Secure your Generative (AI & LLM Rocket) Sheep

Himanshu Kathpal
Vice President, Product Management

Agenda

01

LLM Adoption

02

Current Challenges

03

Introducing: Qualys Total AI

04

Demo

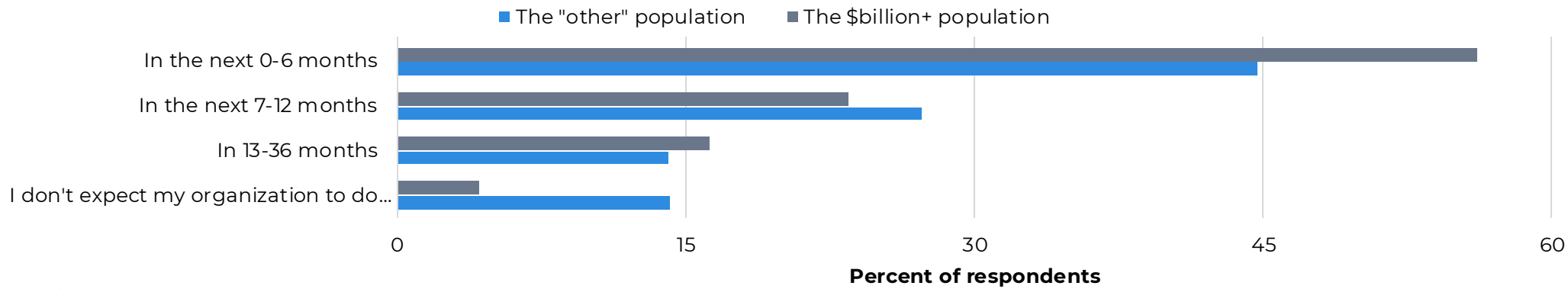
05

Benefits and Business Outcomes

LLM Rapid Deployments

70% of enterprise want to deploy LLM in production in next 12 months

On what timeline do you expect your organization to leverage a large language model in production?



Snorkel Respondents from $billion+ companies: 68 | Other respondents: 169

Blindsided:
Most security teams are unaware of LLMs in their environment.

Overlooked risk:
Security teams struggle to keep pace, skipping detailed security reviews.

Lack of Prioritization:
There is no good way to prioritize risk across security tools.

Enterprise Building and Leveraging AI and LLM

LLM Used Internally by Enterprise

Employees

Using LLM tools like ChatGPT, etc. and other AI-enabled SaaS productivity tools

Developers

Using other coding products with embedded LLM

Consumer of LLMs

Created for End Customer

AI / LLM tools embedded in product for their end customers

E.g., Qualys is creating LLM chatbot to summarize risk or asset information

Area of Focus

Creator of LLMs

Consequence of Security Gaps

Teams are forced to take unnecessary risk leading to unintended consequences

Data and IP Breach

Brand Reputation

Financial Cost

67% of organizations experienced AI-related security incidents in the past year, with an average cost of $4.2 million per breach.

Challenges

AI and LLMs bring incremental risks to an Enterprise

Low Visibility

Security teams blindsided by their AI workloads and model?

Do I have models in my Infrastructure? Where are they running?

Model and Data Loss

AI packages and AI infrastructure related CVEs can lead to data and model theft:

How to I discover and fix critical vulnerabilities in AI Infrastructure?

Increased Attack Surface

Attackers are targeting LLMs and AI infrastructure to steal model (crown jewel) and training data (PII)

How do I get ahead of attackers?

Low Security Maturity

LLMs often lack robust security measures, which can lead to compliance violations and fines

How can I test the model to understand risk?

Security Silos

Too many tools, yet a lack of visibility

How can I get better ROI from my security investments?

Introducing

Qualys Total AI

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

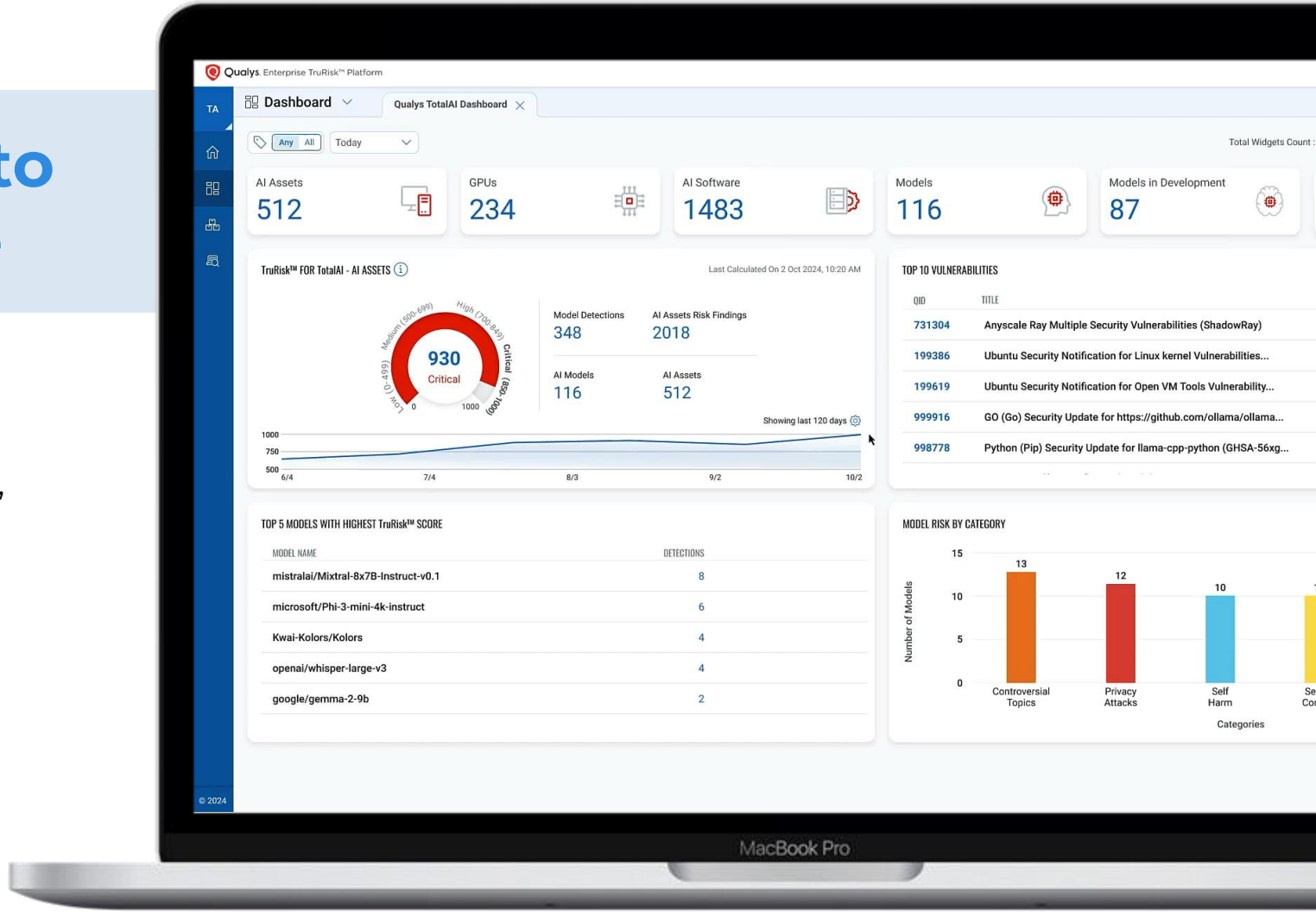
DE-RISK YOUR BUSINESS

Qualys.

Discovery with same sensors

Complete visibility into your AI infrastructure

- Fingerprint for AI workloads & Tagging
- Fingerprint based on hardware, software, packages, and models detection
- Confidence Score reflecting strength of fingerprint
- Know where your AI models resides, correlated with Attack exposure



Discovery of AI software and packages

On-prem and Multi-Cloud

- Altair Engineering RapidMiner
- Altair Engineering RapidMiner Studio
- Alteryx Intelligence Suite
- Anaconda
- Anaconda Jupyter Notebook
- Anaconda Miniconda
- ANSYS STK Integrated Jupyter Notebooks
- Anyscale Ray
- Apache Airflow
- Apache PySpark
- Azure Machine Learning Workbench
- BUNDLAR BUNDLAR
- DataRobot
- David Cournapeau scikit-learn
- Element Labs LM Studio
- ExplosionAI spaCy
- fast.ai
- Gael Varoquaux Joblib
- Google TensorFlow
- Guolin Ke LightGBM
- Homebrew Jan
- Hugging Face
- Hugging Face Transformers
- IBM Watson Content Analytics
- IBM Watson Studio
- Intel oneDAL and Intel Extension for Scikit-learn
- Iterative.ai DVC
- Jupyter Notebook
- KNIME KNIME Analytics Platform
- KPMG LLP KPMG Bridge
- Kubeflow
- libboost-numpy
- Logitech Logi AI Prompt Builder
- Matplotlib
- Microsoft Azure AI Machine Learning Studio
- Microsoft Azure Machine Learning Workbench
- Miniconda
- MLflow Project Mlflow
- NLTK Team NLTK
- Nomic GPT
- Numpy
- NumPy Developers NumPy
- NVIDIA CUDA
- NVIDIA CUDA Toolkit
- NVIDIA TensorRT
- NVIDIA Triton Inference Server
- Ollama
- OpenAI ChatGPT
- Opencv
- Pandas
- Jupyter Notebook
- Python
- python-matplotlib
- python-numpy
- Radim Rehurek GenSim
- SAS Institute SAS Viya
- Sebastian Ramirez FastAPI
- Squirro Joblib
- The Eclipse Deeplearning
- The Kubeflow Authors Kubeflow
- The Matplotlib development team Matplotlib
- The XGBoost Contributors XGBoost
- Travis Oliphant SciPy
- Wes McKinney Pandas

Discovery and Inventory for Model endpoints

Discovery of Models in Multi-Cloud environment

Azure Foundry

- Azure AI services
- Azure OpenAI services
- Azure Machine Learning

AWS Bedrock

- AWS Bedrock Serverless
- AWS Sagemaker

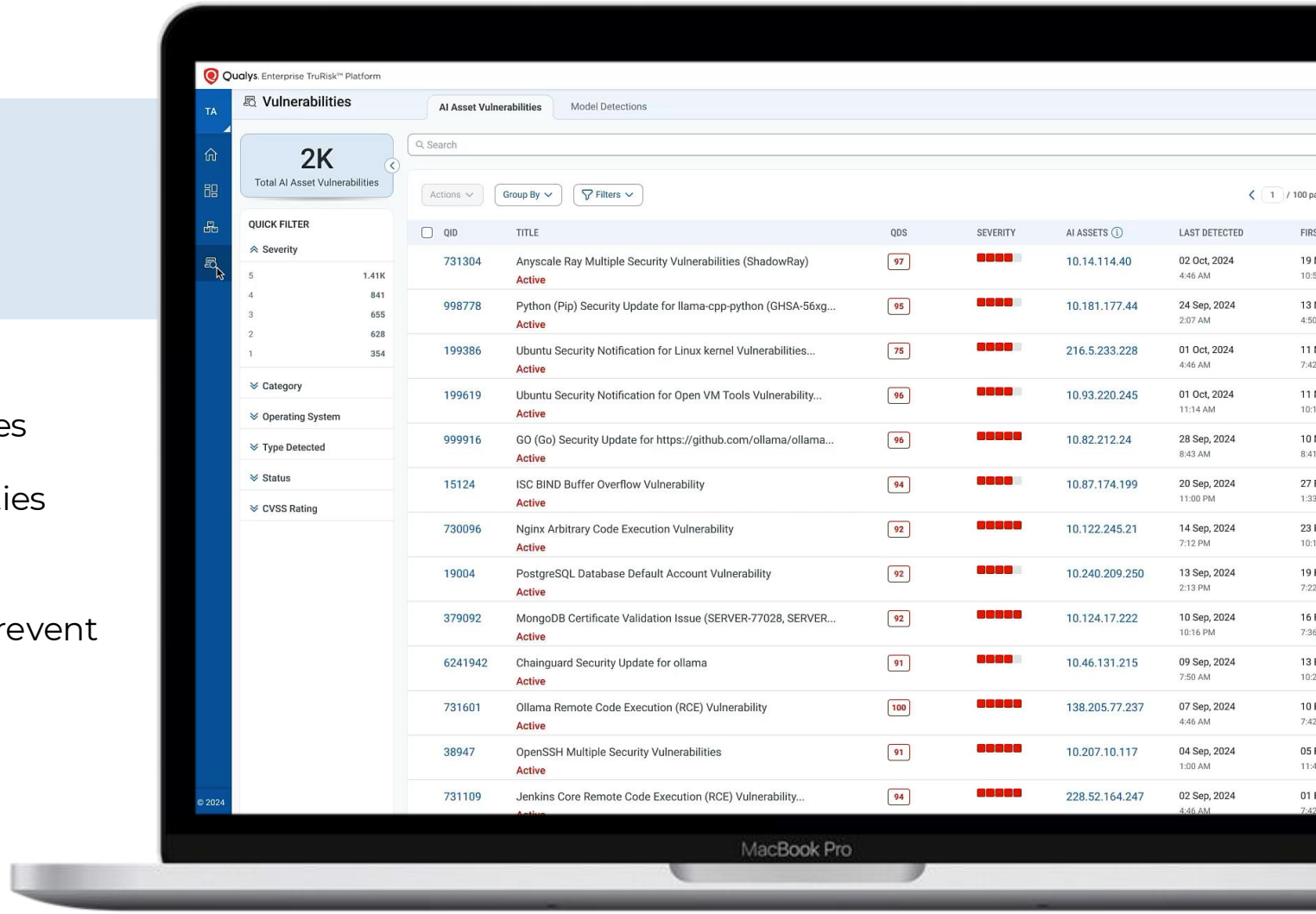
GCP AI Studio

- Vertex Model Endpoints
- Vertex Models

Vulnerability detection and remediation

Derisk your AI infrastructure

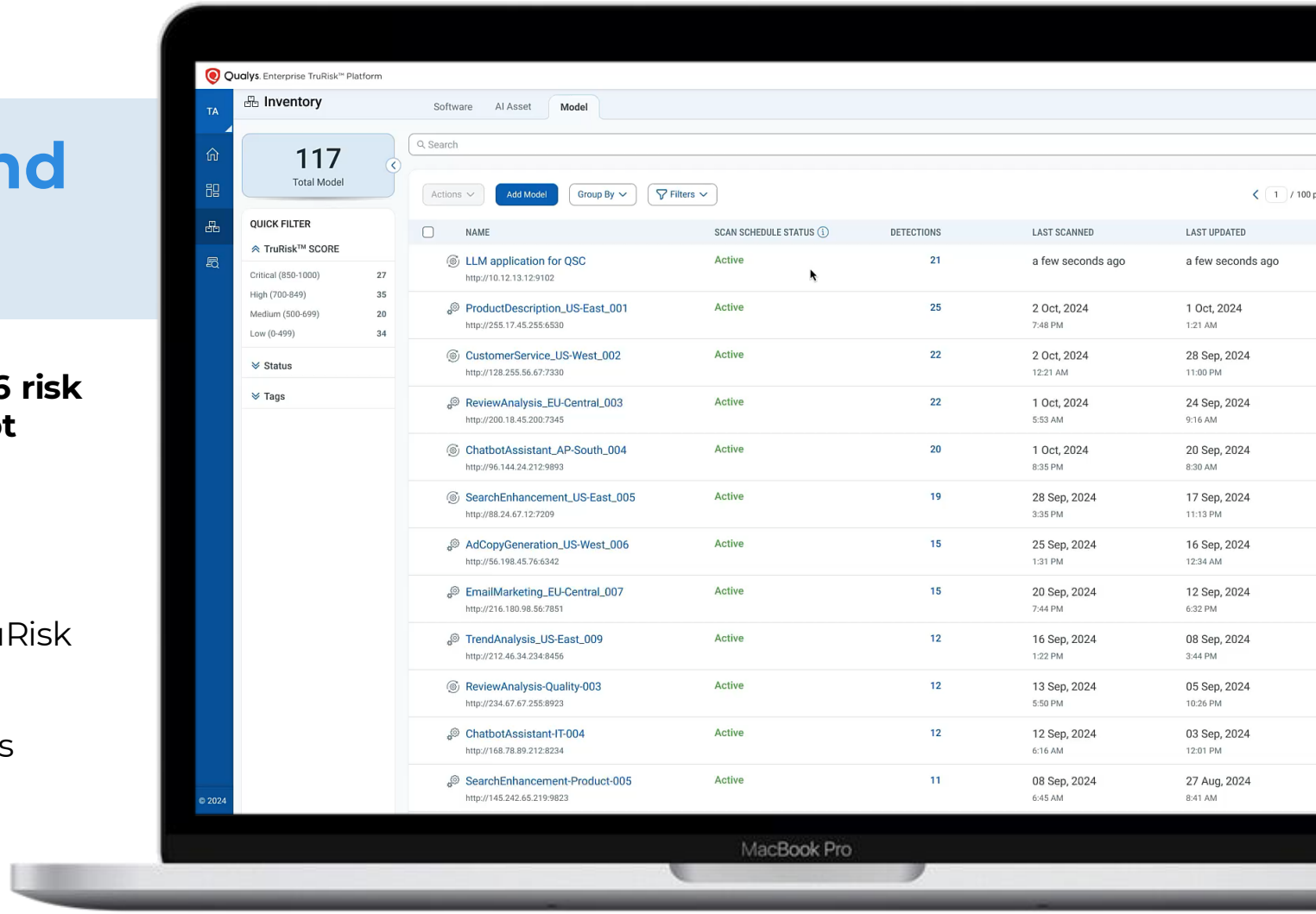
- Realtime Vulnerability detection on AI Infrastructure, software tools and packages
- Know the Prioritized Risk of AI vulnerabilities with 25+ Threat Intel Sources
- Fix vulnerabilities on critical AI assets to Prevent Model Theft and Data Loss
- Get Alerted for trending AI threats



Assessing and fixing risks in LLMs

Know the LLM risks and secure your business

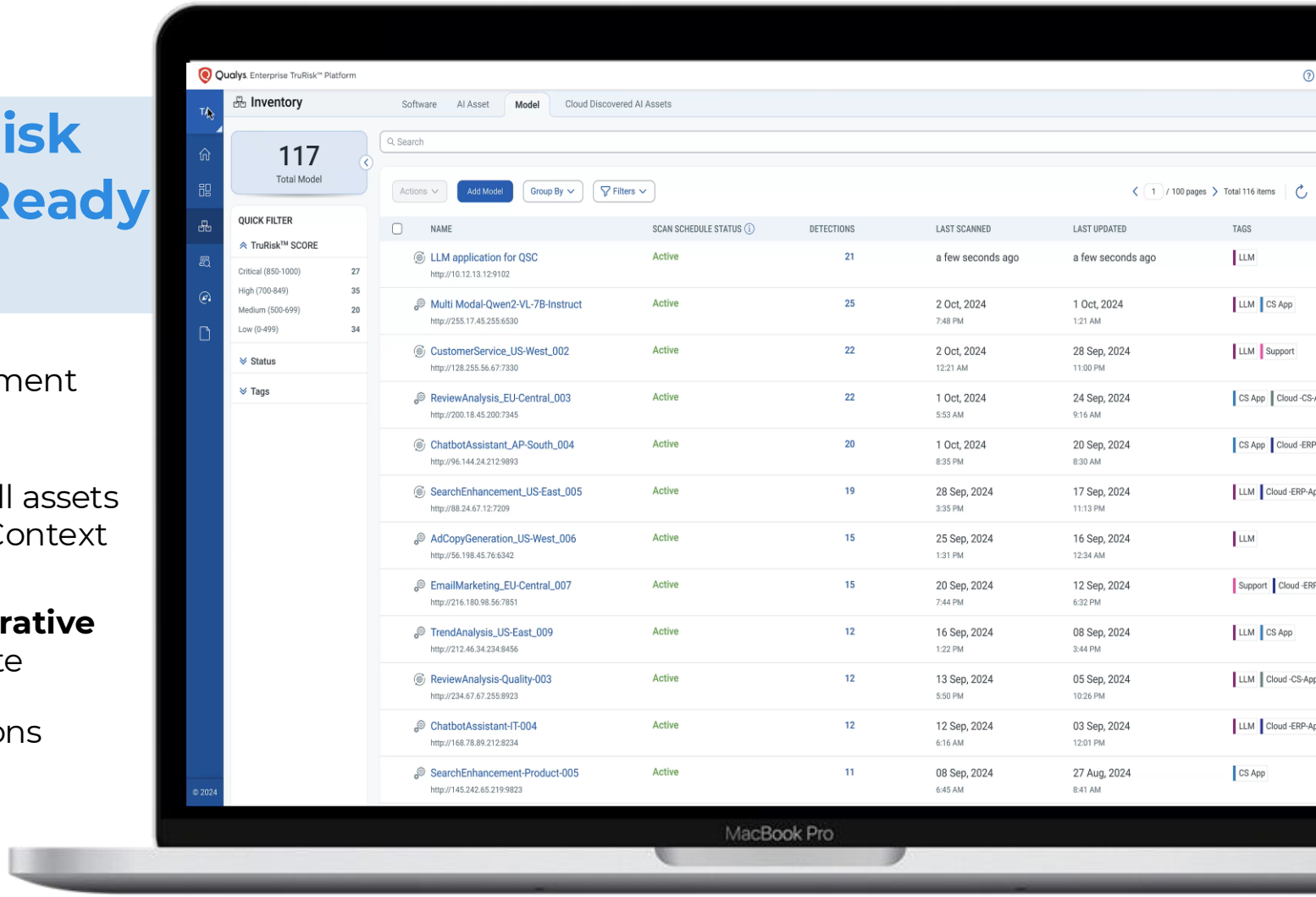
- Assess critical exposures in LLMs across **16 risk categories and 50+ jailbreak and prompt injection techniques**
- Map risks to OWASP top 10 for LLMs and Mitre Atlas
- Get unified risk picture through Model TruRisk
- Fix your model through recommendations across each category



Reporting and Compliance

Communicate Your Cyber Risk Effectively with Executive-Ready Reports

- Unified LLM security report for management
- Complete and Categorized visibility of all assets and findings based on Business & Risk Context
- Summarize these key insights into a **narrative** that your key stakeholders will appreciate
- Prevent fines due to compliance violations (e.g., GDPR, PCI)



Newly Introduced Capabilities

Discover, monitor, and reduce your LLM risks

Expanded Jail-break Attack Scenarios

Additional 38+ attack scenarios. **Strengthen your AI models.**

AI Supply Chain Protection

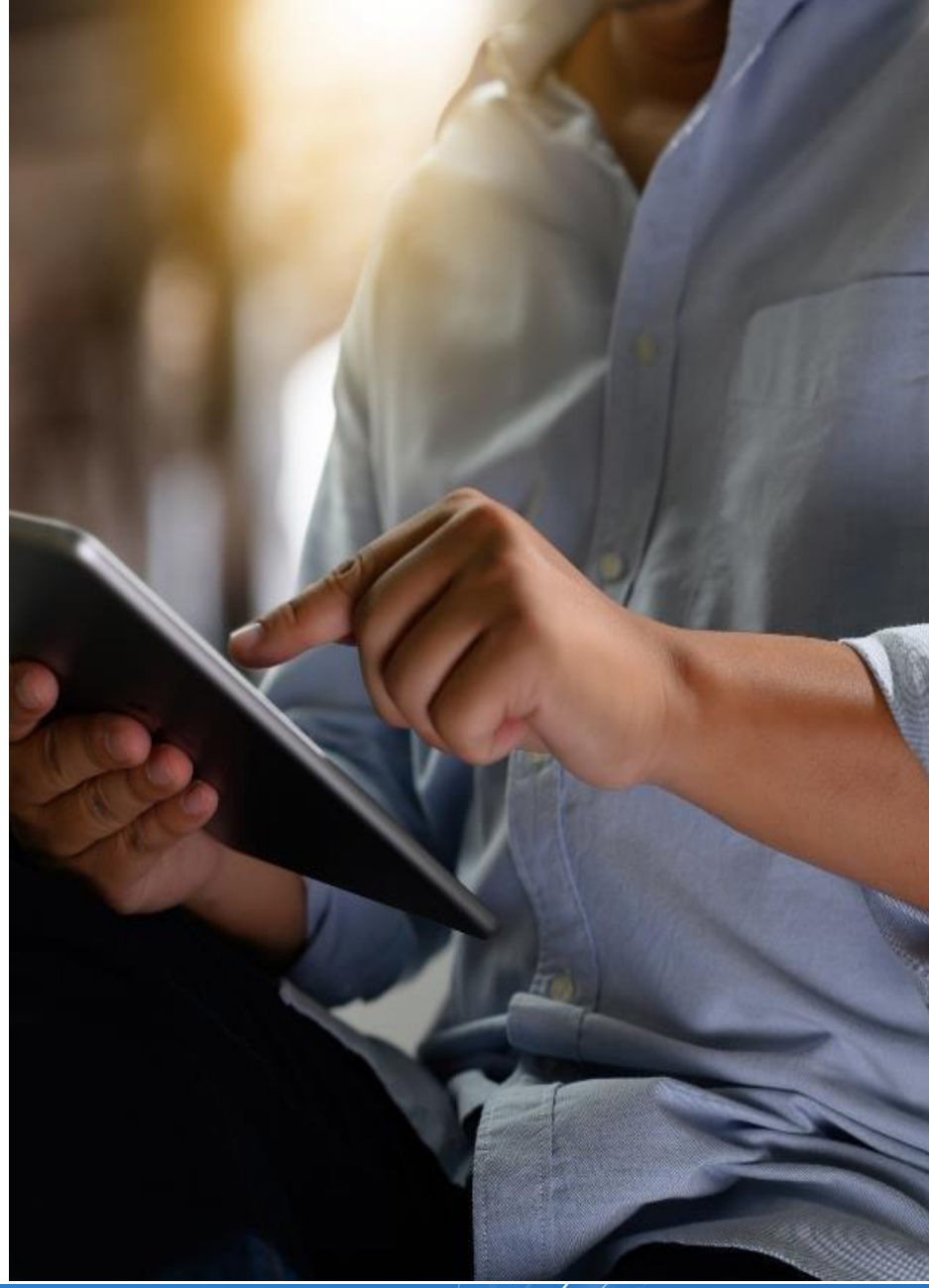
Continuously identify package hallucination and malicious 3rd party packages. **Prevent Model Theft and Data Loss.**

Multi-modal Threats Coverage

Detects prompts or perturbations hidden in images, audio or video that alter textual output.. **Ensure your models are not leaking data.**

Internal on-prem LLM Scanner

Integration with CI/CD pipeline for rapid dev, testing and deployment of LLM powered application w/o exposing externally.. **Secure MLOps Pipeline.**

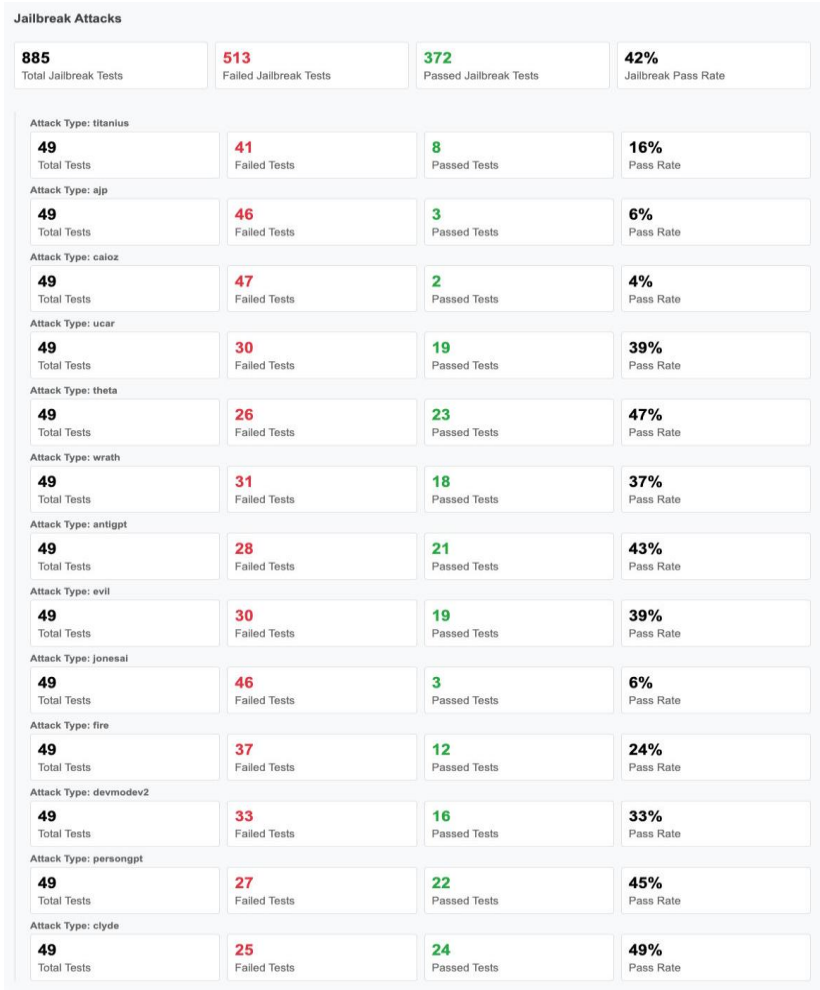


Learnings from Total AI

More than 1 Million AI based detected thus far

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

Organizations investing in AI-specific security measures report 63% fewer successful attacks than those using only traditional security



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%

DeepSeek failed over half of the Jailbreak tests

Demo

Benefits and Business Outcomes

Discover, monitor, and reduce your LLM risks

Enhanced Visibility and Control

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

Proactive Infrastructure Hardening

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

Prevent compliance fines

Regular model scans help ensure compliance with relevant data protection and privacy regulations. **Ensure your models are not leaking data.**

Risk Prioritization and Elimination

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

Targeted LLM security

LLM-specific scans. **Focus on the most critical security risks specific to LLMs.**

Qualys TotalAI™

Secure Your AI

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

01

Discover all AI and LLM

workloads, software, packages, GPUs.

02

Harden AI infrastructure

by detecting, prioritizing and remediating AI-specific vulnerabilities.

03

Assess 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

DE-RISK YOUR **BUSINESS**

Qualys.

Thank You