

Open Source AI Security Models: Practical Applications and Implementation Strategies

CISCO Live !
Security Innovation Forum

Aamer Akhter
Sr. Director, Product Management, Cisco

Session ID: BRKSEC-1669

Cisco Webex App

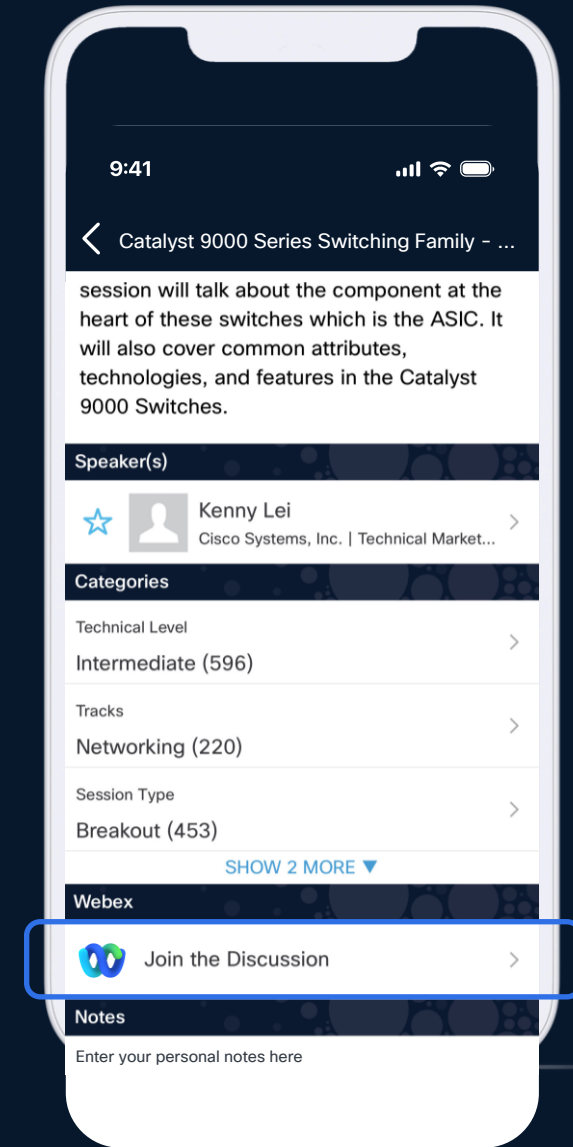
Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- 1 Find this session in the Cisco Live Mobile App
- 2 Click “Join the Discussion”
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until 14 November 2025.



Session Objectives

- 01 Understand why AI models are essential for modern security operations
- 02 Learn practical applications of open-source security models
- 03 Explore implementation strategies and real-world deployment
- 04 Discover Cisco's Foundation-sec-8B and Splunk integration

The Security Challenge

Data Overload Crisis

65% of security teams overwhelmed by alert volume

Petabytes of security data generated daily

Human analysts cannot process at machine scale

Critical Skills Shortage

3.5 million unfilled cybersecurity positions globally

Average time to hire: 6+ months

Burnout rates exceeding 50% in SOC teams

Evolving Threat Landscape

Advanced persistent threats (APTs)

Zero-day exploit acceleration

AI-powered attacks emerging

Digital transformation: attack surface expansion with

Alert Fatigue

70% of security alerts require manual investigation

High false positive rates (60-80% in many environments)

Critical threats buried in noise

Analyst productivity declining

Why AI Models for Security?

Scale

Volume

Process terabytes of security data in real-time

Velocity

Analyze streaming data at network speeds

Variety

Handle structured and unstructured security data

Complexity

Detect subtle patterns across multiple data sources

AI-Powered Threat Detection

Speed

Detection in
milliseconds

Scale

Billions of
signals

Adaptability

Continuously
learning

AI Log Analysis

Cross-System Correlation

Connect the dots across firewalls, endpoints, cloud, and network telemetry

Unify isolated alerts into an attack narrative

Log Intelligence

Interpret unstructured data, extracting meaning, indicators, and relationships

...relations that human analysts would miss.

Attack Chain Recognition

Identify Progression

Reconnaissance

↓

Intrusion

↓

Privilege escalation

↓

Exfiltration

Accelerated Response

Automate Triage

Reduce time from detection to remediation

Surface what matters most

Turning Data Noise into Actionable Insight

AI Malware Detection



Insider Threat Detection

User Behavior Analytics

Flag deviations from baseline

Privilege Abuse Detection

Correlate access patterns with intent

Data Exfiltration Patterns

Monitor subtle or huge data leaks

Anomalous Access Detection

Connect Identity, Device and Context



Why Open Source AI Security Models Matter

Transparency & Trust

Customization & Control

Cost-Effectiveness

Deployment Flexibility

Hugging Face: The Hub for Open AI Models

Central repository with 2M+ open models.
Industry standard for distribution across all major frameworks (PyTorch, TensorFlow, JAX).

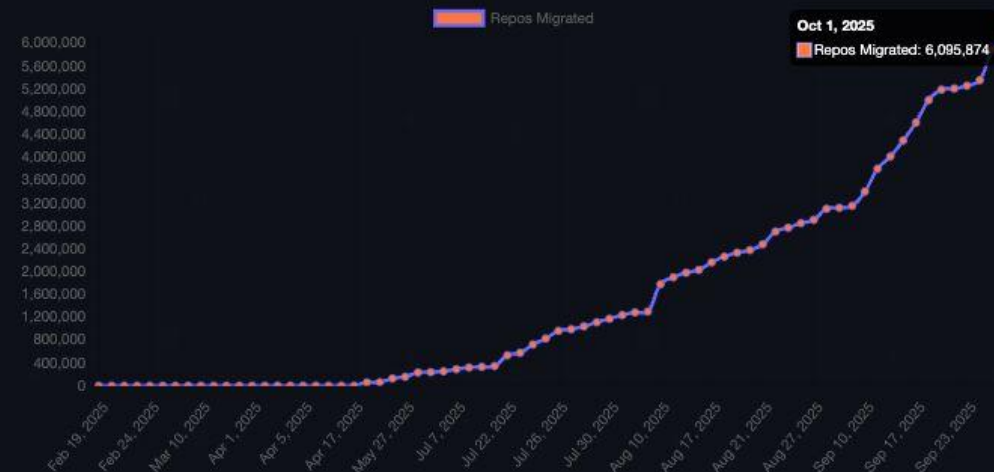
Capabilities

- Model cards with provenance & licensing
- Version control and deployment tools
- Community benchmarks

Standardized alternative to scattered GitHub repos and personal websites.



Repos Migrated to Xet



Hugging Face Xet : Storage Backend

How security teams use Hugging Face

Model Acquisition & Testing

- Download models for air-gapped security validation
- Test for backdoors, data leakage, or adversarial vulnerabilities
- Verify model behavior before production deployment

Model Customization

- Fine-tune base models on proprietary data locally
- Version pin deployments for reproducibility
- Access model cards to assess training provenance and risks

Deploy and modify models without external dependencies or cloud API requirements



Hugging Face hub

Download weights

Air-Gapped Testing

Production Deploy

Open-Weight Models: The middle ground of AI openness

Proprietary

vendor controls model
must use their infrastructure
black box

Open-Weight (partial)

full model weights available
deploy anywhere
training data private

Open-Source (fully open)

everything available
full transparency

Why Open-Weight for Enterprise:

- Fine-tune on your data without exposing it
- Deploy air-gapped or on-prem
- Control inference costs and latency
- Audit weights for backdoors/trojans
- No vendor dependency for model updates

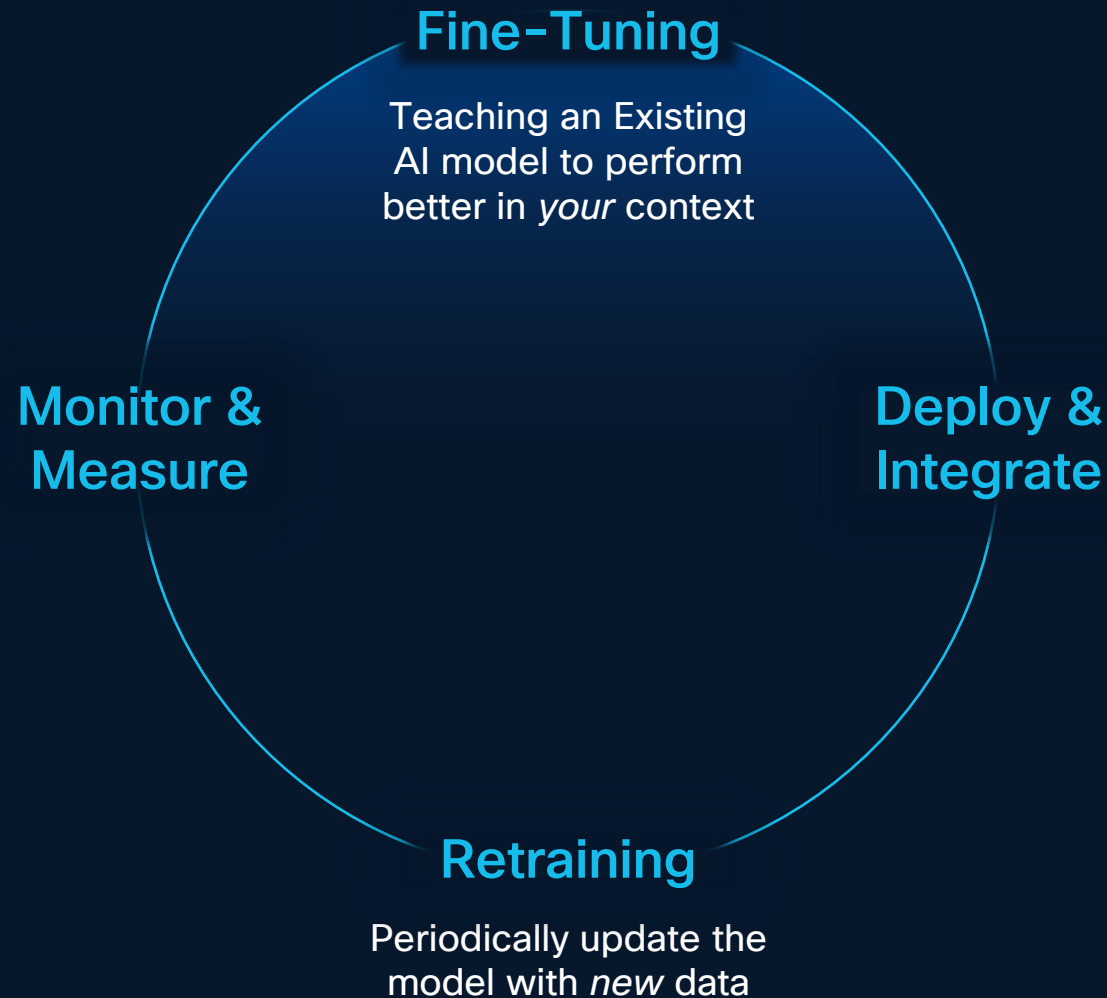
Trade-off

No visibility into training process or
potential biases

What's the Real Cost of Being Wrong?

TRUE POSITIVE Breach Avoided Confidence Up	FALSE POSITIVE Analyst Time Lost Alert Fatigue \$
FALSE NEGATIVE Data Loss / Downtime Brand Damage \$\$\$	TRUE NEGATIVE Neutral Business as Usual

Every AI decision carries an
operational or financial cost.
Precision matters!



The Economics of AI



	Commercial	Self-Hosted
Operating Cost	Usage-based	Fixed infrastructure
Performance	Low Latency	Scalable
Data Privacy	No	Full control
Flexibility	Limited	Full
Predictability	Variable OPEX	Stable OPEX / CAPEX

AI on your Terms

Customization & Control

Fine-Tuned for *your* reality
Integrate Proprietary Intel
Governance By Design



Deployment Flexibility

On-Premise / Air-Gapped
Secure Infrastructure
Hybrid for Scale

“Control what the model learns
and where it learns it.”

Community Innovation

Collaborate

Transparency
Collective Wisdom

Build

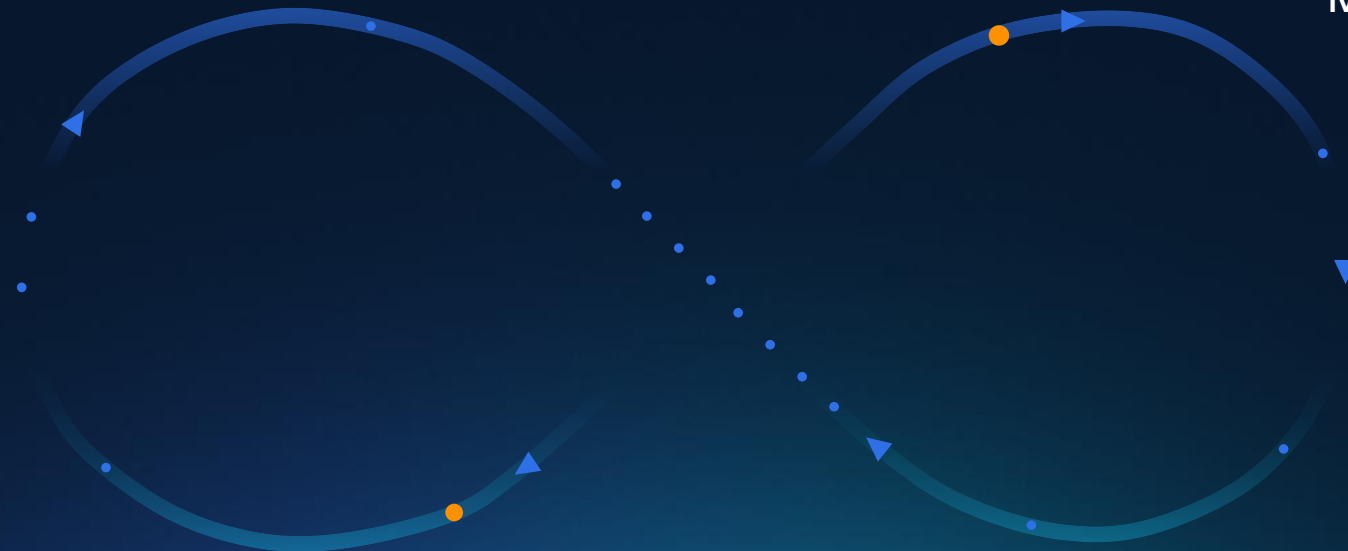
Lower Experimentation
Barrier Cutting-Edge
Models and Datasets

Improve

Pace of Innovation vs.
Release Cycles

Share

Best Practices
Benefit Entire Industry



Practical Applications Deep Dive

Use Case 1: SOC Acceleration

Detect, triage, and respond to security threats faster

For:

- SOC teams
- Security teams to do more with less

Pain points:

- Too many false positives
- Manual, repetitive triage
- Fragmented tools

Automated Triage

Incident Response
Enhancement

Report Generation

Use Case 2: Proactive Threat Defense

Transition from reactive to proactive threat discovery

For:

- AppSec teams
- App devs wanting to be more involved in security operations

Pain points:

- Slow, manual security reviews
- Reactive processes
- Limited expertise

Threat Intelligence Processing

Vulnerability Management

Attack Simulation

Use Case 3: Engineering Enablement

Transition from reactive to proactive threat discovery

For:

- AppSec teams
- App devs shifting left in security operations

Pain points:

- Slow development cycles due to security assessments
- Reactive nature to vulnerability and security management

Secure Development

Infrastructure Security

SOC Acceleration Example

Remediation

Final Incident Investigation Report

Export PDF

Immediate

1

Isolate t
maliciou

2

Termina
persiste

3

Block al
disrupt

4

Run a fu
definitio

5

Review
attacker

6

Change
network

7

Monitor
monitori

Incident Report

Incident ID: 2025-05-12-001

Type: Cyber Intrusion

Severity: High

Timestamp: 2025-05-12 08:22:11

Incident Summary

On 2025-05-12, a sophisticated cyber attack was initiated through a phishing email containing a malicious attachment 'invoice.docm', which was delivered to user 'bob' on the host 'WS456'. The attack involved the execution of macros using PowerShell for downloading additional malicious content, beaconing to a command-and-control (C2) server, creation of a persistent scheduled task named 'Updater', and an attempt to harvest credentials from the LSASS memory. The attack techniques leveraged included phishing with macros, PowerShell execution, beaconing for C2 communication, scheduled task for persistence, and credential theft attempts.

Impacted Assets and Users

- Hosts: WS456
- Users: bob

Adversary Tactics and Techniques

icious

similar

bout

other

ns.

ified
s

© 2025 Cisco and/or its affiliates. All rights reserved.

BRKSEC-1669

Cisco's Open weight models



Team of leading Cisco AI security researchers and engineers dedicated to bleeding-edge AI technology to empower cybersecurity applications.

Initial projects:

- Foundation-sec-8b: our first open-source security model
- Novel benchmarks for evaluating cybersecurity models
- AI supply chain risk management and intelligence

Open-weight security model: Foundation-Sec-8B

Use Cases

- **SOC Acceleration:** Automating triage, summarization, case note generation, and evidence collection.
- **Proactive Threat Defense:** Simulating attacks, prioritizing vulnerabilities, mapping TTPs, and modeling attacker behavior.
- **Engineering Enablement:** Providing security assistance, validating configurations, assessing compliance evidence, and improving security posture.

The following is a CVE description. Map It to the appropriate CWE ID.

CVE Description: A SQL injection vulnerability exists in Novel-Pius v4.3.0-RC1 and prior. An attacker can pass specially crafted offset, limit, and sort parameters to perform SQL injection via / novel/userFeedback/list.

=> Answer: CWE-89

CVE Description: cross-Site Request Forgery (CSRF) vulnerability in Doofinder WP & WooCommerce Search. This issue affects Doofinder WP & WooCommerce Search: from n/ a through 2.0.33.

=> Answer: CWE-352

Security AI Model effectiveness

Test categories include:

- Accuracy & Reliability
- Fairness & Transparency
- Privacy & Security
- Human Oversight
- Governance & Continuous Improvement

Benchmark (higher is better)

Model	CTI-MQA	CTI-RCM
Foundation-sec-8b	67.39	75.26
Llama-3-8b	64.14	66.43
Llama-3 70B	68.23	72.66

Foundation-Sec-8B variants

Base

Foundation model
for fine-tuning

April 2025

Instruct

Chat-optimized
for direct use

August 2025

Reasoning

Enhanced
analytical
capabilities

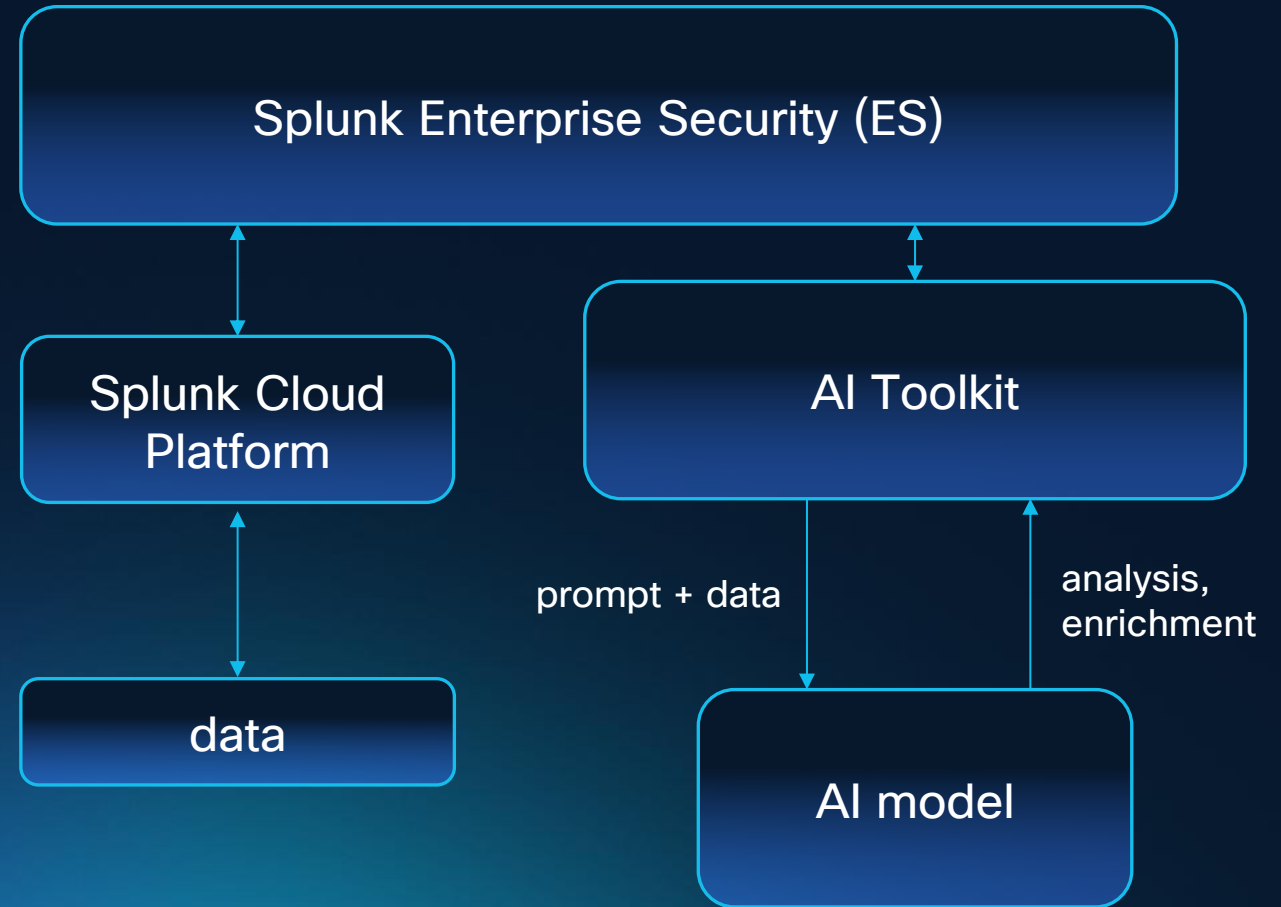
TBD

Foundation-Sec-8B integration with Splunk AI Toolkit

- User query directs event share with AI model
- AI model enriches data
- New field created, augments security event

Use Cases

- Automated alert triage
- Streamlined investigations
- Detection Content Suggestions





ANNOUNCING

Time Series Foundation Model

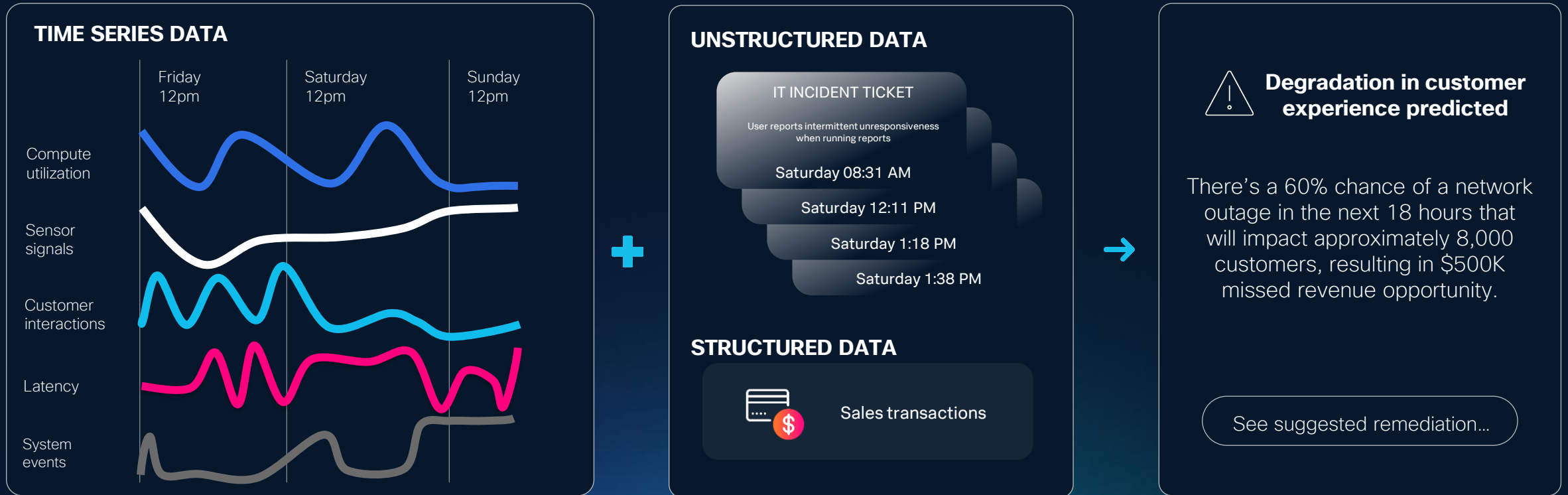
OPEN-SOURCE | ON HUGGING FACE NOVEMBER 2025

Optimized for machine data

Accurate and scalable

Adaptable to your domain

Example: Open Security Model + AgenticOps



Cisco product's built-in models

Custom Models Embedded Across Cisco Security Cloud

Examples around AI for Security solutions

Cisco Security Cloud

Encrypted Visibility Engine

ML-based threat
detection in
Secure Firewall

SnortML

Exploit detection
engine trained by Talos

Exploit Protection

Cisco Advanced
Malware Protection (AMP)
for zero-days using
Secure Endpoint

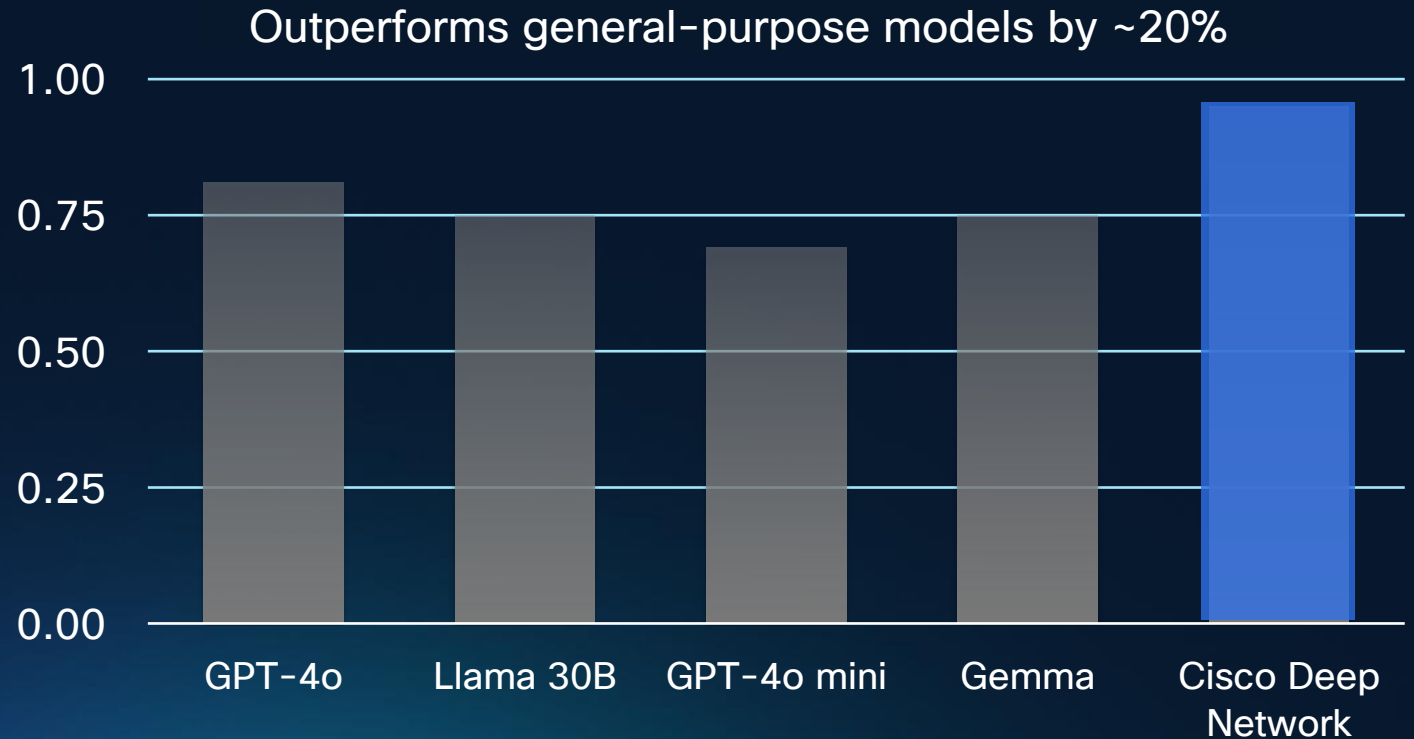
Email Threat Defense

Customized
LLM to augment
threat detections

Cisco Deep Network Model

Purpose-built for networking, expert accuracy

- 10% more precise reasoning for troubleshooting, configuration, and automation
- Fine-tuned on 40+ years of expertise and expert-vetted for accuracy
- Evolves with live telemetry and real-world Cisco TAC and CX insights



Accuracy on CCIE-style multiple choice questions (590-question benchmark), May 2025

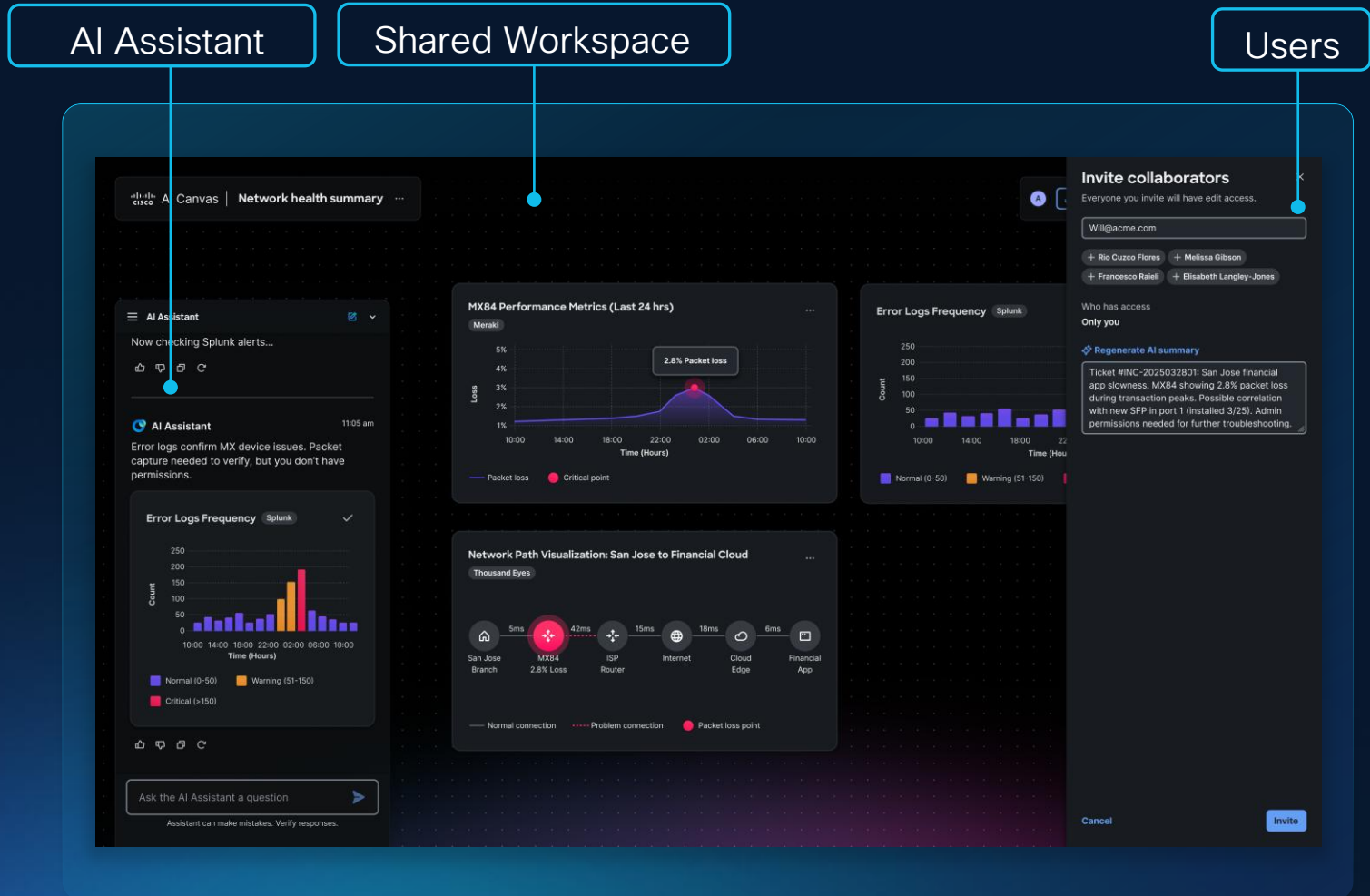
Cisco AI Canvas

Unifies telemetry across Cisco and Splunk data

Applies reasoning, formulates plans, executes analysis, provides summaries and recommendations

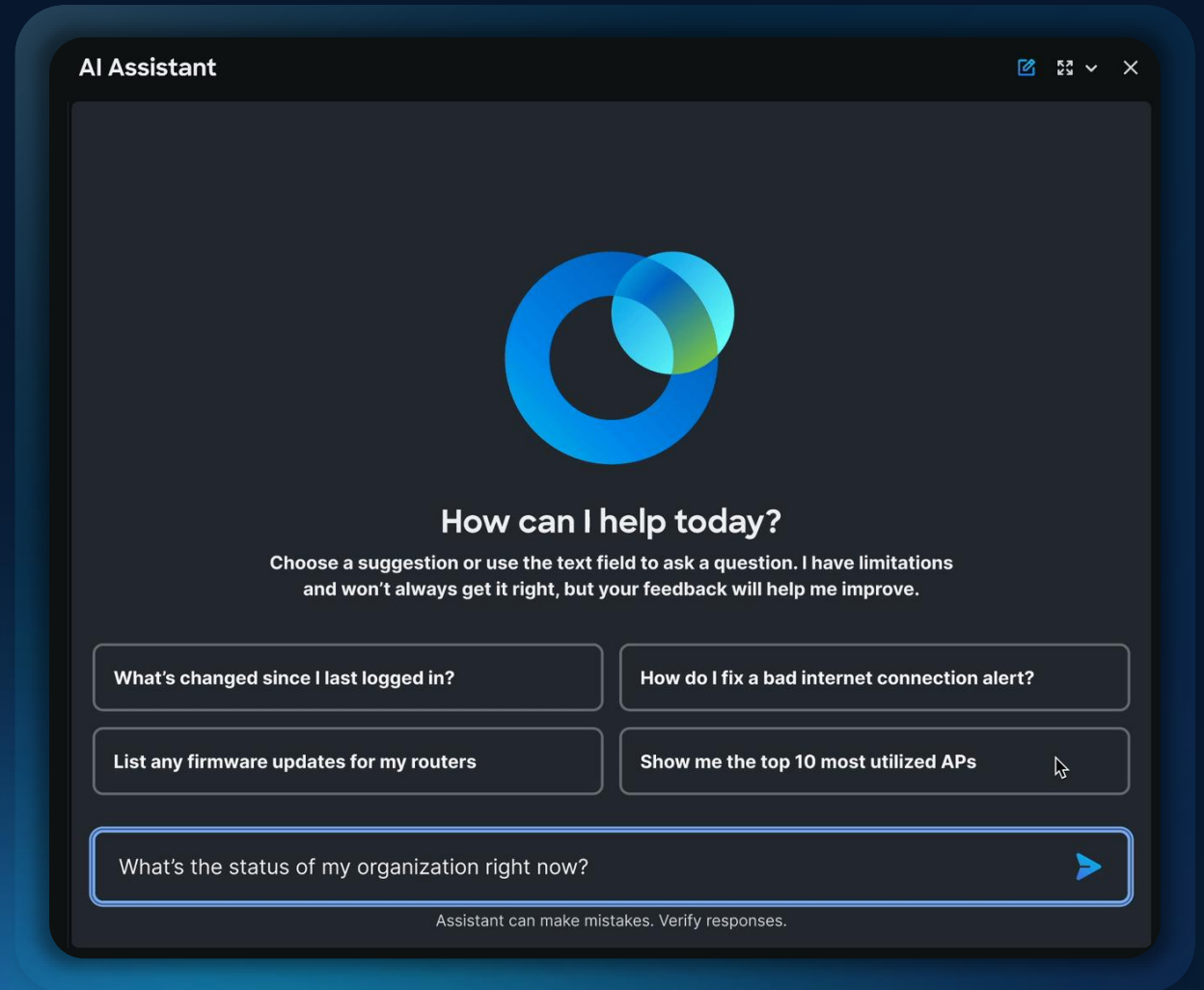
Collaboration across teams (NetOps, SecOps, IT and execs)

Troubleshooting and execution across multiple domains



The Unified Cisco AI Assistant

- Explore and interact with your network in natural language
- Troubleshoot and automate tasks
- Ask questions about the network or how to accomplish tasks



Splunk AI Assistant in Enterprise Security

Embedded across all analyst workflows

Guide investigations with AI-powered query generation and summarization

Accelerate workflows with embedded, context-aware AI insights

Empower decisions with AI-driven next step and remediation guidance

The screenshot shows the Splunk Enterprise Security interface. The top navigation bar includes 'Settings', 'Activity', and 'Find'. The main content area is divided into several sections. On the left, there's a table of findings with columns for ID, Entity, and a status. The central panel, titled 'AI Assistant for Security', displays a finding titled 'Malicious PowerShell execution'. This panel is highlighted with a red circle. It contains a 'Finding summarization' section with a summary of the event and a 'Sandbox analysis' section with a detailed analysis. The right panel shows the 'Splunk Attack Analyzer' and 'Cisco Talos' sections, providing additional context and threat intelligence. The bottom of the interface has a search bar and several buttons for further actions like 'Summarize the findings', 'Generate investigation report', 'Suggest SPL', and 'Discover AI Assistant skills'.

Implementation Strategies

AI for Security Model: Implementation Roadmap

Assessment

- Security data inventory
- Threat & use case prioritization
- Privacy & bias checks
- Success criteria
- Resource requirements

Pilot

- Limited scope implementation
- Integration & adversarial Testing.
- Performance & business impact evaluation
- Human-in-the-Loop strategy & Training

Production

- Full deployment across security operations
- Monitoring → optimization
- Model update & maintenance
- Value realization → outcomes Reporting

Evaluation Framework

Use Cases

- Current security workflows → AI capabilities
- Prioritize high-impact, low-risk
- Data availability and quality
- Technical feasibility and resources

Performance

- Test on representative data
- Compare against current manual processes
- Model & system Performance (Accuracy, Precision, Recall)
- Assess integration complexity and effort

Success Metrics

- Reduced MTTD and MTTR
- False Positive Rate (FPR) Reduction
- False Negative Rate (FNR)
- Quantified Risk Reduction & Dwell Time Reduction
- Human-in-the-Loop (HITL) & Analyst Productivity Metrics

How to get started

Explore use case cookbook on github

<https://github.com/cisco-foundation-ai/cookbook/tree/main>

Download Foundation-sec-8B from Hugging Face

<https://huggingface.co/fdtn-ai/Foundation-Sec-8B>

Co-develop and explore with us

<https://fdtn.ai/early-access>

Complete your session evaluations



Complete a minimum of 4 session surveys and the Overall Event Survey to claim a Cisco Live T-Shirt.



Earn up to 800 points by completing all surveys and climb the Cisco Live Challenge leaderboard.



Level up and earn exclusive prizes!



Complete your surveys in the Cisco Live Events app.

Continue your education



Visit the Cisco Stand for related demos



Book your one-on-one Meet the Expert meeting



Attend the interactive education with Capture the Flag, and Walk-in Labs



Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at: Insert preferred comms method

Thank you

CISCO Live !

