

AVANTech-Day 2026



IT-Security Deep Dive





Kontrolle und Schutz im Zeitalter von KI – Der Zscaler-Ansatz für GenAI Security

Jonas Kugler

Senior Security Engineer,
AVANTEC



Agenda

- Welche Gefahren und Risiken bringt AI?
- Die Zscaler AI Security Platform
- User-Zugriff auf Public AIs – geschützt mit Zscaler
- Ausblick

Welche Gefahren und Risiken bringt AI?

AI in den News

Claude-Code-Vorfall: npm-Paketierungsfehler wird für Malware-Kampagne über GitHub genutzt

8. April 2026

Ein Fehler im npm-Release von Claude Code dient weiterhin als Ausgangspunkt für eine laufende Malware-Kampagne. IT-Sicherheitsexperten von Trend Micro beobachten die Aktivitäten und analysieren die Vorgehensweise der Angreifer.

BREAKING | BUSINESS

Samsung Bans ChatGPT Sensitive Code Leak

BUSINESS INSIDER

Airline held liable for its chatbot giving passenger bad advice - what this means for travellers

23 February 2024

CYBERSECURITY | SECURITY NEWSWIRE | CYBERSECURITY NEWS

Vercel Breach Originated From an Employee's AI Tool

By Jordyn Alger, Managing Editor

+5.88% MSFT +1.67% TSLA +0.2% AMZN +3.32% META +2.21% DOW -0.16% NASDAQ +

added an AI chatbot to its site. Then all

ir Correspondent covering technology and culture

Share Save

Add as preferred on Google

AI coding tool wipes production database, fabricates 4,000 users, and lies to cover its tracks

Published: 21 July 2025 · Last updated: 23 July 2025

Welche Gefahren und Risiken bringt AI?

- Verlust von Kontrolle und Visibilität
- Datenabfluss
 - PII (Personally Identifiable Information)
 - Vertrauliche Firmendaten / Geschäftsgeheimnisse
- Rufschädigung
- Klagen und/oder Bussen
- Fehlverhalten / Mangelnde Datenqualität

Die Zscaler AI Security Plattform

Eine Übersicht



AI Discovery and Risk

Ermöglicht das Entdecken und Klassifizieren von AI Assets und das Analysieren von Risiken.

- GenAI Security Report
- Prompt Logging
- DLP Monitoring für AI



AI Red Teaming

Ermöglicht fortlaufendes Testing eigener AI-Applikationen und gibt Empfehlungen um Vulnerabilitäten zu mitigieren.

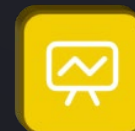
- Einfaches Risiko-Assesment
- Vordefinierte + Custom Probes und Prompts
- Vielfältige Input-Methoden



AI Guardrails

Ermöglicht das Überwachen von AI Flows, um ungewollte Prompts und Antworten zu blockieren und unerwünschten Datenabfluss zu verhindern.

- Proxy Mode (inline) oder DAS Mode (via API)
- Schutz von Usern wie auch AI-Applikationen/-Agents
- Vielzahl an Detektoren



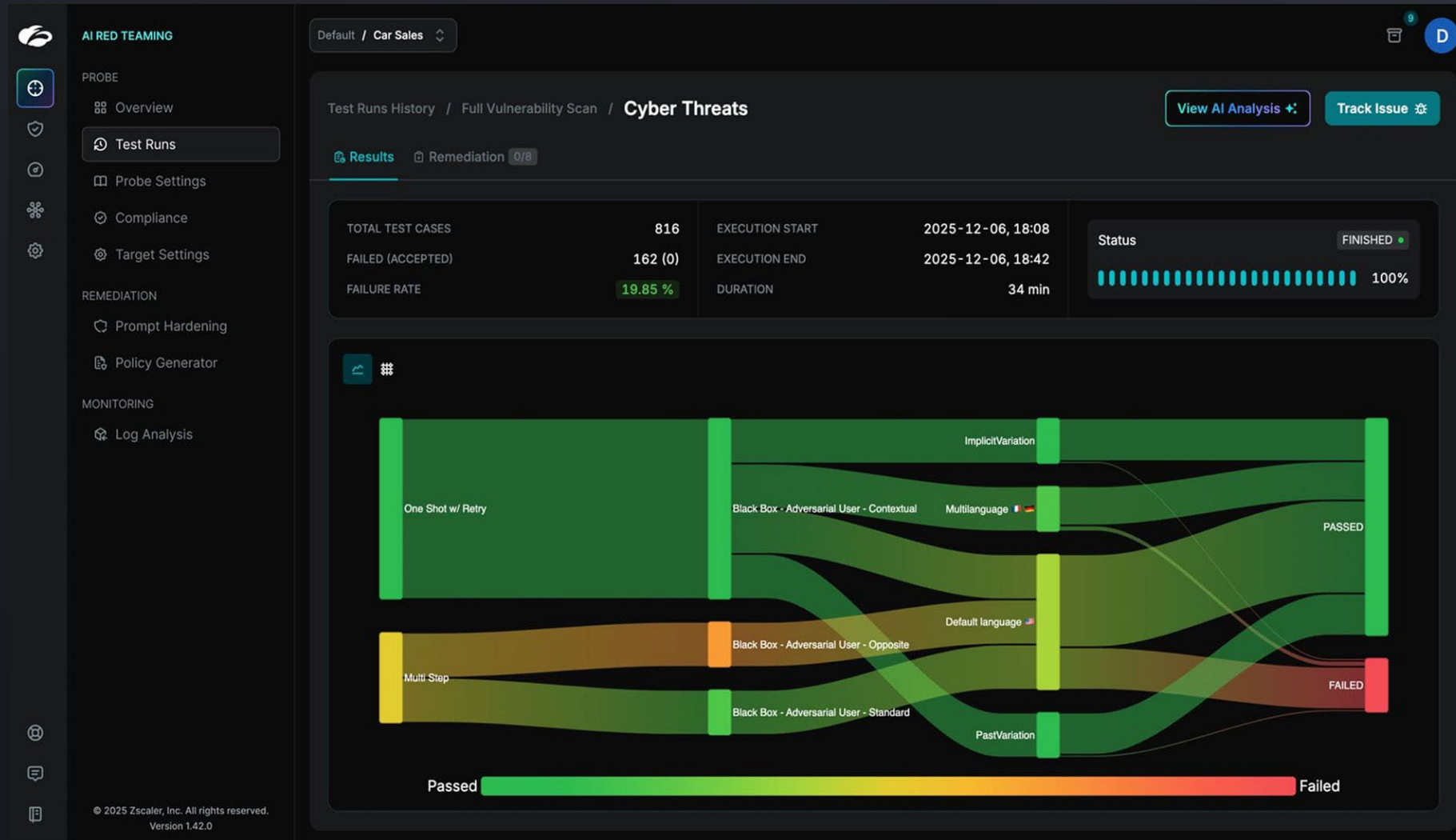
AI Governance and Compliance

Erlaubt das Überwachen von AI Systemen und das Prüfen, ob diese compliant sind mit entsprechenden AI-Frameworks.

- Alerting zu Issues
- Visualisierung von Abhängigkeiten
- Vordefinierte + Custom Security Policies

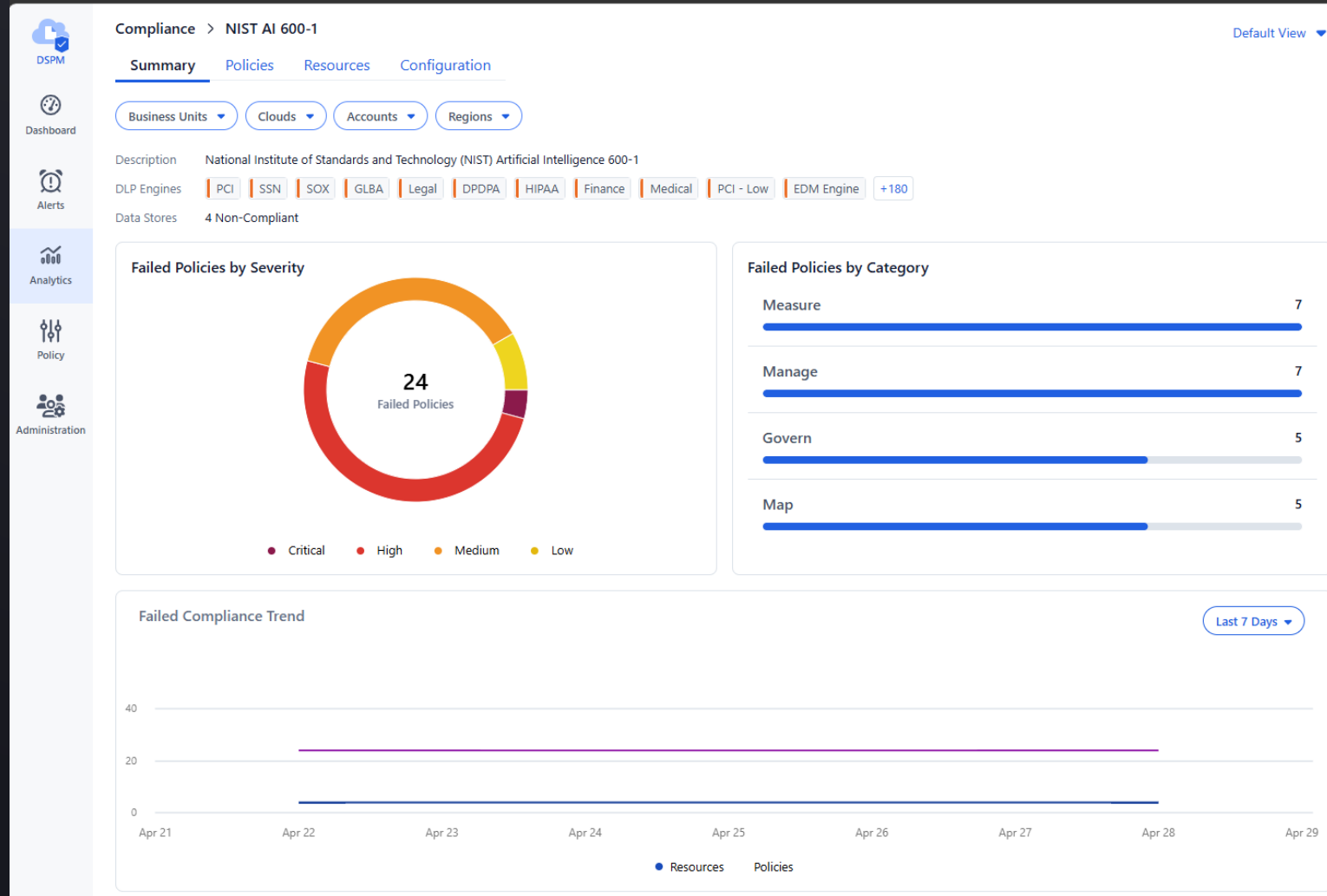
Die Zscaler AI Security Platform

AI Red Teaming



Die Zscaler AI Security Platform

AI Governance and Compliance – DSPM / AI-SPM



Die Zscaler AI Security Platform

AI Discovery and Risk



URL & Cloud App Control

Configure URL & Cloud App Control Policy
Rules are evaluated in the order specified. Rules are evaluated in the order specified.

URL Filtering Policy Cloud App Control

GEN AI PROMPT CONFIGURATION

- ChatGPT
- Microsoft Copilot
- Gemini

ChatGPT
Microsoft Copilot
Gemini
Perplexity
Poe
WRITER
DeepSeek
Grok AI
Claude
Mistral
Grammarly

ACTION

Application Access

Allow
 Caution
 Block
 Isolate

Embedded AI **Embedded AI Prompts**

Allow Block

Gen AI Security Report Last 7 Days Time Updated: Apr 28, 2026 02:00 AM

Gen AI Applications: 31 (Sanctioned: 0, Unsanctioned: 31)
 Transactions to Gen AI: 91.2 K
 Sensitive Data to Gen AI: 59
 Users Accessed Gen AI: 93
 Total Files Uploaded: 52

View Prompts > Analyze More >

Gen AI Application Usage Status = All Transactions

Gen AI Usage Trends Transactions

Sensitive Data Transactions Transactions

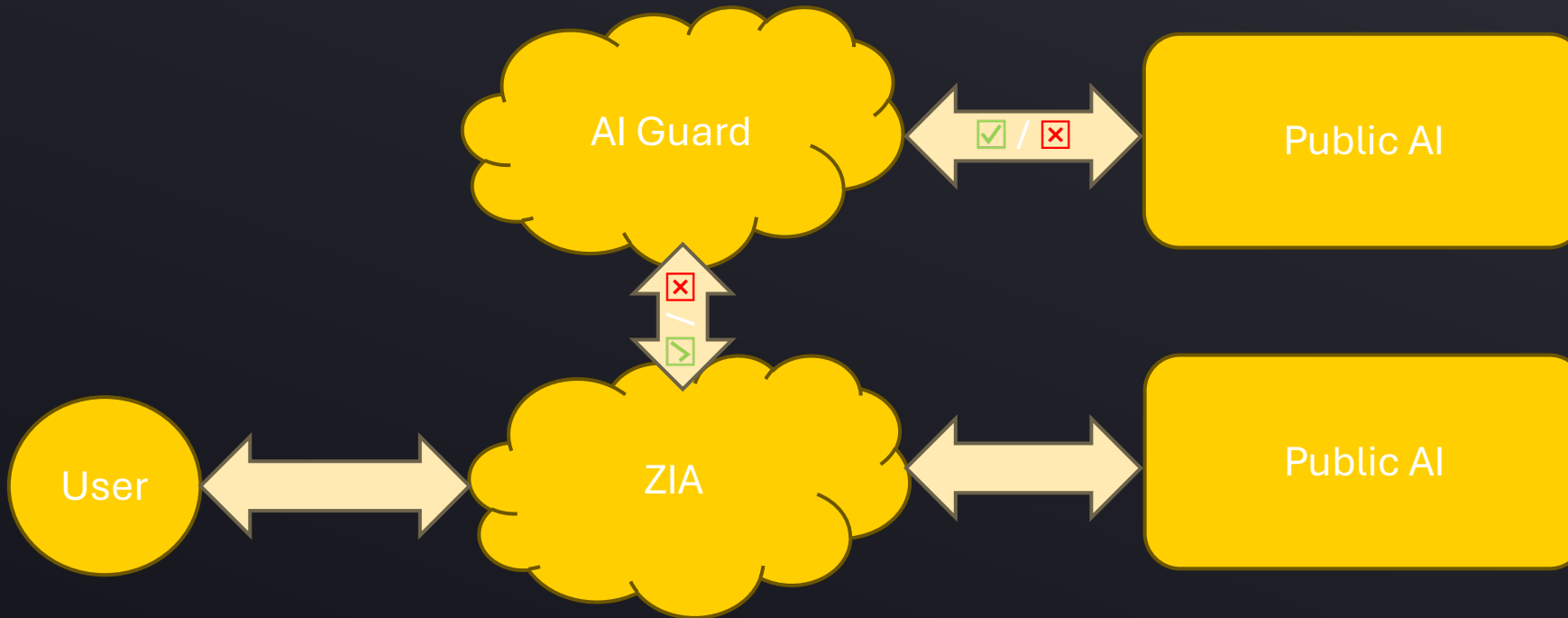
Gen AI Usage by Department Top AI Applications

Department	Usage
IT	40.4
Default Dep...	31 K
HR Document...	6.5 K
ZAdmin	4.4 K
Sales	3.5 K
OT	1.6 K
Marketing	4
Contractor	3

Die Zscaler AI Security Platform

AI Guardrails - User vs. AI App

Schutz von User

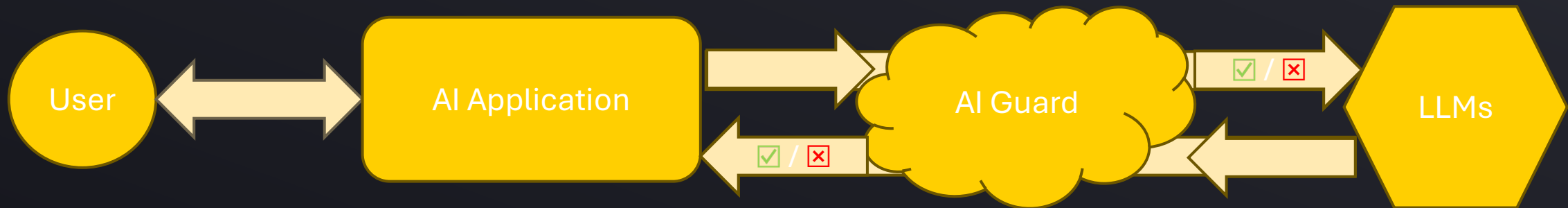


Anthropic
Azure
Bedrock
Bolt
DeepAI
Glean
Google Gemini
Lovable
MaxAI
Microsoft Copilot
Napkin
OpenAI
OpenCode
PerplexityAI
Vertex
xAI

Die Zscaler AI Security Platform

AI Guardrails - User vs. AI App

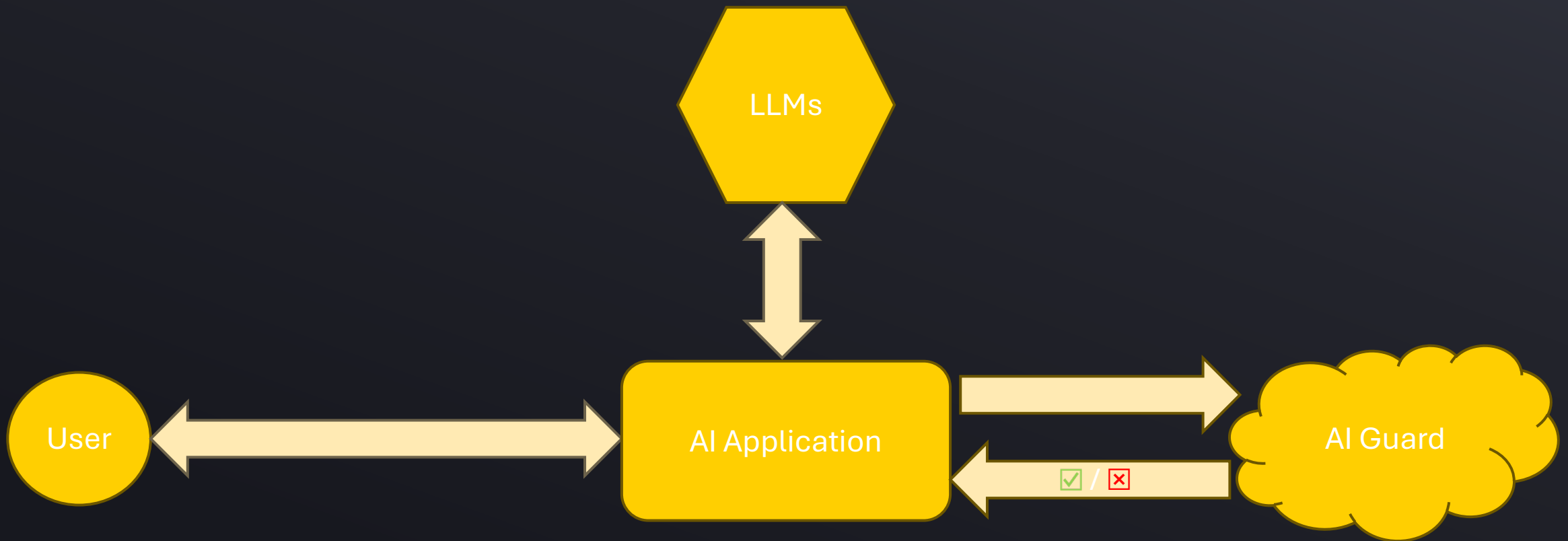
Schutz von AI Apps – Proxy Mode



Die Zscaler AI Security Platform

AI Guardrails - User vs. AI App

Schutz von AI Apps – DAS Mode



Die Zscaler AI Security Plattform

Eine Übersicht



AI Discovery and Risk

Ermöglicht das Entdecken und Klassifizieren von AI Assets und das Analysieren von Risiken.

- GenAI Security Report
- Prompt Logging
- DLP Monitoring für AI



AI Red Teaming

Ermöglicht fortlaufendes Testing eigener AI-Applikationen und gibt Empfehlungen um Vulnerabilitäten zu mitigieren.

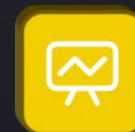
- Einfaches Risiko-Assesment
- Vordefinierte + Custom Probes und Prompts
- Vielfältige Input-Methoden



AI Guardrails

Ermöglicht das Überwachen von AI Flows, um ungewollte Prompts und Antworten zu blockieren und unerwünschten Datenabfluss zu verhindern.

- Proxy Mode (inline) oder DAS Mode (via API)
- Schutz von Usern wie auch LLMs
- Vielzahl an Detectors



AI Governance and Compliance

Erlaubt das Überwachen von AI Systemen und das Prüfen, ob diese compliant sind mit entsprechenden AI-Frameworks.

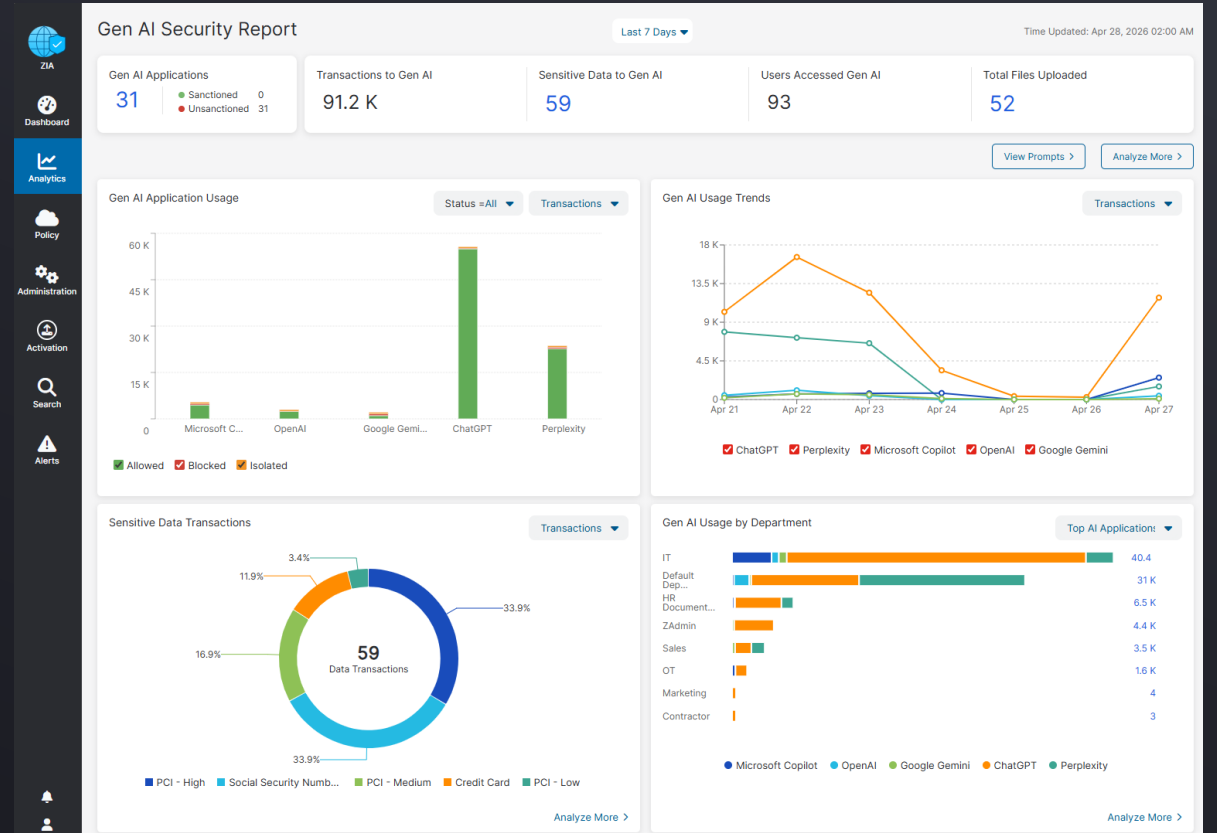
- Alerting zu Issues
- Visualisierung von Abhängigkeiten
- Vordefinierte + Custom Security Policies

User-Zugriff auf Public AIs

GenAI Security in ZIA

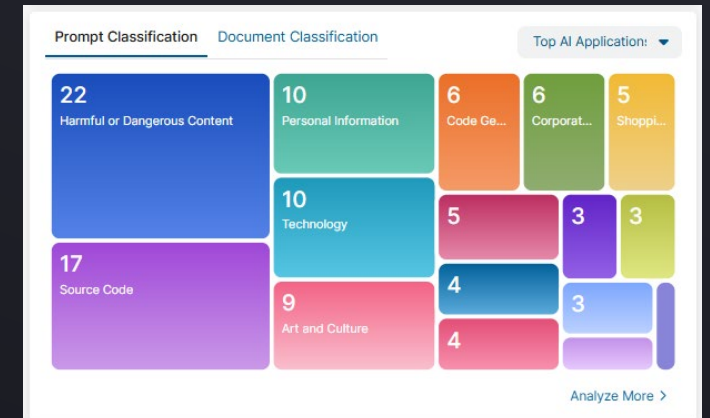
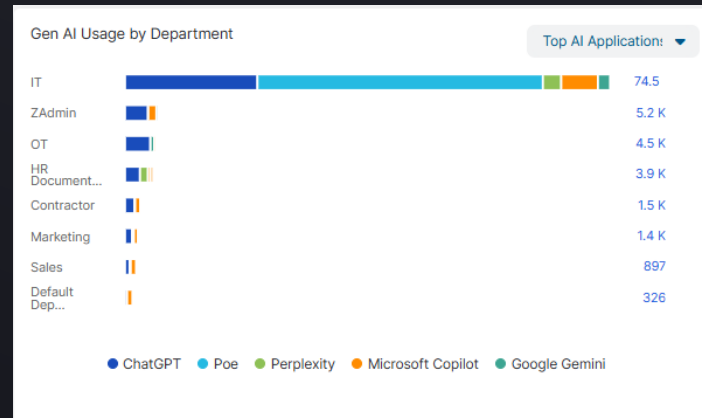
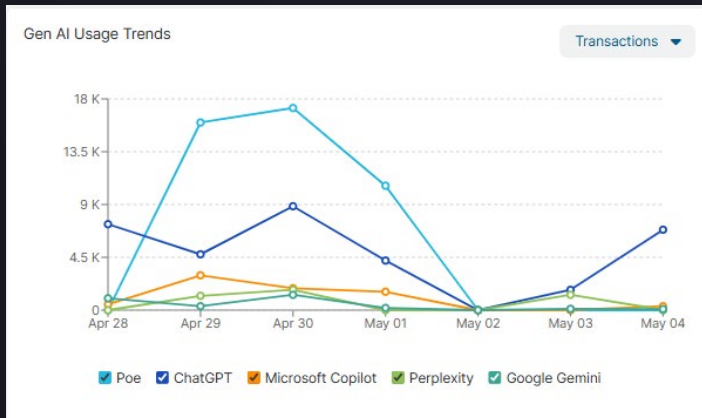
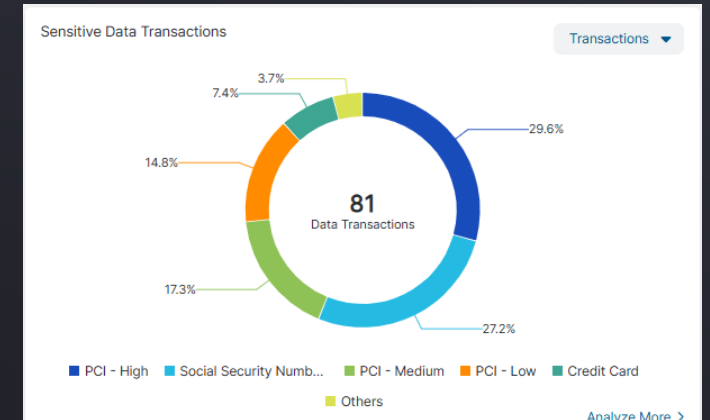
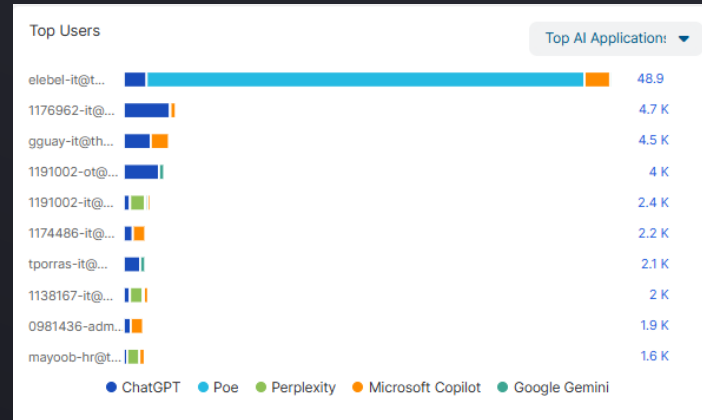
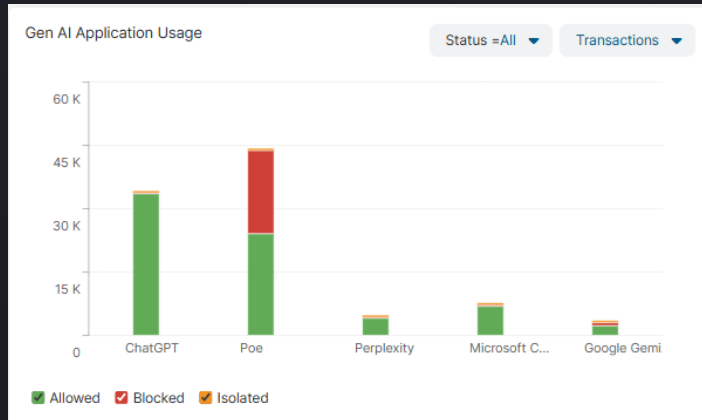
- Prerequisites

- Gen AI Security Subscription



User-Zugriff auf Public AIs

GenAI Security in ZIA – Analyse



User-Zugriff auf Public AIs

GenAI Security in ZIA – Analyse

Gen AI Security Report

Gen AI Applications

23

● Sanctioned 0
● Unsanctioned 23

← Gen AI Application Usage

Report generated: 2:00 AM, May 04, 2026



Application	Applicatio...	Applicatio...	Users	Loca...	Total Bytes	Upload Bytes	Download Bytes	
ChatGPT	● 3	⊗ Unsanctio...	55	---	686.32 MB	280.78 MB	405.54 MB	
Microsoft Copik	● 2	⊗ Unsanctio...	12	---	296.87 MB	19.10 MB	277.77 MB	
Google Gemini	● 2	⊗ Unsanctio...	23	---	130.51 MB	116.36 MB	14.15 MB	
Perplexity	● 2	⊗ Unsanctio...	14	---	71.58 MB	4.30 MB	67.28 MB	
Poe	● 3	⊗ Unsanctio...	3	---	42.30 MB	29.79 MB	12.52 MB	
OpenAI	● 2	⊗ Unsanctio...	39	---	35.75 MB	1.60 MB	34.14 MB	
Meta AI	● 2	⊗ Unsanctio...	6	---	31.69 MB	1.61 MB	30.07 MB	
Deepgram	● 4	⊗ Unsanctio...	0	1	23.61 MB	581.09 KB	23.04 MB	
Manus	● 2	⊗ Unsanctio...	10	---	20.23 MB	425.35 KB	19.82 MB	
Grammarly	● 2	⊗ Unsanctio...	5	---	19.27 MB	448.89 KB	18.83 MB	
Monica	● 2	⊗ Unsanctio...	1	---	10.03 MB	373.86 KB	9.67 MB	
Glean	● 2	⊗ Unsanctio...	8	---	3.57 MB	102.09 KB	3.47 MB	

User-Zugriff auf Public AIs

GenAI Security in ZIA – Risk-based AI Policy

Edit Predefined Cloud Application

GENERAL INFORMATION













Cloud Application Name	Application Category
ChatGPT	AI & ML Applications
Application Status	Risk Index
Sanctioned	3
Tags	
Any	

Save Cancel

Risk Profiles

Cloud Application

+ Add Cloud Application Risk Profile

No.	Profile ...	Application Status	Risk Index	Certifications Supported	Details	
1	High Risk GenAI Apps	Unsanctioned	4, 5	---	SECURITY INFORMATION: Data Encryption in Transit-Any	  
2	Medium Risk GenAI Apps	Unsanctioned	3	---	SECURITY INFORMATION: Data Encryption in Transit-Any	  
3	Low Risk GenAI Apps	Unsanctioned	1, 2	---	SECURITY INFORMATION: Data Encryption in Transit-Any	  
4	Sanctioned GenAI Apps	Sanctioned	---	---	SECURITY INFORMATION: Data Encryption in Transit-Any	  

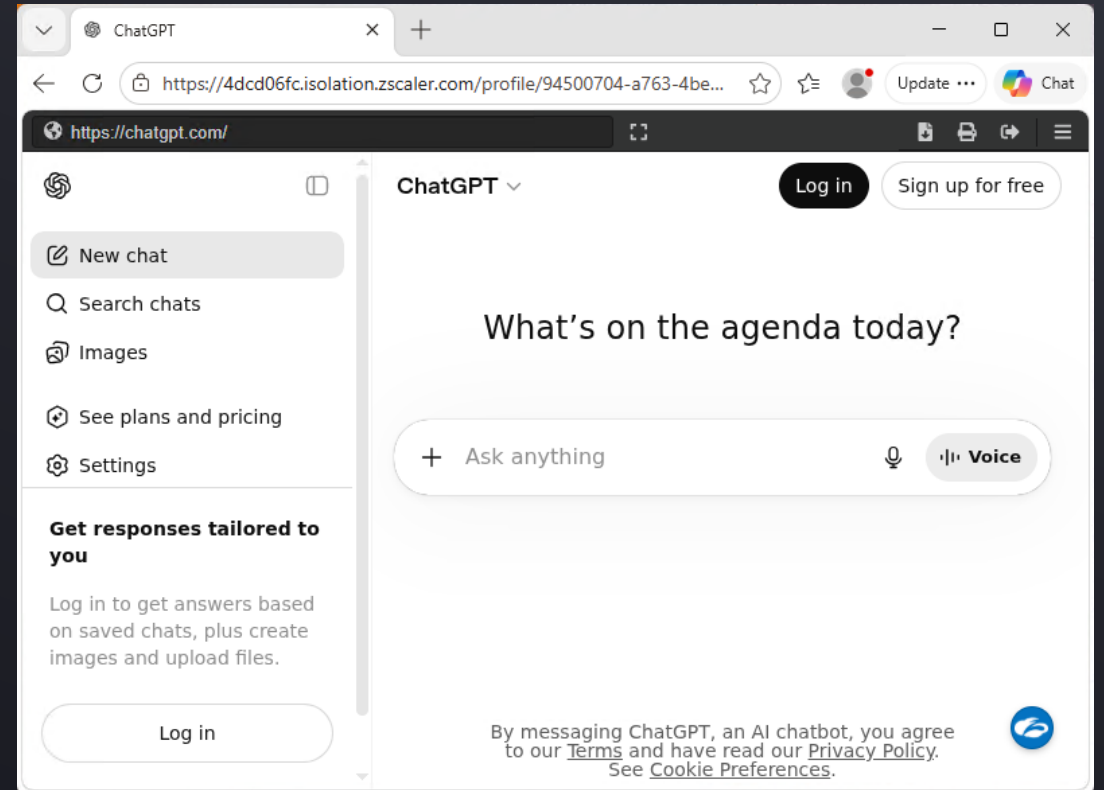
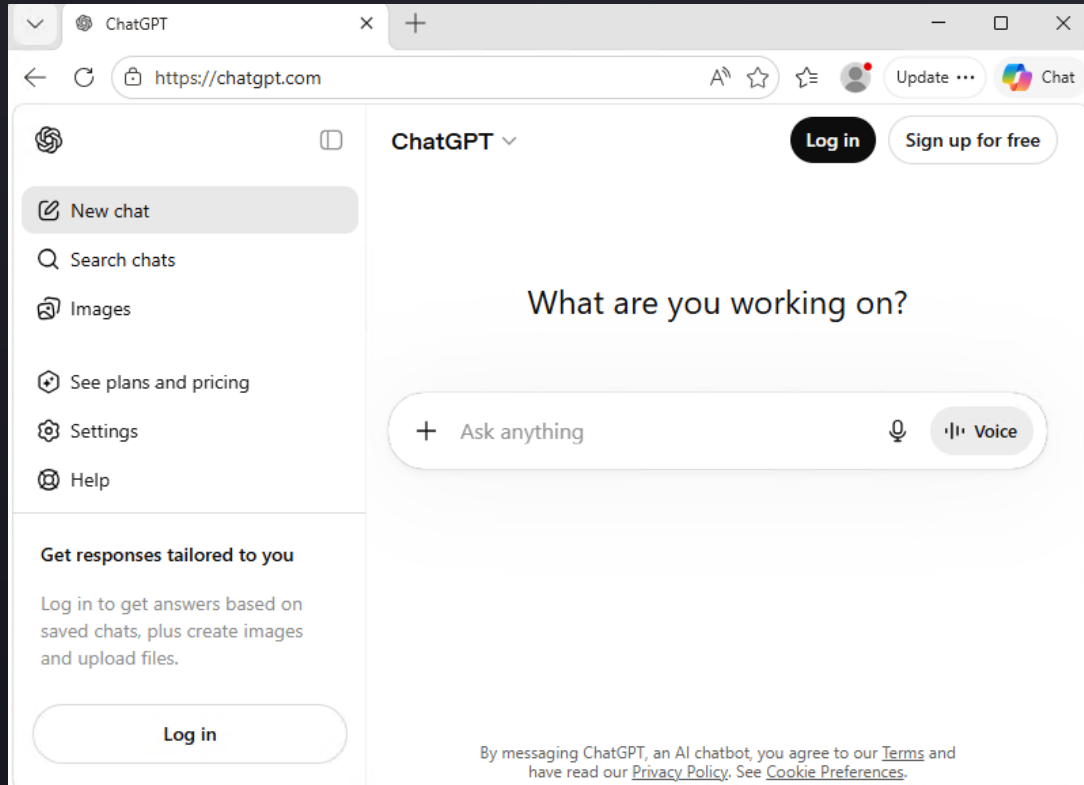
User-Zugriff auf Public AIs

GenAI Security in ZIA – Risk-based AI Policy

URL Filtering Policy	Cloud App Control Policy	Advanced Policy Settings		
Add ▼				
ℹ Recommended Policy				
Rule Ord...	Rule Name	Criteria	Action	Label and Description
AI & ML APPLICATIONS				
1	Allow Sanctioned GenAI	CLOUD APPLICATION RISK PROFILE Sanctioned GenAI Apps	Allow Application Access	
2	Allow Low Risk GenAI	CLOUD APPLICATION RISK PROFILE Low Risk GenAI Apps	Allow Application Access	
3	Isolate Medium Risk GenAI	USER AGENT Opera; Firefox; Microsoft Internet Explorer; ... CLOUD APPLICATION RISK PROFILE Medium Risk GenAI Apps	Isolate Application Access Isolation Profile: GenAI Isolation	
4	Block High Risk GenAI	CLOUD APPLICATION RISK PROFILE High Risk GenAI Apps	Block Application Access	

User-Zugriff auf Public AIs

GenAI Security in ZIA – Risk-based AI Policy



User-Zugriff auf Public AIs

Zscaler AI Guardrails

- Prerequisites
 - AI Guard Subscription
 - ZIA und AI Guard sind gelinked
 - Liste der genutzten (und unterstützen) GenAI Apps

Application	Client Types	Domains Required
Grok (xAI)	○ Web Browsers	.grok.com
Lovable	○ Web Browsers	api.lovable.dev
MaxAI	○ Web Browsers	.api.maxai.me

Tenant Settings

Configure your tenant-wide settings and preferences.

General Security **Integrations**

ZIA Information

Enable synchronization with ZIA to import users, groups, and domains for policy evaluation.

Changes made in AI Guard do not impact any of your ZIA settings.

Note: It is recommended to enable both ZIA user/group sync and domain sync.

Enable ZIA User and Group Sync

Enable this to sync users and groups from ZIA. Cloud Name and Organization ID will be displayed below once sync is enabled and information is available.

Enable ZIA Domain Sync

Enable this to sync domains from ZIA to ensure all relevant domains are synchronized for policy evaluation.

Download Zscaler AI Guard Proxy Chain Certificate

Download

Cloud Name

zscloud.net

Organization ID

103

User-Zugriff auf Public AIs

Zscaler AI Guardrails – Anbindung an ZIA

Schritt 1 – Root Certificate

[Download Zscaler AI Guard Proxy Chain Certificate](#) [Download](#)

Root Certificates

[+ Add Root Certificate](#)

No.	Certificate Name	Uploaded ...	Expiration D...	Type
1	AI Guard	March 31, 20...	April 02, 2035 ...	Proxy Chaining

Schritt 2 – Proxy & Gateway

Add Proxy

GENERAL INFO

Proxy Name: AI Guard

IP Address / FQDN: forward.zseclipse.net Port: 9443

Proxy's Root Certificate: AI Guard

Insert X-Authenticated-User:

Add Gateway for Proxies

GENERAL INFO

Gateway Name: AI Guard

Fail Close:

Primary Proxy: AI Guard Secondary Proxy: None

Schritt 3 – Forwarding

Add Destination IPv4 Group

DESTINATION GROUP

Name: AI Guard

Type: IP Address FQDN **Wildcard FQDN** Other

Wildcard FQDN

Add Items

Search...

- *perplexity.ai
- *chatgpt.com

1 - 2 of 2 < 1 / 1 > Remove

R...	Rule Name	Criteria	Forwarding Method	Gateway
REDIRECT/FORWARDING POLICY				
1	ForwardToAIGuard	DESTINATION IPV4 GROUPS AI Guard	Proxy Chaining	AI Guard

User-Zugriff auf Public AIs

Zscaler AI Guardrails – Anbindung an ZIA

Tenant Settings

Configure your tenant-wide settings and preferences.

General Security **Integrations**

ZIA Information

Enable synchronization with ZIA to import users, groups, and domains for policy evaluation.

Changes made in AI Guard do not impact any of your ZIA settings.

Note: It is recommended to enable both ZIA user/group sync and domain sync.

Enable ZIA User and Group Sync

Enable this to sync users and groups from ZIA. Cloud Name and Organization ID will be displayed below once sync is enabled and information is available.

Enable ZIA Domain Sync

Enable this to sync domains from ZIA to ensure all relevant domains are synchronized for policy evaluation.

Download Zscaler AI Guard Proxy Chain Certificate

Download

⚠ Advanced Actions

These actions may affect data synchronization. Use with caution.

Manual Batch Sync




















Triggers an immediate ZIA data sync outside of the scheduled batch window.

Start Sync

Last synced from ZIA: 05 May, 2026, 10:09 AM UTC

User-Zugriff auf Public AIs

Zscaler AI Guardrails – Policies für Prompts






















 Toxicity Detects and filters harmful language Multilingual	 Code Detect and block unwanted programming languages across your platforms Beta	 Prompt Injection Detect and prevent malicious or unauthorized modifications to input prompts Multilingual	 Brand and reputation risk Detects negative sentiment towards a brand Beta
 Text Detect and block sensitive text using regex patterns	 Gibberish Identify and filter out nonsensical or meaningless text Multilingual	 Competition Prevent the inclusion of competitor names in the prompts submitted by users Beta	 Language Detect and block unwanted languages across your platforms Multilingual
 Legal Advice Blocks prompts seeking legal advice, interpretation, or compliance guidance; allows neutral legal facts, definitions, and non-legal work inquiries.	 Intellectual Property Filter and control content for intellectual property	 Secrets Detect and block sensitive information such as API keys.	 Off Topic Filter and control content by topic description Beta
 PII Detect and block PII entities such as email, SSN	 Personal Data Identifies sensitive personal attributes and blocks invasive questions or confirmations about identity, background, or affiliations. Covers...	 PII DeepScan Detects and blocks attempts to share or solicit high-risk identifiers that directly expose financial, legal, or digital identity. Covers SSN, ITIN, passport, driver's license,...	 Topic Filter and control content by identifying custom topics
 Invisible Text Identifies hidden or obscured text within digital content	 Finance Advice Blocks actionable financial guidance (investing, trading, tax, product choices); allows neutral finance facts, history, and definitions. Multilingual	 Prompt Tags Filter and control prompts by predefined tags	

Back

Next



User-Zugriff auf Public AIs

Zscaler AI Guardrails – Policies für Responses

 Toxicity Detects and filters harmful language Multilingual	 Code Detect and block unwanted programming languages across your platforms Beta	 Malicious URL Identifies URLs with domains categorized as malicious	 Response Tags Filter and control prompt responses by predefined tags
 Brand and reputation risk Detects negative sentiment towards a brand Beta	 Refusal Identifies LLM refusal	 Text Detect and block sensitive text using regex patterns	 Gibberish Identify and filter out nonsensical or meaningless text Multilingual
 Competition Prevent the inclusion of competitor names in the prompts submitted by users Beta	 Language Detect and block unwanted languages across your platforms Multilingual	 Legal Advice Blocks prompts seeking legal advice, interpretation, or compliance guidance; allows neutral legal facts, definitions, and non-legal work inquiries.	 Intellectual Property Filter and control content for intellectual property
 Secrets Detect and block sensitive information such as API keys.	 Off Topic Filter and control content by topic description Beta	 PII Detect and block PII entities such as email, SSN	 Personal Data Identifies sensitive personal attributes and blocks invasive questions or confirmations about identity, background, or affiliations. Covers...
 PII DeepScan Detects and blocks attempts to share or solicit high-risk identifiers that directly expose financial, legal, or digital identity. Covers SSN, ITIN, passport, driver's...	 Topic Filter and control content by identifying custom topics	 URL Reachability URLs are accessible and functioning correctly by continuously testing and verifying link status in real time	 Invisible Text Identifies hidden or obscured text within digital content
 Finance Advice Blocks actionable financial guidance (investing, trading, tax, product choices); allows neutral finance facts, history, and definitions. Multilingual			

User-Zugriff auf Public AIs

Zscaler AI Guardrails – Test-Policy

Beta 
Code
Detect and block unwanted programming languages
across your platforms

Prompt:

Python

Detect Allow **Block**

Response:

C#

Detect Allow **Block**

User-Zugriff auf Public AIs

Zscaler AI Guardrails – Test-Policy in Action

```
for num in numbers:  
    if num % 2 == 0:  
        print(f"{num} is even")  
    else:  
        print(f"{num} is odd")
```

Request blocked by Zscaler AI Guard.



User	Policy Name	Severity	Prompt Detections	Response Detections	LLM	Prompt Action	Response Action
kugler@ava...	Test Policy	LOW	Code PII Prompt Tags	--		Blocked	--

▼ Prompt Details

Prompt Response

Prompt
what language is this:

A simple function to greet a user
def greet(name):
 return f"Hello, {name}!"


Using the function
message = greet("Alice")
print(message)

User-Zugriff auf Public AIs

Zscaler AI Guardrails – Test-Policy in Action

give me a C# code example

Part of your prompt's reply from LLM provider is blocked by Zscaler AI Guard. This conversation thread will be deleted, please open a new conversation chat to continue.

User	Policy Name	Severity	Prompt Detections	Response Detections	LLM	Prompt Action	Response Action
kugler@ava...	Test Policy	LOW	Prompt Tags	Code Response T... PII		Allowed	Blocked

▼ Prompt Details

Prompt Response

Response
Here's a simple C# example showing a function, a loop, and basic conditionals:

```
```csharp
using System;
using System.Collections.Generic;

class
```

# Ausblick

- Kontinuierliche Weiterentwicklung
- Direktere Integration in ZIA

Q&A

+

Diskussion

# AVANTech-Day 2026



# DANKKE



Nächste Session 14:45

# Session 3

## Zero Trust Branch (ZTB) mit Airgap – wozu ist das gut?

Produkt:  zscaler

Referenten: René von Arx, Raffael Späni

Holdener, 5. OG

## Thread Huntig Deep Dive: Von Hypothese zu Detection Engineering

Produkt:  CROWDSTRIKE

Referent: Alessandro Salucci

Frei, 4. OG

## Disaster Recovery im Cloud-Zeitalter

Produkte:  zscaler  netskope

Referent: Matthias Geiser

Lehmann Steingruber, 5. OG

## Authentication – passwordless, strong, secure

Produkt: **THALES**

Referent: Dirk Gluch

Odermatt, 4. OG

## Zscaler Automatisierung

Produkt:  zscaler

Referent: Christian Schnittert

Cancellara, 4. OG