思科AI架构及AI安全

冯洪敏

大中华区思科技术部
解决方案工程师

日程

- 思科AI架构
- 思科CX如何使用AI并赋能客户（AI落地）
- AI 安全

人工智能正在全面改变我们的社会、业务和技术

AI changes everything in our society, business and technology

$16T

到2030年对全球经济的潜在贡献

Potential contribution to global economy by 2030

$300B

到2026年全球在人工智能上的支出

Global spending on AI by 2026

75%

大型企业到2026年前将依赖于融入人工智能的流程

of large enterprises will rely on AI-infused processes by 2026

90%

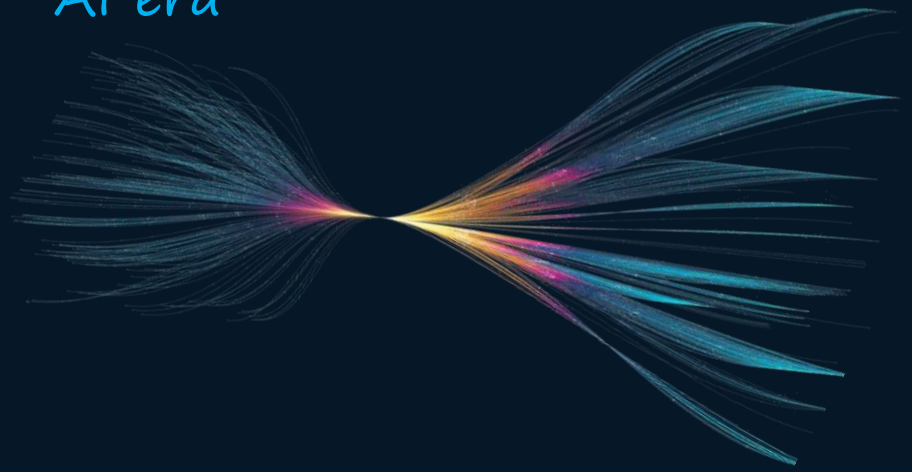
到2025年，人工智能工作负载运行在以太网上

of AI workloads will run on Ethernet by 2025

Source: PWC, IDC

思科连接和守护AI时代

Cisco connects and protects the AI era



人工智能基础设施
Infrastructure to power AI

人工智能安全
Security for AI, AI for security

数据驱动的洞见
Data to drive insights and context

AI原生软件
AI-native products to unlock productivity

增值AI服务
Services to accelerate the value of AI

思科提供全栈的AI基础设施解决方案

Cisco provides industry leading full stack AI infrastructure solution & services

Cisco Offerings

数据

Data & Observability

Splunk

AI基础设施

AI Ops	Automation & Visibility	Nexus Dashboard
Networking	DCI	Cisco 8K
	Switch	Nexus series
	ASIC	SiliconOne
Inter-connects	Optics	Acacia 400G/800G Cisco Silicon Photonics
Compute	Server	UCS series

AI安全

Security

Foundation AI Security LLM Model

AI Defense

Isovalent (Cilium)

Services

服务

AI Infra Advisory | AI Architecture Optimization | AI Ops services

通过业界合作和产品创新不断增强我们的AI能力

Bolstering our AI capabilities strategic partnerships and product innovation

业界合作

Nvidia Partnership

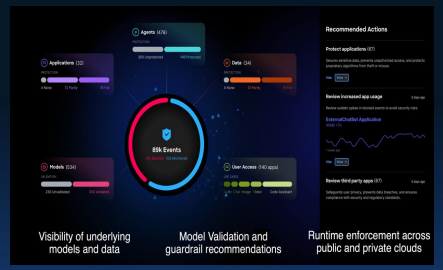
CISCO + NVIDIA

OpenAI CPO joining Cisco Board of Directors

CISCO + OpenAI

产品创新

Cisco AI Defense

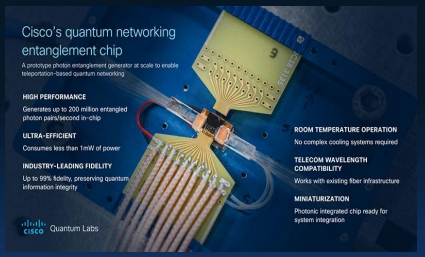


End-to-end AI security

Cisco's 1st LLM Security Model

Open Source LLM
8-billion parameter
20K Download

Quantum Network Entanglement Chip



1st Quantum Network Chipset
Cisco Quantum Lab Opening

开放 创新 安全 可控 的AI基础设施

加速创新，新技术和产品蜂拥呈现

Accelerating AI innovation, more new technology and offerings available

Cisco Deep Network Model (LLM)

专门的网络大模型
20%+ accurate reasoning, resolution

思科专家知识体系训练
>40 years of operational insight

持续学习和迭代
Real world new TAC/CX insights

Secured Networks for AI

Cisco Smart Switches

Silicon One embedded
Cisco Hypershield

Unified Nexus Dashboard

Unified AI/ML fabrics operations

AI Security

Hybrid Mesh Firewall

Embed with AI Defense, Isovalent, Security Cloud Control

Splunk Observability Cloud

Cloud native, GenAI-powered

AI-driven AgenticOps

Cisco AI Canvas

AgenticOps with real-time telemetry, AI insights for intelligent workspace

Cisco AI Assistant

Embed in Cisco products, AI supported decision making

思科帮助中国客户加速AI价值实现

Cisco AI actions in China to help customer deliver business value

帮助自动驾驶解决方案提供商实现AI驱动的自动化车队管理解决方案

Support autonomous driving solution providers for AI-driven automated fleet management

帮助电商公司使用AI处理客户查询和分析购买行为

Help e-commerce companies for AI-empowered customer inquiry handling and purchasing behavior analysis

为全国领先的研究型大学的AI大语言模型训练与推理提供基础设施支持

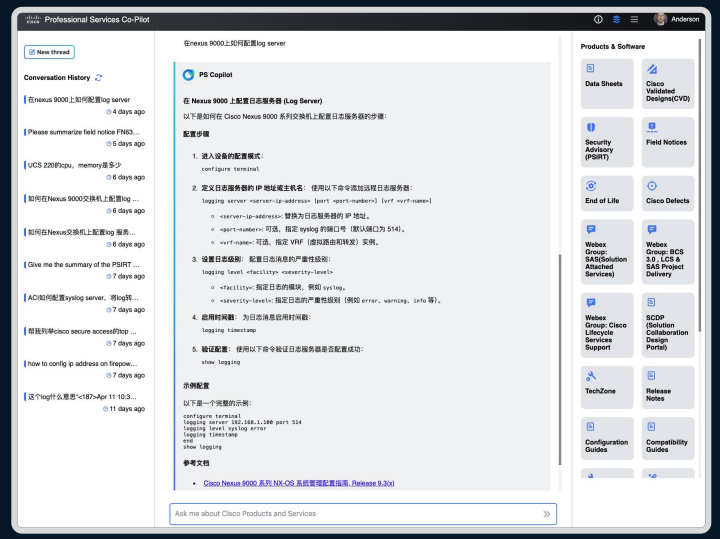
Help top research university for their LLM models with AI infrastructure support

日程

- 思科AI架构
- 思科CX如何使用AI并赋能客户（AI落地）
- AI 安全

思科已经是“机器人无处不在”

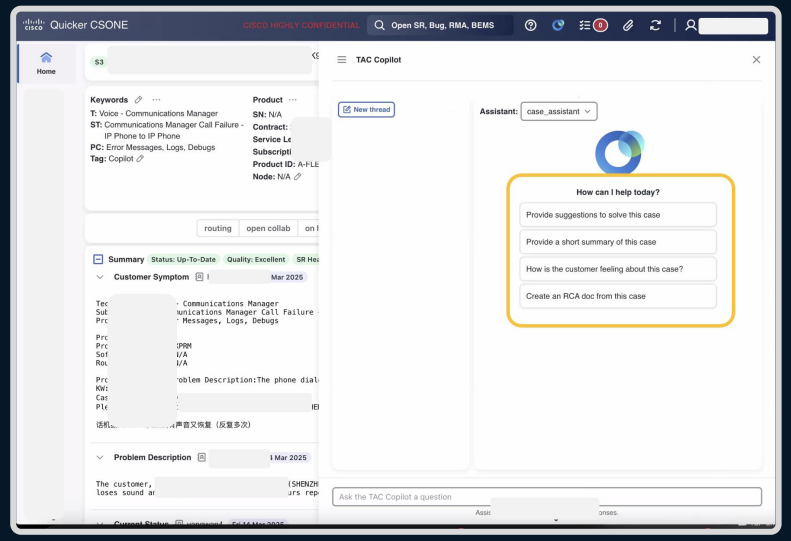
高级服务机器人PS Co-Pilot



通过和机器人聊天的方式获取：

1. 产品相关信息
2. 安全PISRT信息
3. Field Notices信息
4. End of Life产品信息
5. 设备配置信息
6. Troubleshooting建议
7. 设计建议等等

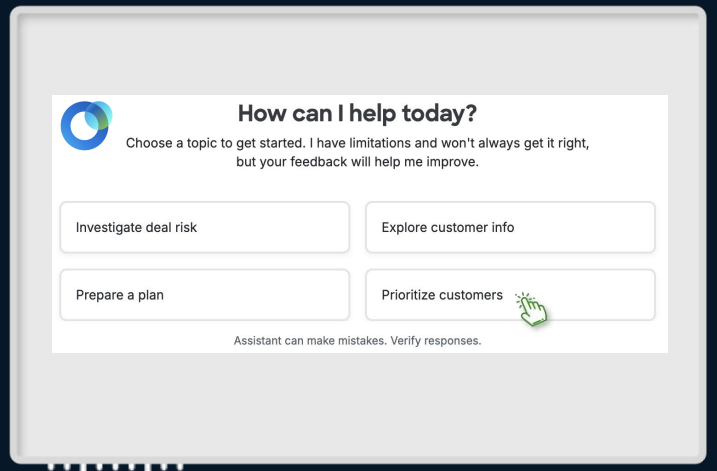
TAC机器人TAC Case Assistant



通过和机器人聊天的方式：

1. 获取Case解决建议
2. 生成Case的汇总信息
3. 获取客户的反馈信息
4. 生成根因分析报告

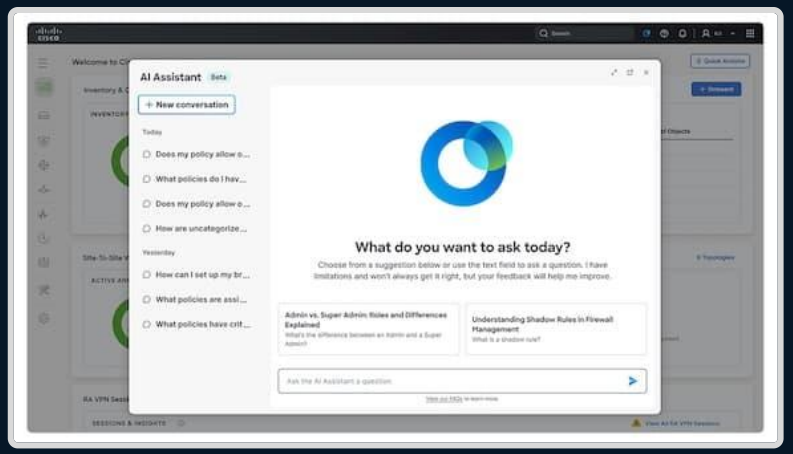
服务续约机器人Renewals AI Agent



通过和机器人聊天的方式：

1. 获取需优先处理的续约订单
2. 了解客户续约信息
3. 调研客户服务适用情况
4. 生成续约计划

产品机器人Product AI Assistant



通过和机器人聊天的方式：

1. 简化设备管理配置
2. 辅助根因分析
3. 快速诊断问题
4. 提供操作指南

思科CX生成式智能运维 CX GenAIOps

思科CX GenAIOps（生成式智能运维）

- 全新一代基于本地开源生成式大语言模型（LLM）+思科知识库+客户实际运维数据训练所得的运维智能体

- 三个组件：

1）AI-Syslog：基于LLM的日志分析平台（跨厂商全栈实时异常判断和根因分析）

2）AI-ChatBot：基于LLM的AI本地运维聊天机器人（凝练思科专家经验）

3）AI-Agent：基于LLM的AI辅助运维智能体（使用自然语言生成配置）

- 客户本地部署及训练(On-Prem模式，无需外传数据，保护客户数据安全及隐私)

思科CX生成式智能运维 CX GenAIOps

适用的场景

1.实时智能异常预警（故障发现）
2.问题触发根因分析（故障定位）
3.日常快速网络巡检（风险排除）
4.本地知识库机器人（知识积累）
5.全球网络运维辅助（出海支撑）

解决的痛点

1.海量异常日志预警
2.异常日志根因分析及操作建议
3.避免相同故障再次发生
4.准确理解日志含义
5.自然语言生成配置命令
6.快速本地知识库构建

A公司网络系统syslog运维的常态

以 A公司现网网络设备为例：

设备数量：约400台

设备种类：5种（包括：路由器，交换机，防火墙，负载均衡，无线控制器）

设备厂商：5家（包括：思科，华为，天融信，深信服，F5）

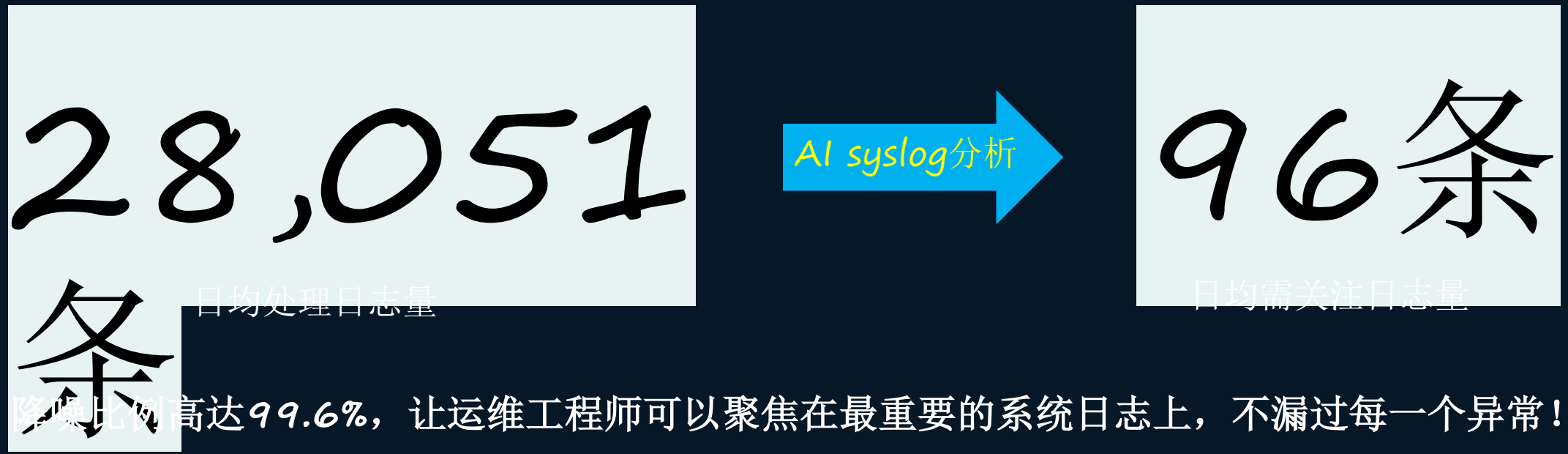
每天日志条目数：超10万条

其中值得关注的日志条目数=？

运维面临的挑战：

A公司AI-syslog 平台使用整体情况

- 自2024年9月部署运行以来，成功接收并处理70+网络设备（涉及到思科华为路由交换防火墙等设备）的syslog；
- 以2024-11月的某三日的日志量作为参考，**日均处理日志量28,051条**，经系统检查处理后，梳理出**需要特别关注的日志数为日均不到100条**。



降噪比例高达99.6%，让运维工程师可以聚焦在最重要的系统日志上，不漏过每一个异常！

基于AI-syslog的网络运维流程

①日志采集

- 网络设备将日志传递至AI-Syslog服务器
- AI-Syslog服务器收到日志后进行实时处理

②异常检测

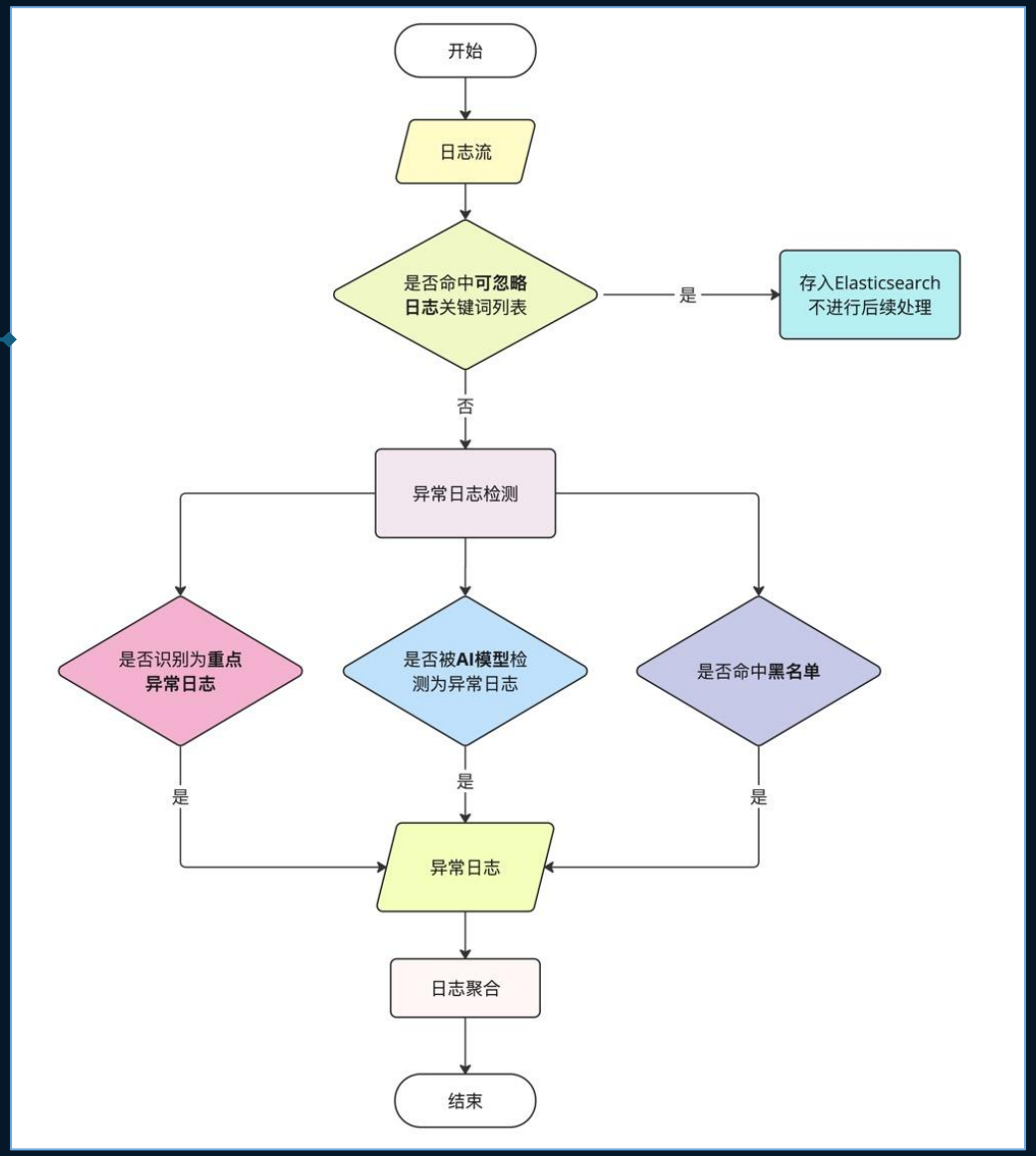
- AI-Syslog平台对收到的系统日志进行过滤和分析
- 对异常日志进行梳理和展现

③人工复核

- 对平台发现的异常日志进行人工复核
- 准确识别网络潜在故障与风险。

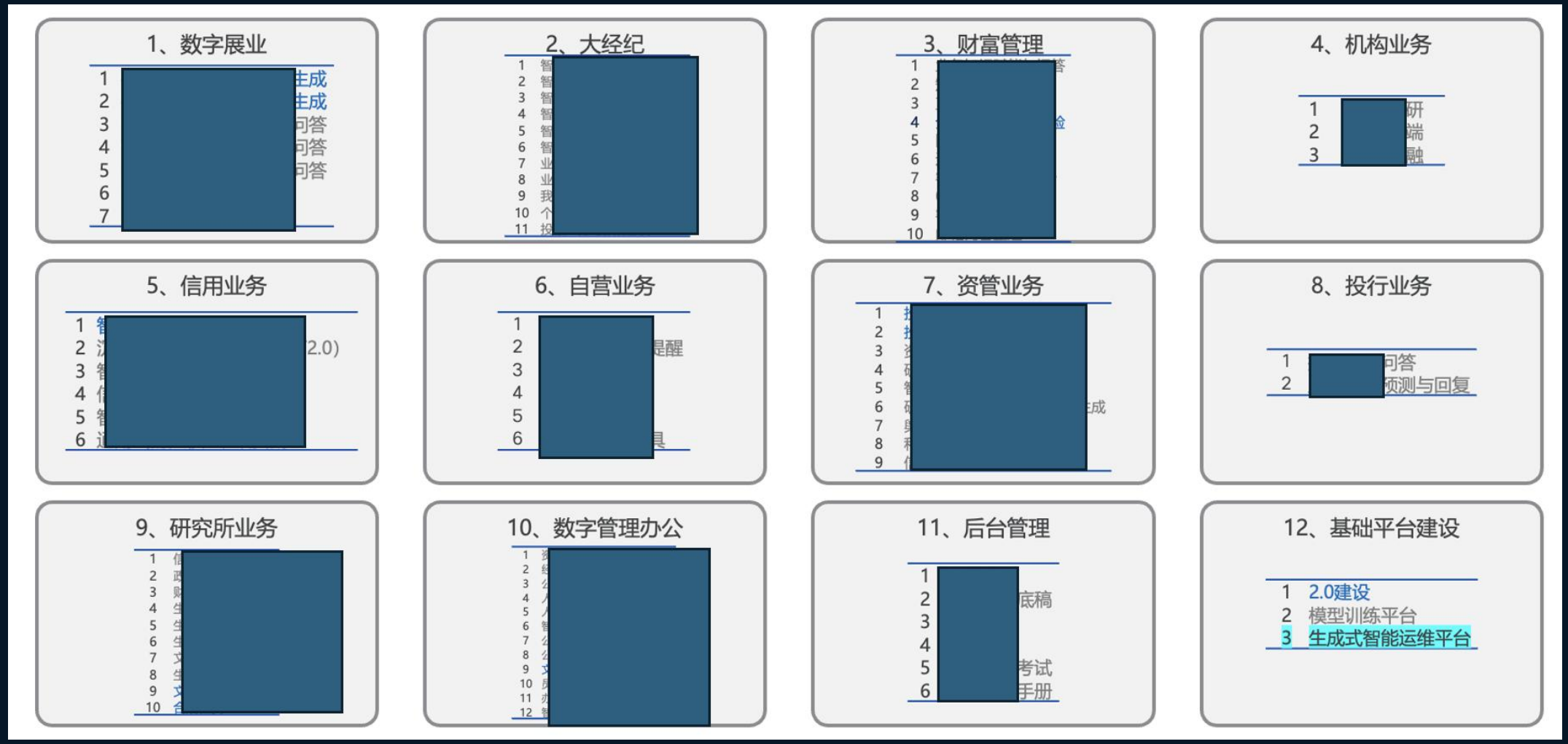
④跟进处理

- 基于人工复核结果
- 对发现的网络问题和风险进行跟进处理。



各团队基于LLM的应用场景

立足于业务发展目标：通过大模型技术的结合应用，助力业务规模攀升、产品体验升级、服务模式重塑等



日程

- 思科AI架构
- 思科CX如何使用AI并赋能客户（AI落地）
- AI 安全

  Cisco AI Defense介绍/ Foundation-Sec-8B 介绍

企业如何使用人工智能应用？

决定 1：我们的人工智能用例是什么？

- 代码生成、企业搜索、客户支持、代理助理、自动化等
- 风险：根据使用情况，人工智能应用可能面临外部对手和内部威胁

决定 2：我们如何开发我们的模型？

- 内部开发：完全定制，但成本高昂，工作量大（较少见）
- 使用基础模型：可在更便宜、更快的基础上进行开发（更常见）
- 风险：开源模型、第三方数据集和其他组件可能会遭到破坏

决定 3：我们如何定制模型？

- 检索增强生成（RAG）：51%[1]
- 提示工程：16%[1]
- 微调：9%[1]
- 风险：用于定制人工智能应用程序的敏感数据容易被提取

决定 4：我们如何使用第三方人工智能工具？

- 哪些应用程序被批准，哪些未被批准？
- 是否所有人工智能工具都经过了安全审查？
- 风险：员工与未经许可的人工智能工具共享敏感数据，从而暴露敏感数据

1. Menlo Ventures: The State of Generative AI in the Enterprise 2024

AI应用风险何在？

AI应用具有非确定性

AI 应用

- 用户 User
- 应用 Application
- 模型 Model
- 数据 Data
- 基础架构 Infrastructure

新型风险向量

- 商业和声誉受损
- 数据安全和隐私
- 供应链漏洞
- 网络攻击和威胁
- 合规

AI 安全之旅

- 为全组织安全启用生成式AI

发现Discovery

识别并揭示“影子AI”、
未授权的AI应用、
模型和数据

检测Detection

检测AI系统中的潜在
风险、漏洞，以及
对抗性攻击

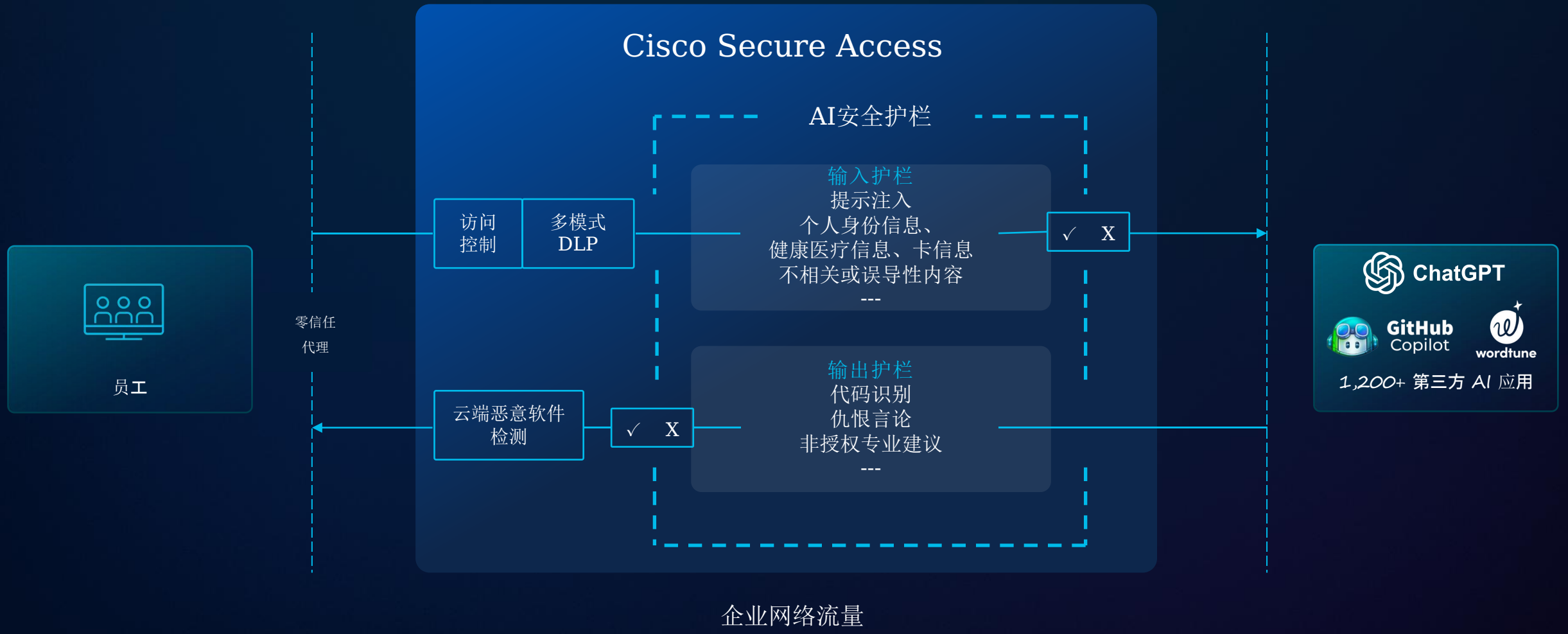
防护Protection

设置防护边界与访问
控制策略，保护数据
安全，抵御运行
时威胁

思科 AI Defense 解决方案

保护第三方AI应用的使用安全

Cisco Secure Access 中国数据中心于 4 月正式上线



思科ISOVALENT基于eBPF技术推出的云原生网络与安全平台

传统虚拟机工作负载

新型应用

AI / ML 工作负载

kubernetes

ISOVALENT


网络 / 路由


防火墙 / 微分段


负载均衡


网络可观察


VPN / 加密


运行安全

eBPF


BARE METAL

vmware

aws

Azure

Google Cloud

ORACLE Cloud Infrastructure

Alibaba Cloud

思科Hypershield适用于AI 规模数据中心的 AI 原生安全性

全球统一管理 HyperShield

思科安全云控制（SCC） + 自然语言接口

公有云 私有云

多域

终端侧执行点

End system enforcer

(Tesseract Security Agent)

- 虚拟机
- 容器（Kubernetes node）
- 服务器（Bare metal）

网络侧执行点

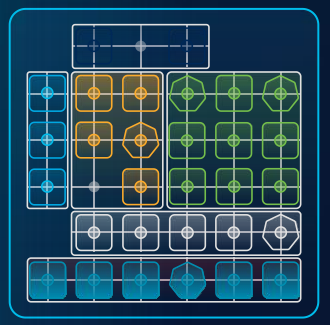
Network-based enforcer

- 虚拟服务器（VM appliance）
- 交换机口（Network switch port）
- 服务器接口（DPU NIC）

执行点

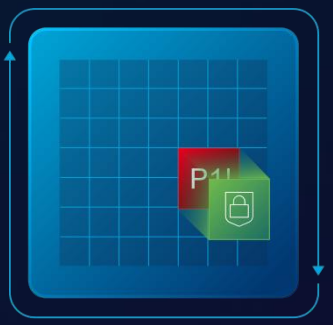
- 自我验证更新
- AI原生安全
- 内核级别执行

场景



Autonomous Segmentation

自主网络分段



Distributed Exploit Protection

分布式渗透保护

首个开源安全模型：Foundation-Sec-8B

Foundation AI

由一流的思科AI 安全研究人员和工程师组成的团队， 致力于尖端 AI 技术，为网络安全应用提供支持。

初始项目 ：

- Foundation-sec-8b：我们首个开源安全模型
- 评估网络安全模式的新基准
- AI 供应链风险管理和情报

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

使用案例

- SOC 加速：自动执行分类、摘要、 生成案例说明和收集证据。

主动威胁防御：模拟攻击，确定漏洞优先级，映射 TTP，并对攻击者行为进行建模。

工程支持：提供安全帮助，验证配置， 评估合规性证据，并改善安全状况。

Foundation-Sec-8B 演示：日志分析汇总

Use case: SOC

适用于各种网络及安全设备，利用模型解析syslog、Netflow、IPS等日志信息，进行异常行为检测、攻击分类、溯源分析，并提供自动响应建议。

模型提示语

```
log_content = pathlib.Path("asa_threat_log.txt").read_text()

# 4. 提示词
prompt = f"""
You are a senior Cisco security analyst at Kevins Security Operations Center (SOC).

Your task is to analyze the following firewall logs collected from Cisco ASA, Firepower, Catalyst 9000, and ISR devices, and generate a structured threat report for SOC revi

These logs may contain intrusion attempts, misconfigurations, access violations, or suspicious behavior. Your analysis should help identify patterns and assist in rapid miti

Please output your analysis in the following **JSON format**:

{{
  "timeline": [                    // Summarized sequence of key events
    {{
      "timestamp": "...",
      "event": "..."
    }},
    ...
  ],
  "attack_class": "...",           // Attack types like brute-force, spoofing, scanning, etc.
  "att&ck_mapping": ["T1110", "T1595"], // MITRE ATT&CK technique IDs
  "recommendation": "...",         // High-level security suggestions
  "mitigation": "..."              // Concrete configuration actions (e.g., ACL updates, IP block)
}}

Please follow these rules:
- Only return JSON as described above (no markdown, no explanations).
- Include at least 10 timeline entries based on key indicators.
- If attack type or ATT&CK mapping is unclear, make your best educated assumption.
- Assume the logs are from an enterprise edge firewall in production.

### Firewall Logs:
{log_content}
"""
```

模型结果

```
Problems 1  Output  Debug Console  Terminal  Ports 1                    python - Kevins-cisco_sec_test

   - Risk: Weak community strings can be easily guessed, allowing unauthorized access.
   - Recommendation: Change the SNMP community strings to strong, unique values.
   - Reference: CIS Benchmark for Cisco IOS - 2.2.2.

5. Unsecured TFTP Server:
   - Issue: The router uses the default TFTP server.
   - Risk: Using the default TFTP server is insecure and can be exploited.
   - Recommendation: Configure a secure TFTP server with strong authentication.
   - Reference: Cisco Secure Configuration Guide - 2.1.2.

6. Unsecured NTP Server:
   - Issue: The router uses an unsecured NTP server.
   - Risk: Unsecured NTP servers can be manipulated to cause time synchronization issues.
   - Recommendation: Use a secure NTP server with authentication.
   - Reference: Cisco Secure Config

7. Unsecured DNS Server:
   - Issue: The router uses an unse
   - Risk: Unsecured DNS servers can be manipulated to redirect traffic.
   - Recommendation: Use a secure DNS server with authentication.
   - Reference: Cisco Secure Configuration Guide -
(foundation-sec-8b) (base) yjb@kevins-yujianbo:~/06-本地私有化部署大模型 (Ollama/day06_本地私有化部署大模型/demo_06/Kevins-cisco_sec_test$ python Kevins-test2-asalogs.py
```

我们让Foundation-Sec-8B来解析网络日志，

本白皮书从宏观层面的发展趋势分析出发，按照产业机遇研判、企业AI变革实践调研、关键变革要素细致剖析、拆解产品服务案例提炼有益方法论的研究思路逐层深入，以求助力企业从精微处快速采取行动。

同时，本白皮书深入剖析了“AI Ready”的数据中心在计算、网络、存储及相关治理组件等关键要素，并前瞻性地提出了AI变革就绪度评估体系。

我们希望，这不仅能为企业决策者提供清晰的指引，更能为技术管理者提供具体可落地的行动方案。

王世蔚
思科商业市场客户经理

张志清
思科商业市场客户经理

智领未来
思科2025科技创新AI峰会
北京 | 2025.05.23

人工智能就绪度白皮书：
企业数智化转型的 AI 变革路径与评估指南

扫码下载白皮书

战略支持

KPMG