

# Build vs Buy: Open-Source AI Guardrails and Evaluation vs Commercial Platforms

*Open-source tools cover guardrails, red teaming, bias testing and LLM tracing well. Commercial platforms add inventory, workflow, evidence and support. Here is how to decide.*

Updated October 29, 2026

You can build much of an AI governance stack from open-source parts. You cannot easily build all of it. This guide shows what open source covers well, where commercial products add value, and how to decide.

## What open source covers well

These projects are active on GitHub. We checked each repository and its license on 29 October 2026.

| Job | Project | License | | --- | --- | --- | | Run-time guardrails | [NeMo Guardrails](#) (NVIDIA) | Apache 2.0 | | Run-time guardrails and validation | [Guardrails](#) (Guardrails AI) | Apache 2.0 | | LLM vulnerability scanning | [garak](#) (NVIDIA) | Apache 2.0 | | Red teaming framework | [PyRIT](#) (Microsoft) | MIT | | Evaluation and red teaming | [promptfoo](#) | MIT | | Bias and fairness testing | [Fairlearn](#) | MIT | | Bias and fairness testing | [AI Fairness 360](#) | Apache 2.0 | | ML and LLM monitoring | [Evidently](#) | Apache 2.0 | | LLM tracing and evaluation | [Langfuse](#) | MIT, except enterprise folders | | LLM tracing and evaluation | [Phoenix](#) (Arize) | Elastic License 2.0 |

Two license notes:

- The Elastic License 2.0 says you may not offer the software to others as a hosted or managed service. Internal use is fine. Check with your legal team.
- Langfuse puts code in its ee folders under a separate license. The rest is MIT.

PyRIT moved from the Azure GitHub organization to `microsoft`. The old repository is archived. Use the new one.

## What open source does not cover well

Open-source projects focus on technical checks. These parts are harder to find or build:

- **AI inventory** across teams, vendors and SaaS features
- **Risk workflow**: intake, questions, approvals, sign-off
- **Regulatory mapping** to the EU AI Act, ISO/IEC 42001 and the NIST AI RMF, kept current as rules change
- **Audit evidence** that pulls results from many tools into one report

- **Support, SLAs and security reviews** that procurement can sign off

Rules change. The EU changed the AI Act's high-risk dates in 2026 ([Article 113](#)). US bank regulators replaced SR 11-7 with SR 26-2 on 17 April 2026 ([Federal Reserve](#)). If you build, someone on your team must track these changes.

## Many vendors sell both

Several vendors offer an open-source core and a paid tier. [promptfoo](#) has a free Community version and custom-priced Enterprise plans ([promptfoo](#)). [Langfuse](#) offers self-hosting and a paid cloud ([Langfuse](#)). [Arthur](#) has an open-source evals engine and paid plans ([Arthur](#)). You can start with the open-source part and buy later.

## How to decide

| Choose build when | Choose buy when | | --- | --- | | You have ML platform engineers with time to spare | Your team is mostly compliance and risk staff | | You run few AI systems, mostly built in-house | You have many AI systems, including vendor AI | | You need deep custom tests | You need audit-ready reports soon | | Data must never leave your network | A vendor can deploy in your cloud or on-premise | | You already have a GRC tool for the records | You have no system of record for AI risk |

## Count the full cost

For build, count:

- Engineers to integrate, host and upgrade each tool
- Someone to write and maintain tests and policies
- Someone to track rule changes and update mappings
- Hosting and model costs for evaluation runs

For buy, count:

- License fees. Most governance vendors do not publish them. See [how AI governance software is priced](#).
- Integration time, which is never zero
- Usage fees for traces, evaluations or screened tokens

## A common pattern

1. Run open-source tests in CI on every model or prompt change.
2. Run guardrails in front of production traffic.
3. Send results to one system of record.
4. Buy that system of record if you lack one. Build it only if you have a team to own it.

Use the [LLM deployment checklist](#) to see which checks you need.

## Sources

- GitHub repositories and license files for [NVIDIA-NeMo/Guardrails](#), [guardrails-ai/guardrails](#), [NVIDIA/garak](#), [microsoft/PyRIT](#), [Azure/PyRIT](#), [promptfoo/promptfoo](#), [fairlearn/fairlearn](#), [Trusted-AI/AIF360](#), [evidentlyai/evidently](#), [langfuse/langfuse](#), [Arize-ai/phoenix](#), GitHub, accessed 2026-10-29
- [Pricing](#), promptfoo, accessed 2026-10-29
- [Pricing](#), Langfuse, accessed 2026-10-29
- [Pricing](#), Arthur, accessed 2026-10-29
- [Article 113](#), artificialintelligenceact.eu, accessed 2026-10-29
- [SR 26-2: Revised Guidance on Model Risk Management](#), Federal Reserve Board, 17 April 2026, accessed 2026-10-29

## Frequently asked questions

### Is open-source AI tooling free to use in production?

The license may be free, but you pay in engineering time, hosting and upkeep. Also read each license. Most tools in this guide use Apache 2.0 or MIT, but Arize Phoenix uses the Elastic License 2.0, and Langfuse keeps some enterprise folders under a separate license.

### Can open-source tools make us compliant with the EU AI Act?

They can produce test results and logs. Compliance also needs an inventory, risk decisions, approvals and records. You can build that, buy it, or keep it in an existing GRC tool.

### What is a good mixed approach?

Many teams run open-source tests and guardrails in their pipelines and send the results to one governance record. That record can be a commercial platform or an internal system.