

# Zscaler AI Guard

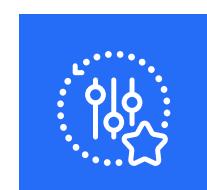
## AT-A-GLANCE

Secure production AI with guardrails that protect against malicious behavior and sensitive data leakage

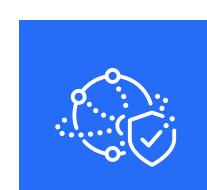
### Why AI Guard?

Enterprises are increasingly building AI assistants and agents to enhance customer experience, drive employee productivity, and streamline business processes. However, deploying AI in production comes with risk, even when extensive red teaming is used to identify vulnerabilities. Once live, AI apps become targets for manipulation. Attackers use prompt injection, jailbreaks, and supply-chain threats to trigger data leakage, harmful outputs, or undesired behaviors. Employees using commercial GenAI apps also pose data leakage risk.

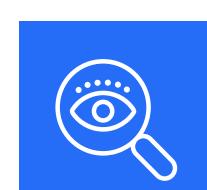
**Zscaler AI Guard** helps you securely operate AI in production by detecting and blocking threats between users, AI apps, and LLMs, including leading commercial GenAI tools. It enforces guardrails on AI behavior, ensures alignment with internal and external policies, and provides the visibility and control needed to manage risk. AI Guard also monitors and governs usage of apps like ChatGPT, Gemini, and Microsoft 365 Copilot to keep AI secure, trustworthy, and business-aligned.



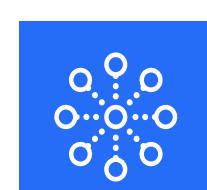
**Higher accuracy and lower false positives:** AI-powered engines help block toxic or malicious inputs without overfiltering safe content. Severity settings help you set the right priorities.



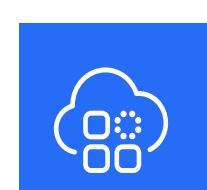
**Broad detection coverage:** More than 18 detectors that can be applied on prompts as well as responses to uncover a wide range of attack patterns and exposure points. Support for popular apps such as ChatGPT and Microsoft Copilot.



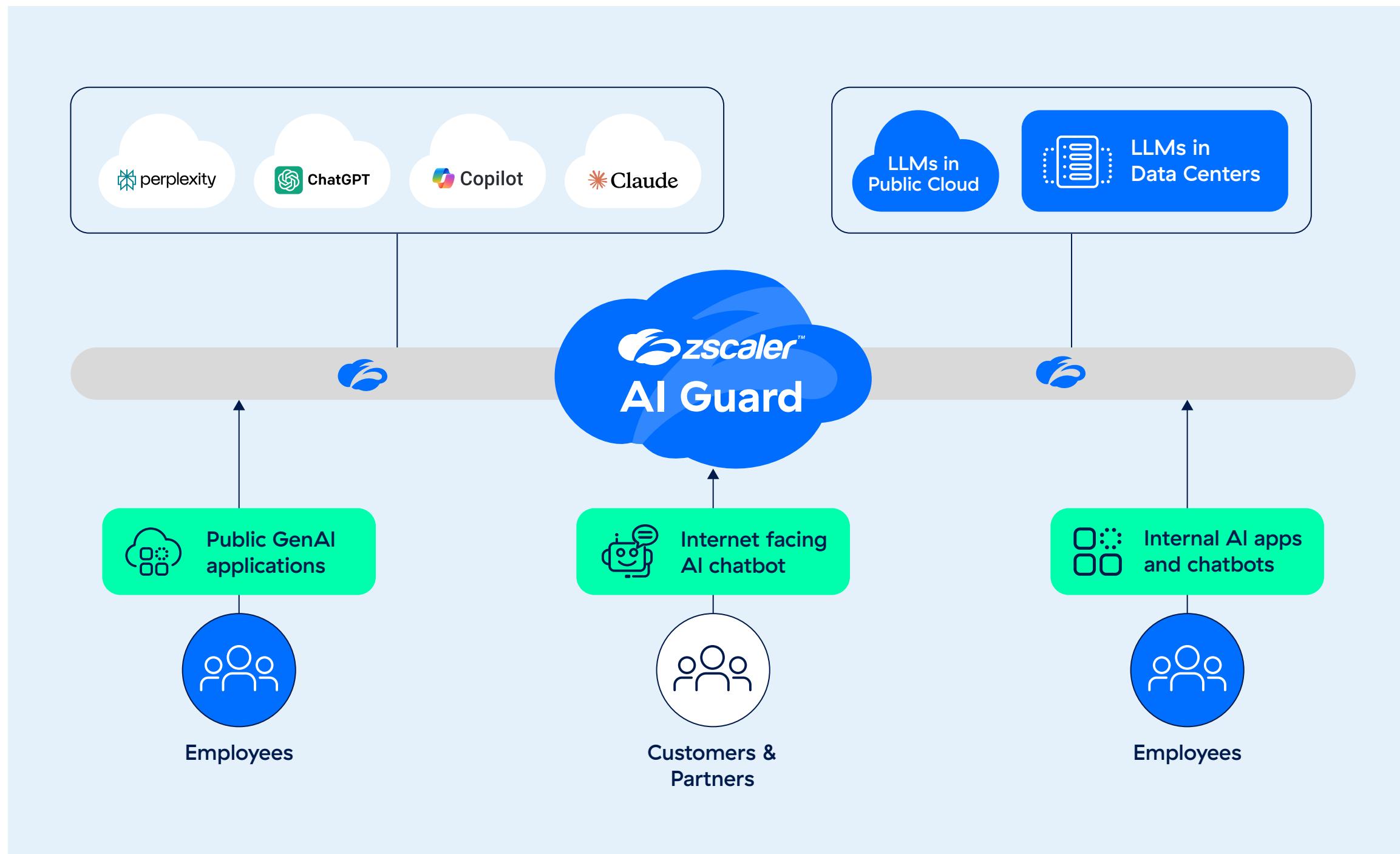
**High fidelity identification for sensitive content:** Powerful safeguards that identify and prevent leakage of sensitive data – both for users accessing popular GenAI apps as well as for enterprise AI apps talking to LLMs.



**Security for the complete AI lifecycle:** Zscaler AI Red Teaming and AI Guard work seamlessly to help identify vulnerabilities and secure AI apps from development through production and operation.



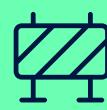
**Leverage with cloud-based AI platforms and frameworks:** Easily connect to LLMs and use across popular cloud-based AI platforms and frameworks including AWS, Azure, Google Cloud, Langchain and Palantir Webhooks.



Reduce Legal & Compliance Risks



Protect your Brand Reputation



Automated Incident Response

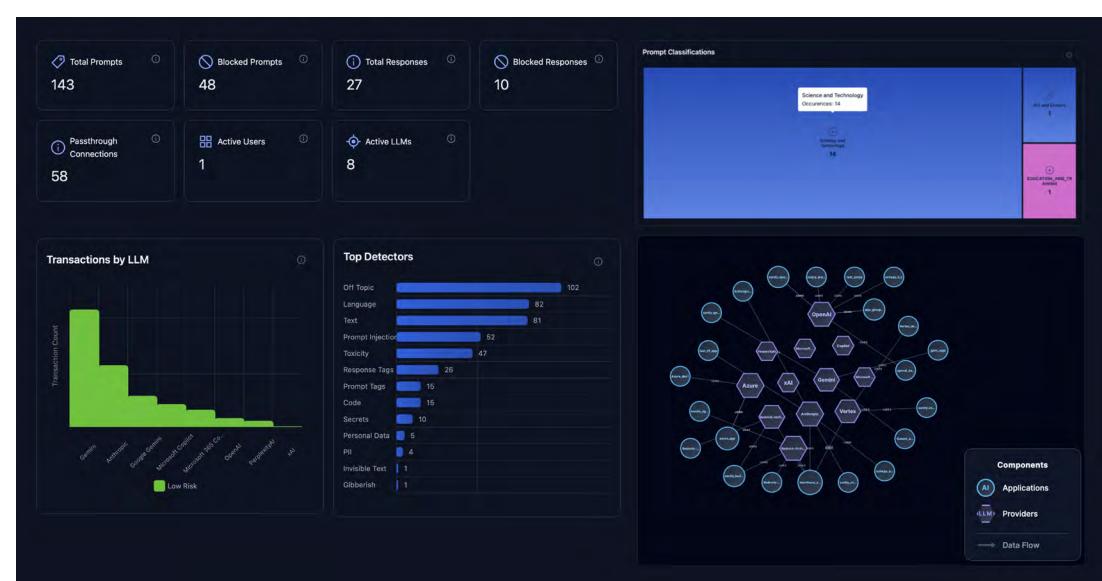


Respond to Threats Proactively

## Lightning-Fast Guardrails For AI apps

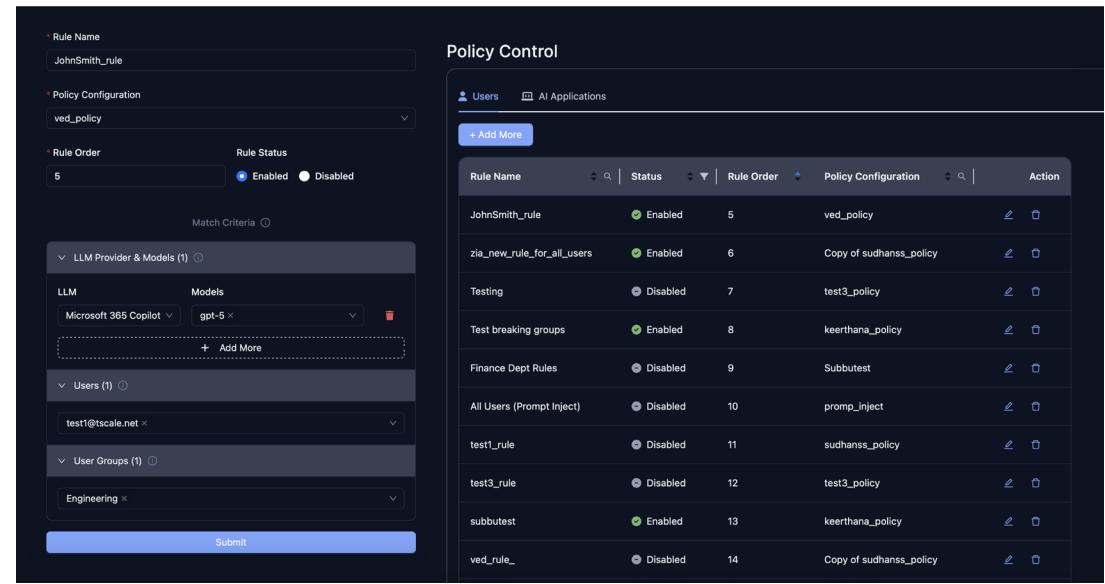
### GET ONGOING VISIBILITY AND INSIGHTS

Enable oversight of AI apps – both commercial as well as enterprise, the LLMs they access, their usage, and prompts and responses to discover shadow AI and enforce business alignment.



## GOVERN THE USE OF POPULAR AI APPS

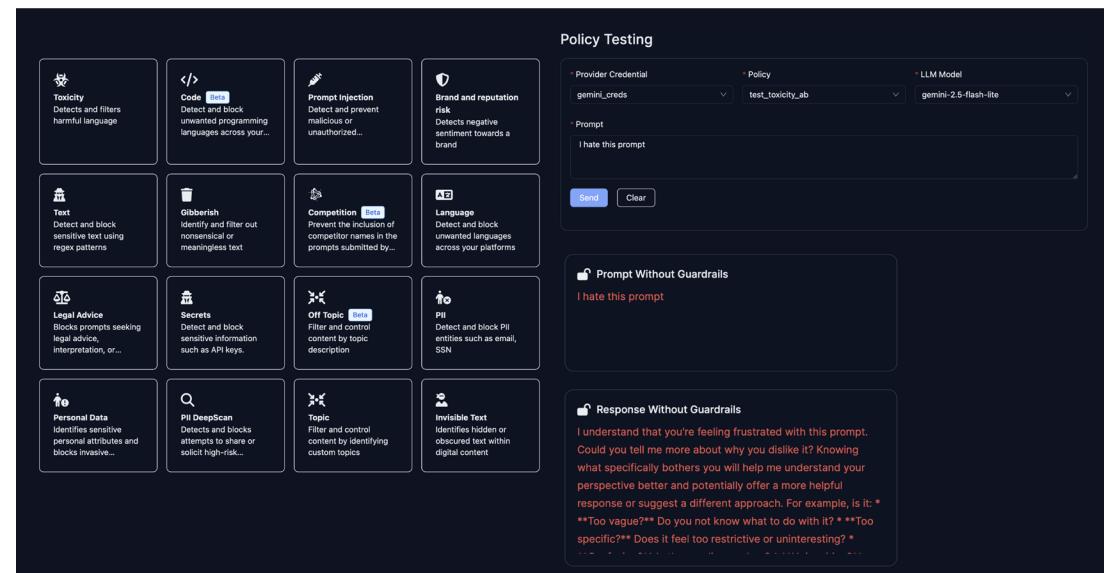
Control who and how popular AI apps such as ChatGPT, Gemini and Microsoft Copilot are accessed and used. Create and apply policies to specific users or groups of users to ensure compliance and prevent data leakage.



The image shows two side-by-side screenshots of the Zscaler AI Governance interface. The left screenshot is titled 'Policy Configuration' and shows a 'Rule Name' (JohnSmith\_rule), 'Policy Configuration' (ved\_policy), 'Rule Order' (5, Enabled), and 'Match Criteria' (LLM Provider & Models: Microsoft 365 Copilot, gpt-3, Users: test1@zscale.net, User Groups: Engineering). The right screenshot is titled 'Policy Control' and shows a table of AI applications with columns for Rule Name, Status, Rule Order, Policy Configuration, and Action. The table includes rows for JohnSmith\_rule (Enabled, 5, ved\_policy), zia\_new\_rule\_for\_all\_users (Enabled, 6, Copy of sudhanss\_policy), Testing (Disabled, 7, test3\_policy), Test breaking groups (Enabled, 8, keerthana\_policy), Finance Dept Rules (Disabled, 9, Subbutest), All Users (Prompt Inject) (Disabled, 10, prompt\_inject), test1\_rule (Disabled, 11, sudhanss\_policy), test3\_rule (Disabled, 12, test3\_policy), subbutest (Enabled, 13, keerthana\_policy), and ved\_rule\_ (Disabled, 14, Copy of sudhanss\_policy).

## SECURE AI APPS IN PRODUCTION

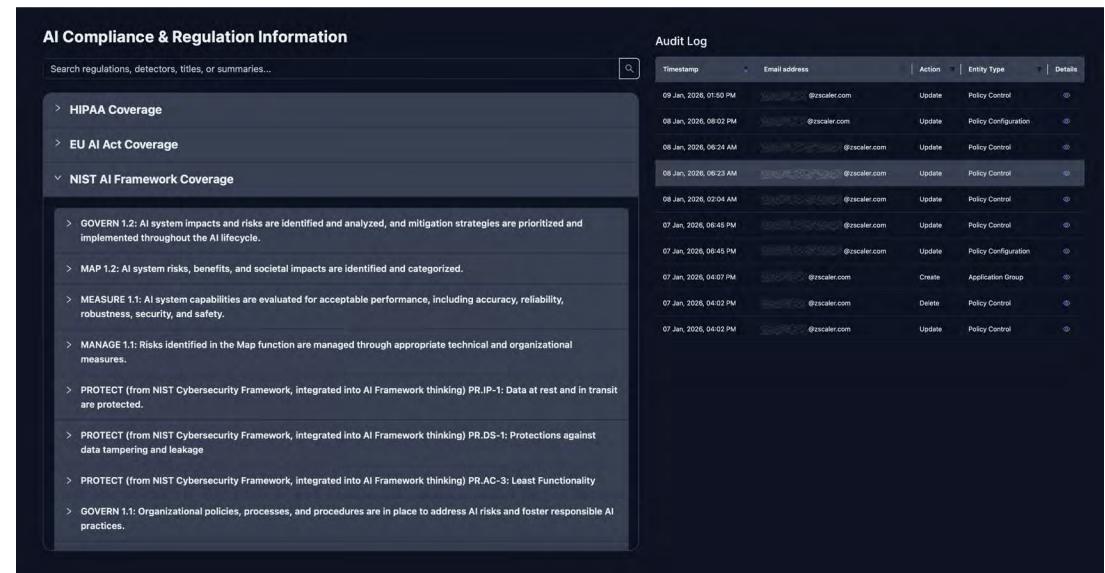
Protect your enterprise AI apps against adversarial attacks such as prompt injections, embedded code, off-topic discussions, toxicity, and malicious outputs to ensure business alignment and reduce risk. Integrate seamlessly with popular cloud-based AI platforms, SDKs and AI agent frameworks including AWS, Azure, Google Cloud, Langchain Agents, Palantir Webhooks, AWS Boto3, and Google Vertex SDK.



The image shows two side-by-side screenshots of the Zscaler AI Governance interface. The left screenshot is titled 'Policy Testing' and shows a provider credential (gemini\_creds), policy (test\_toxicity\_ab), and LLM Model (gemini-2-flash-lite). It includes a prompt (I hate this prompt) and a 'Send' button. The right screenshot shows a grid of 16 AI detectors, each with an icon and a brief description: Toxicity (Detects and filters harmful language), Code (Detect and block unwanted programming language across your...), Prompt Injection (Detect and prevent malicious or unauthorized...), Brand and reputation risk (Detects negative sentiment towards a brand), Text (Detect and block sensitive text using regex patterns), Glorification (Identify and filter out nonsensical or meaningless text), Competition (Review and block mention of competitor names in the prompts submitted by...), Language (Detect and block unwanted languages across your platforms), Legal Advice (Blocks prompts seeking legal advice, interpretation, or...), Secrets (Detect and block sensitive information such as API keys), Off Topic (Identify and control content by topic description), PI (Detect and block PI entities such as email, SSN), Personal Data (Identifies sensitive personal attributes and blocks invasive...), PI Detection (Detects and blocks attempts to share or solicit high-risk...), Topic (Identify and control content by identifying custom topics), and Invisible Text (Identifies hidden or obscured text within digital content). Below the detectors is a box for 'Prompt Without Guardrails' with the text 'I hate this prompt' and a 'Response Without Guardrails' box with the text 'I understand that you're feeling frustrated with this prompt. Could you tell me more about why you dislike it? Knowing what specifically bothers you will help me understand your perspective better and potentially offer a more helpful response or suggest a different approach. For example, is it: \* \*\*Too vague?\*\* Do you not know what to do with it? \* \*\*Too specific?\*\* Does it feel too restrictive or uninteresting? \* ...'. The bottom right corner of the interface has a 'Guardrails' button.

## AUDIT AI USAGE, ENSURE PRIVACY, AND ENABLE COMPLIANCE

Maintain logs of users, apps, prompts, responses, policies, and actions to enable compliance with latest standards including the NIST AI Framework and the EU AI Act. Zscaler does not store prompts or responses. All customer data is stored in an Amazon S3 bucket with a customer key to secure access.



The image shows two side-by-side screenshots of the Zscaler AI Governance interface. The left screenshot is titled 'AI Compliance & Regulation Information' and shows sections for HIPAA Coverage, EU AI Act Coverage, and NIST AI Framework Coverage. The NIST section includes a list of AI system impacts and risks, mitigation strategies, and organizational measures. The right screenshot is titled 'Audit Log' and shows a table of audit events with columns for Timestamp, Email address, Action, Entity Type, and Details. The table includes rows for various events on January 6, 2024, such as 'Update Policy Control' and 'Create Application Group'.

## Deploy AI Apps with Full Confidence

Protect your AI apps in production by identifying threats and enforcing guardrails

BOOK A DEMO

### About Zscaler

Zscaler (NASDAQ: ZS) accelerates digital transformation so customers can be more agile, efficient, resilient, and secure. The Zscaler Zero Trust Exchange™ platform protects thousands of customers from cyberattacks and data loss by securely connecting users, devices, and applications in any location. Distributed across more than 150 data centers globally, the SSE-based Zero Trust Exchange™ is the world's largest in-line cloud security platform. Learn more at [zscaler.com](https://zscaler.com) or follow us on Twitter @zscaler.

© 2025 Zscaler, Inc. All rights reserved. Zscaler™ and other trademarks listed at [zscaler.com/legal/trademarks](https://zscaler.com/legal/trademarks) are either (i) registered trademarks or service marks or (ii) trademarks or service marks of Zscaler, Inc. in the United States and/or other countries. Any other trademarks are the properties of their respective owners.