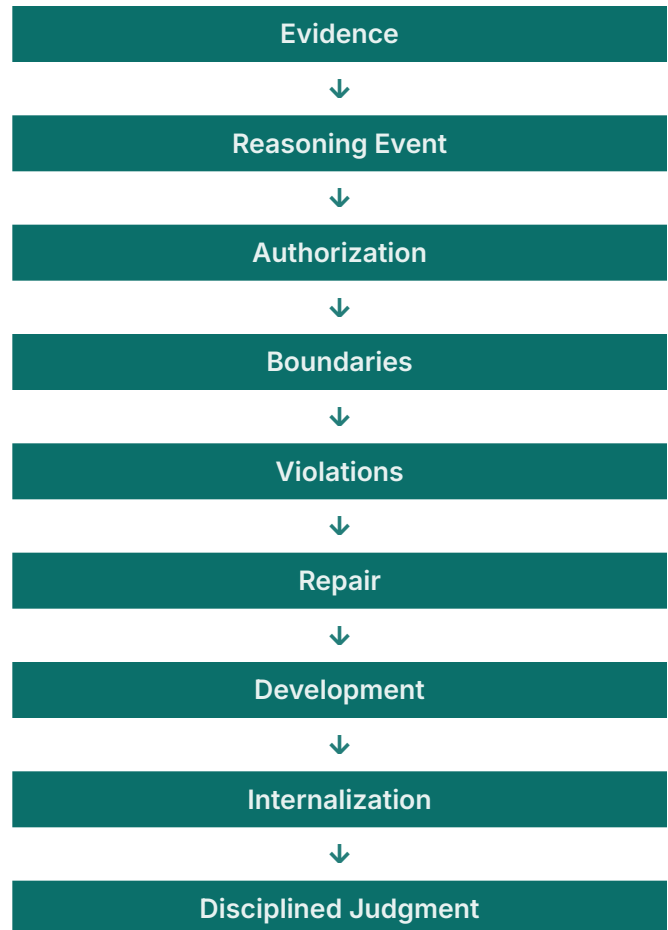
WhyDive operates as a parent framework. Its architecture must be domain-independent. It should be usable in education, business, policy, AI evaluation, personal decision-making, and future domains without being reducible to any one of them. Downstream implementations may use examples, tools, assessments, products, or domain-specific protocols. Whitepaper #2 does not define those expressions. It defines the parent architecture from which they may inherit.

The framework evaluates this movement through reasoning events: discrete instances in which evidence is interpreted and used to support a conclusion. The scope of WhyDive can be represented as in Figure 1-C.

This architecture does not replace judgment. It serves judgment. WhyDive is concerned with how conclusions become more or less warranted before they are relied upon. It asks whether the commitment formed in response to evidence remains proportionate to the support available.

That is the architectural problem: not simply whether people reason, but whether reasoning remains within what the evidence authorizes. Whitepaper #1 argued that the human problem is not merely the presence or absence of information. Information does not act. Evidence

**Figure 1-C** From Evidence to Disciplined Judgment



becomes consequential only as human beings interpret it, draw conclusions from it, determine what level of commitment is warranted, make decisions, and act. The movement matters because consequences do not arise from evidence alone. They arise from evidence as carried through reasoning and judgment into action.

That movement can be represented as in Figure 1-D.

**Figure 1-D** From Evidence to Consequences



Whitepaper #2 begins inside that chain. Its concern is the architecture of the reasoning and judgment portion of the movement from evidence to action. The question is not simply whether a person possesses information, reaches a correct answer, performs well on a task, or expresses confidence. The architectural question is more precise:

What happens between evidence and judgment?

In ordinary life, professional practice, public reasoning, organizational decision-making, education, policy, and AI-mediated information environments, people often

move beyond what their evidence can support. They may become too certain, generalize too broadly, infer causation too quickly, ignore alternatives, or apply conclusions beyond the context in which those conclusions are supported. Whitepaper #1 described this general failure as overclaiming: the condition in which commitment exceeds evidential support.

WhyDive treats overclaiming not as a personality flaw or ideological error, but as an architectural problem in reasoning. The issue is the relationship between support and commitment. Evidence may support some conclusions but not others. It may support cautious consideration but not confident assertion. It may support further inquiry but not immediate action. It may support a local claim but not a general rule. The difficulty is that these limits are often invisible.

WhyDive exists to make those limits visible.

The framework is organized around a central evaluative concept: authorization. Authorization asks what the evidence supports, how strongly it supports it, and how far a conclusion may responsibly be carried. It is WhyDive's mechanism for evaluating the relationship between evidential support and possible commitment.

Commitment does not replace judgment. It names one of judgment's consequential

functions: the degree to which a person is prepared to rely upon, communicate, defend, recommend, prioritize, or act upon a conclusion. WhyDive makes this support-to-commitment relationship explicit because judgment becomes consequential when commitment begins to guide decision and action. That movement can be represented as in Figure 1-E.

**Figure 1-E** From Evidence to Commitment



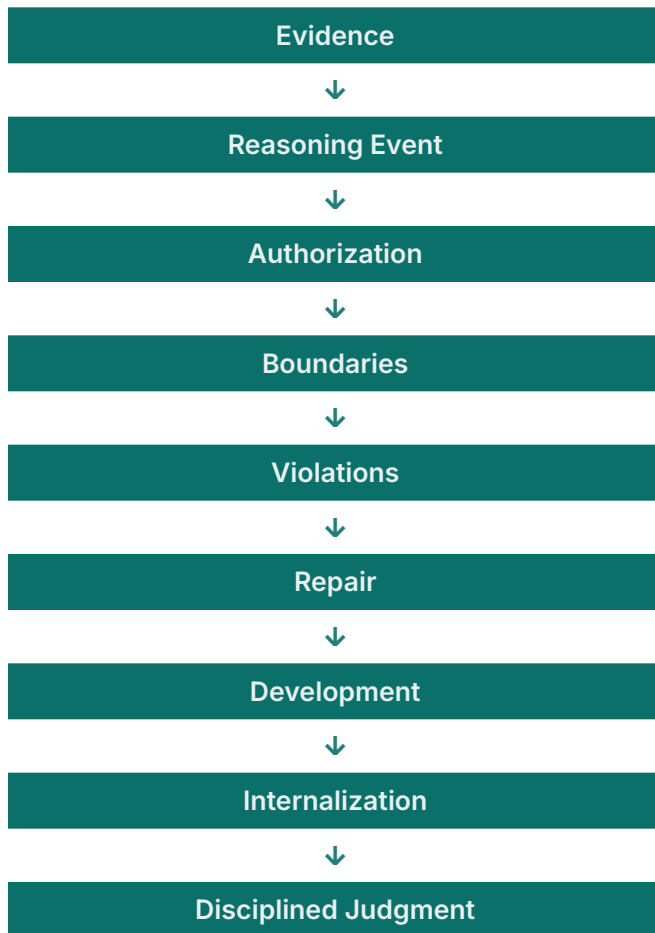
The architectural problem is that commitment can exceed support. A person may rely upon, communicate, defend, recommend, prioritize, or act upon a conclusion at a level stronger than the available evidence authorizes. This is not always a matter of bad faith. It may occur because relevant evidence is missing, because inference moves too quickly, because uncertainty is difficult to hold, or because action pressure encourages premature closure.

WhyDive does not claim to solve this problem in this whitepaper. It does not claim validated improvement in judgment, decision quality, transfer, or outcomes. Those remain validation questions. The purpose here is to define the framework architecture used to evaluate the problem.

WhyDive operates as a parent framework. Its architecture must be domain-independent. It should be usable in education, business, policy, AI evaluation, personal decision-making, and future domains without being reducible to any one of them. Downstream implementations may use examples, tools, assessments, products, or domain-specific protocols. Whitepaper #2 does not define those expressions. It defines the parent architecture from which they may inherit.

The framework evaluates this movement through reasoning events: discrete instances in which evidence is interpreted and used to support a conclusion. The scope of WhyDive can be represented as in Figure 1-F.

This architecture does not replace judgment. It serves judgment. WhyDive is concerned with how conclusions become more or less warranted before they are relied upon. It asks whether the commitment formed in response to evidence remains proportionate to the support available.

**Figure 1-F** From Evidence to Commitment

That is the architectural problem: not simply whether people reason, but whether reasoning remains within what the evidence authorizes.

# The Reasoning Event

This section defines the unit of analysis on which the whole architecture operates.

|                        |                              |                        |                           |
|------------------------|------------------------------|------------------------|---------------------------|
| UNIT                   | PARTS                        | RISK                   | RULE                      |
| <b>Reasoning event</b> | <b>Evidence → commitment</b> | <b>Collapsed steps</b> | <b>Evaluate the event</b> |

WhyDive evaluates reasoning events, not people.

# The Reasoning Event

WhyDive evaluates reasoning events.

This distinction is foundational. The framework does not evaluate people as whole persons. It does not evaluate identity, intelligence, ideology, group membership, political affiliation, or moral worth. It evaluates discrete instances in which evidence is used to support a conclusion.

A reasoning event is the primary unit of analysis within the WhyDive framework. More precisely, it is the fundamental evaluative unit on which the architecture operates: authorization, boundary analysis, violation detection, repair, development, and internalization all begin with a reasoning event. Figure 2-A demonstrates the reasoning event parts.

Each part matters.

**Evidence** is the information available to support, challenge, constrain, or inform a conclusion. It may include observations, measurements, charts, testimony, records, documents, research findings, or AI-generated outputs. Evidence is input. It is not automatically meaning, conclusion, judgment, or decision.

**Figure 2-A** Reasoning Event Parts



**Interpretation** is the process of assigning meaning to evidence. A chart may show that sales increased by 12 percent. That observation does not interpret itself. One person may interpret it as evidence that performance improved. Another may interpret it as a temporary fluctuation. A third may ask whether the comparison period was unusual. Interpretation begins the movement from information toward meaning.

**Inference** is the movement from interpreted evidence toward a proposed conclusion. If the evidence is that sales rose 12 percent, and the interpretation is that performance improved, an inference

# The Reasoning Event

WhyDive evaluates discrete reasoning events, not persons, identities, intelligence, ideology, or moral worth.

| INPUT                  | QUESTION                  | PRIMARY RISK           | GOVERN RULE              |
|------------------------|---------------------------|------------------------|--------------------------|
| <b>Reasoning Event</b> | <b>What is evaluated?</b> | <b>Person-labeling</b> | <b>Event, not person</b> |

## Evidence

Input available to support, challenge, constrain, or inform a conclusion.



## Interpretation

The evidence is assigned meaning.



## Inference

Movement from interpreted evidence toward conclusion.



## Conclusion

The claim produced through reasoning.



## Commitment

The degree to which the conclusion is relied upon or used.

might be that a new marketing strategy contributed to the improvement. That inferential move may or may not be supported. Inference bridges evidence and conclusion, but it is not proof. It may preserve alternatives or close them prematurely. It may stay within what evidence supports or move beyond it.

**Conclusion** is the claim produced through reasoning about what the evidence supports. Conclusions may be tentative, moderate, strong, qualified, or uncertain. A conclusion may exist before a person is ready to rely on it strongly.

**Commitment** is the degree to which the conclusion is relied upon, communicated, defended, recommended, prioritized, or used to guide action. Commitment is where proportionality becomes visible. Multiple people may reach the same conclusion while holding different levels of commitment to it. One may treat “the treatment may help” as a tentative possibility; another may treat the same conclusion as grounds for strong recommendation. WhyDive is especially concerned with whether the degree of commitment remains proportionate to evidential support.

This structure matters because many reasoning failures occur when these elements collapse into one another. Evidence is treated as if it already contains interpretation. Interpretation is treated as

if it were conclusion. Inference is treated as proof. Conclusion is treated as decision. Confidence is treated as authorization. WhyDive separates these movements so they can be evaluated.

Consider the following abstract pattern:

- **Evidence:** A relationship appears between A and B.
- **Interpretation:** A and B may be connected.
- **Inference:** A may influence B.
- **Conclusion:** A is likely one contributing factor.
- **Commitment:** Further investigation is warranted.

This reasoning event is cautious. It preserves uncertainty and limits application.

Now compare a different pattern:

- **Evidence:** A relationship appears between A and B.
- **Interpretation:** A and B are connected.
- **Inference:** A causes B.
- **Conclusion:** A is the reason B occurred.
- **Commitment:** Immediate broad action is required.

The same initial evidence may now be carrying a much stronger conclusion and a much stronger action commitment. Whether that movement is authorized depends on the evidence and the support it provides. WhyDive evaluates that relationship.

Reasoning events allow the framework to remain structurally disciplined. A person may reason well in one event and poorly in another. A person may make a cautious inference in one domain and an overextended inference in another. WhyDive does not need to reduce the person to a label. It can evaluate the event.

This also protects the parent-framework nature of WhyDive. In education, the reasoning event may involve a student's interpretation of a chart. In business, it may involve a team interpreting customer data. In public policy, it may involve officials interpreting an evaluation report. In AI evaluation, it may involve a model's generated explanation. The domain changes. The unit of analysis remains the reasoning event.

Once the reasoning event is visible, WhyDive can ask its central question:

What does the evidence authorize?

## 03

# Authorization

This section names the central evaluative mechanism of the framework.

|                      |                                      |   |                                 |
|----------------------|--------------------------------------|---|---------------------------------|
| MECHANISM            | QUESTION                             | RISK                                      | RULE                            |
| <b>Authorization</b> | <b>What does evidence authorize?</b> | <b>Truth/<br/>confidence<br/>collapse</b> | <b>Support, not omniscience</b> |

Authorization evaluates whether conclusion and commitment remain proportionate to support.

# Authorization

Authorization is the central evaluative mechanism of WhyDive.

It asks:

What does the evidence authorize?

Authorization is the evaluation of whether a conclusion and the commitment attached to it remain proportionate to available evidential support. It concerns support, scope, confidence, and application. It evaluates reasoning events, not people.

The authorization question is distinct from several nearby questions. It is not the same as asking whether a conclusion is true, whether a person agrees with it, whether it is popular, whether experts believe it, or whether the reasoner feels confident. Those questions may matter in other contexts, but they are not the WhyDive question. WhyDive asks what conclusions, and what levels of commitment to those conclusions, are proportionate to the support available.

Authorization evaluates the relationship among evidence, conclusion, and commitment. That relationship is shown in Figure 3.

**Figure 3** Parts Authorization Evaluates



Authorization evaluation is not binary in a simplistic sense. Evidence may authorize a cautious possibility but not a confident assertion. It may authorize a limited conclusion but not a general rule. It may authorize further investigation but not implementation. It may authorize action in one context but not another.

This is why authorization operates across dimensions.

First, authorization concerns conclusion strength. A claim of possibility requires less support than a claim of probability. A claim of near certainty requires stronger support still. The stronger the conclusion, the stronger the support required.

Second, authorization concerns scope. Evidence from one setting may support

# What Does the Evidence Authorize?

Authorization is the framework's central evaluative mechanism: it evaluates the relationship among evidence, conclusion, and commitment.

## NOT THE SAME AS

**Truth, agreement, confidence, popularity, or correctness**

Those questions may matter elsewhere, but they are not the WhyDive question.

## WHYDIVE QUESTION

**What conclusions and commitments are proportionate to available support?**

Evidence may authorize a cautious possibility but not a confident assertion.

FRAMEWORK  
ANCHOR

**A conclusion may later prove true and still have been unauthorized when first made.**

a conclusion about that setting without supporting a conclusion about all settings. Evidence from one classroom does not automatically authorize a claim about all schools. Evidence from one market does not automatically authorize a claim about all markets.

Third, authorization concerns confidence. Confidence should remain proportionate to support. Preliminary evidence may authorize tentative confidence. It may not authorize certainty. Stronger evidence may authorize stronger confidence, but confidence itself is not authorization.

Fourth, authorization concerns application. Evidence may authorize discussion, inquiry, monitoring, or further investigation without authorizing major action. Application carries its own authorization burden because action magnifies consequences.

Authorization must also be distinguished from truth. A conclusion may eventually prove true even if it was not authorized when first made. A lucky guess may be correct. Correctness after the fact does not prove the original reasoning was proportionate to the evidence available at the time. Conversely, a conclusion may be responsibly authorized and still later turn out incomplete or mistaken because the available evidence was limited.

Authorization is also distinct from confidence. A person can be highly

confident with weak support, or cautious despite strong support. WhyDive does not treat confidence as self-validating. It evaluates whether confidence is warranted by support.

Authorization is distinct from agreement. Consensus does not automatically determine authorization. A popular conclusion may be unsupported. An unpopular conclusion may be supported. WhyDive remains evidence-centered.

Authorization is related to justification, but the two are not identical. Justification often concerns why a person believes something. Authorization concerns what the evidence supports. A person may have reasons for believing a conclusion, but WhyDive asks whether those reasons provide support strong enough for the commitment being made.

This distinction is essential to the framework's discipline. WhyDive does not claim access to final reality. It does not claim that authorization proves truth. It claims that reasoning events can be evaluated by whether their conclusions and commitments remain within the support available.

Authorization creates the next architectural need. If evidence authorizes some conclusions and not others, then support has edges. Those edges are boundaries.

# 04

## Boundary System

This section shows how authorization creates visible limits.

|                   |                                |                    |                                 |
|-------------------|--------------------------------|--------------------|---------------------------------|
| SYSTEM            | QUESTION                       | RISK               | RULE                            |
| <b>Boundaries</b> | <b>Where does support end?</b> | <b>Unseen edge</b> | <b>Boundaries reveal limits</b> |

Boundaries identify where evidential support ends.

# The Boundary System

Authorization asks what the evidence supports. Boundaries identify where that support ends.

A boundary is not a restriction. It is a recognition of the point beyond which a conclusion, confidence level, inference, or application is no longer authorized by available evidence. Boundaries do not prevent reasoning. They discipline reasoning by making support visible.

Without authorization, no boundary exists. Without boundaries, no violation can be identified. The boundary system operationalizes authorization by showing where support exists, where it weakens, and where it stops.

WhyDive currently recognizes four primary boundary classes. These are architectural categories, not a claim that empirical research has proven these are the only possible boundaries:

- Evidence Boundary
- Inference Boundary
- Confidence Boundary
- Application Boundary

The Evidence Boundary asks:

What information is actually available?

This boundary is created by the available information itself. In that sense, it is foundational: before a framework can ask what follows from evidence, it must ask what evidence is actually present. It distinguishes what is known from what is not known. A reasoning event crosses the Evidence Boundary when it treats absent, assumed, incomplete, or unexamined information as though it were present. For example, a single survey may provide useful information. It does not automatically provide replication, longitudinal evidence, or evidence from alternative populations. The Evidence Boundary prevents reasoning from smuggling absent evidence into the support base.

The Inference Boundary asks:

What follows from the evidence?

# Where Support Ends

Boundaries identify the point beyond which a conclusion, confidence level, inference, or application is no longer authorized by available evidence.

## BOUNDARY 1

### Evidence Boundary

What information is actually available?

## BOUNDARY 2

### Inference Boundary

What follows from the evidence?

## BOUNDARY 3

### Confidence Boundary

How certain should we be?

## BOUNDARY 4

### Application Boundary

How far may this conclusion be carried?

## GOVERN RULE

**Boundaries do not restrict thought. They make support visible.**

This boundary limits the conclusions that can reasonably be derived from available evidence. Evidence may show that two variables are related. That does not automatically authorize a causal claim. Evidence may make one explanation plausible. That does not automatically eliminate alternatives. The Inference Boundary separates supported conclusions from unsupported inferential leaps.

The Confidence Boundary asks:

How certain should we be?

This boundary limits the level of certainty authorized by available support. A preliminary observation may authorize a tentative conclusion. It may not authorize near certainty. This is not primarily a psychological claim about how confident a person feels. It is an evidential claim about how much confidence the support warrants. This boundary is adjacent to calibration research, but WhyDive uses it as part of its own authorization architecture. The question is not merely whether confidence matches later outcomes, but whether confidence is proportionate to present support.

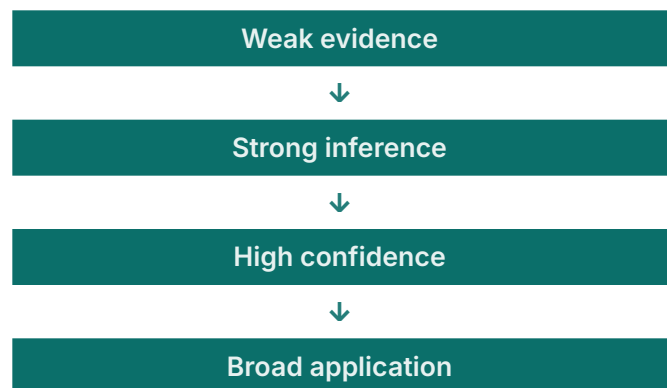
The Application Boundary asks:

How far may this conclusion be carried?

This boundary limits the actions, recommendations, decisions, or contexts to which a conclusion may be extended. Evidence may authorize further investigation but not immediate large-scale change. Evidence may authorize a local recommendation but not a national policy. The Application Boundary matters because judgment becomes consequential through decisions and actions.

Boundaries often interact. A reasoning event may begin with weak evidence, move through a strong inference, attach high confidence, and then support broad application. In that event, multiple boundaries may be crossed at once, as illustrated in Figure 4.

**Figure 4** Boundary Interactions



This could involve an Evidence Boundary problem, an Inference Boundary problem, a Confidence Boundary problem, and an Application Boundary problem simultaneously.

Boundaries are not always cliff edges. Support may weaken gradually. A future implementation may distinguish clearly authorized zones, questionable zones, boundary zones, and unauthorized zones. Whitepaper #2 does not validate such a classification. It identifies the architectural need: support has limits, and disciplined reasoning requires those limits to remain visible.

Boundary visibility is central to WhyDive. Many reasoning failures occur because people notice evidence and conclusion while overlooking where support stopped.

WhyDive restores visibility to the edge between what is supported and what is not.

The purpose of the Boundary System is not to eliminate uncertainty. The purpose is to ensure that uncertainty remains visible. When uncertainty remains visible, reasoning becomes more disciplined, judgment becomes more proportional, and commitment becomes more accountable.

When a reasoning event crosses a boundary, the framework can identify a violation.

## 05

# Violation Taxonomy

This section translates boundary crossing into a disciplined vocabulary of structural failures.

| CONDITION                            | ORGANIZER               | RISK                   | RULE                          |
|--------------------------------------|-------------------------|------------------------|-------------------------------|
| <b>Commitment &gt; authorization</b> | <b>Boundary crossed</b> | <b>Moralized error</b> | <b>Classify the structure</b> |

A violation is a structural feature of a reasoning event.

# Violation Taxonomy

A violation occurs when commitment exceeds authorization through one or more boundary crossings.

Formally:

Commitment > Authorization

A violation is not disagreement. It is not ideological difference. It is not moral failure. It is not low intelligence. It is not even necessarily incorrectness. A conclusion may later prove correct while still having been unauthorized by the evidence available at the time. In WhyDive, a violation is a structural feature of a reasoning event.

The Violation Taxonomy organizes failures by the boundary crossed. A boundary is the limit of support; a violation is the reasoning event's movement beyond that limit.

- Evidence Violations
- Inference Violations
- Confidence Violations
- Application Violations

**Evidence Violations** involve misuse, absence, distortion, or assumption of evidence. Evidence Substitution occurs

when assumed evidence replaces available evidence. Evidence Inflation occurs when limited evidence is treated as extensive evidence. Evidence Omission occurs when relevant evidence that materially affects authorization is ignored.

**Inference Violations** involve unsupported movements from evidence toward conclusion. An Inference Leap occurs when a conclusion exceeds what follows from the evidence. Unsupported Causation occurs when association, sequence, or relationship is treated as proof of causation. Explanation Inflation occurs when one plausible account is elevated into the definitive or only explanation without sufficient support. Premature Closure occurs when evaluation stops before reasonable alternatives are considered.

**Confidence Violations** involve certainty that exceeds support. Overcertainty occurs when confidence exceeds authorized evidential support. False Precision occurs when uncertainty is presented with more exactness than evidence allows. Certainty Escalation occurs when confidence increases without a corresponding increase in evidence.

# Overclaiming Is a Family of Structural Failures

A violation occurs when commitment exceeds authorization through one or more boundary crossings.

## FAMILY 1

### Evidence Violations

Substitution, inflation, omission.

## FAMILY 2

### Inference Violations

Leap, unsupported causation, explanation inflation, premature closure.

## FAMILY 3

### Confidence Violations

Overcertainty, false precision, certainty escalation.

## FAMILY 4

### Application Violations

Overgeneralization, application drift, premature implementation, institutional

FORMULA

**Violation = Commitment > Authorization**

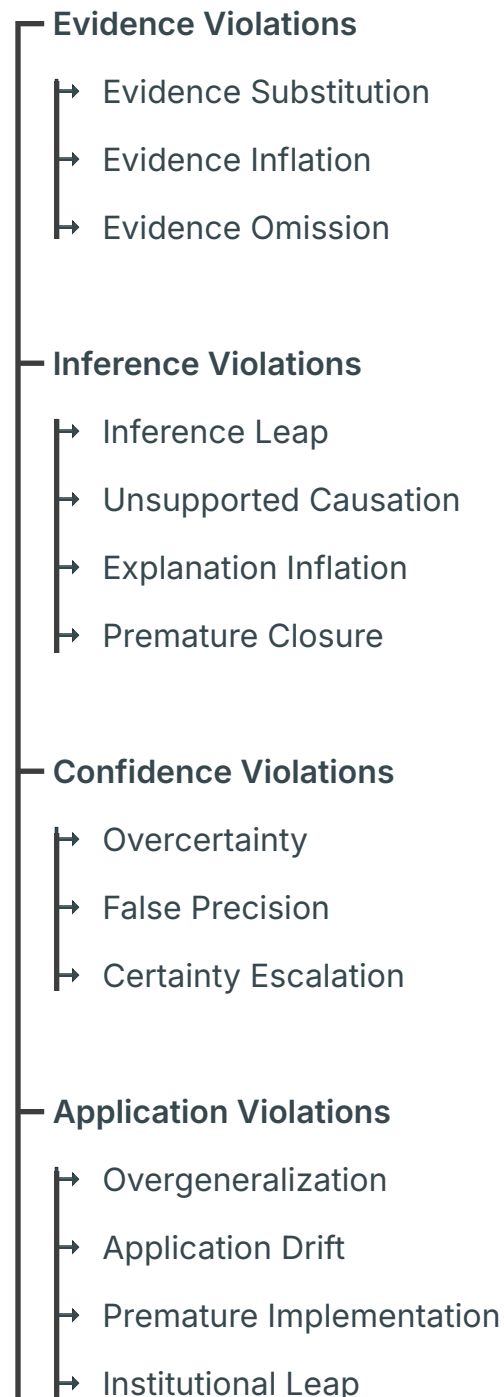
**Application Violations** involve carrying conclusions farther than evidence authorizes. Overgeneralization occurs when a conclusion is applied beyond the population, context, or conditions supported by evidence. Application Drift occurs when a conclusion is used in contexts not authorized by the available support. Premature Implementation occurs when evidence supports further inquiry but is used to justify stronger action. Institutional Leap occurs when limited findings are moved into broad organizational, policy, or system-level action without sufficient support.

The taxonomy can be represented as in Figure 5-A.

Overclaiming is therefore not one thing. It is a family of structural failures in which commitment exceeds support. The taxonomy allows WhyDive to describe the form of the failure rather than merely saying that reasoning went wrong.

Most real-world violations are compound. A person may inflate weak evidence, make a strong inference, express high confidence, and apply the conclusion broadly. The framework may identify multiple violations within the same reasoning event. This matters because repair depends on the type of violation. A confidence violation is not repaired in exactly the same way as an evidence violation. An application violation is not

**Figure 5** Violation



repaired in exactly the same way as an inference violation.

WhyDive should not initially classify violations as minor, moderate, or severe without reliability evidence. Severity may become part of a later validated assessment system. At the architectural level, the first task is to identify violation type. This preserves the framework's research posture and avoids pretending that unvalidated scoring categories already exist.

The distinction between violation and correctness is especially important. If a person makes an unsupported guess and the guess later proves correct, the original reasoning event may still contain a violation. WhyDive evaluates whether the conclusion was authorized by the evidence available when the reasoning occurred. It does not treat later luck as evidence of earlier proportionality.

Violations are not endpoints. They are signals that a reasoning event exceeded authorized support. The purpose of identifying them is repair.

# 06

## Repair

This section shows how WhyDive moves from diagnosis into development.

|                         |  |                                |                                       |
|-------------------------|--|--------------------------------|---------------------------------------|
| ENGINE<br><b>Repair</b> | SEQUENCE<br><b>Recognize → stabilize</b> | RISK<br><b>Assessment only</b> | RULE<br><b>Repair restores bounds</b> |
|-------------------------|--|--------------------------------|---------------------------------------|

Repair restores proportionality between support and commitment.

# Repair

The purpose of violation identification is not punishment. The purpose is repair.

If WhyDive stopped at classification, it would be a diagnostic framework. It could identify where reasoning exceeded support, but it would not explain how reasoning becomes more disciplined. Whitepaper #2 defines WhyDive as a Bounded Reasoning Development System. Development requires a pathway from violation to improved capability. That pathway is repair.

Repair is the process through which a reasoning event is restored to authorized boundaries and future reasoning becomes more disciplined.

The current repair sequence contains five stages. This sequence is architectural rather than a validated claim that every repair process must unfold in exactly the same order as shown in Figure 6-A.

Recognition asks:

What happened?

The first step is to identify the violation. Instead of treating the problem merely as a wrong answer or bad conclusion, WhyDive names the structural issue. The

**Figure 6-A** Repair Stage Order



event may involve overgeneralization, unsupported causation, overcertainty, evidence omission, or another violation type. Recognition gives the problem form.

Reflection asks:

Why did it happen?

Reflection is not self-criticism. It is reasoning examination. It asks what evidence was present, what evidence was missing, what alternatives remained possible, where certainty increased, and where application expanded. Reflection helps the reasoner see how the reasoning

# Violations Are Not Endpoints

Repair restores a reasoning event to authorized boundaries and supports future reasoning discipline.

## Violation

The reasoning event exceeds authorized support.



## Recognition

What happened?



## Reflection

Why did it happen?



## Correction

How should the reasoning change?



## Reevaluation

Does the repaired reasoning remain authorized?



## Stabilization

Can the repaired reasoning be demonstrated consistently?

event crossed a boundary.

Correction asks:

How should the reasoning  
change?

Correction restores the reasoning event to authorized boundaries. Evidence violations may require adding or revisiting evidence. Inference violations may require revising the conclusion or reopening alternatives. Confidence violations may require reducing certainty. Application violations may require narrowing scope or changing the recommended action.

For example:

**Original:**

This proves the program works.

**Corrected:**

This provides preliminary support that the program may be effective.

The correction does not merely soften language for politeness. It restores proportionality between support and commitment.

Reevaluation asks:

Does the repaired reasoning  
remain authorized?

Repair is not simply changing words. A revised statement must be checked

against the evidence, authorization, and boundaries. The reasoning event is repaired only if the conclusion, confidence, scope, and application now remain within support.

Stabilization asks:

Can the repaired reasoning be  
demonstrated consistently?

One correction is not development. A person may repair one reasoning event under guidance and still repeat the same violation later. Stabilization is the bridge from event repair to capability growth. It asks whether the repaired pattern can be demonstrated again, under relevant variation, until the discipline becomes less dependent on the original prompt or evaluator. Stabilization moves from isolated correction toward durable capability.

Repair pathways differ by violation family. Evidence violation repair may identify missing evidence, review available evidence, reexamine the conclusion, and reassess authorization. Inference violation repair may examine the inferential move, consider alternatives, revise the conclusion, and reassess authorization. Confidence violation repair may evaluate certainty, examine support, adjust confidence, and reassess authorization. Application violation repair may examine scope, narrow context, revise recommendation, and reassess

authorization.

Repair transforms the framework from evaluation to development as shown in Figure 6-B.

**Figure 6-B** From Evaluation to Development



This is also where WhyDive remains distinct from assessment alone.

Assessment identifies a current state. Repair supports a future state. WhyDive should never be reduced to assessment, scoring, or reporting. Those may become domain-specific tools, but the parent architecture is developmental.

Repair is not the same as making a conclusion correct. A repaired conclusion must be authorized. A person may move from unauthorized but correct to authorized and correct, or from unauthorized certainty to authorized uncertainty. Both may represent

successful repair if the revised reasoning event better matches the evidence.

Violations are therefore developmental opportunities. The purpose of repair is not merely to restore a single event. It is to increase the individual's ability to remain within authorized reasoning boundaries in future reasoning events.

The bridge from event repair to capability development is the next architectural step.

## 07

# Development

This section shows how WhyDive moves from diagnosis into development.

|                                   |                                      |                                 |   |
|-----------------------------------|--------------------------------------|---------------------------------|---|
| PROGRESSION<br><b>DB-1 → DB-8</b> | QUESTION<br><b>How consistently?</b> | RISK<br><b>Age/status proxy</b> | RULE<br><b>Capability is demonstrated</b> |
|-----------------------------------|--------------------------------------|---------------------------------|---|

Development is demonstrated boundary discipline across increasingly complex evidence environments.

# Development

WhyDive is not only a reasoning evaluation system. It is a development system.

Development, in WhyDive, is the progressive strengthening of an individual's ability to remain within authorized reasoning boundaries across increasingly complex evidence environments.

The developmental question is:

How consistently can a person remain within authorized reasoning boundaries?

This is different from asking how much information a person knows, how many correct answers a person produces, how intelligent a person is, or how confident a person feels. A person may know many facts and still overclaim. A person may occasionally produce correct answers without stable reasoning capability. A person may feel confident while repeatedly exceeding support.

WhyDive development rests on five governing principles.

First, development is earned. Progression is based on demonstrated capability, not

age, grade, experience, title, credential, or status alone. This principle helps protect the framework from assuming that social category or institutional placement automatically indicates reasoning discipline.

Second, development is gated. Higher capabilities depend on lower capabilities. A person cannot reliably manage complex uncertainty while consistently failing to recognize missing evidence. A person cannot sustain self-repair while violations remain invisible. The gating principle is informed by developmental traditions, but the WhyDive band architecture remains a WhyDive proposition.

Third, development is non-compensatory. One strength does not erase missing foundations. Strong interpretation cannot compensate for persistent overcertainty. Strong evidence awareness cannot compensate for repeated application drift. Reasoning events are structurally interconnected. Missing foundations create recurring instability.

Fourth, development requires stability. The framework rejects:

One success = Mastery

Peak performance is not the same as stable capability. This distinction is informed by the Fischer/Siegler neighborhood while remaining part of the WhyDive architecture. A person may demonstrate disciplined reasoning once under ideal conditions. Development requires more reliable performance across multiple reasoning events.

Fifth, development leads toward internalization. The endpoint is not classification. The endpoint is self-regulating judgment.

The framework currently uses DB-1 through DB-8 as developmental bands. These bands are WhyDive constructs. They are not externally validated developmental stages. This limitation must remain visible. The bands provide an architectural language for describing increasing boundary visibility, authorization discipline, repair capability, and self-regulation.

At a high level:

**DB-1:** Reasoning largely reacts to visible information. Boundary awareness is minimal. External support is required.

**DB-2:** The person begins recognizing support and contradiction but remains heavily dependent on guidance.

**DB-3:** The person can identify some limitations. Violations become more visible after prompting.

**DB-4:** Independent boundary recognition begins. Authorization judgments become more deliberate.

**DB-5:** The person can evaluate competing interpretations and manage uncertainty with greater discipline.

**DB-6:** Authorization discipline becomes more consistent. Repair is increasingly self-directed.

**DB-7:** Self-monitoring is strong. Boundary recognition is frequently self-initiated, though not assumed to be complete or universal.

**DB-8:** Reasoning is highly disciplined across the kinds of evidence environments the framework is evaluating. Authorization, boundary awareness, and repair are largely self-regulating, while still remaining subject to error and future repair.

Whitepaper #2 should not fully operationalize every band. That belongs in later specification and validation work. The purpose here is to explain the developmental architecture.

The bands are not meant to float above the rest of the architecture. They describe progression across the same capabilities the framework has already introduced. At this architectural level, development occurs across several dimensions: evidence awareness, inference discipline, confidence calibration, boundary recognition, repair capability, and self-

# Development Is Increasing Boundary Discipline

The DB bands are WhyDive constructs for describing progression. They are not externally validated developmental stages.

|   |   |   |  |
|---|---|---|--|
| <b>DB-1</b><br>External support required.                         | <b>DB-2</b><br>Support and contradiction begin to appear. | <b>DB-3</b><br>Limitations visible after prompting.     | <b>DB-4</b><br>Independent boundary recognition begins.        |
| <b>DB-5</b><br>Competing interpretations and uncertainty managed. | <b>DB-6</b><br>Repair becomes more self-directed.         | <b>DB-7</b><br>Self-monitoring is frequently initiated. | <b>DB-8</b><br>Boundary discipline is largely self-regulating. |

**BOUNDARY NOTE**

**DB bands are architecture propositions, not validated developmental stages.**

regulation. These dimensions are not merely academic categories. They describe the capabilities required for a person to remain within authorized reasoning boundaries when evidence becomes more complex, uncertain, contested, or consequential.

Development is also defined by a person's relationship to violations. Lower developmental levels may involve less visible boundaries and externally driven repair. Higher developmental levels involve more visible violations and increasingly self-directed repair. Development is not the absence of mistakes. It is a changing relationship to mistakes, boundaries, and repair.

Repair is the engine of development, as shown in Figure 6:

**Figure 6** Repair Engine



Without repair, no developmental movement is assumed.

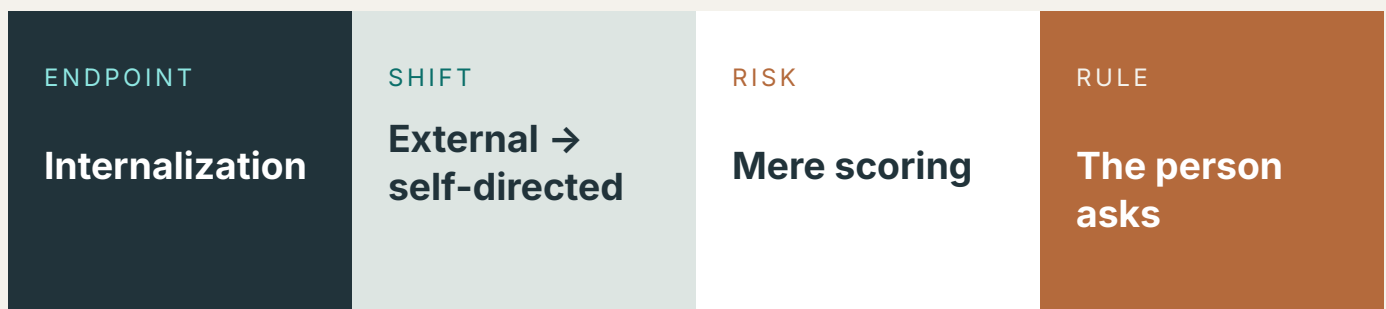
The architecture may support cross-domain expression, but cross-domain transfer remains a validation question. WhyDive may hypothesize that disciplined authorization, boundary recognition, and repair can appear across domains. It should not claim that transfer has been established. That distinction is essential to the research posture of the framework.

Development culminates in a shift from external boundary recognition to internal boundary recognition, and from external repair to self-repair. That transition leads to internalization.

## 08

# Internalization

This section shows how WhyDive moves from diagnosis into development.



At advanced stages, the individual asks the authorization questions.

# Internalization

Internalization is the endpoint toward which the WhyDive Development System is directed.

The core question is:

What happens when authorization, boundary awareness, and repair no longer require external prompting?

At early stages, the framework asks the questions. At advanced stages, the individual asks the questions. That transition defines internalization.

Internalization is the state in which authorization awareness, boundary discipline, and repair habits become self-regulating rather than externally directed. It is the internal adoption of the discipline; self-regulating judgment is the judgment posture that results when those disciplines begin operating without external prompting. Internalization does not mean memorizing vocabulary, complying with procedures, agreeing with someone else's conclusions, or reasoning perfectly. Internalized thinkers still make mistakes, miss evidence, cross boundaries, and require repair. What changes is how

violations are recognized and addressed.

The framework does not exist merely to produce scores, reports, classifications, or band placements. These may be useful in downstream contexts, but they are not the destination. The destination is a change in how people approach evidence, conclusions, judgment, and action.

The internalization shift can be described in stages as shown in Figure 8-A.

In early development, questions come from outside:

- What evidence supports that?
- What evidence is missing?
- Are you too certain?
- Does that actually follow?

In transitional development, these questions begin appearing with prompting. The person increasingly recognizes uncertainty, alternatives, and boundary concerns, but still requires support.

In advanced development, the questions arise without prompting:

- What does the evidence support?
- What does it not support?

**Figure 8-A** Internalization Stages

- Where does support stop?
- What uncertainty remains?

In internalized development, the questions become habitual. They become part of the person's reasoning process.

WhyDive currently identifies several internalized habits. Evidence awareness asks what evidence is present and what evidence is absent before strong conclusions form. Boundary awareness asks where support stops before conclusions expand. Alternative consideration asks what else might explain the evidence before premature closure. Confidence discipline asks how certain one should be rather than how certain one feels. Application discipline asks what action the evidence actually supports. Self-repair asks whether the reasoning

event exceeded evidence before an external evaluator must intervene.

Internalization changes the relationship to violations:

#### Early stages:

Violation → External recognition → Repair

#### Internalized stages:

Emerging violation → Self-recognition  
→ Self-repair

The framework becomes increasingly internal.

Internalization also clarifies the purpose of developmental bands. The bands describe progression. Internalization describes destination. DB-8 is not the endpoint by itself. It is the closest architectural neighbor to internalization, but the deeper goal is the self-regulating judgment that internalization makes possible.

Internalization requires stable capability, not occasional performance. A person who sometimes demonstrates disciplined reasoning has not necessarily internalized the framework. Internalization implies consistency across reasoning events. This remains aligned with the developmental emphasis on stability while avoiding the claim that internalization has been empirically validated.

# From External Questions to Self-Regulating Judgment

Internalization occurs when authorization awareness, boundary discipline, and repair habits become self-regulating rather than externally directed.

## 01 External Questions

The framework or evaluator asks the questions.

## 02 Guided Reflection

The reasoner responds to prompts and begins seeing boundaries.

## 03 Independent Reflection

The questions begin to appear without constant prompting.

## 04 Self-Monitoring

Emerging violations are noticed before external intervention.

## 05 Internalization

The discipline becomes habitual, though still fallible.

### ANCHOR CLAIM

**At early stages, the framework asks the questions. At advanced stages, the individual asks the questions.**

Internalization does not automatically establish cross-domain transfer. The framework may hypothesize that internalized reasoning disciplines can be expressed across evidence environments. Whether that occurs, how reliably it occurs, and under what conditions it occurs remain validation questions.

Whitepaper #1 argued that reasoning exists in service of judgment. Internalization extends that principle. The ultimate goal is not merely better reasoning events. The goal is more disciplined judgment, because judgment influences decisions, actions, and consequences.

The framework should also be careful about human flourishing. WhyDive may say that internalization seeks to support more disciplined judgment. It should not claim

that internalization produces flourishing. Human flourishing involves values, character, purpose, relationships, culture, and community beyond the scope of this framework.

WhyDive ultimately seeks more than correct answers. It seeks the development of individuals who habitually ask what evidence supports, what evidence does not support, where support ends, what uncertainty remains, and what commitment is justified before judgment becomes action.

When those disciplines become self-regulating rather than externally imposed, the framework's developmental purpose has been realized.

# Research Posture and Validation Agenda

This section protects proportionality by separating architecture from proof.

| STATUS                   | LIMIT                | NEXT WORK                     | RULE                 |
|--------------------------|----------------------|-------------------------------|----------------------|
| <b>Research-informed</b> | <b>Not validated</b> | <b>Reliability / validity</b> | <b>Same standard</b> |

Research-informed is not the same thing as validated.

# Research Posture and Validation Agenda

A framework committed to proportionality must apply proportionality to itself.

Whitepaper #2 therefore ends by distinguishing what WhyDive may responsibly claim from what remains unvalidated. This is not a secondary concern. It is part of the architecture. If WhyDive exists to discipline the relationship between evidence and commitment, then WhyDive must not claim more for itself than the evidence supports.

Three categories must remain visible.

First, WhyDive is a research-informed framework. It is informed by neighboring literatures concerned with reasoning, judgment, decision-making, epistemic cognition, calibration, developmental progression, skill development, self-regulation, and transfer. These literatures provide important context and support for the problems WhyDive addresses. They also help locate WhyDive within a broader intellectual landscape.

Second, WhyDive contains framework propositions. Authorization, the Boundary System, the Violation Taxonomy, the Repair Protocol, the Developmental Bands, and the Internalization Model

are WhyDive constructs. They may be informed by research, but they should not be presented as established findings from those literatures. They are architectural propositions.

Third, WhyDive has validation questions. These questions must remain open until empirically examined.

Examples include:

- Does WhyDive improve judgment?
- Does WhyDive improve decision quality?
- Does WhyDive improve educational outcomes?
- Does WhyDive transfer across domains?
- Do DB-1 through DB-8 accurately model developmental progression?
- Can evaluators apply the framework reliably?
- Does internalization produce measurable changes in judgment quality?

Whitepaper #2 must not claim that these questions have already been answered.

It must not claim that WhyDive works, improves judgment, improves decision quality, improves outcomes, transfers across domains, or has validated developmental bands. Those claims would exceed the current architecture and violate the framework's own proportionality discipline.

This distinction matters for scholarly credibility. A skeptical reader may accept the architectural clarity of the framework while still asking whether it can be reliably applied, whether it predicts growth, whether it transfers, and whether it improves outcomes. WhyDive should welcome those questions rather than avoiding them. They are part of the validation agenda.

The same distinction matters for parent-framework purity. WhyDive Education, Data Literacy Standards, Chart-Ed, business applications, policy applications, AI evaluation, and other future domains may generate empirical programs of their own. Those programs may test specific implementations, populations, use cases, and outcomes. But domain validation should not be retroactively treated as validation of every claim in the parent framework. Evidence must remain proportionate to the claim being made.

Whitepaper #2 is therefore not a validation study. It is not a product proof. It is not a claim of demonstrated effectiveness. It is

the architecture document for a research-informed framework.

Its legitimate claim is:

**Here is the conceptual architecture WhyDive proposes for evaluating and developing reasoning under evidence constraints.**

Its illegitimate claim would be:

**This architecture has been proven to improve judgment.**

The first claim belongs in Whitepaper #2. The second belongs only after future validation.

This posture does not weaken WhyDive. It strengthens it. The framework's credibility depends on the same discipline it asks of others: distinguish support from aspiration, proposition from proof, architecture from validation, and possibility from established outcome.

The validation agenda should therefore ask, at minimum:

**Reliability:** Can different evaluators apply authorization, boundary, and violation categories consistently?

**Development:** Do developmental bands describe stable differences in reasoning capability?

**Repair:** Does repeated repair improve future boundary recognition and authorization discipline?

**Internalization:** Can self-monitoring and self-repair be observed over time?

**Transfer:** Do WhyDive disciplines appear across domains, and under what conditions?

**Outcomes:** Do improved authorization and boundary discipline affect judgment, decision quality, or action?

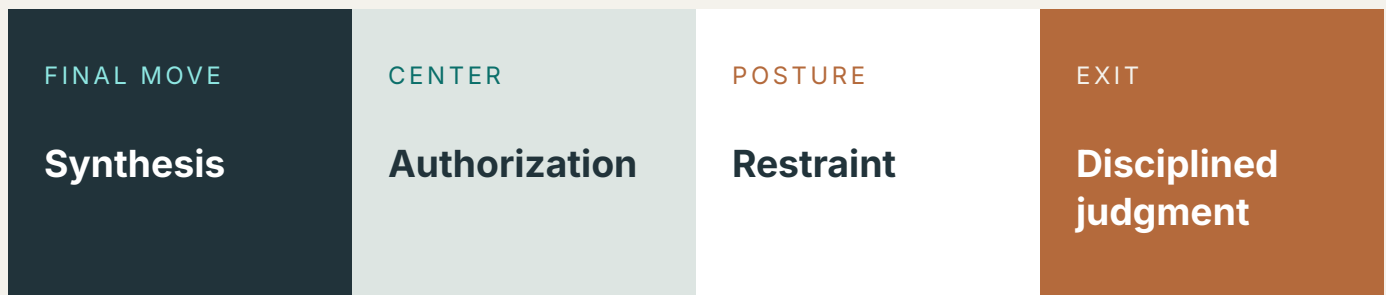
Until such questions are answered, WhyDive remains a research-informed framework with defined architecture and open validation questions.

That is the appropriate posture for Whitepaper #2.

# 10

## Conclusion

This final movement closes the architecture without adding new claims.



What does the evidence authorize?

# Conclusion

Whitepaper #1 answered why WhyDive exists. It argued that human beings move from evidence to action through reasoning, judgment, decision, and consequence, and that this movement becomes dangerous when commitment exceeds support.

Whitepaper #2 has answered a different question:

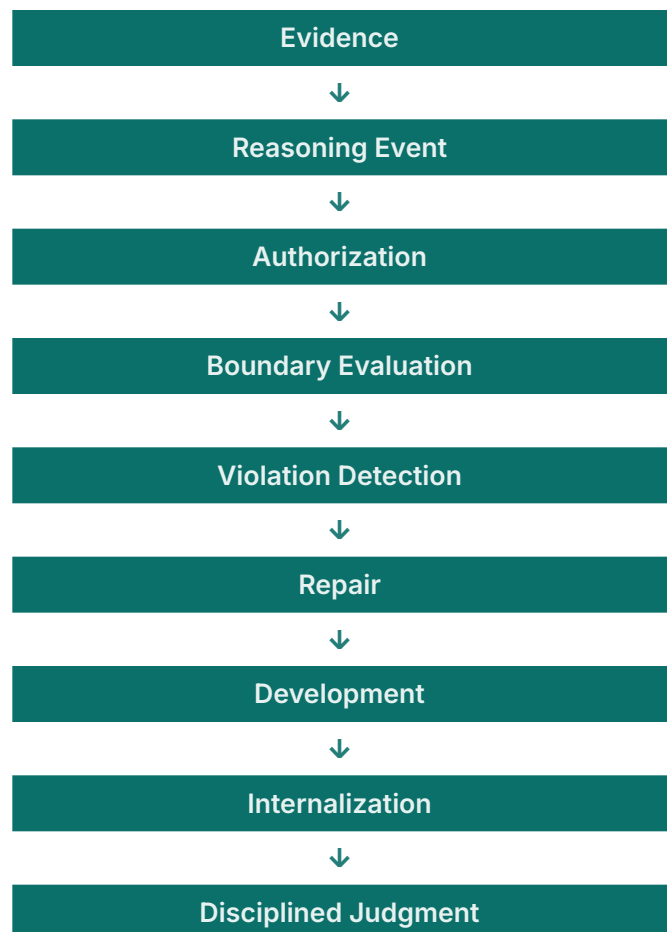
How does WhyDive work?

It works by treating reasoning events as the unit of analysis. It asks what available evidence authorizes. It identifies boundaries where support ends. It classifies violations when commitment crosses those boundaries. It repairs reasoning events by restoring proportionality. It models development as increasing boundary discipline across more complex evidence environments. It defines internalization as the point at which authorization awareness, boundary recognition, and repair become self-regulating.

The architecture can be summarized as in Figure 10.

This is the Bounded Reasoning Development System.

**Figure 10** Bounded Reasoning Development System Architecture



The framework is not a decision system, policy system, worldview system, educational implementation, data literacy initiative, product platform, or validation study. Those may become downstream expressions or future applications.

Whitepaper #2 defines the parent architecture.

The goal of this paper is not to convince the reader that the machine has been proven to work. That would exceed the evidence currently available. The goal is to make the machine understandable.

A reader should now be able to answer:

- What is WhyDive?
- What is a reasoning event?
- What is authorization?
- What are boundaries?
- What is a violation?
- What is repair?
- How does development work?
- What is internalization?
- What remains unvalidated?

If those questions are answerable, the paper has done its work.

WhyDive ultimately seeks disciplined judgment under evidence constraints. It begins with the recognition that strong conclusions require strong evidence. It develops that principle into an architecture for authorization, boundary recognition, violation repair, developmental progression, and internalization.

The work ahead is validation. The work of this whitepaper is architecture.



**WhyDive™** disciplines the movement from evidence to judgment by evaluating what the evidence authorizes.

Strengthening Judgment Under Evidence Constraints  
Living Spiral Studio LLC / [whydive.org](http://whydive.org) / June 2026