Aurascape

Technical Brief

Aurascape Application Risk Methodology

Purpose

Aurascape evaluates third-party AI applications to help security teams assess risk and define access policies. This document explains the core methodology we use to assess risk, what you will see in the product, and how it helps you control app access based on risks.

How We Generate the Risk Score: The Core Methodology

Aurascape AI Agent & Crawler

Our custom AI agent scans the application's website, locates its policy documents (e.g., privacy policy, terms, pricing, etc.), and subscribes to feeds for security breaches, CVEs, and vulnerabilities.

The AI agent extracts relevant facts and converts them into boolean and date-based attributes. The agent then generates a general recommendation on what attributes, modes and subscriptions should be blocked or allowed via policy, based on the observed risk factors.

LLM Ensemble (Task-Specific Models)

The AI agent uses a variety of large language models (LLMs), each chosen based on the specific task. Different models are best suited for the task based on the ability to quickly extract the relevant data with accuracy such as fact extraction, and summarization. This combination of models ensures accurate and efficient analysis.

Third-Party Human Research (as needed)

For certain applications where additional details are not publicly available, we may engage third-party researchers to enhance the dataset. This step helps ensure the completeness of the risk analysis where automation might fall short.

Key Areas We Evaluate To Determine Risk Score

This allows us to turn vendor policies and public incident data into seven clear risk attributes with evidence and plan-specific notes. The result is an explainable basis to set access policy—allow, limit, or block—per app and with context of risk.

Attribute	What we look for
Trains on customer data	Whether vendor policies state that customer-provided data may be used to train models.
Retains customer data	Whether the vendor retains customer data and the stated retention behavior.
Full rights for generated content	Whether the customer keeps full rights over model outputs versus vendor license claims.
Recently launched (≤90 days)	Whether the application or major capability is new/early-stage.
Recent security breaches (12 months)	Publicly reported incidents affecting the vendor, with date and references.
May generate toxic content	Evidence the app has produced or enables NSFW/offensive/discriminatory content, or lacks safeguards.
Sanctioned app (customer setting)	Your organization's governance setting for the app used to align access policies.

Additional attributes for diligence and filtering

Beyond the seven core risk attributes, Aurascape also displays additional traits that do not affect the overall risk score but are useful when defining policy. These include Data Privacy & Protection, Security & Risk Management, Compliance & Legal and App Design & Operations.

How This is Expressed in the Aurascape Interface

When using the Aurascape platform, users will interact with the risk score results through an intuitive interface:

Application Risk Score (0–100)

Low Risk (0–30)
Medium Risk (31–70)
High Risk (71–100)

Application Risk Attributes

These traits describe how an app operates—such as its handling of sensitive content, data encryption, security practices, compliance with standards, and potential risks like generating toxic content. They help assess the app's behavior and guide you in making policy decisions.

How these attributes are displayed
• Top-Level Labels: Each risk attribute will display a simple Yes/No/Undisclosed label.
• Expandable Panel: Clicking the label expands a panel showing:
o A brief rationale (e.g., Why this attribute was marked Yes/No/Undisclosed).
o Relevant excerpts from vendor policies or publicly available documentation.
o Plan-Specific Differences: If risk behavior differs by plan (e.g., Free vs. Enterprise), the panel will highlight those nuances, ensuring your policy decisions are aligned with your specific contract terms.

How the Results Can Be Used: Risk-Based Policy

You can now use the risk score to establish clear policies for application access based on the evaluated attributes and their associated risks. Let your security and operations teams have full control over who gets access to the application and under what conditions.

Learn More at aurascape.ai