

AI risk does not stop at deployment. Neither should evaluation.

What continuous, consequence-aware evaluation asks of a governance committee

Vishal Kumar Srivastava, PhD | 29 October 2026 | Proposal; not validated; no NIST endorsement

The issue in one paragraph

A deployment approval rests on evidence about a particular configuration at a particular time. When the model, prompt, tools, data or permissions change, the evidence may no longer describe the system, and nothing signals it. At the same time, monitoring and independent validation are limited, so someone is deciding, explicitly or not, which systems and which failures receive attention.

What the proposal changes

Question	What R-BED v2.1 asks for
How long is an approval good for?	Evidence carries an owner, a configuration, an expiry and invalidation triggers. The approval cannot outlive the evidence it relies on.
Do the safeguards work?	A safeguard relied on for a high-consequence risk is credited only after it is shown to work under challenge, ideally verified by someone other than its builders.
Where does scarce attention go?	Monitoring effort is directed by the consequence of failure (severity, reversibility, exposure) and by the strength of the evidence, in tiers, with a random audit so that low-ranked items are still sampled.
Who decides?	The approving authority renews, restricts, remediates or stops. It is not the party that operates the system or the monitor.

Decisions this puts in front of a committee

1. Which high-consequence systems or pathways are **unfunded** for monitoring or independent validation this period, and is that acceptable?
2. Which authorizations rely on evidence that is **expiring, past review or invalidated**?
3. Which credited safeguards are **not independently verified**?
4. Which approvals should be **restricted or stopped** until evidence is renewed?

The portfolio report lists these without reducing them to a score.

What it does not do

It does not replace model risk management, security or privacy assessment, or enterprise governance. It is not a monitoring product, a certification or a standard. It prescribes no budget split between pre-deployment validation and monitoring.

What has and has not been shown

- **Specified:** the method, the records and their checks, and a reference implementation whose tests show that the code follows the text.
- **Not shown:** nothing in v2.1 has been empirically validated; in particular, that it improves decisions, that organizations can sustain the effort, or that it is better than alternatives. Cadences and thresholds are illustrative defaults to be justified locally.
- **Testable:** one hypothesis (H7) says allocating by consequence avoids more harm than allocating by frequency. The plan names situations in which it is expected to do *worse*, for example when consequence is misjudged.

A reasonable first step

Choose one deployed system with a high-consequence pathway. Write down which evidence its authorization depends on and when that evidence expires; identify the safeguards the risk argument relies on and whether anyone independent has tested them; and decide what would invalidate the approval. This costs little and shows whether the gaps exist.

Detail: R-BED Public Framework and CSE Specification v2.1. © 2026 Vishal Kumar Srivastava; rights as stated in the Public Framework. **Contact:** vshfrm@gmail.com