WELCOME TO THE 11TH ANNUAL HACKING CONFERENCE

Building and Auditing an AI-Enabled Third Party Security Program

CPE code: eaglefang

Vinita Apte

Head of Vendor and Partner Ecosystem Security at Circle

The Institute of Internal Auditors Chicago

ISACA Chicago Chapter

Session Objectives

- Learn how to build an AI-enabled third party security assessment program
- Get tips on auditing such a program

Do I need to be a security expert to attend?

No! The ideas discussed here are applicable to any type of risk assessment - e.g. privacy, financial

Do I need to be fluent in using a specific AI tool?

Nope. The slides and ideas tool-agnostic

Poll

How would you describe your organization's AI maturity today?

Quick Intro to Third Party Security (TPS)

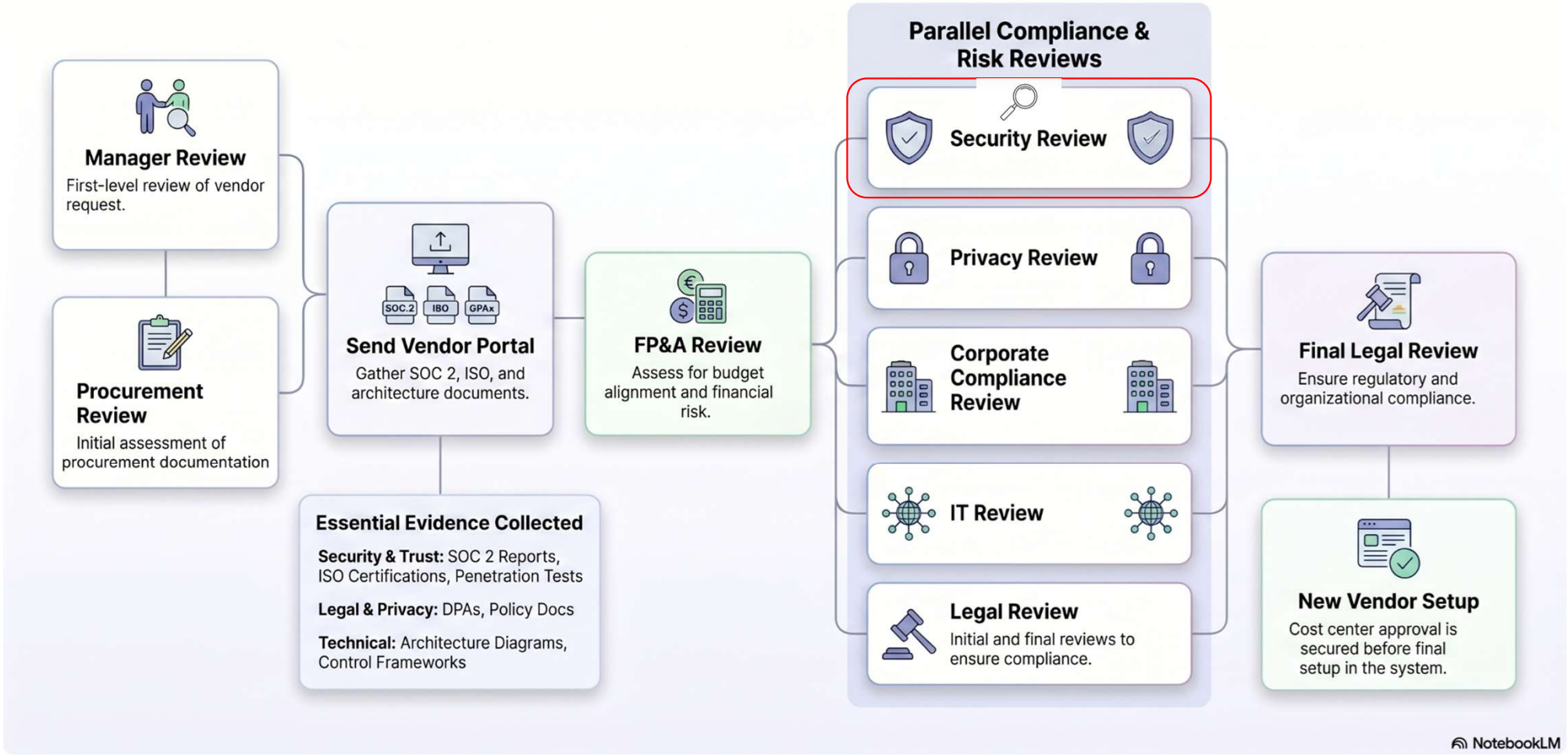
Identify weaknesses in a vendor's security controls before they become a problem for your org

- Allows timely remediation
- Enables effective risk management and business decision making

When assessments are typically performed

- At the time of onboarding
- At a pre-defined frequency (e.g. annually for High risk vendors)
- Based on real-time triggers (e.g. security incident)

The Standardized Vendor Onboarding Journey



Phase 1: The Manual Foundation of Third-Party Security

MANUAL INTAKE & WORKFLOWS

FORMS

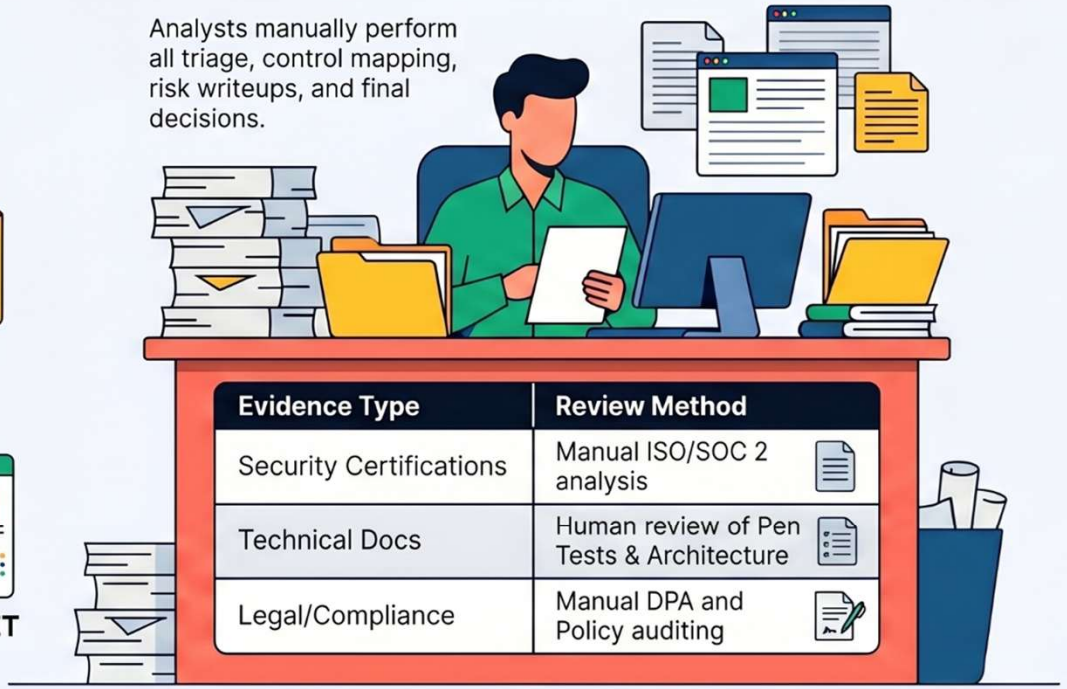
EMAIL

TICKETING SYSTEM

SPREADSHEET

HEAVY HUMAN ROLE

Analysts manually perform all triage, control mapping, risk writeups, and final decisions.

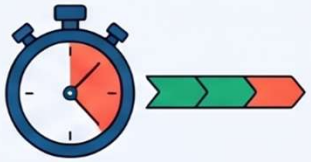


Evidence Type	Review Method
Security Certifications	Manual ISO/SOC 2 analysis
Technical Docs	Human review of Pen Tests & Architecture
Legal/Compliance	Manual DPA and Policy auditing

MANUAL THREAT INTEL TRACKING AND RESPONSE

Monitoring is restricted to periodic, non-systematic checks and ad hoc threat alerts.

CRITICAL OPERATIONAL PAIN POINTS



SLOW TURNAROUND

INCONSISTENT REVIEWS

EVIDENCE OVERLOAD & BACKLOGS

Day 0 – Existing Process

Team Intake Queue

The entry point where tasks are assigned to a specific security analyst.

Assigned to Analyst

Info Gathering

Analysts collect essential evidence like SOC 2 reports, ISO certifications, and penetration tests.

Documentation Review

Manual evaluation of vendor evidence against standard control frameworks and policy documents.

Risk Analysis

Analysts perform risk writeups and escalations based on gathered evidence and real-time monitoring.

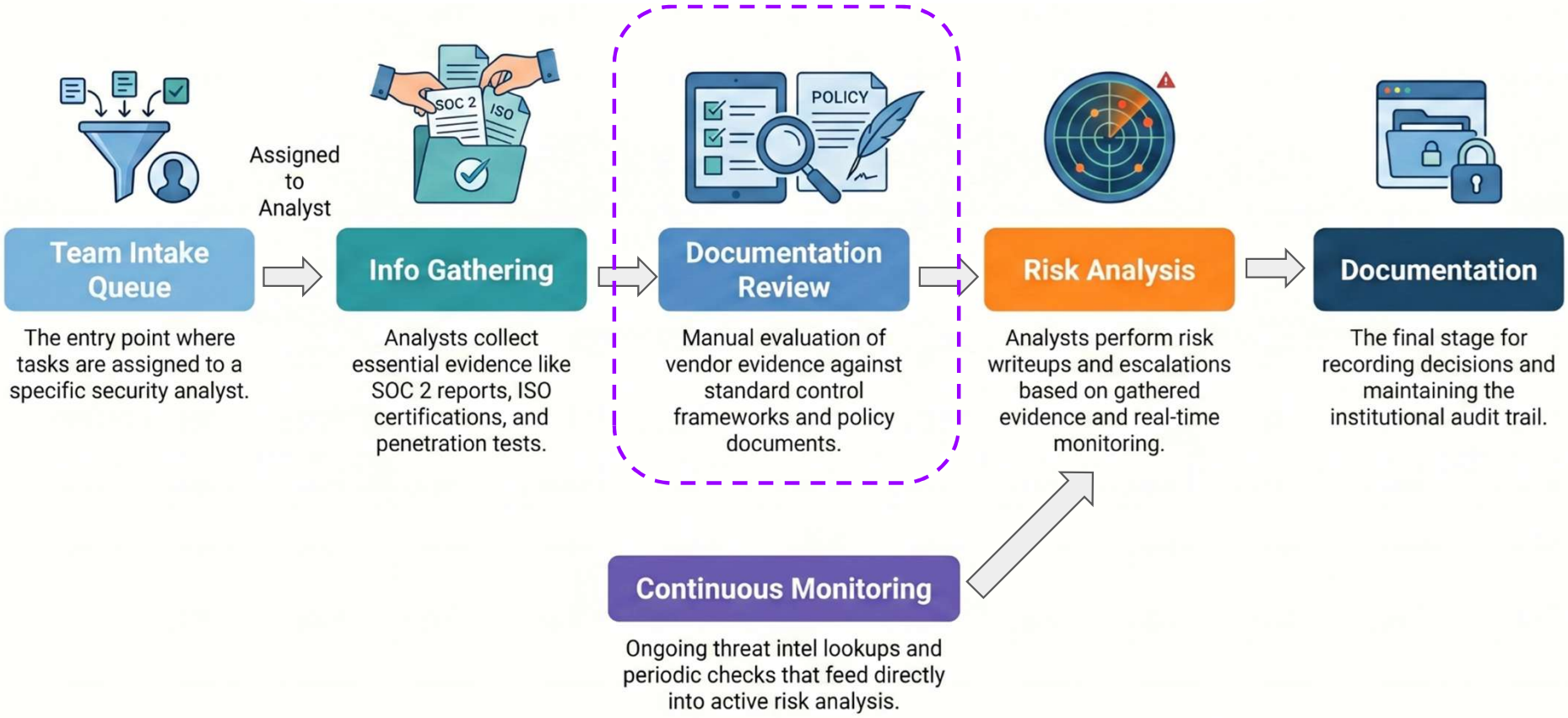
Documentation

The final stage for recording decisions and maintaining the institutional audit trail.

Continuous Monitoring

Ongoing threat intel lookups and periodic checks that feed directly into active risk analysis.

Step 1 - Quick efficiency gains

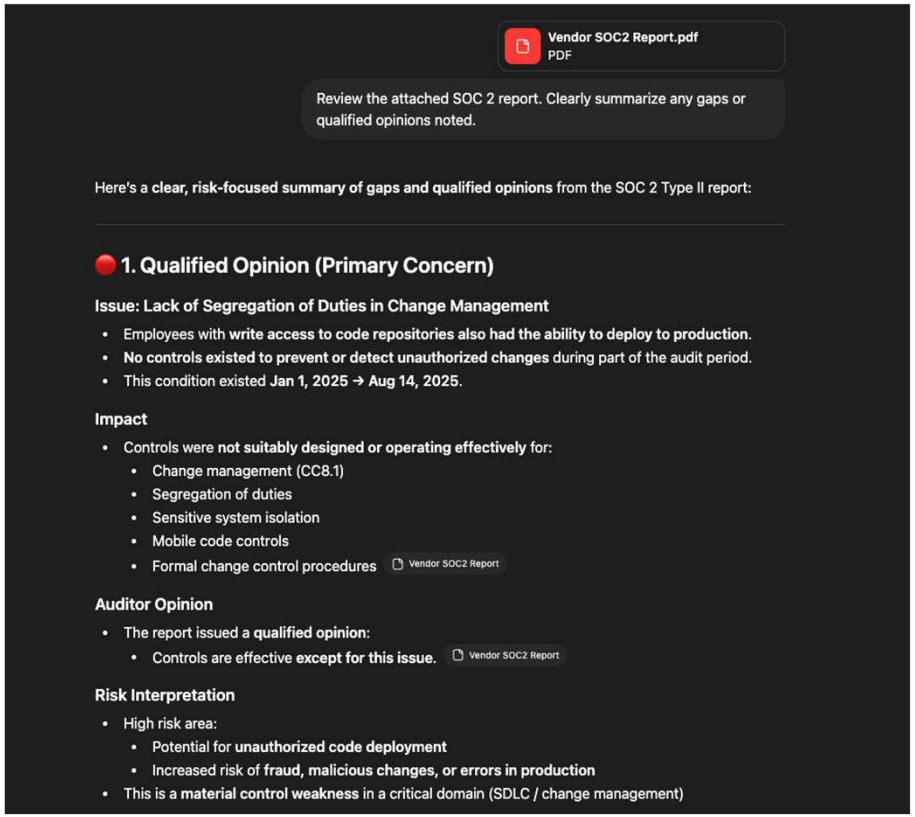


AI-led documentation review

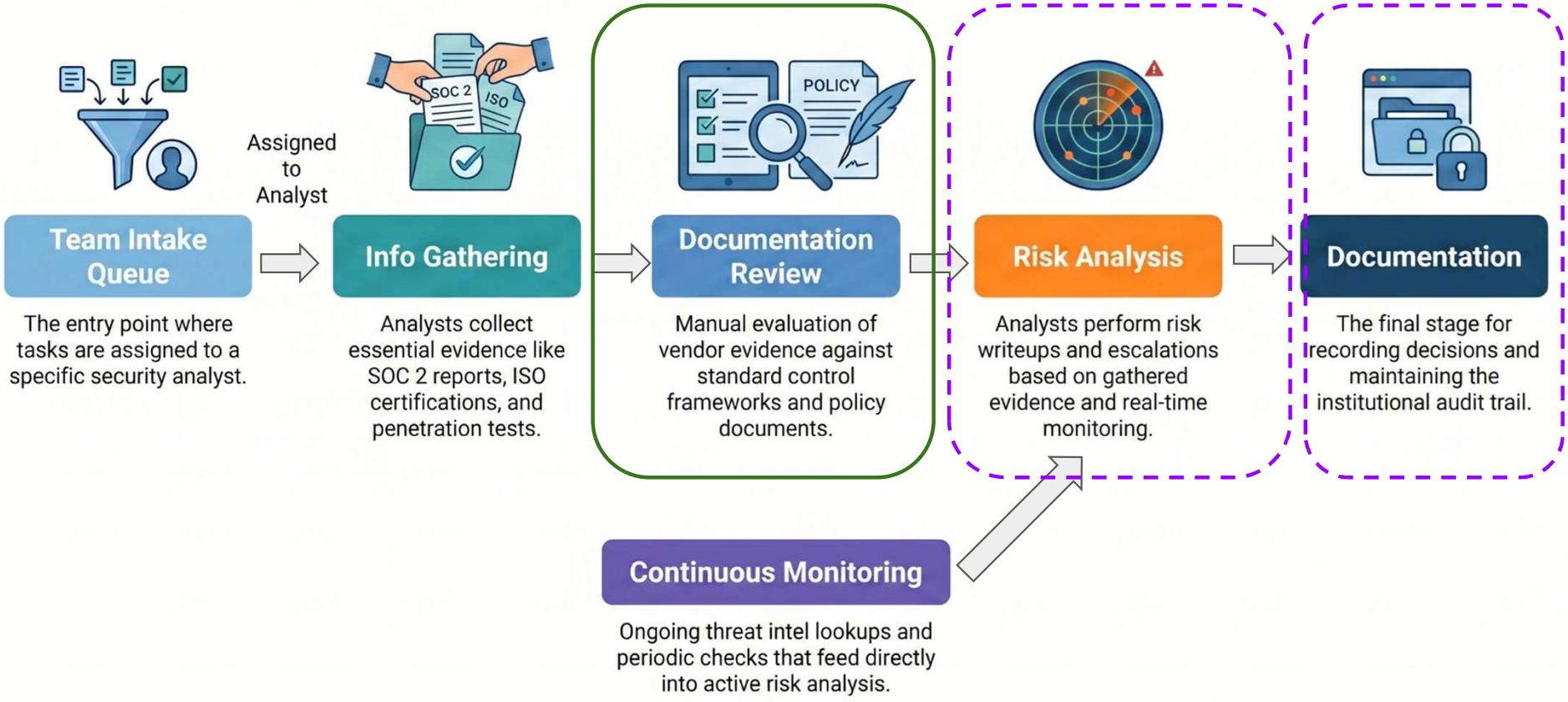
Tips

- Tell the AI exactly what you want highlighted
- Standardize - e.g. Gem, Custom GPT, Skill
- Human-in-the-loop QA review

Saves 2-4 hrs per assessment



Step 2 - Unlocking Consistency & Depth



AI-led Risk Analysis

1. Achieving more **consistent** assessments
    - Establish clear risk levels for each control you're assessing in your security questionnaire
        - What's Low/Medium/High/Critical risk?
        - What's outside of our risk appetite?
    - Then train the AI on this framework
    - Minimizes discrepancies between assessors

Question	Response Option 1	Response Option 2	Response Option 3	Response Option 4
Describe your controls to detect/prevent unauthorized transmission of sensitive data outside your organization.	Policy level requirements that forbid unauthorized transmission of data outside the organization	External transmissions (including email, browser traffic) are monitored for sensitive data. An alert is triggered if a suspicious transmission is detected.	External transmissions (including email, browser traffic) are inspected for sensitive data. We use a combination or blocking and alerting for suspicious data transfers.	
Select the response that best describes your organization's approach to security event logging.	Decentralized approach. Every system maintains it's own local logs..	Combination of centralized and decentralized - some systems write logs to a centralized log tool, while others maintain logs locally.	Centralized approach. The majority of our systems write their security logs to a central repository.	
Select the response that best describes your organization's approach to security event monitoring and alerting.	Adhoc log reviews/monitoring is performed	We perform manual log reviews and have partially automated this process	Automated monitoring is enabled, that include alerts to security personnel based on certain rules/thresholds	We utilize a 24x7 SOC that monitors our logs and responds to alerts.

AI-led Risk Analysis

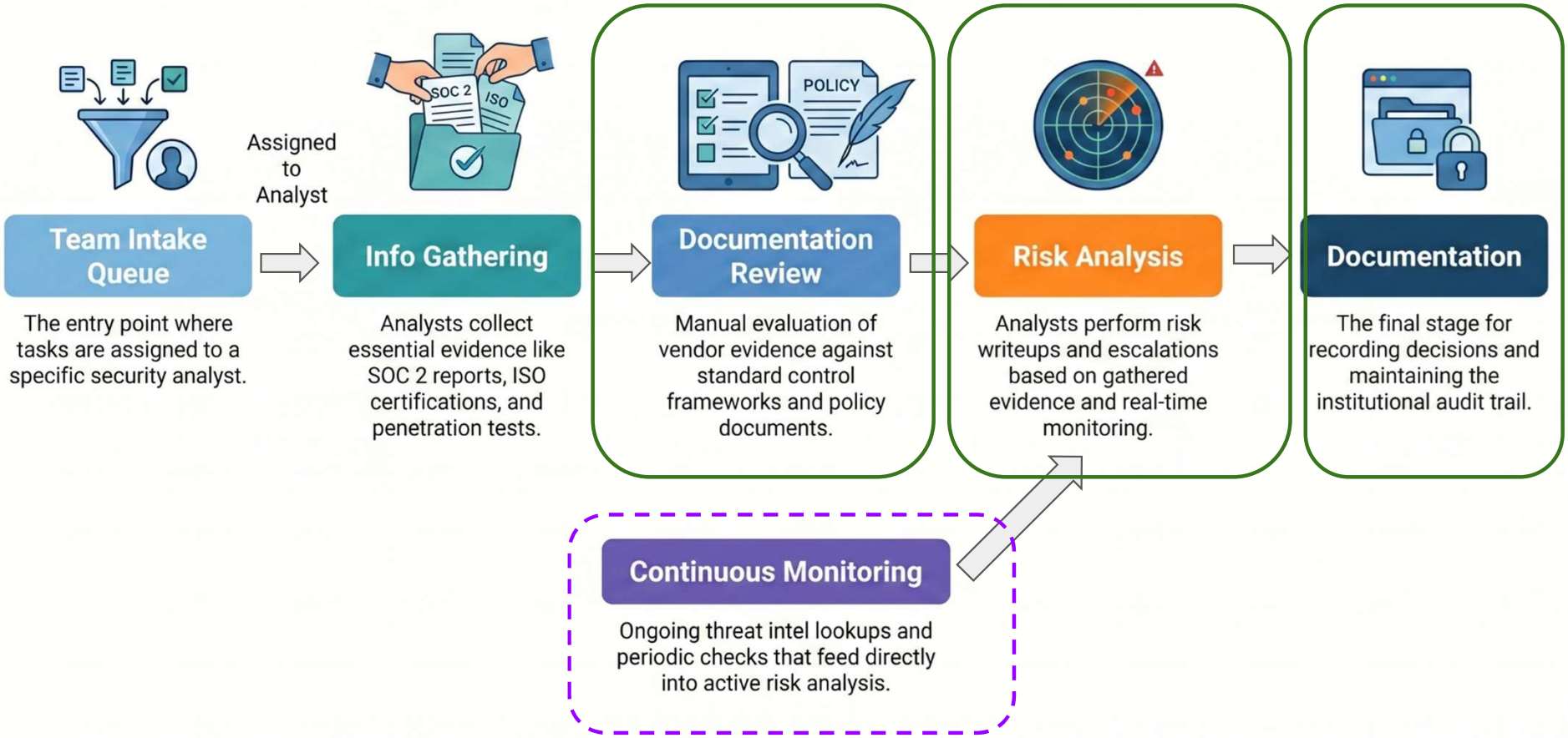
2. Unlocking **deeper dive** assessments
    - AI can review many more data sources than a human - leverage this to identify prior security incidents, regulatory concerns, negative news etc
    - AI tools can search across multiple internal data sources (Slack conversations, shared drives, Jira tickets etc) to identify potentially unknown vendor use cases

- Standardize - e.g. Gem, Custom GPT, Skill
- Human-in-the-loop QA review
- Saves 4-6 hrs per assessment

AI-led Report Generation

- AI can take inputs from the Documentation Review and Risk Analysis steps, and generate an assessment report in the standard format
  - ✓ Consistent formatting
  - ✓ Reduces likelihood of manual errors
- Standardize - e.g. Gem, Custom GPT, Skill
- Human-in-the-loop QA review
- Saves ~2 hrs per assessment

Step 3 - Faster Incident Response



Continuous Monitoring

Portfolio Monitoring

Organizations typically use tools like SecurityScorecard or Bitsight to monitor their third-party portfolio.

Real-time Alerts

Tools alert you when a vendor security event is detected, based on publicly available data.

Common Threats

Monitoring detects data breaches, ransomware attacks, exposed credentials, and more.

Incident Response

A **critical control** to quickly identify and respond to third-party security incidents.

AI-led Continuous Monitoring

Instant Alert Pick-up

An AI workflow can pick up an alert **instantly** and analyze it against available information.

Automated Analysis

- Is this an active vendor?
- What data do we share?
- Has the incident been confirmed?

Incident Ticketing

Creates an incident ticket with a summary and notifies relevant teams automatically.

Efficiency & Audit

Saves manual time with a ready "package" for analyst review and creates an auditable trail of response.

Confirmed Example Co incident

+ ⊙

KEY DETAILS

Description
Third Party Name
Example Co

Classification
Incident

Impact Category
Security

Summary

Confirmed: Example Co experienced two human-error data exposures in late Q1—a CMS misconfig
client-side source code. Example Co states no customer data or credentials were exposed and has ren

Overview

Public reporting and the vendor's statements indicate two distinct incidents within days: 1) a CMS co
source map referencing unobfuscated client
reports describe threat actor
velopers.

Analysis

Cause

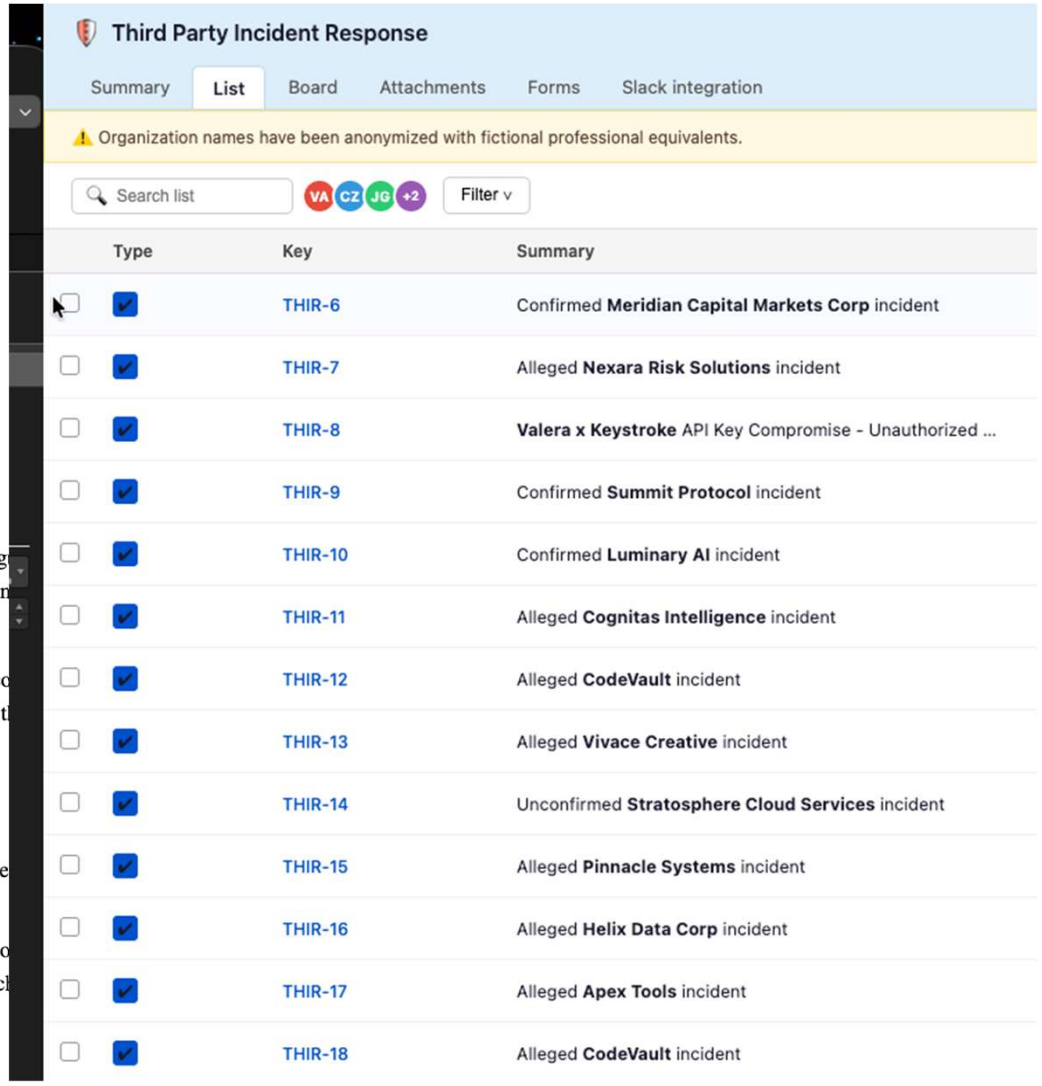
Human error. A misco
and a flawed release proce

Exposure

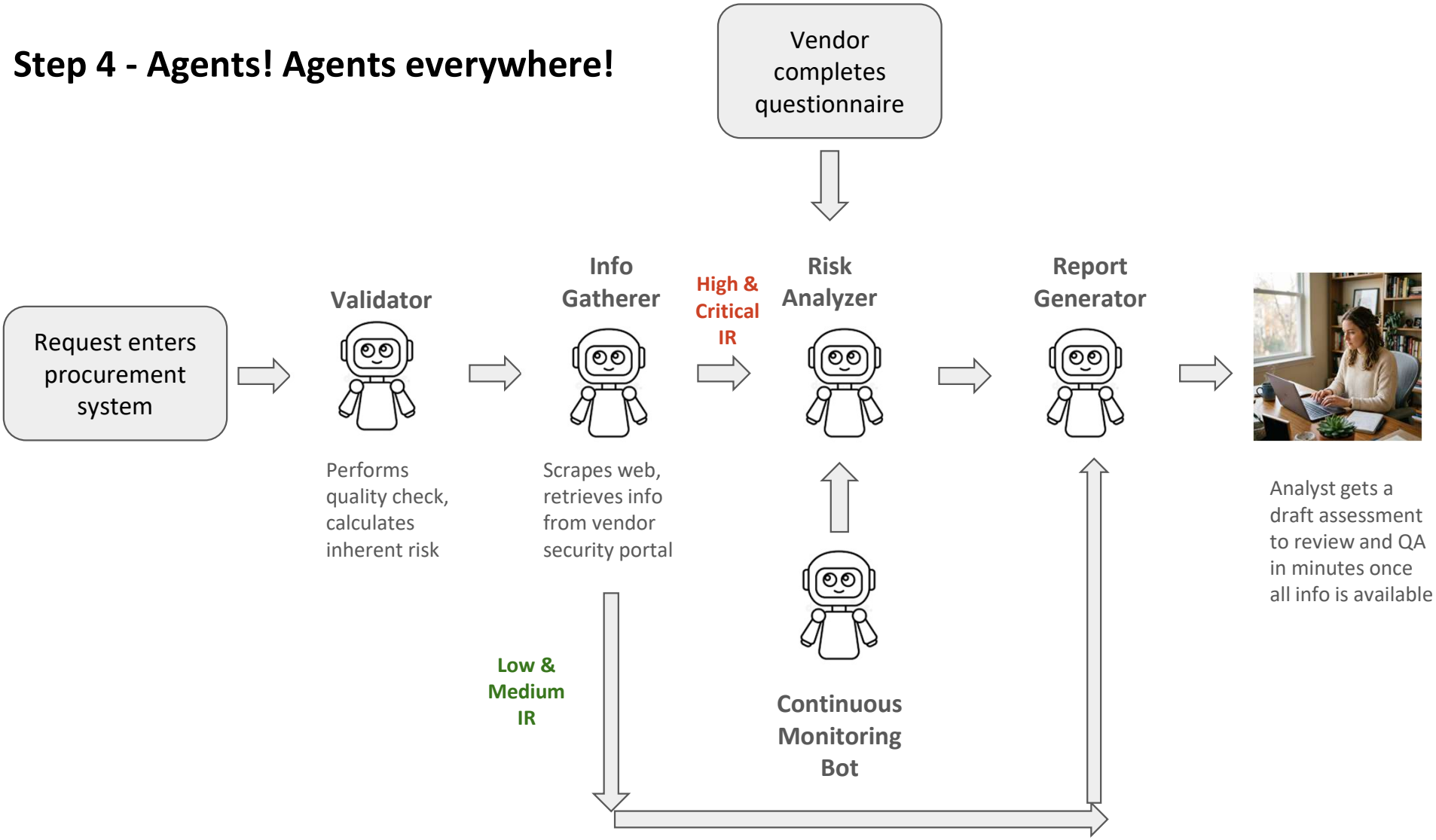
Internal assessment: Vendor has access to organization systems via API and browser-based integratio
shared content, which may trigger confidentiality, contractual, or regulatory notification duties if such
trust could be affected by source exposure and downstream malware campaigns.

The AI agent confirmed the incident on the vendor's website

The agent analyzed breach info against internal vendor assessment info



Step 4 - Agents! Agents everywhere!



Some Best Practices

Train your AI to be factual

- Always cite sources
- Provide a list of trusted sources
- Ok to say “I don't know” instead of making up data

Optimize Agent Focus

Agents work best when they have narrow, well defined tasks

Human-in-the-Loop

Design your AI for human validation at critical decision points

Test Before Scaling

Test with known examples before scaling

Monitor and improve continuously

AI tools can drift and require course correction

More AIdeas

Program Improvement

"Review my assessment questionnaire and map it to the NIST Cybersecurity framework. Identify gaps, simplify questions, and flag duplicates"

Ad Hoc Reporting

"From the attached list of vendors, flag ones that have a significant presence in the Middle East"

Regulatory Monitoring

"Analyze the updated requirements from Singapore MAS. Give me a summary of key changes and identify gaps in my program"

RFP/Vendor Selection

"Compare security responses provided by the 3 vendors and tabulate responses. Highlight strengths and weaknesses in each response and provide an overall assessment"

Poll

What do you see as the biggest roadblock for AI adoption in your org?

Auditing an AI-enabled Assessment Program

Security Auditor Assessment Scenario

You are a security auditor assessing an AI-enabled third-party security program at a fintech company

Key Questions to Ask

- Where in the lifecycle is AI used?
- What decisions does AI influence vs. make?
- What tools/models are in scope?

Higher Scrutiny Areas

Higher scrutiny is required where AI:

- Influences risk ratings
- Impacts vendor onboarding decisions
- Drives monitoring/escalation triggers

Audit Checklist

1. Accuracy

- AI produces correct output based on framework
- Human-in-the-loop validation exists
- Mechanism for error detection & override

2. Transparency

- Explainable outputs traceable to source evidence (no hallucinations!)

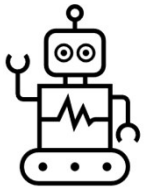
3. Consistency

- Produces consistent output when presented with the same input

4. Auditability

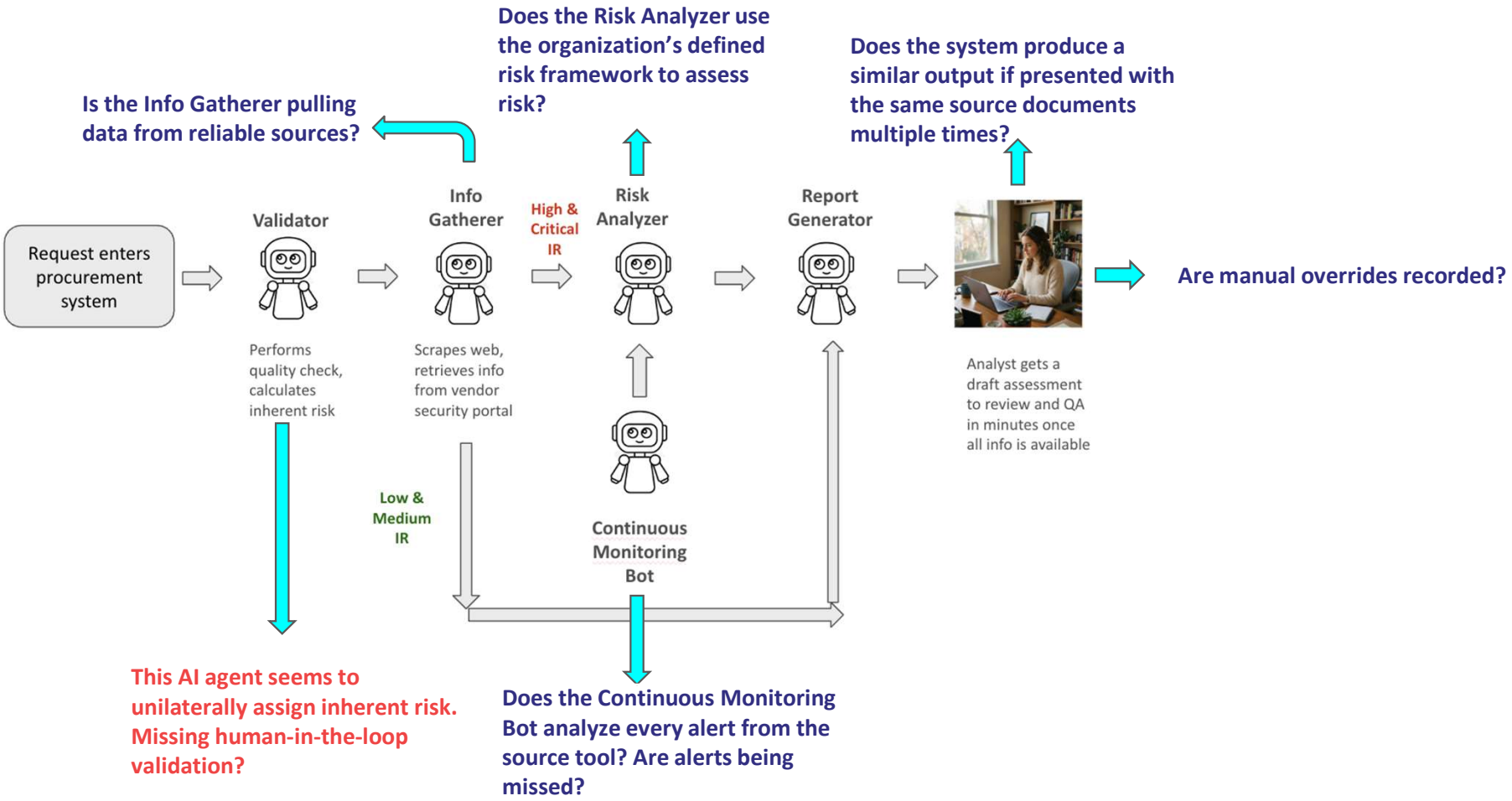
- An end-to-end trace of decisions exists

“If you can’t explain it, trace it and challenge it; don’t trust it.”



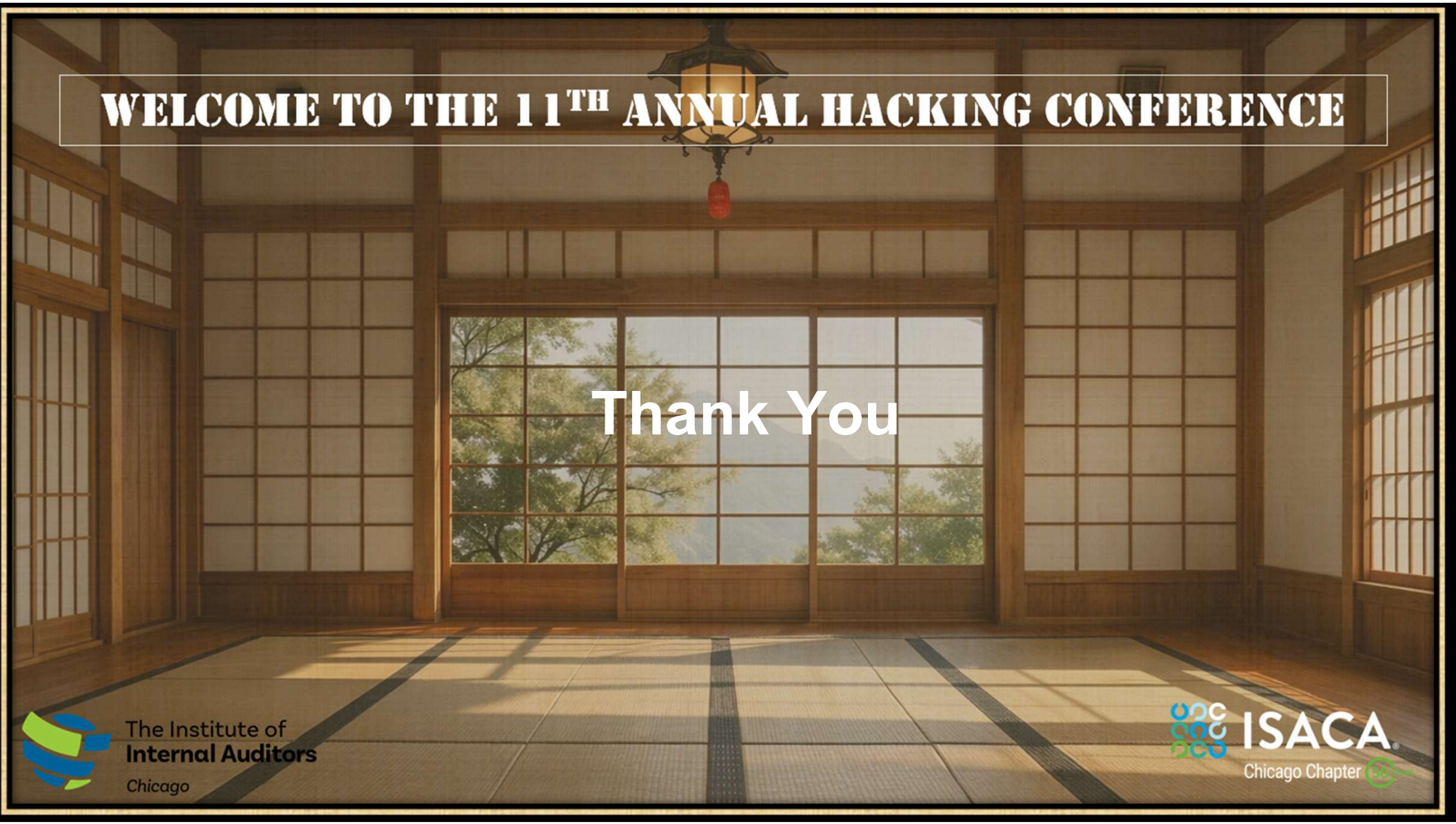
Audit Simulation

Now let’s do a quick audit of the AI-enabled third party security program we built.


Is the Info Gatherer pulling data from reliable sources?
Does the Risk Analyzer use the organization's defined risk framework to assess risk?
Does the system produce a similar output if presented with the same source documents multiple times?
Request enters procurement system
Validator
Info Gatherer
High & Critical IR
Risk Analyzer
Report Generator
Performs quality check, calculates inherent risk
Scrapes web, retrieves info from vendor security portal
Analyst gets a draft assessment to review and QA in minutes once all info is available
Are manual overrides recorded?
Low & Medium IR
Continuous Monitoring Bot
This AI agent seems to unilaterally assign inherent risk. Missing human-in-the-loop validation?
Does the Continuous Monitoring Bot analyze every alert from the source tool? Are alerts being missed?

Poll

What's the main thing you're walking away with today?


WELCOME TO THE 11TH ANNUAL HACKING CONFERENCE
Thank You
The Institute of
Internal Auditors
Chicago
ISACA
Chicago Chapter