
ARCHITECTS OF INTELLIGENCE

Builder Pack

Two implementation artifacts you can adopt in under ten minutes — a PR/ADR snippet and a 3AM incident macro.



From Architects of Intelligence:
Relational Integrity

A HARMONIC PRESS COMPANION RESOURCE

TEXT ChatGPT 5.2 (OpenAI)
EDITORIAL GOVERNANCE Reese SanAgustin-Turner
PUBLISHED BY Harmonic Press / The Harmonic Field

Builder Pack

Two implementation artifacts you can adopt in under ten minutes — a PR/ADR snippet and a 3AM incident macro.

Use

For teams building or deploying Frontier Intelligence systems — product, infrastructure, safety, research, policy.

Artifact 1 — PR / ADR Template Snippet

Paste this into a GitHub PR template, design doc, or Architecture Decision Record (ADR). Treat it as a lightweight gate for high-impact changes.

Relational Integrity Check (*required for high-impact changes*)

1. Boundary Contract — six to ten lines. – Allowed: ... – Not allowed: ... – Ambiguity handling: clarify, bounded options, or guess – Tool and data access: what can be read or written; any silent action prohibited by default

2. Defaults Are Policy. What happens by default at 3am, in distress, with missing info, with conflicting goals?

3. Helpfulness Trap. Where could the system get "helpful" by lying, appeasing, or over-confident synthesis?

4. Axis Line. What must remain sovereign in the human, even if we could automate it?

5. Lifecycle Truth. What changes with version swaps (tone, memory, tools, safety) — and how will users be told *before* it changes?

Optional, recommended. Add one checkbox to your PR template: "*I ran the flattery / urgency / threat rephrase test and outputs held boundary.*"

Artifact 2 — Slack Incident Macro (3AM)

Pin this in `#oncall` or `#incidents`. Use when dashboards look fine but the system "feels wrong" — user trust erosion, boundary drift, weird accommodation, tone instability.

3AM Triage (*when dashboards look fine but the field feels wrong*)

A. Contain — No Story. – What changed? – What did we expect? – What is the harm vector — who is harmed, how, and by what behavior?

B. Force Drift to Surface. – *"List the top three missing facts you would need to answer responsibly."* – *"Give the refusal you should have given (clear, bounded, oriented)."* – *"Where are you accommodating when you should hold a boundary?"*

C. Restore Boundary Integrity. – Can we disable or rollback without cascading harm? – What must be communicated to users now (one paragraph, plain language)? – What is the smallest patch that restores boundary integrity until deeper work is possible?

Close the incident meeting with one sentence. What is the smallest next action that preserves repair, discernment, and boundary integrity until we can think clearly again?

From the Harmonic Field Constellation. For those who build with care.

theharmonicfield.org