WHITEPAPER

AI Transparency Datasheet

Scope: Sekoia Defend, Roy, AI Cases, Elevate, and behavioral detection features

August 2026 · Version 1.0 · Public

Sekoia

Table of contents

1. What AI does in the Sekoia platform

The AI features built into the Sekoia SOC platform have one precise purpose: help security teams detect, qualify, and investigate incidents faster, without replacing their judgment. In practice, these features:

- reduce repetitive tasks
- speed up information search and correlation
- improve alert comprehension
- apply analyst investigation methods at scale
- produce traceable, verifiable results

AI assists. The analyst decides. Results can be reviewed, corrected, or rejected at any time by an authorized user.

2. AI-powered features

Sekoia's AI capabilities are grouped below by their primary role in the SOC. The examples reflect features available at the publication date and may evolve as new capabilities are introduced.

Feature	**AI role**
Roy	Natural language search, query building assistance, event summarization, and access to Cyber Threat Intelligence
AI Cases	Alert correlation and incident summary generation
Elevate	Agentic alert investigation, evidence collection, correlation, scenario evaluation, verdict proposal and runbook generation
Behavioral detection	Behavioral baseline establishment and deviation or anomaly detection

The exact scope of available capabilities depends on the subscription, configuration, and permissions assigned to the customer. Feature-specific data flows, permissions, and processing details are described in the applicable product and contractual documentation.

New AI capabilities may be introduced over time without changing the general principles and controls described in this datasheet.

3. Technologies used

Sekoia combines several complementary approaches:

- generative models for comprehension, summarization, and investigation
- machine learning models and statistical methods for behavioral detection
- detection rules, runbooks, and deterministic mechanisms
- Cyber Threat Intelligence data and platform search capabilities
- agentic orchestration to structure investigations

These components serve different purposes. Generative models and statistical detection models are not interchangeable.

4. Architecture and data processing

AI features that process customer security data run within Sekoia's applicable service environment. Depending on the feature, processed data may include:

- security alerts and events
- logs and query results
- user, machine, and asset information
- Indicators of Compromise (IoCs)
- Cyber Threat Intelligence data
- investigation rules and procedures
- queries or instructions entered by users

AI access to data is constrained by user permissions, active connectors, tenant configuration, applicable rules and runbooks, and platform access controls. Users

have no direct access to the underlying models. Inputs pass through Sekoia's orchestration and control layers before reaching the relevant component.

5. Models and technical evolution

Sekoia uses different model categories and hosting patterns depending on the feature, the sensitivity of the data processed, the service region, and the applicable security and performance requirements.

At the date of publication, the main model categories and hosting patterns are:

- Elevate: agentic generative AI and large language models, hosted on Sekoia infrastructure
- AI Cases: generative AI combined with rules and correlation logic, hosted on Sekoia infrastructure
- Roy: conversational large language models hosted on Sekoia infrastructure for natural language search, query assistance, summarization, and CTI access
- Behavioral detection: statistical and machine learning models integrated into the Sekoia platform

Sekoia's product doesn't rely on external large language models.

Sekoia may use different model families within these categories, including self-hosted open-weight models for features that process sensitive customer security data. The model families and versions used may evolve over time.

Sekoia may update or replace a model for reasons of security, performance, availability, compliance, or resilience. Such changes don't affect commitments relating to confidentiality, isolation, or customer data use. The model categories and hosting patterns described in this section are current as of the publication date and may be updated as the service evolves.

6. Use of data for training

Customer data is not used to train Sekoia's generative foundation models. This covers all shared-model training, non-contracted fine-tuning, continued pre-training, and third-party model training.

Some behavioral detection features use tenant data to establish a baseline specific to that environment. This baseline is used solely to detect anomalies within the relevant tenant. It isn't shared with other customers and doesn't constitute training of the generative model.

User feedback may improve prompts, runbooks, orchestration, the interface, or product quality. It doesn't serve to retrain foundation models.

7. External AI providers and AI-related data flows

AI features run within Sekoia's applicable service environment. Sekoia's product doesn't rely on external large language models, and customer data isn't sent to third-party AI providers. Detailed information on AI-related data flows is available in the contractual documentation, the DPA, the Sekoia Trust Center, or through confidential exchanges.

8. Context retrieval and vectorized data

Elevate conducts investigations by querying data available on the platform and authorized Cyber Threat Intelligence sources. Investigations rely on queries run against security data, rules and runbooks, context information, and relationships between events and entities.

Elevate doesn't store a vectorized representation of customer data. The agent queries data directly from the platform. Results are tied to the queries and elements observed during the investigation.

9. Governance and design principles

Sekoia's AI features are built to be secure, traceable, and controllable.

- **Security and robustness:** AI components are integrated into the platform's development, review, testing, deployment, and incident management processes.
- **Confidentiality protection:** Data accessible to AI is limited to the authorized scope. No sharing between tenants.
- **Traceability:** Actions and results are logged to support operational review and audit.

- **Human oversight:** Analysts retain control over results. They can correct, replace, or reject any AI-generated decision.
- **Transparency:** Sekoia documents the general operation, limitations, and commitments applicable to each feature.
- **Continuous improvement:** Components and controls are updated on an ongoing basis.

Information on applicable certifications, frameworks, and compliance commitments is available in the Sekoia Trust Center and contractual documents.

10. Isolation, encryption, and data protection

AI features follow Sekoia's platform multi-tenant isolation model. Key controls include:

- logical separation of customer environments
- permission controls and agent-accessible connector controls
- no data sharing between tenants
- logging of significant actions
- encryption in transit and at rest in accordance with applicable platform controls
- privileged access and key management following Sekoia's security processes

One customer's data is never used to produce another customer's results.

11. Data location and retention

Data location depends on the service region and contractual configuration. Three dimensions apply: the region where data is stored, the region where processing runs, and any external services a given feature relies on. Telemetry data is subject to the storage, retention, and deletion policies applicable to the platform and relevant features. Exact retention periods vary by subscription, region, feature, and contractual parameters. They're specified in the service documentation and applicable contractual documents.

12. Autonomy and human oversight

AI features operate at different levels of autonomy:

- **Roy** provides assistance and information to the analyst.
- **AI Cases** correlates alerts and generates incident summaries.
- **Elevate** can automatically conduct an investigation and propose a verdict.
- **Behavioral detection** produces signals or scores intended for analyst review.

In all cases, analysts can review results, verify evidence, examine queries run, correct a verdict, replace an AI-generated decision, or add a comment. High-impact actions require human validation according to platform configuration.

13. Explainability and traceability

For relevant investigations, analysts can access:

- the detection that triggered the investigation
- runbook questions and queries run
- collected results and evidence
- false positive scenarios examined
- the confidence level and any evidence gaps
- actions taken by the agent
- corrections made by the analyst

Actions and results are logged to support operational review and audit of investigations.

14. Performance, limitations, and error handling

AI systems can produce incomplete, incorrect, or inadequate results. Sekoia's features are designed to reduce those risks, not eliminate them. Controls in place include: use of security data and authorized sources, tying results to queries and evidence, displaying confidence levels, analyst validation, testing on representative cases, tracking errors and corrections, performance monitoring, and fallback mechanisms in case of unavailability. Residual risks include novel threats with no known precedent, insufficient or inaccurate data, false positives and false negatives, and human interpretation or validation errors. **AI-generated results are an aid to analysis. They should be verified before any critical decision.**

15. Bias and fairness

Sekoia's AI features are designed to analyze technical events, alerts, and behaviors related to information system security. They don't assess individuals or groups. Technical identifiers present in the data (usernames, machine names, IP addresses, service accounts) are treated as evidence in a security investigation, not as individual attributes. These features make no decisions relating to employment, access to services, health, credit, or any other situation involving individual evaluation.

16. Security and misuse prevention

Controls in place limit abusive or out-of-scope use: permission-based access control, per-tenant consumption limits, no direct model access, request handling through orchestration, protection against prompt injection attempts, action logging, and volume monitoring. Agents cannot modify their own objectives, policies, runbooks, or execution logic.

17. Restricting AI access and degraded mode

Access to AI features is controlled through permissions. To restrict access across a workspace or community, administrators must remove the relevant AI permissions from all roles.

This restriction doesn't affect the platform's core functions, including alert and case management, event search, detection mechanisms, playbooks, and workflows that don't depend on AI.

If an AI component becomes unavailable, fallback mechanisms may help maintain core functions, depending on the feature.

18. Frequently asked questions

Is customer data used to train the models? No. Customer data isn't used to train Sekoia's generative foundation models. Some behavioral features may use tenant data to establish a baseline specific to that environment, solely for internal detection purposes.

Is customer data sent to third-party AI providers? No. Customer data isn't sent to third-party AI providers. AI features run within Sekoia's applicable service environment. Detailed information on AI-related data flows is available in the

contractual documentation, the DPA, the Sekoia Trust Center, or through confidential exchanges.

Which model categories are used? Sekoia uses different model categories depending on the feature and the type of data processed. The main categories and hosting patterns are detailed in section 5 of this document. Exact model families and versions may change over time.

Does Elevate make decisions on its own? Elevate can run an investigation and propose a verdict. The analyst retains control of the outcome and can correct or replace it. High-impact actions require human validation.

Are results explainable? Yes. Investigations can show the queries run, data observed, evidence, confidence level, and actions taken by the agent.

Is a vector database containing customer data used? No. Elevate doesn't store a vectorized representation of customer data. The agent queries data directly from the platform.

Can access to AI features be restricted? Yes. Access to AI features is controlled through permissions. To restrict access across a workspace or community, administrators must remove the relevant AI permissions from all roles. Core platform functions remain available, but AI-dependent workflows may not.

What if the AI gets it wrong? The analyst can review the evidence, correct the verdict, and flag an incorrect result. AI features are a decision-support tool, not a source of irreversible decisions.

Can models change? Yes. Sekoia may update models for security, performance, availability, or compliance reasons. Commitments relating to confidentiality, isolation, and data use remain applicable.

Where can I find detailed information? In the Sekoia Trust Center, the Data Processing Agreement, product documentation, service region documentation, subprocessor documentation, and through confidential exchanges tailored to the customer's scope.

19. Scope and further information

This document covers the general operation of Sekoia's AI features and the main applicable commitments. The exact scope of processing depends on the active feature, service region, subscription, and customer configuration.

For a given project, Sekoia can provide a scope sheet specifying the active AI features, categories of data processed, processing environment, relevant vendors or subprocessors, applicable retention rules, oversight and traceability mechanisms, and the models or model families used.

Detailed information is provided, based on sensitivity, through the Sekoia Trust Center, product documentation, contractual documents, or confidential exchanges with the customer.