Gremlin Foresight AI — Security & Data Usage FAQ

Summary

Foresight AI adds an AI layer to the Gremlin platform. It answers questions about your reliability posture in natural language, generates reliability reports, analyzes your environment, and recommends tests and remediations.

Adding AI capabilities to a service that collects data about your production systems understandably raises questions, and this document is designed to help answer them.

We do not use your data to train AI models. Instead, Foresight uses a pre-trained model to analyze the data Gremlin already collects as context.. You also retain complete control over Foresight. AI features can be enabled and disabled by your administrator, and every test that runs or every change that is applied is approved by a person on your team. Foresight makes recommendations and enables your team, but the operator controls what actions are taken.The whole AI layer can be switched off in one setting, with the core platform continuing to work exactly as before.

The following FAQs dig into these core concepts with more detail. If your question isn't answered here, just ask us!

What Foresight is

What is Foresight AI?
Foresight is an optional add-on to the Gremlin platform. It provides a chat interface for asking questions about your environment, scheduling narrative reliability reports, suggesting priority actions, and getting recommendations for tests and remediations. It is a layer over the Gremlin platform you already use, not a separate product with its own data.

How does Foresight work?
Foresight pairs a pre-trained third-party large language model with your own Gremlin data. The model is supplied with context drawn from your Gremlin environment, our product documentation, the deterministic rule-based analysis Gremlin already performs (Reliability Intelligence), and the Failure Atlas — Gremlin's proprietary record of how systems fail in practice, distilled by Gremlin engineers from more than a decade of resilience engineering.

Which AI model does Foresight use?

Anthropic Claude. Any change of model or provider is subject to security review and approval before use.

Does Gremlin train or fine-tune its own model?
No. Foresight uses a pre-trained model consumed as a service. None of our data or your data is incorporated into model weights. Because there is no training pipeline, there is nothing for your data to flow into.

Does Foresight change my systems?
No. Foresight's agents are supervised. They recommend tests, analyze results, and deliver fixes as concrete artifacts like configuration patches and infrastructure-as-code diffs. Your team reviews and approves what runs and what merges. Nothing is applied automatically.

Do I have to use the Gremlin interface?
The full Foresight experience, including the chat interface, is delivered through the Gremlin UI. Any work product that Foresight generates is available through the Gremlin API and MCP server, and subject to the same permissions.

How Foresight uses your data

What data does Foresight use?
Foresight uses the operational and infrastructure data the Gremlin platform already collects as context, including service definitions, dependencies, test runs and results, detected risks, health checks, and reliability scores. Foresight introduces no new data collection and no new agent permissions.

Do you use our data to train AI models?
No. This commitment covers our own systems, any third-party AI service we use, and all current and planned AI capabilities on the platform. Your data is processed to deliver analytics, insights, and platform functionality back to you. Nothing else.

What stops the LLM provider from retaining or training on our data?
Our agreements with AI providers grant no training rights and no data ownership, restrict use to providing the contracted service, prohibit any secondary use, and require vendor retention and training features to be disabled. These terms are a precondition of approving any AI tool for use at Gremlin.

Is personally identifiable information sent to the AI?
No personally identifiable information is sent to the AI with one exception: a user's email address may be transmitted for user identification when AI features are enabled. Everything

else processed by our AI systems is operational and infrastructure data that contains no personal information.

Where are prompts and completions stored?
Prompts and completions are stored in Gremlin's AWS environment in the United States, encrypted at rest, under our standard Data Retention Policy. They are not written to any customer-controlled system, are not copied to any additional third-party store, and are not retained by the LLM provider.

Where does AI processing occur?
All AI processing for Gremlin occurs exclusively within the United States.

Is our data separated from other customers' data?
Yes. We follow the practice of strict tenant isolation to segregate your data from every other Gremlin customer, and Foresight operates only within the boundaries of the requesting user's existing permissions.

How long is data retained?
Data is retained as per our Data Retention Policy, available upon request. Customer-provided data is classified High Risk and is retained no longer than 90 days from the point at which it no longer serves a purpose. At that point, it is deleted using secure deletion. Contractual or legal retention obligations are documented as exceptions to this policy.

Controls you have over AI and your data

How do we turn AI off?
Foresight can be turned off with a single company-level "Allow LLM access" setting. When it’s disabled, no data is sent to the AI layer, Foresight is unavailable, and the core Gremlin platform continues to operate normally.

Can we enable AI for some teams but not others?
No. The setting is company-wide and all-or-nothing. There is no user-level or team-level AI toggle.

However, team membership and user roles do govern what Foresight can access. Foresight operates within the requesting user's existing role-based access controls and permission boundaries, so each user can only see the teams and services they are already entitled to see. What roles and teams do not control is whether the AI layer is available at all. That is a single company-level decision.

What else do we control?
As always, customers control where Gremlin agents are deployed and what permissions they hold; which tests may run, where, and on what schedule; who may run them; and which recommendations are passed onward to people or to other tooling. Individual recommendations can be dismissed, and a dismissed recommendation is not suggested again.

Are users told that they're interacting with AI?
Yes. The capability is presented as an AI feature, delivered through a labelled AI chat interface and clearly identified AI-generated reports. Enabling it is an explicit administrative action.

Accuracy and safety

How likely is Foresight to be wrong, and what happens if it is?
Foresight's responses are constrained by what it is given, not by what the model happens to know. Four inputs shape every recommendation:

- Your own environment: service definitions, dependencies, test runs, test results, detected risks, and reliability scores drawn from your Gremlin data. Access to these is limited to the teams and services the requesting user is already entitled to see.
- Deterministic analysis: Gremlin's rule-based Reliability Intelligence, which diagnoses failed tests without AI and is a primary input to recommendations.
- The Failure Atlas: Gremlin's own curated record of how systems fail in practice, distilled by Gremlin engineers from more than a decade of resilience engineering.
- Gremlin product documentation.

Recommendations are then validated empirically rather than asserted. A recommendation is a testable claim: run the test, apply the fix, re-run the test, and establish whether the claim holds. That loop is the same fault injection Gremlin has always performed. The AI proposes what to test and interprets the result, but the evidence comes from your running systems.

The realistic cause for an incorrect response is a stale recommendation rather than an invented one. Most often this results from an issue that was fixed after a test was run, but the test has not been re-run, and thus a lower-priority or unnecessary test is proposed. Re-running the relevant test refreshes the evidence and supersedes the outdated finding.

Even when there is a wrong suggestion, the impact is bounded by design. Outputs are advisory; a person approves every test that runs and every remediation applied; any test that does run is contained by blast-radius controls and automatic halt-and-rollback; and the underlying data is technical telemetry about infrastructure, not business or customer data.

The only realistic outcome of a "wrong response" is engineering effort spent on a lower-priority test that still improves reliability.

How do you limit AI hallucination?
We limit hallucinations by grounding Foresight with context. Responses are constrained to your own environment data, to deterministic Reliability Intelligence analysis, and to the Failure Atlas rather than to the model's general knowledge. Recommendations are also re-validated empirically. Every recommendation is a testable claim, and re-running the relevant test establishes whether the fix successfully addressed the risk.

How do you correct incorrect output?
Recommendations and remediations are based on the context data. When there's an incorrect output, the solution is to revise the context and the instructions. When you re-run the relevant test, it refreshes the underlying evidence and supersedes a stale finding, which will provide a different output. But the core instructions and context will also be improved over time. Changes to that material, to the Reliability Intelligence rules, or to supporting code follow Gremlin's formal change management process: peer review by someone other than the author, automated security scanning, security-team review before promotion to production, acceptance testing, and a defined rollback path.

The goal is to always provide accurate, trustworthy recommendations that can be verified by re-running the failed test.

What guards against prompt injection and adversarial input?
Foresight is protected by four layers of security. AI capabilities are tested for prompt injection and for data leakage using adversarial techniques before release, while inputs and outputs are validated and sanitized at runtime. Additionally, grounding responses in your data and in curated reference material narrows the surface available to manipulation.

Most importantly, Foresight operates strictly within the requesting user's existing permissions, so a successful injection cannot escalate privilege, cross team boundaries, or reach another tenant. No change is applied without explicit human approval, which keeps Foresight contained.

Is Foresight tested for bias or unfair discrimination?
Foresight makes no decisions, predictions, or inferences about people, which means it's never in a position to introduce bias or unfair discrimination. Foresight operates entirely on technical telemetry about applications and infrastructure. It ingests no consumer, credit, demographic, or protected-class data, and only produces engineering recommendations about the resilience of software systems. There is no individual or population against which disparate outcomes could arise. This allows us to focus our testing efforts on accuracy, safety, and robustness.

Who is responsible for validating output?

Responsibility for output validation is shared by Gremlin and the user. Gremlin validates AI capabilities before release, which includes confirming that output is not harmful, inappropriate, or undesirable, and testing for hallucination or adversarial manipulation. Our policy holds the requesting user accountable for the output of AI tooling and requires normal review and test processes to be applied to it, which is why recommendations are surfaced for approval rather than acted on independently.

Security, compliance, and availability

Is Foresight covered by your SOC 2 report?
Customer data reaching Foresight is accessed through the Gremlin API, which has been within the scope of our annual SOC 2 Type II examination for the past seven consecutive years, covering Security, Availability, and Confidentiality. The current report is available upon request.

How is data protected in transit and at rest?
Data follows our Cryptography Policy and NIST guidance, and is protected by TLS 1.2 or higher in transit and AES-256 or stronger at rest. Any keys protecting customer data are managed in AWS KMS and may not be extracted.

What governs AI use at Gremlin?
Gremlin follows an AI Data Usage Policy, which is reviewed at least annually and is part of mandatory security awareness training. Decision authority over all AI models, tools, and implementations sits with the Head of Security in collaboration with the CTO and CEO. Every AI model and tool requires security review and approval before use and re-approval when usage patterns change. The Head of Security monitors compliance, reviews AI tool usage and controls, and addresses violations. Our customer-facing AI Data Usage Policy is published to our Trust Share portal, where customers are also notified of material changes.

How is the AI provider vetted?
AI providers are vetted under the same vendor security requirements as any other vendor handling our data: documented security review and approval before use, annual re-review, encryption in transit and at rest, MFA, RBAC, and a current SOC 2 or ISO 27001 report with no findings above low severity. Any additional AI provider would be subject to the same review before use.

What happens if the AI layer fails or degrades?
Foresight is an advisory layer, not a dependency. It does not sit in the path of any production transaction and does not change your systems on its own, so an AI failure does not propagate into your environment. The core Gremlin platform (reliability testing, scores, reports, detected risks, and API and MCP access) continues to operate without it, and deterministic Reliability

Intelligence analysis remains available. The AI layer can also be disabled immediately at the company level while you continue using the platform.

What are your recovery objectives?
RTO does not exceed 24 hours, and RPO is instantaneous. We use continuous replication to achieve instantaneous RPO for up to 30 days, so recovery to any point within that window is possible. The plan is tested end to end at least annually and after any functional change to the plan, procedures, or responsible personnel.

How and when are we notified of a security incident?
As per our security policies, customers are notified within 24 hours of a confirmed security incident involving your services or data, and within 72 hours of a suspected but unconfirmed incident. Incidents are managed under our Security Incident Response Policy, which covers evidence preservation, forensics, customer notification, and tracking through to remediation.

Who do we contact?
For support and product issues, including inaccurate or unexpected output, please contact support@gremlin.com, or your account and Customer Success contacts. For any concern involving a security dimension, such as suspicious or harmful output, suspected data exposure, or evidence of prompt injection, please contact security@gremlin.com. For privacy matters, please reach out to privacy@gremlin.com.