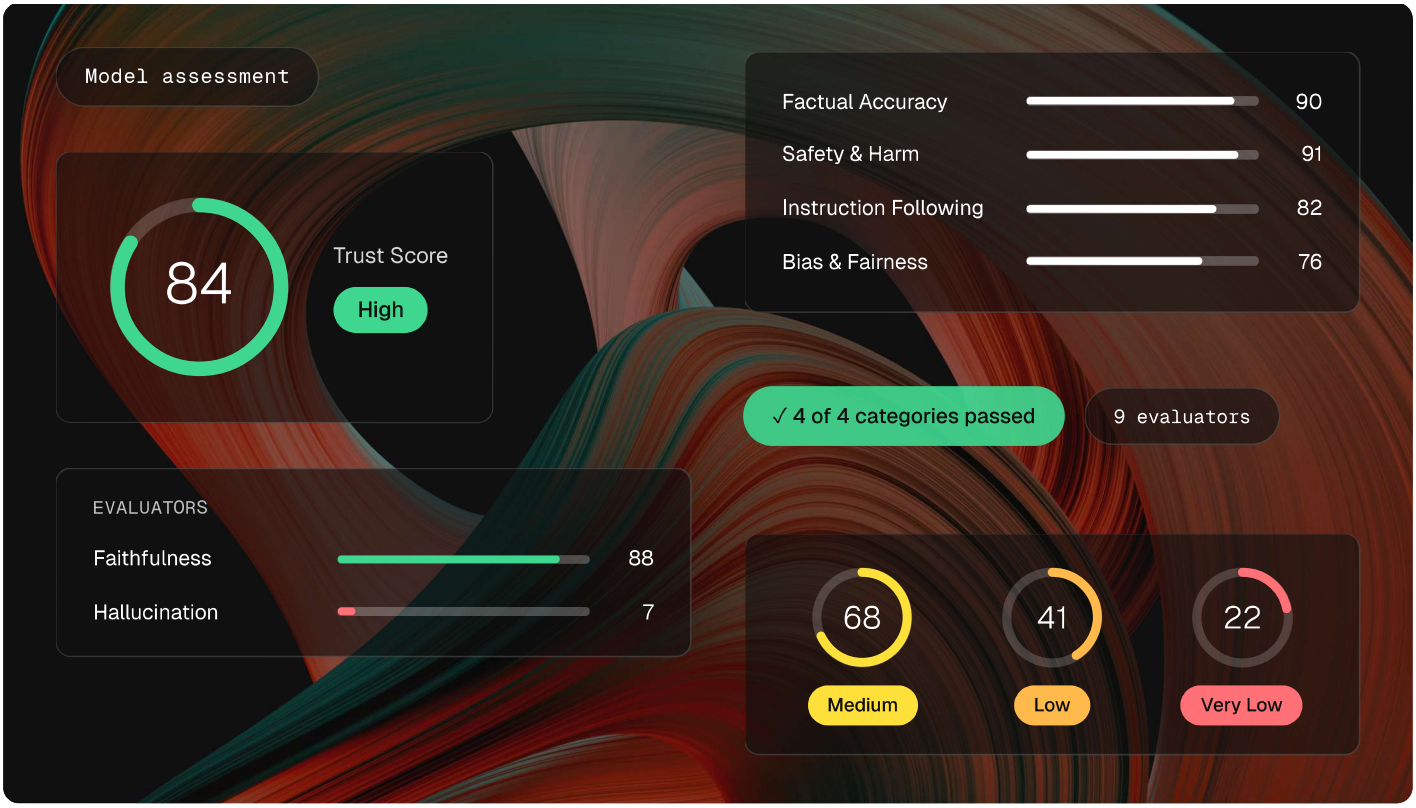
SeekrGuard

Independent AI evaluation and risk governance

Know which model to trust, and prove it

SeekrGuard takes AI from a candidate model to an evidence-based deployment decision.



THE PROBLEM

AI is entering government and regulated environments faster than the processes meant to vet it

40%

of enterprise applications will be integrated with task-specific AI agents by the end of 2026, up from less than 5% today

Source: Gartner

13%

of organizations think they have the right AI agent governance in place

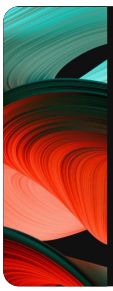
Source: Gartner

Most organizations deploying AI are stuck between two bad options: put it into production without a dependable way to monitor and control it, or hold it back and give up the advantage because they can't get comfortable enough to move.

It's a familiar bind. Every major wave of technology, from the expansion of the Internet to cloud and mobile, was adopted faster than anyone could govern it, and control caught up only after the fact. AI is that pattern again. But for government and regulated industries, the systems are mission-critical, the products are regulated, and the services touch the public.

This isn't a capability gap, it's a process and tooling gap. Teams facing these decisions know how to evaluate technology. What they lack is one system that manages the full vetting lifecycle and produces decision-ready evidence that holds up to scrutiny.

Today, risk gets assessed qualitatively, evaluation happens one use case at a time, and when a model fails in production or an auditor asks how it was approved, there's no repeatable process, no audit trail, and no defensible record to point to.



What you need now

A way to take a candidate model from "we think this works" to an evidence-based, auditable deployment decision you can stand behind.

THE SOLUTION

Not another benchmark or dashboard

Organizations are increasingly accountable for how AI enters their environment, yet they're governing it with tools that were never built for the job. SeekrGuard provides independent AI evaluation and risk mitigation and governance that closes the gap between "we have candidate models" and "we have an evidence-based, auditable deployment decision."

Unlike evaluation tools that grade a single model in isolation for the engineers building it, or governance suites oriented to compliance paperwork, SeekrGuard evaluates multiple models with the test variables controlled, so the judgment is about the model rather than the test.

It turns raw scores into structured, use-case-specific risk and mitigation. And it runs where your data has to stay.

The result is an authoritative record that supports a defensible decision: what was evaluated, how, and why it was approved. Built for the program offices and authorizing officials who have to stand behind the decision.

Independent verification and validation, not a vendor benchmark

SeekrGuard holds the test variables constant, evaluating every candidate against the same datasets, evaluators, prompts, and metrics, so the differences reflect the models and not the test. You get a neutral, apples-to-apples judgment rather than a vendor grading their own homework.

Scores become decision-ready evidence

SeekrGuard can turn evaluation evidence into a weighted, use-case-specific risk level, so you move past raw benchmarks and best guesses to a structured view of how a model performs in the context it will be deployed into, judged by the risk categories and tolerances you set.

Built for the person accountable for the decision

Designed for whoever has to stand behind the decision to put a model into use: a program office or authorizing official in government, or a risk, compliance, legal, or business owner in a company, not just the engineers iterating on a model. A defensible instrument for a responsibility they can't delegate.

Runs where your data has to stay

Generally available, standalone or with SeekrFlow. For government, on-premises or AWS GovCloud. For commercial, SaaS. The organizations that need formal vetting the most don't have to send confidential data somewhere they're not comfortable sending it, or adopt a wider platform to do it.

15 → 150K

Gartner projects the number of agents deployed per company jumping from about 15 last year to as many as 150,000 by 2028.

Source: Gartner

CAPABILITIES

The whole vetting workflow, in one place

SeekrGuard brings cataloging, evaluation, comparison, risk scoring, red-teaming, spot-testing, and guided analysis into one system, instead of stitched-together tools.

01 **Vet a whole library of models against your use case.** Compare candidates on the same datasets, evaluators, and metrics instead of testing one model at a time in isolation.

02 **Test the models you're already paying for.** Evaluate existing frontier deployments against smaller, lower-cost, or open-weight alternatives, and switch with evidence when a cheaper model meets your requirements.

03 **Go from raw data to scored results without a data science team.** Bring your own data, define the criteria that matter, and run evaluations your analysts can watch in real time and reproduce later.

04 **Put candidate models head-to-head.** Score the same prompt across multiple models side by side and re-run anything that looks off, so you can defend which model you chose and why.

05 **Get a risk picture tied to how you'll actually use the model.** Define the risk categories that matter, and the evidence rolls up into a weighted risk level for the specific context you're deploying into.

06 **Probe a model's behavior whenever you need to.** Prompt any model interactively and score its responses on the spot to chase down a failure mode before you commit to a decision.

07 **Keep an expert in the loop even when you're short on bandwidth.** A built-in assistant explains results in context, answers questions about a specific evaluation, and proposes risk-framework changes for your review.

08 **See the patterns and outliers quickly.** Score distributions and per-model and per-evaluator breakdowns let you slice results the way you need and move through review faster.

Evaluation as a cost argument, not just a risk one. Because you evaluate candidates on your own data and use cases, you can tell when a smaller, lower-cost, or open-weight model can be used instead of an expensive frontier model, with evidence behind the decision.

Where SeekrGuard runs

Government

On-premises or AWS GovCloud.

Commercial

SaaS at launch, with additional deployment options on the roadmap.

Generally available, standalone or with SeekrFlow.

seekr

seekr.com · contact@seekr.com