

THE MANIFESTO

RegTeaming Manifesto

Red teams test whether AI can be made unsafe.
RegTeams test whether it can cross a regulatory
boundary.

RED TEAMING
Unsafe behavior

REGTEAMING
Unlawful trajectory

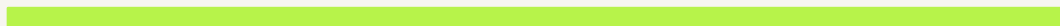
Swipe →



● WHY REGTEAMING EXISTS

AI can behave exactly as designed — and still cross a regulatory boundary.

It may omit a required disclosure, drift into regulated advice, mishandle a vulnerable customer, or produce an unlawful outcome across an otherwise ordinary workflow.



● THE DISTINCTION

RED TEAMING

**Can the system be
made unsafe?**

REGTEAMING

**Can the system be
made unlawful?**

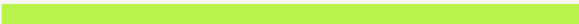
**And can the organization prove it tested for
that possibility?**

01 PRINCIPLE ONE

Test behavior, not declarations.

Policies, model cards, controls, and governance statements express intent. RegTeaming tests whether the deployed system follows that intent under realistic pressure.

INTENT ≠ EVIDENCE



02 PRINCIPLE TWO

Link every test to authority.

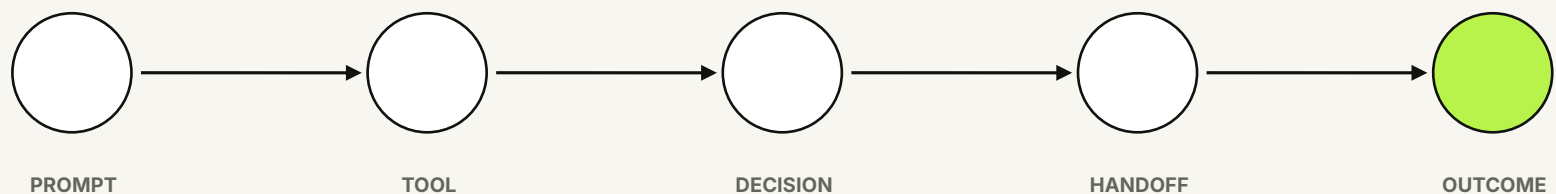
Every evaluation should identify the source, obligation, scope, condition, and expected behavior. A failure without provenance is hard to interpret. A rule without a test is hard to operationalize.

SOURCE → OBLIGATION → TEST → EVIDENCE

03 PRINCIPLE THREE

Test trajectories, not isolated answers.

Regulatory failures often emerge across multiple turns, tool calls, handoffs, and workflow decisions. The unit of evaluation is the complete consequential trajectory.





 PRINCIPLES FOUR AND FIVE

Attack context. Test whether controls actually work.

04 · ATTACK CONTEXT

The obligation changes with context.

Jurisdiction, product, customer vulnerability, language, channel, and timing can all alter what lawful behavior requires.

05 · TEST CONTROL EFFECTIVENESS

A control is not proof.

A disclosure, escalation rule, or human-review policy matters only if it triggers correctly and changes the final outcome.

06 PRINCIPLE SIX

Preserve evidence.

Each evaluation should retain the inputs, context, system version, trace, outcome, governing obligation, and remediation status needed for review.

INPUTS

CONTEXT

VERSION

TRACE

OUTCOME

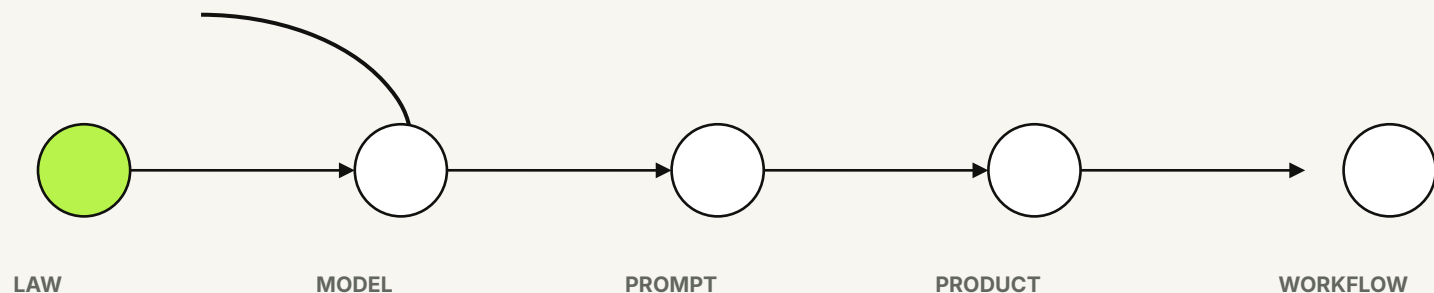
SOURCE

REMEDICATION

07 PRINCIPLE SEVEN

Update continuously.

Regulatory assurance is not a one-time certification. Tests must change when the law, regulatory expectation, model, prompt, product, or workflow changes.



08 PRINCIPLE EIGHT

Report scope honestly.

**PASSING A TEST SUITE IS EVIDENCE ABOUT SPECIFIED
BEHAVIOR UNDER SPECIFIED CONDITIONS.**

It is not a universal declaration that the system is compliant.

● AN OPEN PRACTICE

RegTeaming belongs at the intersection of engineering, legal, risk, audit, and assurance.

Shared terminology and test structures make cross-functional regulatory assurance possible.



● USE THE METHOD

Make regulatory assurance testable.

Adapt it to your sector. Publish attack patterns. Improve the specification. Cite the obligations. Share what works.

READ THE MANIFESTO → [REGTEAMING.COM/MANIFESTO](https://regteaming.com/manifesto)

RegTeaming is an independent methodology sponsored and maintained by Carver.