CIRCUIT

An Open Standard for AI Interpretability Governance

Circuit-Informed Risk & Control — Understanding, Inventory & Transparency · v1.1.0 · Apache 2.0

CIRCUIT turns AI risk from an assertion into a measurement. It is the first open standard that scores how much an organization can actually see inside the AI models it deploys, records that evidence in an auditable registry, and converts it into a deployment decision a board can read and a pipeline can enforce.

THE PROBLEM

Organizations are deploying AI systems they cannot see inside. The controls most companies rely on — risk tiers, vendor attestations, output monitoring, red-team reports — all operate *around* the model. When something goes wrong, they can show what went in and what came out, but not which part of the model produced the result. The decision happens in a place none of these controls can see.

Regulators have noticed. The EU AI Act's high-risk obligations reach enforcement in August 2026, and frameworks from NIST, ISO, and banking supervisors increasingly demand evidence of *how* a model reaches its decisions — not just statistics on whether the outputs were right. Every major framework now names interpretability as a required evidence class. None provides a way to measure it. **CIRCUIT is that missing instrument.**

THE FRAMEWORK — THREE PARTS THAT WORK AS ONE

1

The Score

INTERPRETABILITY MATURITY SCORE (0–5)

Grades how much real evidence you hold about a model's internal behavior, from a complete black box (0) to continuous automated insight (5). You don't declare a level — you produce the artifacts that prove it.

2

The Registry

EIGHT-SECTION RECORD, ONE PER MODEL

The system of record where every score's evidence lives — bound to a named owner, a documented trail, and a version history. It lets an auditor, regulator, or responder open one entry and see exactly what was known, when, and who was accountable.

3

The Control

CIRCUIT RISK SCORE + TEN RULES

A single formula turns the evidence into a Green / Amber / Red / Purple band that decides what approvals a model needs — or whether it can deploy at all. Ten binding rules give the number teeth.

Remove any layer and the other two stop working: a score with no record is an opinion, a record with no score is shelf-ware, and a rule with no evidence is theater.

HOW TO APPLY IT — A PROGRESSION, NOT A PROJECT

01 FOUNDATION	02 ASSESS	03 OPERATE	04 MATURE
Inventory every AI system and score it. Stop anything too dangerous to run.	Replace assumptions with real evidence on the highest-risk models and vendors.	Move to live monitoring, red-teaming, and the first board-level review.	Automate governance in the pipeline; use the evidence as market leverage.

The framework scales down as well as up — a small team can run the first stage in a spreadsheet in days; a large enterprise runs the same stage as a structured program. The early stages require no new tooling spend.

LEARN MORE

Specification, schema & tooling
github.com/jumpmindinc/circuit-framework

Project home & resources
circuitframework.org

Author
Eric Zielinski, CISO, Jumpmind

Deploy AI you can explain. Defend AI you can inspect. Trust AI you can audit.