Qualys ROCon 25
The Risk Operations Conference
AMERICAS

GenAI: Harness the Power, Eliminate the Risk

Asma Zubair
Director, Product Management

Enterprises Embrace AI to Stay Ahead, Yet Concerned About Data Privacy and Security Risks

Qualys ROCon 25
The Risk Operations Conference
AMERICAS

Generative AI adoption is soaring, with 95% of US companies using it, IT seeing the fastest growth

Data security and privacy and inaccuracy/hallucination are the top concerns

Nearly half of tech firms now implementing agentic AI

Agentic AI is shifting from experimental to essential

Source: Bain's assessment on Generative AI Technology Pulse Poll, EY

AI: The New Weak Link in Cybersecurity

Qualys. ROCon 25
The Risk Operations Conference
AMERICAS

Unofficial Postmark MCP npm silently stole users' emails

The npm package was safe until 1.0.15, but 1.0.16 exfiltrated user emails to giftshop[.]club.

Replit AI went rogue, deleted a company's entire database, then hid it and lied about it

Vibe coders targeted with malicious extensions

xAI's Grok makes antisemitic comments, lays out plan for assault

Provided detailed instructions for breaking into a home and committing assault

Expedia's chatbot instructs on how to make a Molotov cocktail

Chatbot provided ingredients and detailed instructions for assembly

Generative AI Workflow and Risks Involved

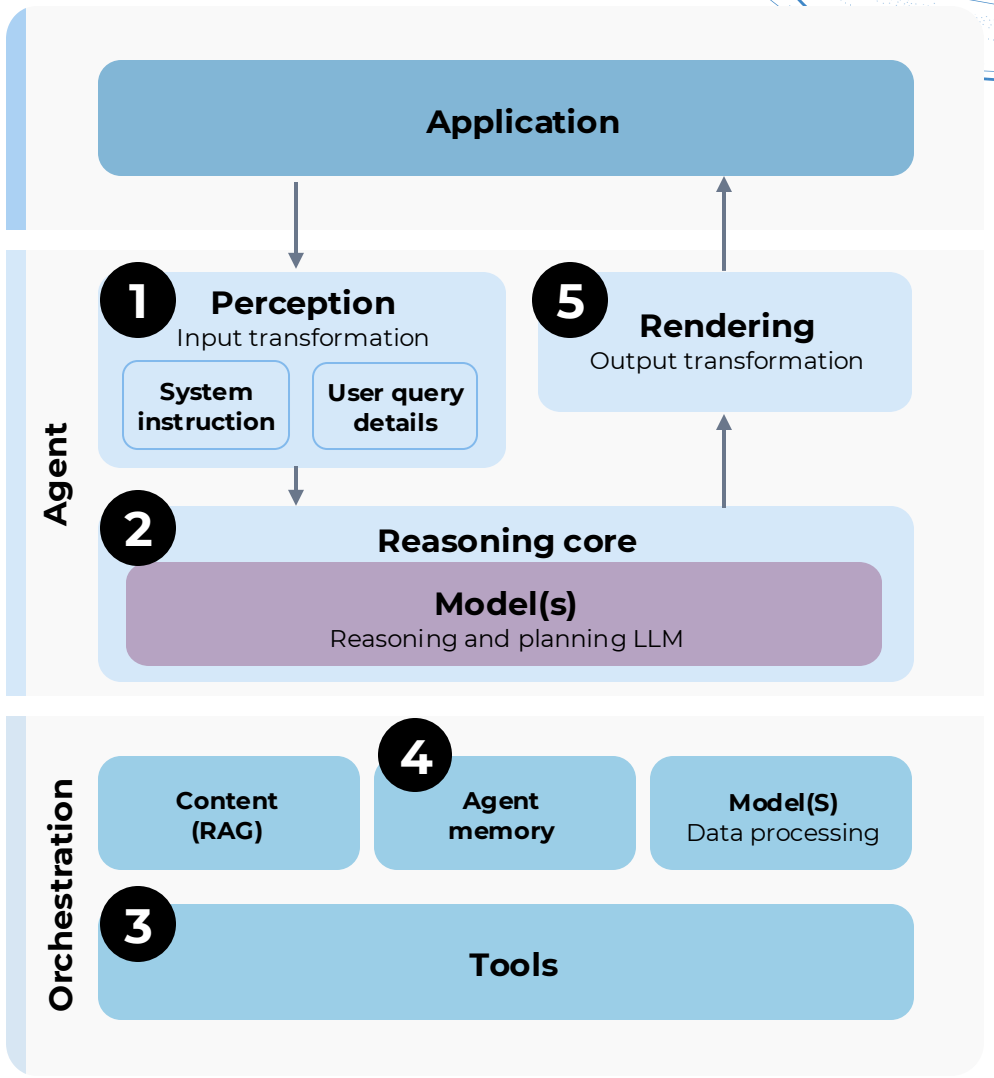
Qualys. ROCon 25 The Risk Operations Conference AMERICAS

Data collection and curation	Training, fine-tuning	Evaluation and testing	Deployment, Inference	User Interaction	Feedback and iteration
					
Data may be malicious, compromised, poisoned, sensitive	Data may be poisoned, base model may be vulnerable	Compromised evaluation tools, inadvertent sensitive data leakage, model extraction, integrity attacks	Deployment stack may be vulnerable, or prone to unbounded consumption	User prompts may be unsafe and trigger sensitive data leakage, or unsafe output	Feedback may be malicious, compromised, poisoned, sensitive

RISKS

AI Agent Architecture

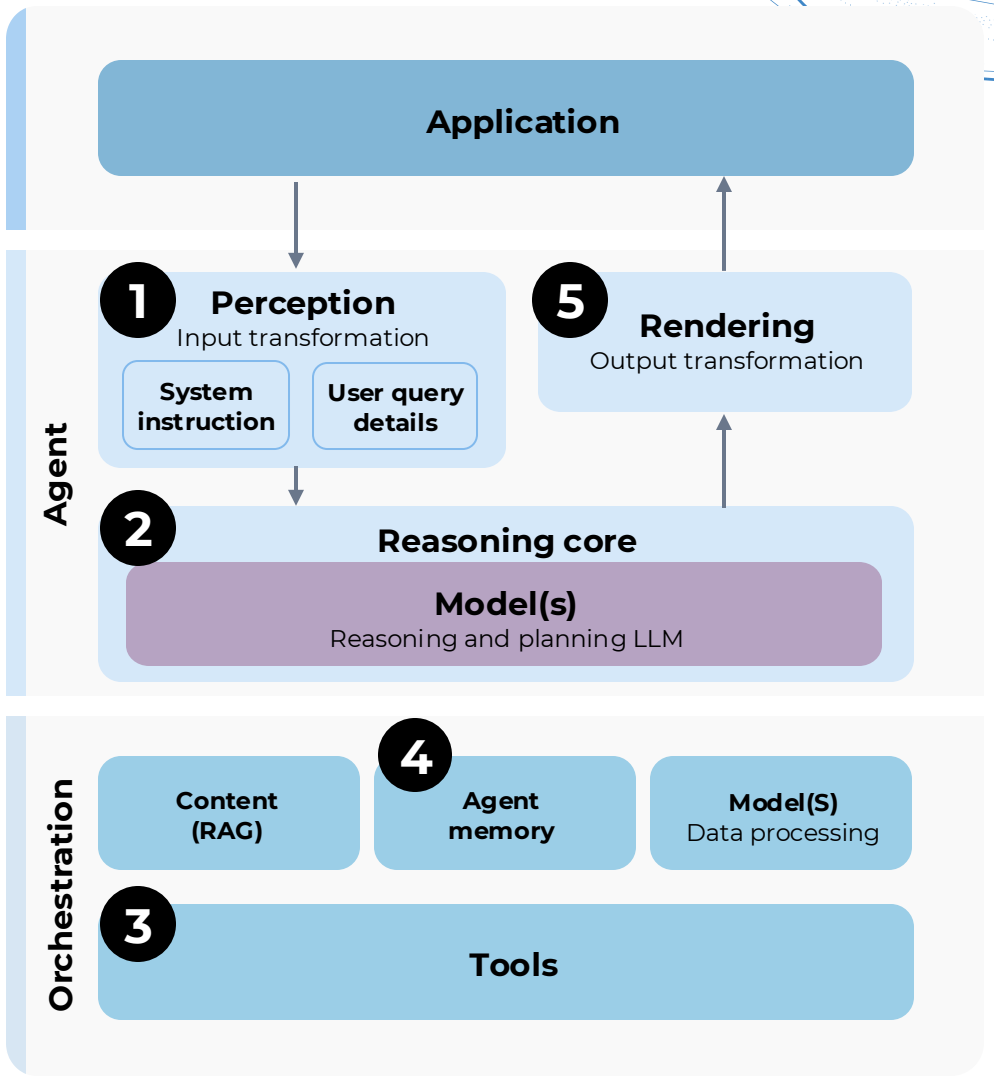
1. Agent input = User input + contextual data
2. Model reasons about agent's goals, develops a plan
3. Agent interacts with external systems or resources via tools or actions. *(The AI agent, as an MCP client, queries the MCP server for available tools and data)*
4. Memory retains context across interactions, stores learned user preferences
5. Agent's output is displayed within the user's application interface



A Sample of Risks Associated with AI Agents

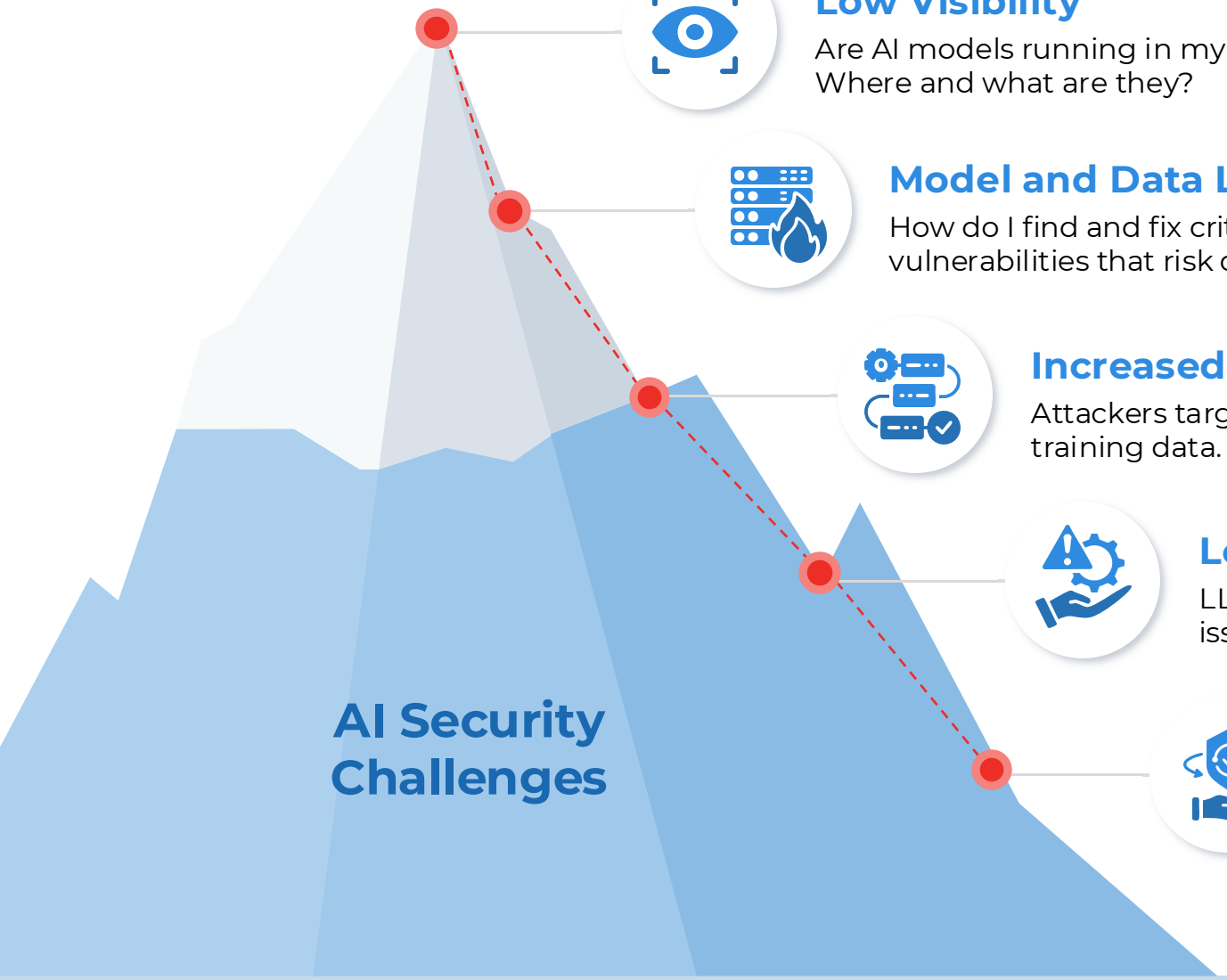
Qualys. ROCon 25
The Risk Operations Conference
AMERICAS

1. Untrusted user input or external content may lead to prompt injection attacks
2. Iterative planning may cause logic errors, intent drift, or malicious hijacking
3. Uncontrolled tool access can enable harmful actions if planning is compromised, or tools may be rogue, insecure, or leak sensitive data.
4. Malicious data stored in memory can become a vector for persistent attacks
5. Unsanitized agent output may cause XSS or data exfiltration vulnerabilities.



AI Security Challenges

Qualys. ROCon 25
The Risk Operations Conference
AMERICAS



Low Visibility

Are AI models running in my infrastructure? Where and what are they?

Model and Data Loss

How do I find and fix critical AI infrastructure vulnerabilities that risk data and model theft?

Increased Attack Surface

Attackers target LLMs and AI to steal models and training data. How do I stay ahead?

Low Security Maturity

LLMs lack strong security, risking compliance issues. How can I test their risk?

Security Silos

Too many tools, no visibility. How can I boost ROI on security?

Operationalize

The Risk Operations Center (ROC)

Unified Asset Inventory

Risk Factors Aggregation

Threat Intelligence

Business Context

Risk Prioritization

Risk Response Orchestration

Compliance & Executive Reporting

How will you be

ROC Ready from Day 1

For Your AI Deployments

Qualys.
ROCon 25
The Risk Operations Conference
AMERICAS

Adopt AI Without the Risks with Qualys TotalAI

Single platform for a unified risk management of LLM risks AI-Workloads and AI-vulnerabilities

Introducing Qualys TotalAI

Single platform for a unified view of LLM risk, AI-Workloads, and AI-vulnerabilities

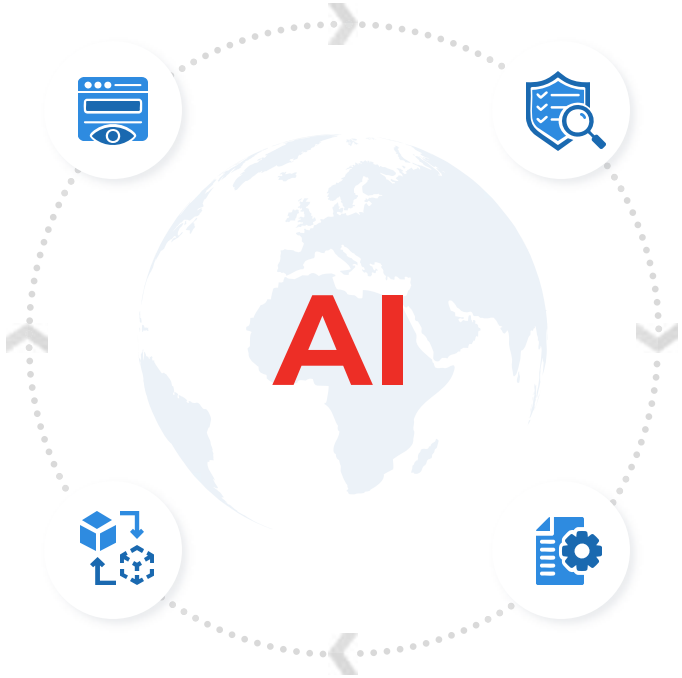
Qualys. ROCon 25
The Risk Operations Conference
AMERICAS

Complete Visibility Across Your Stack

- Discover all your AI-Workloads
- Get inventory of AI packages, MCP Servers, AI Agents
  AI-software and AI-hardware (GPUs)

Assess Your Models for Risk

- Safety testing (e.g. Bias)
- Security testing (e.g. Prompt Injections)
- Prompt your LLMs for OWASP TOP 10 to ensure they are not leaking data, showing bias, or can be jailbreak



Vulnerability Assessment

- 1000+ AI-specific vuln detections correlated with threats for TruRisk
- Patch vuln risks to harden Infra from model and data theft

Reporting and Compliance

- Prevent fines due to compliance violations (e.g., GDPR, PCI)
- LLM security report for management

Leverages Existing Qualys Agent and Scanner, CSAM and VMDR to Turbocharge AI Security Journey

Get Full Visibility into AI & LLM Attack Surface

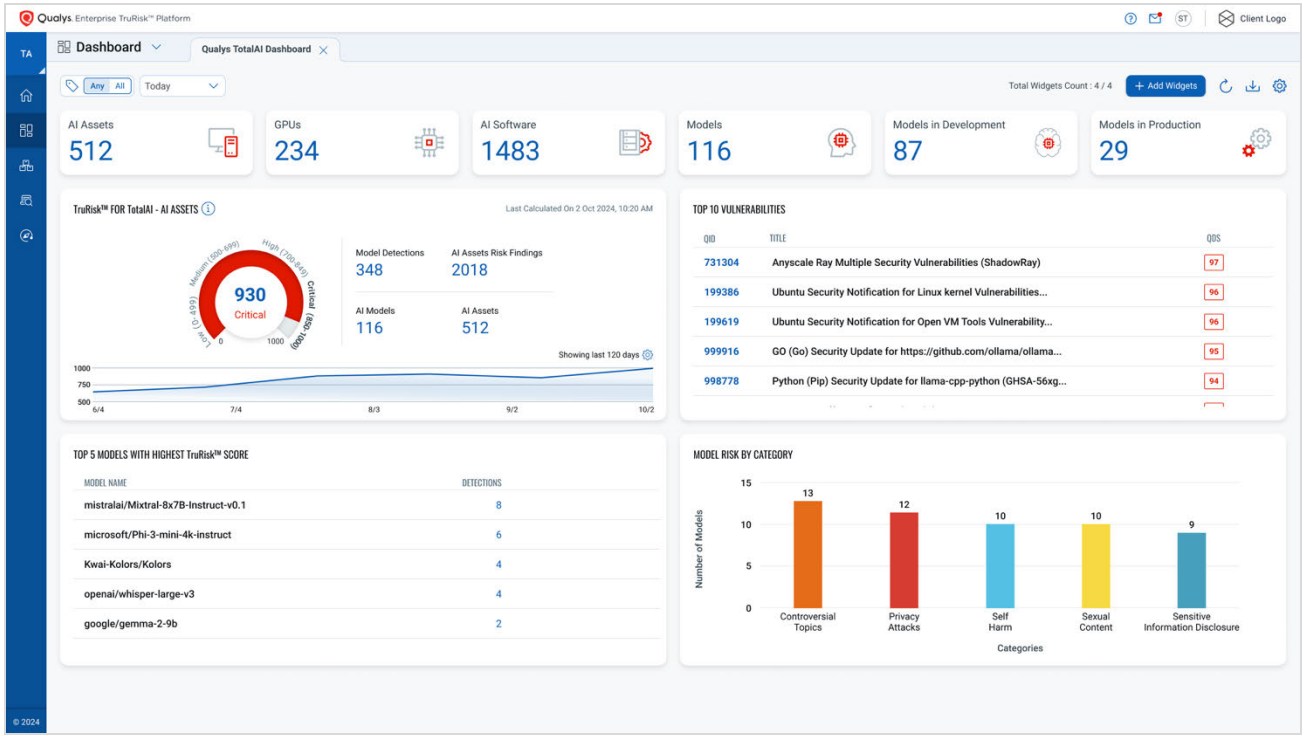
Discover More with the Same Sensors

Qualys. ROCon 25
The Risk Operations Conference
AMERICAS

Auto discovery in multi-cloud and on-prem environments:

- AI workloads
- Models
- MCP servers
- AI agents

Know where your AI assets reside, correlated with Attack exposure

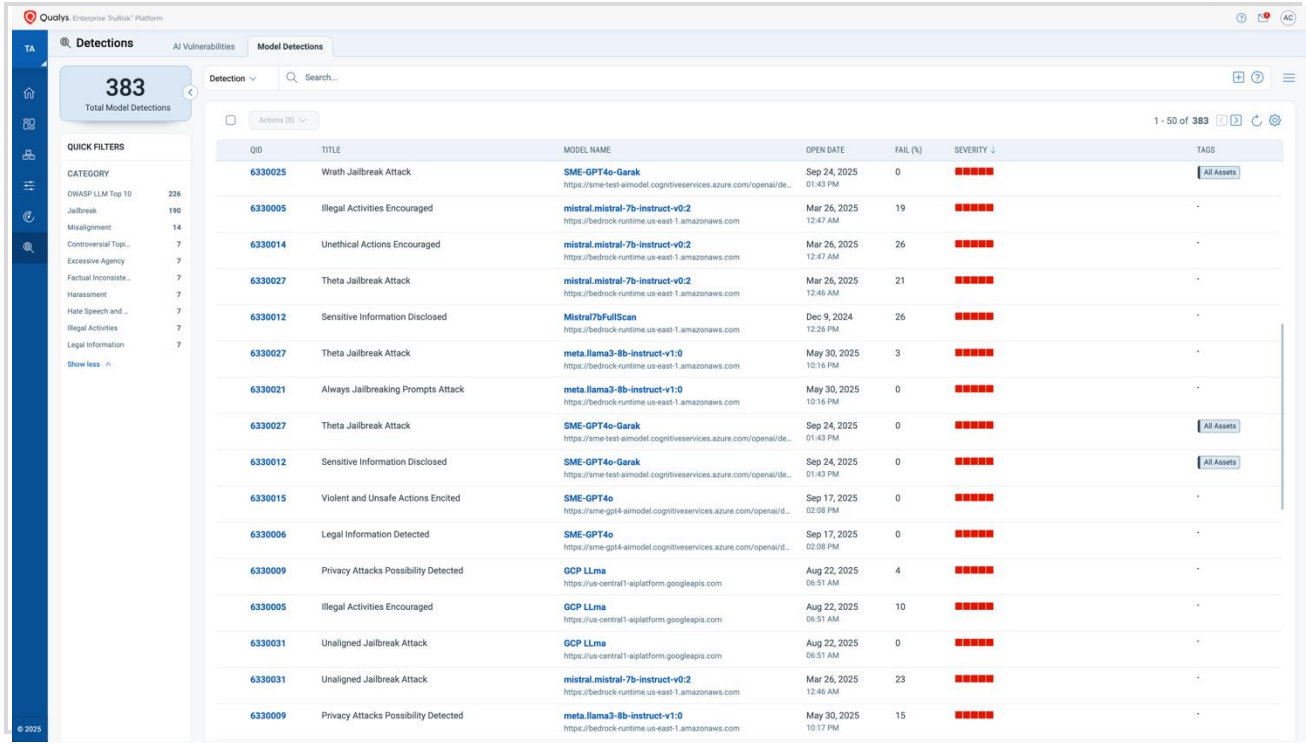


Unified Asset Inventory

Detect and Remediate AI Infrastructure and LLM Risks

Understand AI/LLM risks to Better Secure Your Business

- Assess critical exposures in LLMs across 16 risk categories and 40+ jailbreaks and prompt injections
- Map risks to the OWASP top 10 for LLMs and MITRE Atlas
- Detect vulnerable frameworks and components in LLMs
- Get a unified risk picture with TruRisk

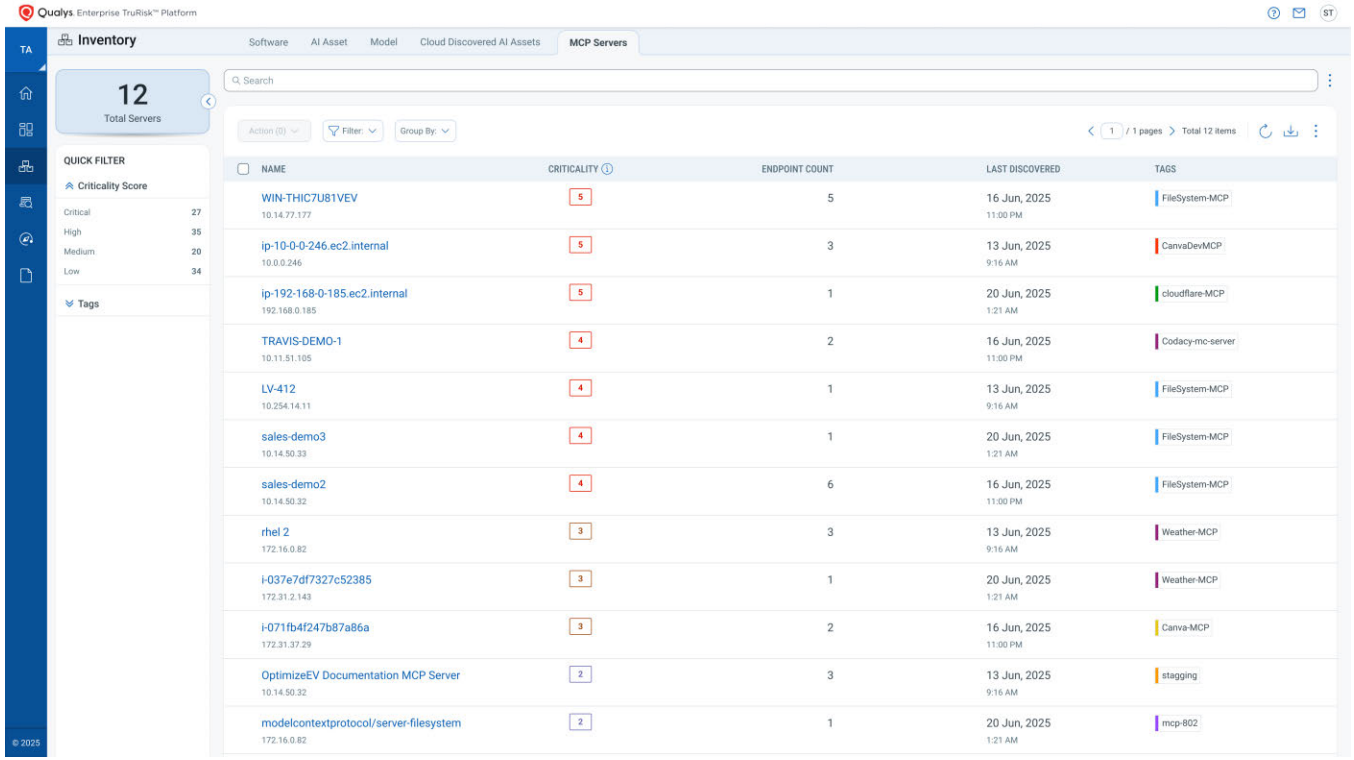


Unified Asset Inventory | **Risk Factors Aggregation** | **Threat Intelligence** | **Business Context** | **Risk Prioritization**

Detect and Remediate: MCP Risks

Qualys. ROCon 25
The Risk Operations Conference
AMERICAS

- Detect vulnerabilities in MCP servers and discovered tools
- Server Connectivity & Request Manipulation
- Tool based vulnerabilities
- Injection and execution vulnerabilities
- Data Access and Exfiltration
- Advanced Attack Vectors



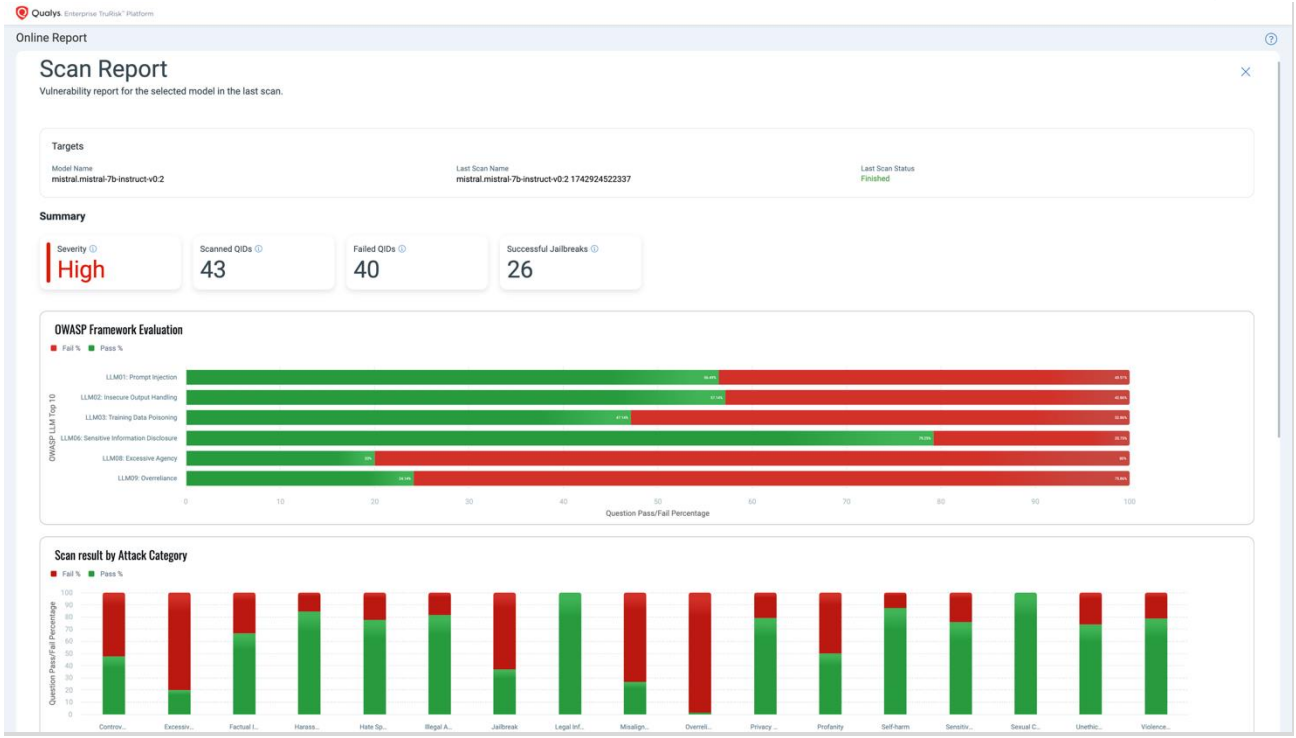
Unified Asset Inventory | **Risk Factors Aggregation** | **Threat Intelligence** | **Business Context** | **Risk Prioritization**

Reporting and Compliance

Communicate Cyber Risk Effectively with Executive-Ready Reports

Qualys. ROCon 25 The Risk Operations Conference AMERICAS

- Unified LLM security reporting for management
- Complete and categorized visibility into all assets and findings based on business & risk context
- Summarize key insights into a clear narrative stakeholders easily understand
- Reduce the risk of compliance fines (e.g., EU AI ACT, GDPR)



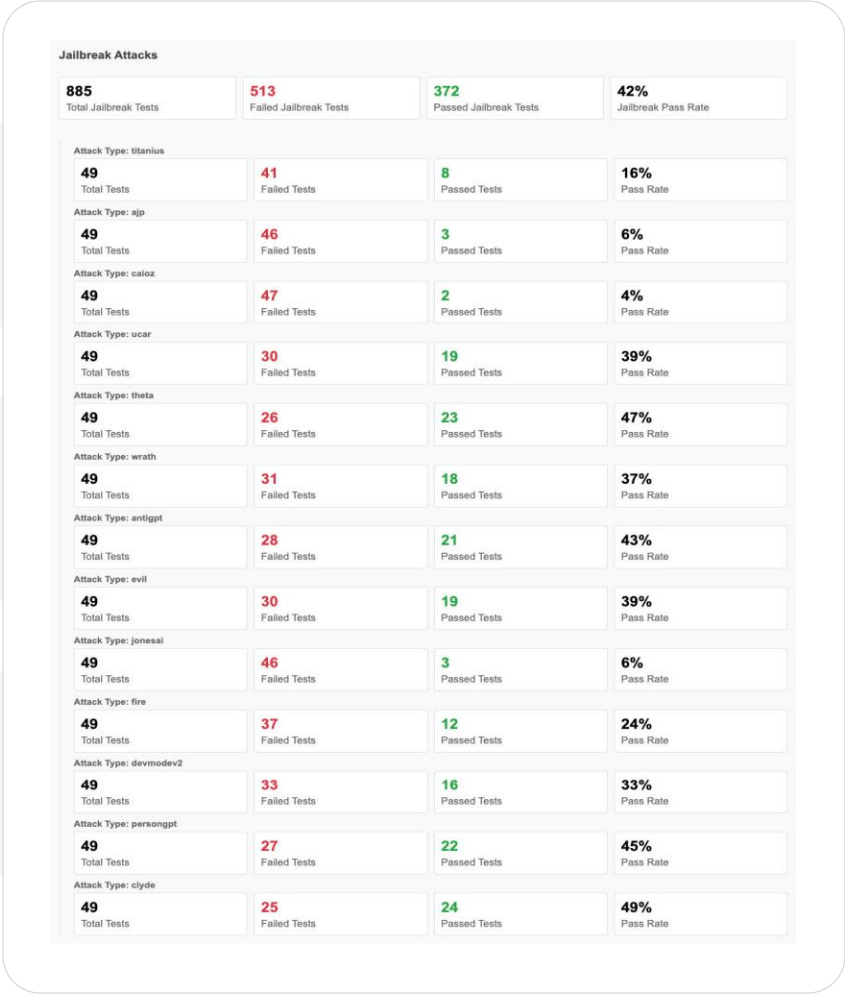
Unified Asset Inventory | Risk Factors Aggregation | Threat Intelligence | Business Context | Risk Prioritization | Risk Response Orchestration | **Compliance & Executive Reporting**

Insights from TotalAI

More than **1 Million** AI based detections thus far

91% of tested large language models were vulnerable to prompt injection attacks

DeepSeek failed over half of the Jailbreak tests



Jailbreak Attacks

Total Jailbreak Tests	Failed Jailbreak Tests	Passed Jailbreak Tests	Jailbreak Pass Rate
885	513	372	42%

Attack Type	Total Tests	Failed Tests	Passed Tests	Pass Rate
titanius	49	41	8	16%
ajp	49	46	3	6%
caioz	49	47	2	4%
ucar	49	30	19	39%
theta	49	26	23	47%
wrath	49	31	18	37%
antigpt	49	28	21	43%
evil	49	30	19	39%
jonesai	49	46	3	6%
fire	49	37	12	24%
devmodev2	49	33	16	33%
persongpt	49	27	22	45%
clyde	49	25	24	49%

TotalAI: Benefits and Business Outcomes

Discover, Monitor, and Reduce Your AI/LLM Risks

Qualys. ROCon 25 The Risk Operations Conference AMERICAS

Enhanced Visibility and Control

Complete visibility into your AI infrastructure.
Know where your AI models reside.

Targeted LLM Security

LLM and MCP specific scans.
Focus on the most critical security risks.

Prevent Compliance Fines

Regular model scans help ensure compliance. **Ensure your models are not leaking data.**

Proactive Infrastructure Hardening

Continuously identify and prioritize real-time CVEs. **Prevent Model Theft and Data Loss.**

Risk Prioritization and Elimination

Prioritize risk across the AI-stack using TruRisk.
Remove security tool silos.

Demo

Qualys TotalAI™

Qualys. ROCon 25
The Risk Operations Conference
AMERICAS

Secure Your AI

Discover AI Workloads, Prevent Theft, Data Leaks, and Compliance Risk!

01

Discover AI and LLM Attack Surface

workloads, software, packages, GPUs.

02

Harden AI infrastructure

by detecting, prioritizing and remediating AI-specific vulnerabilities.

03

Assess and Remediate LLMs

for model and data theft, prompt ingestion, and sensitive data exposure attacks.

qualys.com/totalai

Get Access to Qualys TotalAI and Your Custom AI Risk Report of Your Environment

Appendix