FAQ

- **Q&A:** There will be time for questions at the end of the live presentation. If you have a question during the live presentation, please submit it through the Q&A feature. After the presentation, visit RSAC™ Membership to discuss with peers.
- **Recording & Slides:** You will receive an email within 24 hours containing a link to the video replay and the presentation deck.
- **CPE Credits:** Following the live webcast, you will receive an email containing your attendance certificate.
- **Cybersecurity Learning Program:** Share your expertise by contributing to a webcast, podcast, seminar, or blog. Visit RSAConference.com/BecomeAContributor for more information.

RSAC

#RSAC

RSAC™
Cybersecurity, Together™

What's in Your AI?
The Case for AI Bills of Materials

Allan Friedman, PhD
Senior Technical Advisor
Institute for Security & Technology
allan@securityandtechnology.org

This Webcast is Sponsored By: WIZ

Webcast

Allan Friedman

- Researcher and advocate for supply-chain transparency
- Former U.S. government official; helped lead SBOM efforts at NTIA and CISA
- SBOM was kinda my fault
- Important: SBOM was not my idea
- Equally important: AIBOM is also not my idea
- *Note: This presentation is based on my own research and perspective*

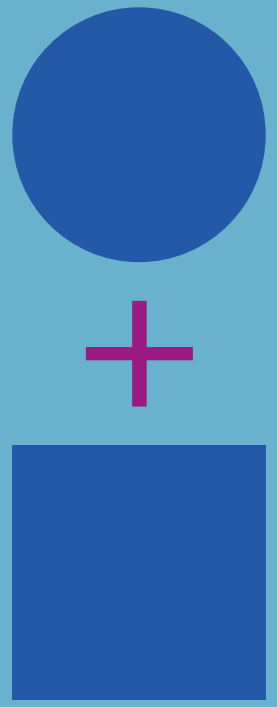
RSAC

- AI systems have supply chains, and risk attaches to identifiable components and relationships.
- An AIBOM provides structured inventory and becomes valuable when joined with external risk information and operational context.
- Progress towards AI transparency requires a useful, available, and consistent baseline supported by both suppliers and customers.

Some combinations are inevitable

Cybersecurity + Acronyms

Conferences + Lanyards



RSAC + Swag Bags

Phones + Chargers

RSAC

Some combinations are inevitable

Cybersecurity + Acronyms

Conferences + Lanyards

SBOM-style Transparency

+

AI supply-chain risks

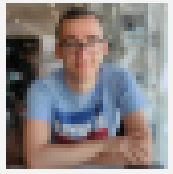
RSAC + Swag Bags

Phones + Chargers

RSAC

Scary security stories around AI and supply chain

Data Scientists Targeted by Malicious Hugging Face ML Models with Silent Backdoor



SHARE:

By David Cohen, JFrog Senior Security Researcher

February 27, 2024

RSAC

Fake OpenAI Privacy Filter Repo Hits #1 on Hugging Face, Draws 244K Downloads

Ravie Lakshmanan May 11, 2026 Supply Chain Attack / Threat Intelligence

Kyle Wiggers

TechCrunch

OpenAI rolls back update that made ChatGPT 'too sycophant-y'

OpenAI CEO Sam Altman said Tuesday the company is "rolling back" the latest update to the default AI model powering ChatGPT, GPT-4o, after complaints about strange behavior, in particular extreme sycophancy.

"[W]e started rolling back the latest update to GPT-4o last night," Altman wrote in a post on X. "[I]t's now 100% rolled back for free [ChatGPT] users and we'll update again when it's finished for paid users, hopefully later today. [W]e're working on additional fixes to model personality and will share more in the coming days."

RSAC

OpenAI's ChatGPT back up after brief outage

By **Reuters**

November 8, 2024 9:34 PM EST · Updated November 8, 2024

ChatGPT logo is seen in this illustration taken, March 11, 2024. REUTERS/Dado Ruvic/Illustration/File Photo Purchase Licensing Rights

Nov 8 (Reuters) - Microsoft-backed OpenAI's popular chatbot ChatGPT was back up after an outage affected thousands of users, the company said on Friday.

Is Claude Down? Anthropic Says It Fixed the Latest AI Outage

Anthropic identified issues affecting multiple Claude models in the early hours, but now says all of its systems are fixed.

+ Follow AI

OUR EXPERT

James Peckham

Reporter

With over a decade of experience reporting on consumer technology, James covers mobile phones, apps, operating systems, wearables, AI, and more.

Updated August 24, 2026

Largest Dataset Powering AI Images Removed After Discovery of Child Sexual Abuse Material

The model is a massive part of the AI-ecosystem, used by Stable Diffusion and other major generative AI products. The removal follows discoveries made by Stanford researchers, who found thousands instances of suspected child sexual abuse material in the dataset.

SAMANTHA COLE · DEC 20, 2023 AT 7:00 AM


Supply chain is only part of a complete AI security breakfast
It will not solve every problem involving model behavior, access, misuse, or application security.

What can go wrong

- **Compromise:** poisoned dataset, backdoored model, vulnerable software
- **Change:** silent model updates, shifting datasets, substituted components
- **Dependency:** hidden services, tools, suppliers, concentration
- **Provenance:** sensitive or improperly sourced data, licensing, inherited risk
- **Uncertainty:** information unavailable, outdated, unknown, or withheld

AI systems are software systems, so standard software supply chain risks also apply.

RSAC

Software Dependencies

Hosted Foundation Model

Agent Orchestration

Fine-tuned adapter

Embedding Model

Modern AI systems have many components

Agent Tools

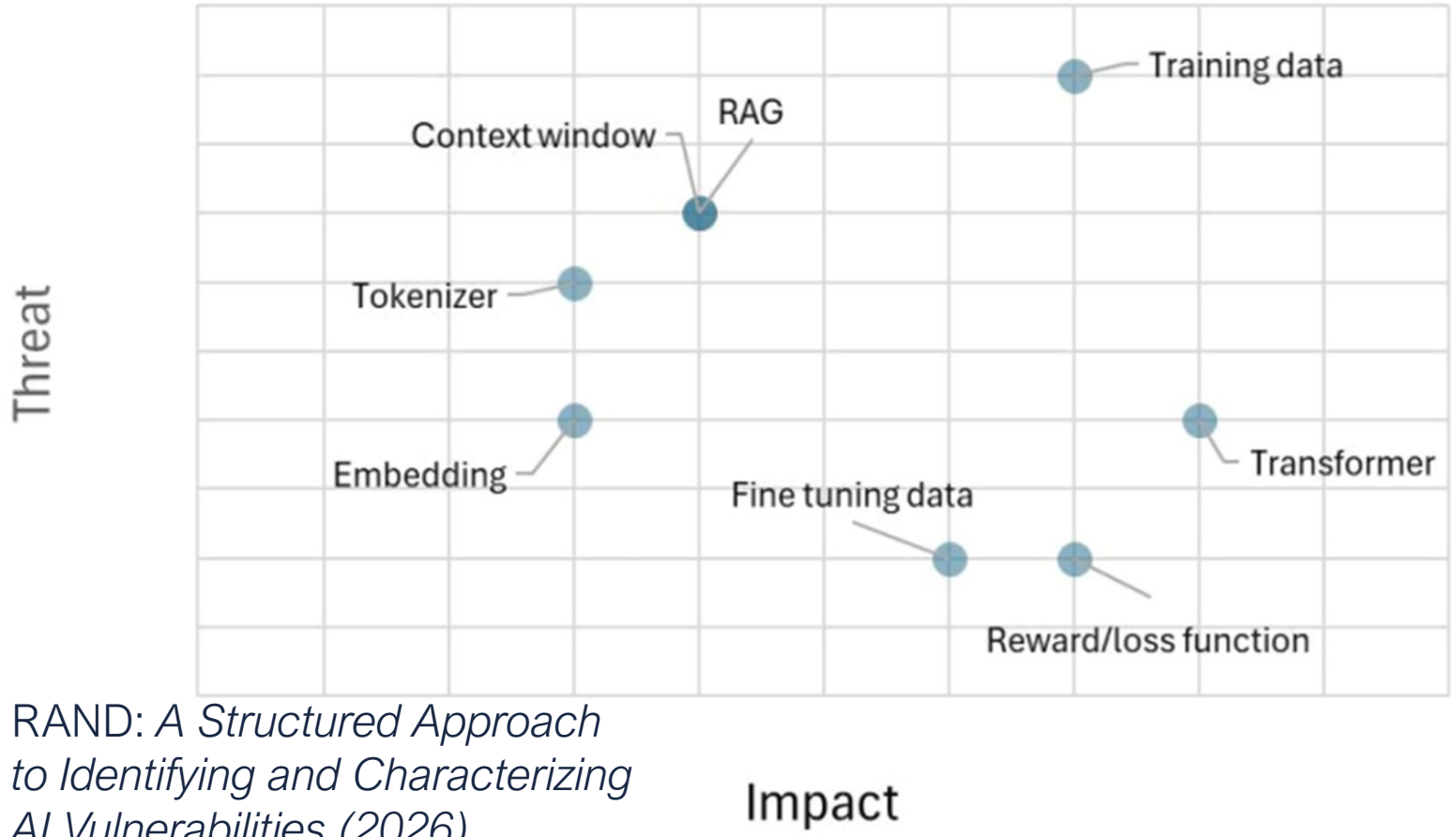
Training or fine-tuning dataset

Internal Knowledge Base

RSAC

Risks vary across different components

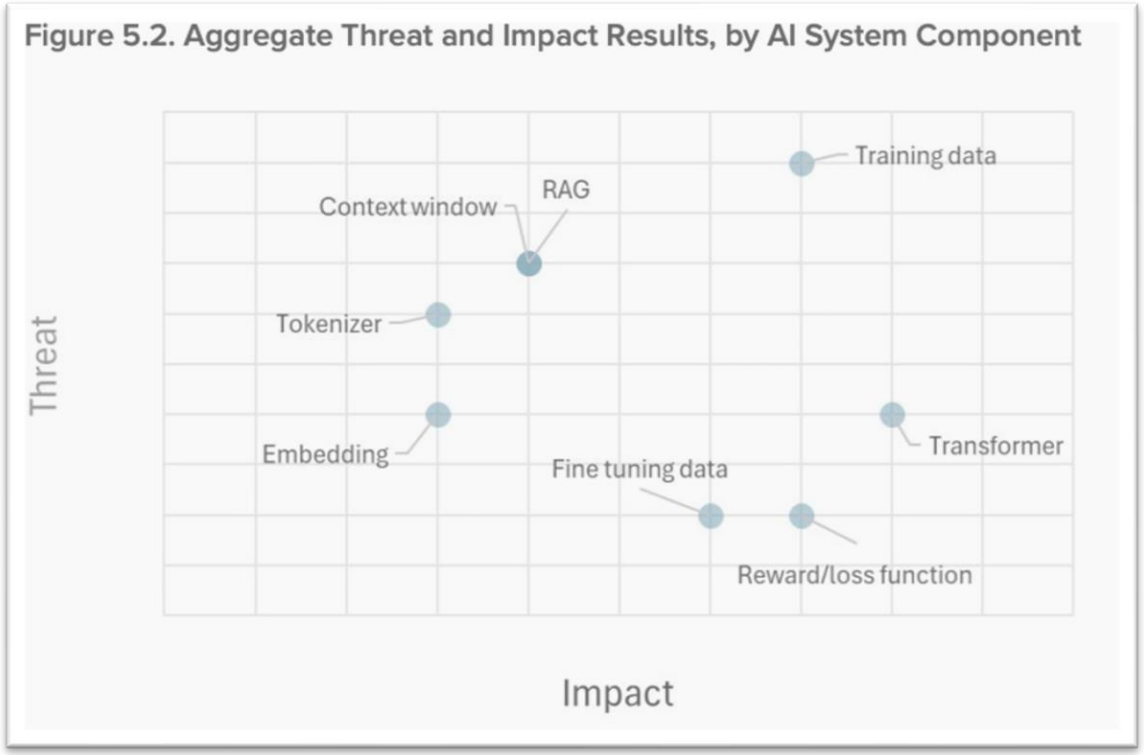
Figure 5.2. Aggregate Threat and Impact Results, by AI System Component



RAND: *A Structured Approach to Identifying and Characterizing AI Vulnerabilities (2026)*

https://www.rand.org/pubs/research_reports/RRA4983-1.html

Risks vary across different components

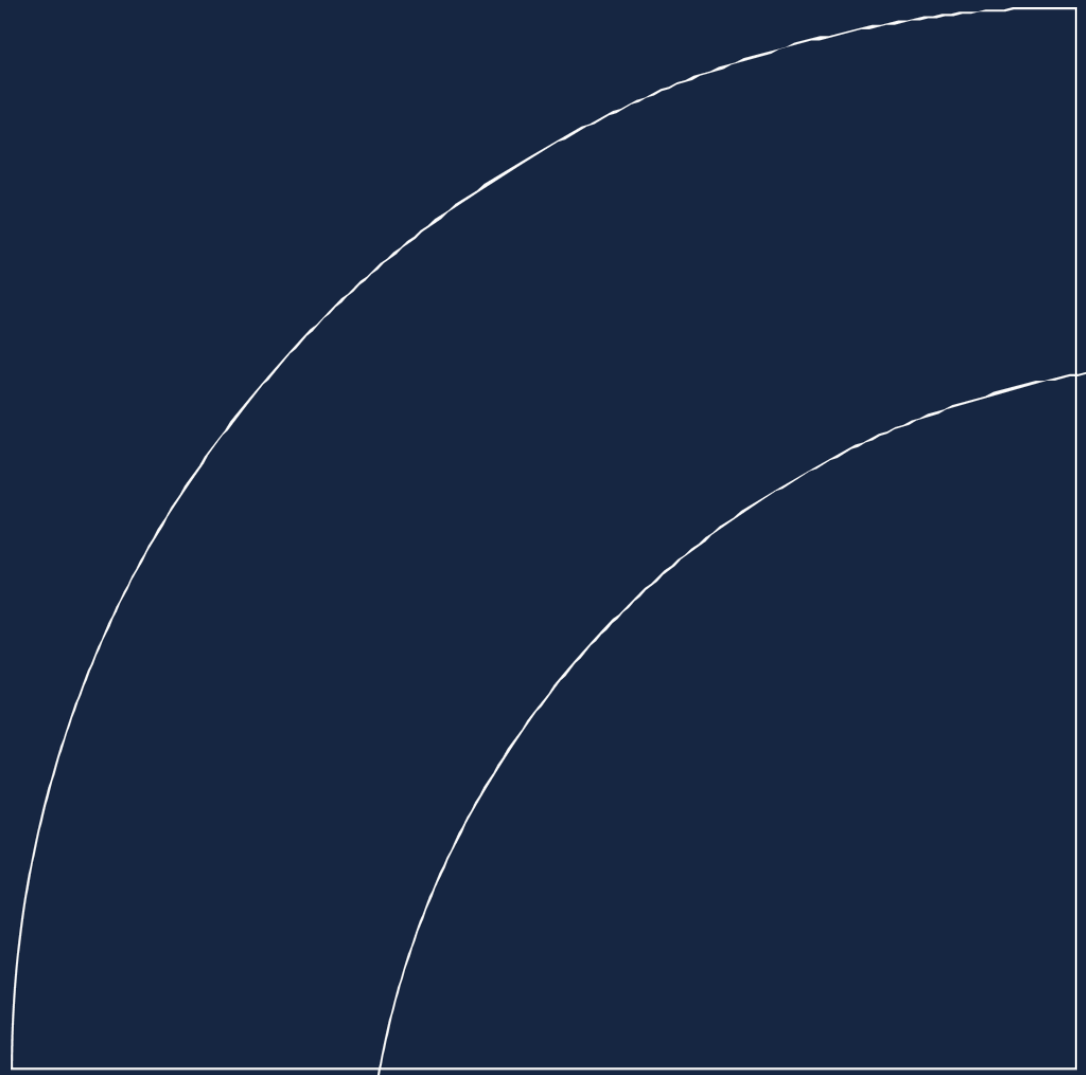


RAND: *A Structured Approach to Identifying and Characterizing AI Vulnerabilities (2026)*

- **AI risk attaches to identifiable components.**
- **Different components produce different risk profiles.**

https://www.rand.org/pubs/research_reports/RRA4983-1.html

AI and Transparency



LunchBot

Hosted foundation model

Fine-tuned adapter

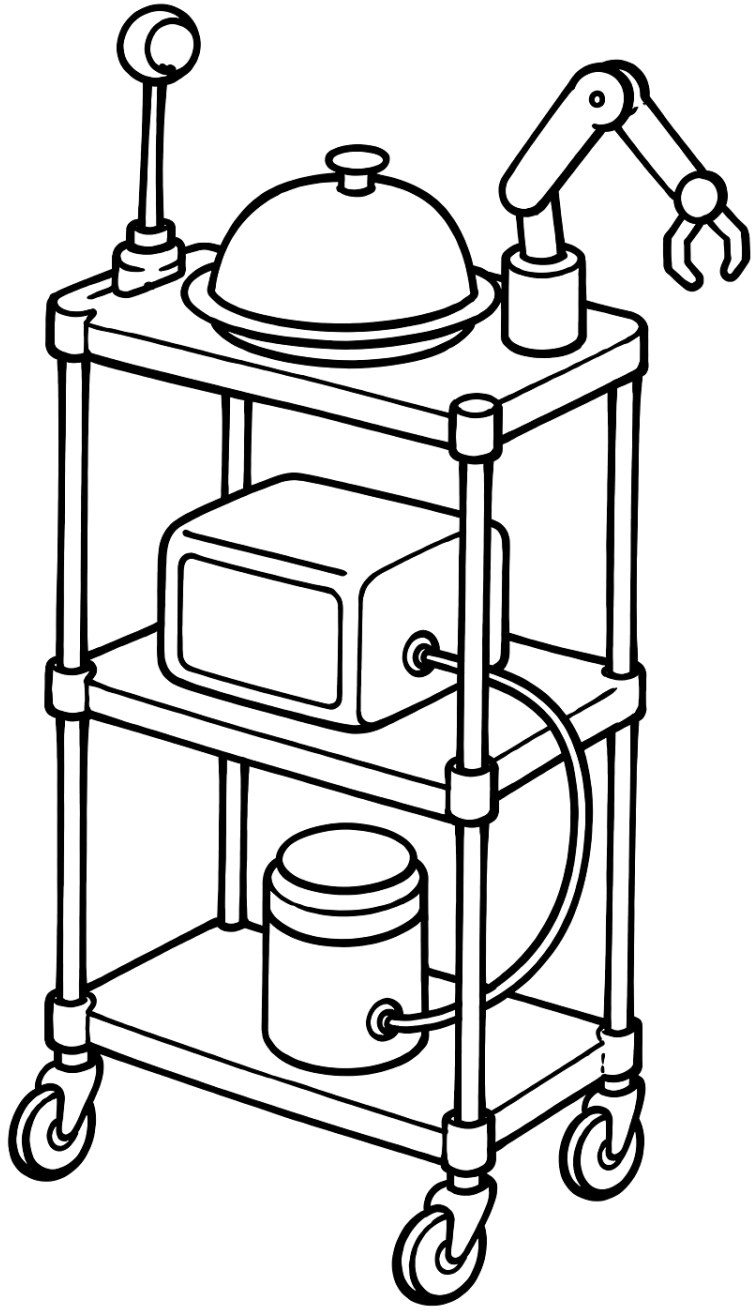
Embedding model

Internal knowledge base

Agent orchestration

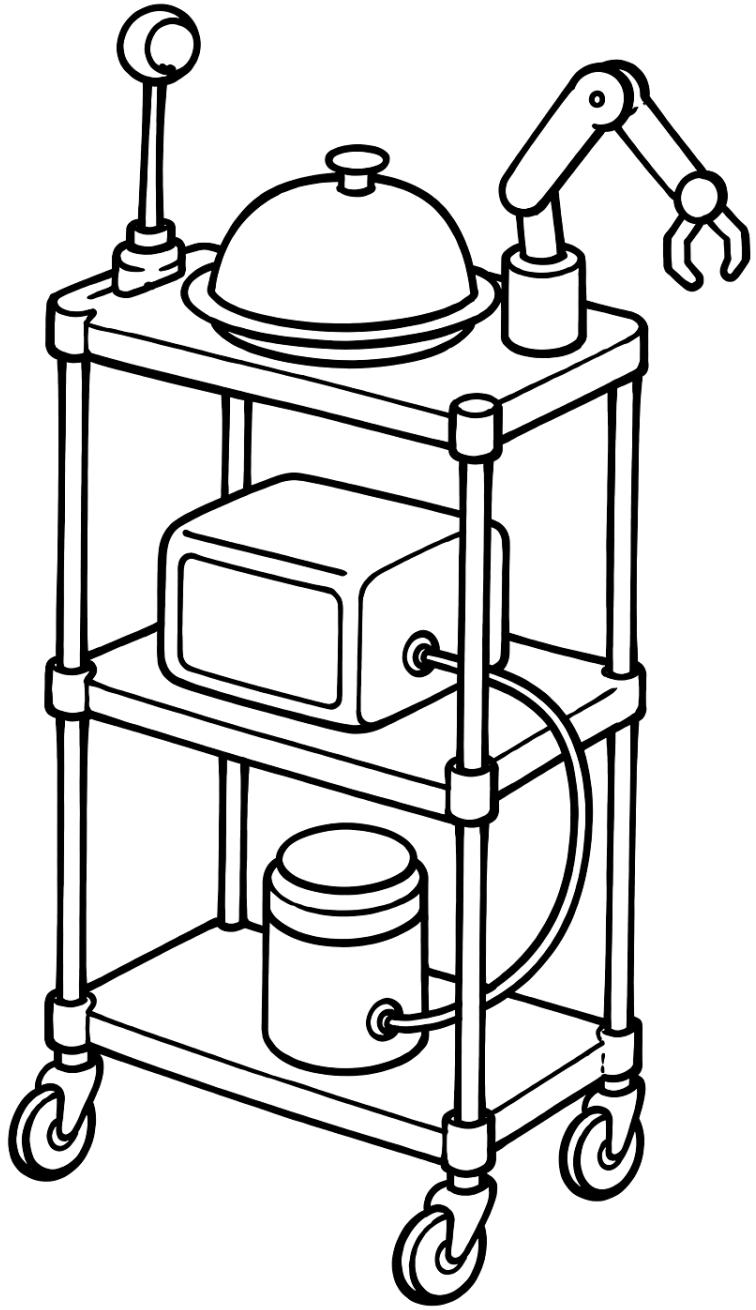
Agent tools

Software dependencies



LunchBot

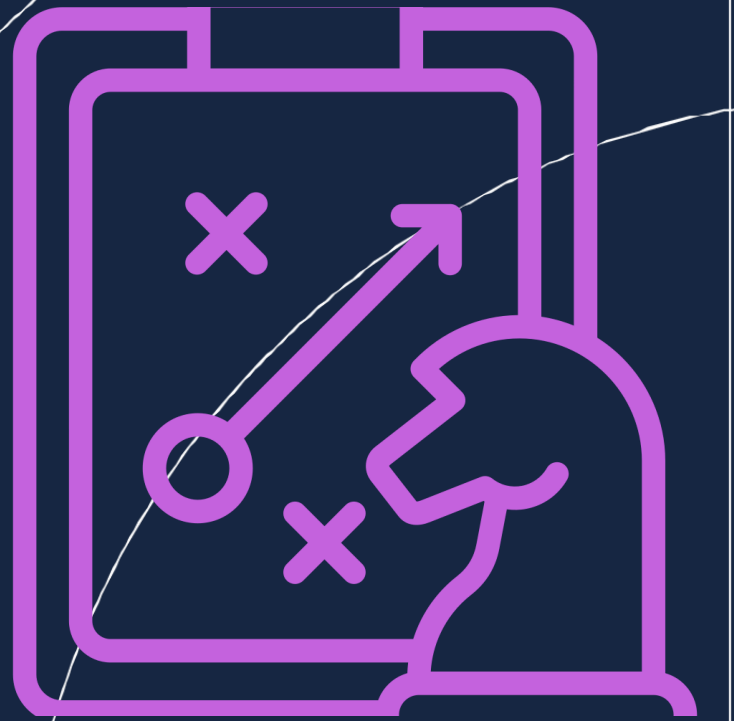
Hosted foundation model	MegaFrontierLab Nickname 2.1
Fine-tuned adapter	LoRA trained to interpret preferences
Embedding model	Embed-a-Lot 1.7 for encoding
Internal knowledge base	Expense Policies, dietary preference
Agent orchestration	Coordination for reasoning, retrieval, etc
Agent tools	Menu search and lunch ordering APIs
Software dependencies	Packages for auth, APIs, logging



RSAC

Why do we care about what's in an AI system?

- Strategic – understanding critical dependencies and concentrations
- Economic – improve purchasing decisions and supplier incentives
- Operational – find affected systems and respond quickly

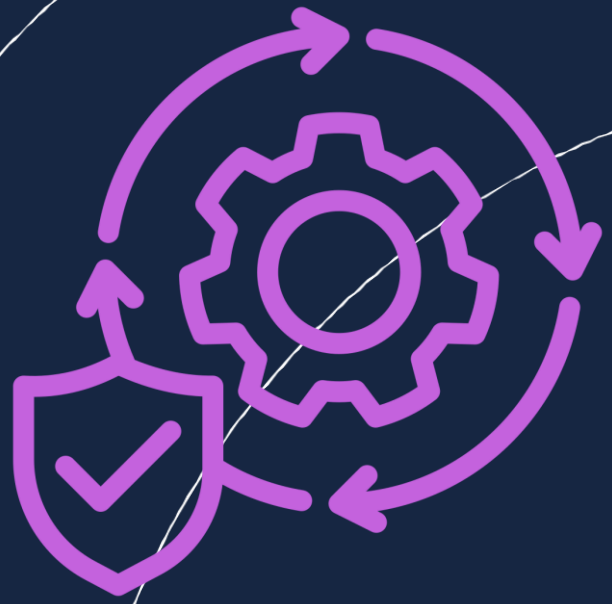


Why do we care about what's in an AI system?

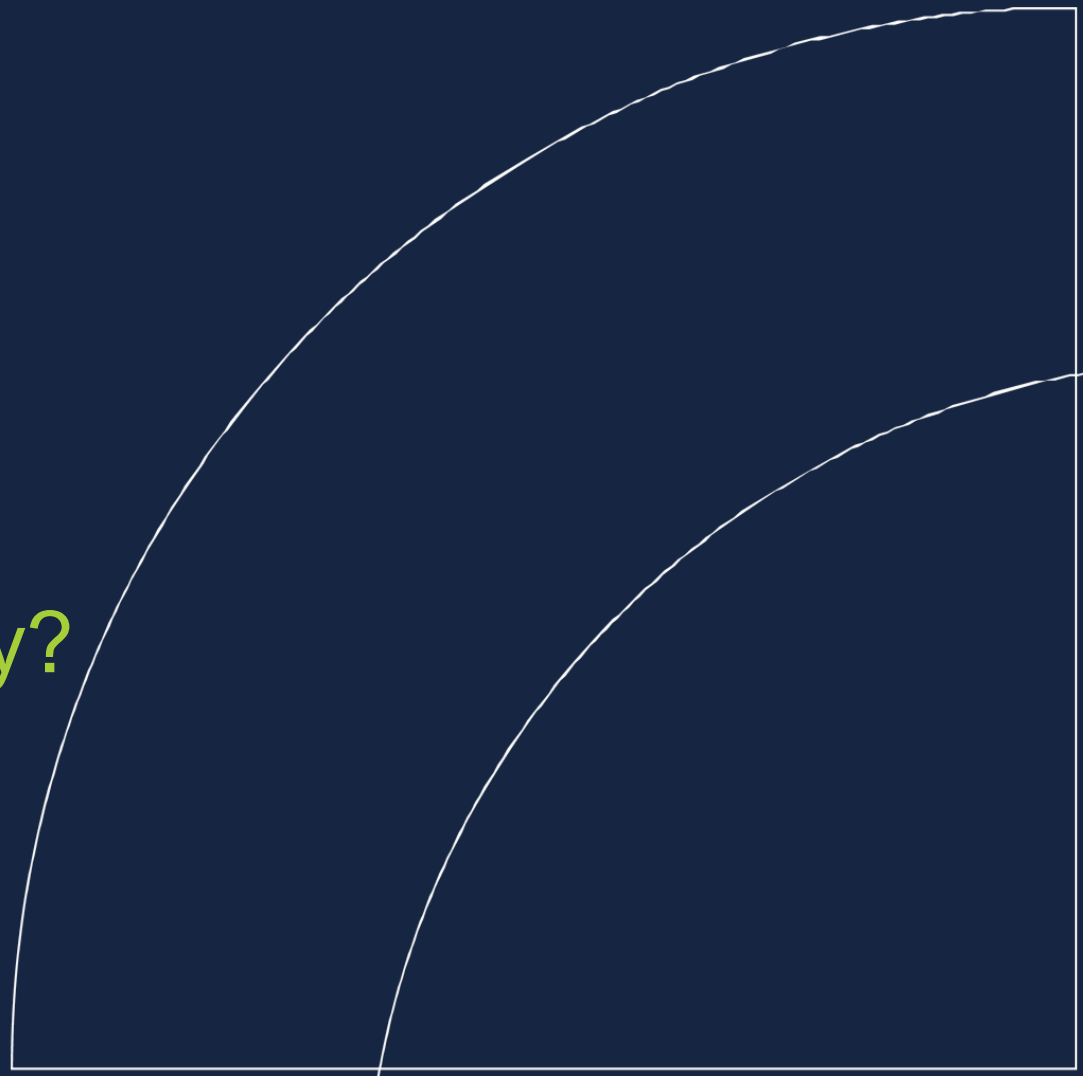
- Strategic – understanding critical dependencies and concentrations
- Economic – improve purchasing decisions and supplier incentives
- Operational – find affected systems and respond quickly

Why do we care about what's in an AI system?

- Strategic – understanding critical dependencies and concentrations
- Economic – improve purchasing decisions and supplier incentives
- Operational – find affected systems and respond quickly



So… what is an AIBOM, already?



“expanding on the SBOM to include the documentation of algorithms, data collection methods, frameworks and libraries, licensing information, and standard compliance”

- Linux Foundation (2024)

“A practical and standardized way to improve transparency, security, and trust across AI systems”

- CISA AI SBOM ‘Tiger Team’ (2024)

A structured record of details and supply chain relationships for the various components used in building an AI system.

- The G7 (2025)

“Empowering organizations to gain visibility into their AI model components.”

- OWASP AIBOM Initiative, Gen AI security project

“a structured machine readable inventory of AI components”

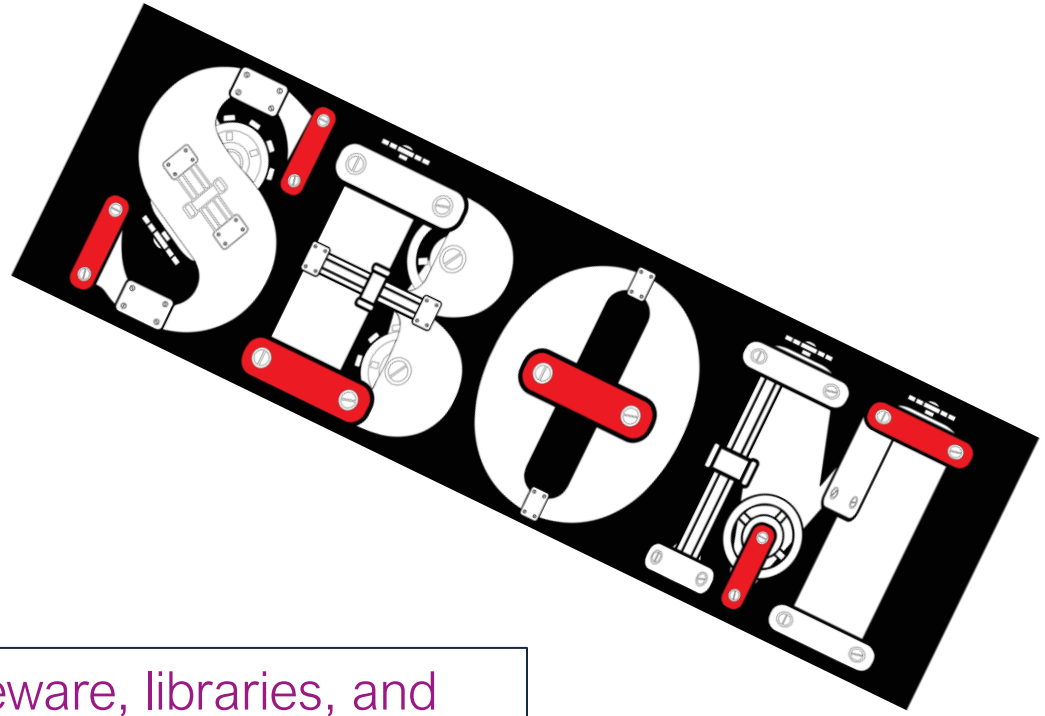
- OWASP AIBOM project

RSAC

An AIBOM must include or be paired with a full SBOM

"Since AI is software, AI models – and their dependencies, including data – should be captured in software bills of materials."

- CISA (2023)



Orchestration frameworks, agent infrastructure, middleware, libraries, and tool-connectivity layers do not stop mattering because a model is present.

RSAC

Simple Fields

Complex Fields

RSAC

Simple Fields

- Relatively static
- Easy to characterize consistently
- Usually available to supplier
- Automatable assembly and interpretation

Complex Fields

- More dynamic or contextual
- Open-ended definitions & implementation
- May be unavailable or undisclosed
- More difficult to automate
- Agents: permissions, routing, config

RSAC

Simple Fields

- Relatively static
- Easy to characterize consistently
- Usually available to supplier
- Automatable assembly and interpretation

Complex Fields

- More dynamic or contextual
- Open-ended definitions & implementation
- May be unavailable or undisclosed
- More difficult to automate
- Agents: permissions, routing, config

Identity is Hard

RSAC

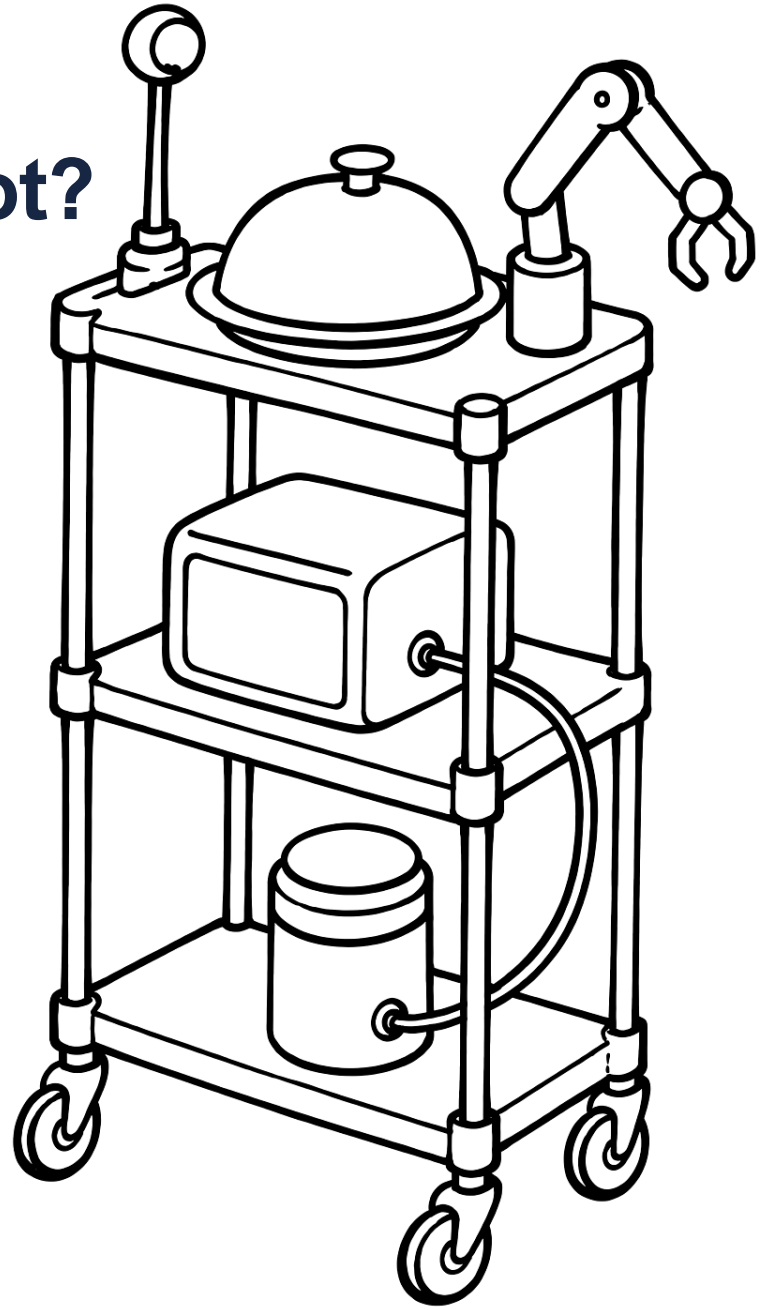
Questions that AIBOMs should be able to help with

- Where is a particular model, dataset, tool, or service used?
- Which systems send data to external AI providers?
- What components with newly identified security, licensing, privacy, or integrity concerns used?
- Which agents can invoke a particular tool or service?
- What changed before an incident?
- Who supplied each component, and what evidence supports what we know about it?

RSAC

How does this help us manage LunchBot?

- The allergy model has been found unreliable. Is LunchBot using it, or an adapter derived from it?
- The menu-search tool has been compromised. Which LunchBot deployments can still call it?
- LunchBot sends employee dietary information somewhere. Which external providers actually receive it?
- LunchBot suddenly starts making terrible choices. What changed?
- What do we actually know about each component?



RSAC

Inventories vs External Risk Data

Advisory

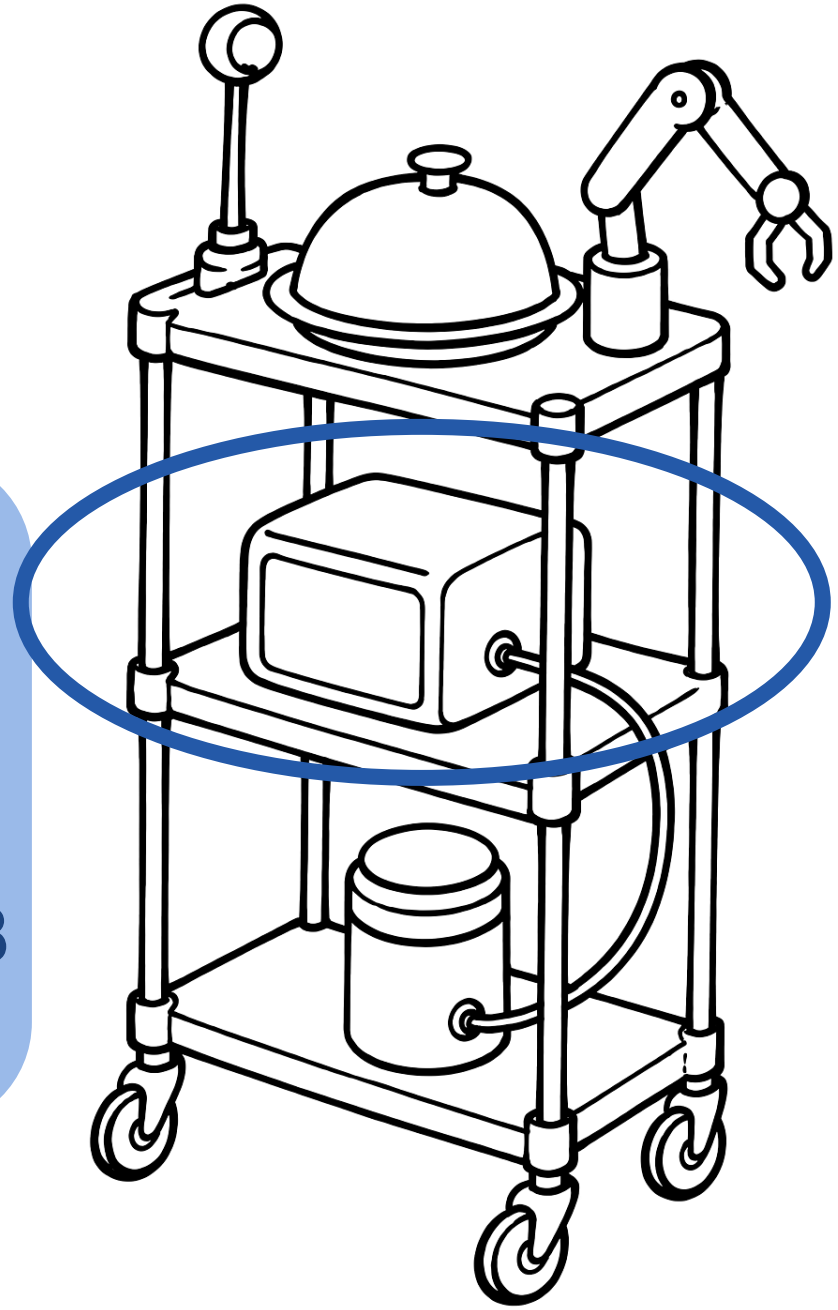
False negatives for tree-nut allergens

AIBOM

AllergyCheck 2.3

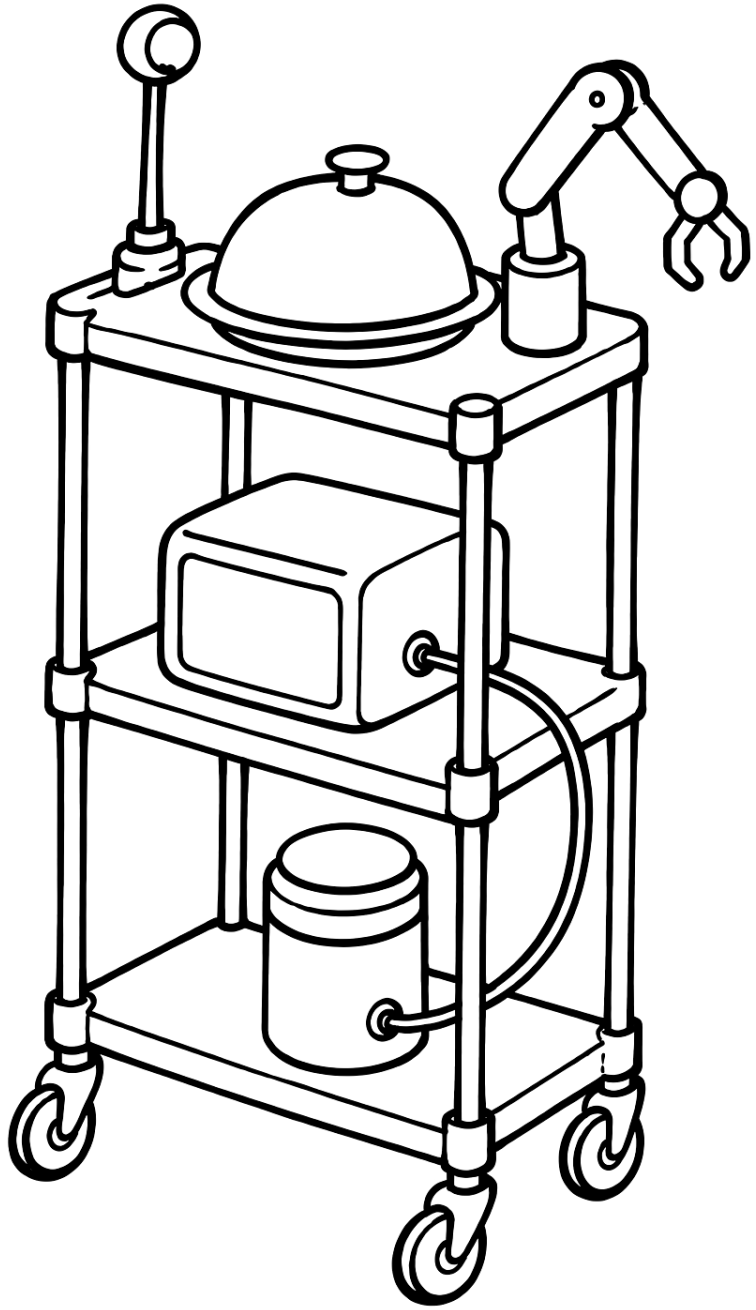
Derived from

FoodSafe Base 1.8



BOMs vs External Risk Data

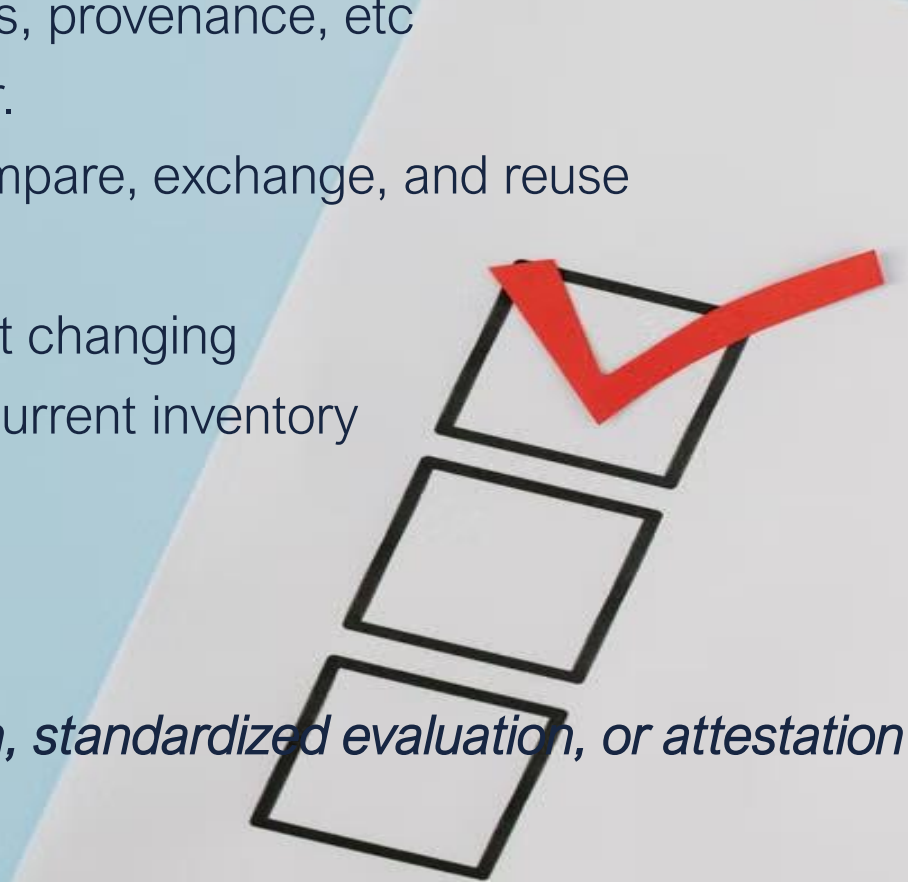
- Vulnerability disclosures and security advisories
- Model or dataset incident reports
- Threat intelligence and supplier-risk information
- Licensing, consent, and legal findings
- Evaluation and testing results
- Deprecation, recall, and end-of-support notices



Caution – don’t turn a BOM into an assessment

- BOMs work best for inventory facts: identity, relationships, provenance, etc
- Evaluated claims depend on method, scope, or evaluator.
- Embedding these judgments makes BOMs harder to compare, exchange, and reuse

- Judgments can change over time without the component changing
- This risks leaving stale conclusions inside an otherwise current inventory



However: A BOM may identify and reference a certification, standardized evaluation, or attestation

Image by Tara Winstead from Pexels: https://www.pexels.com/photo/red-check-mark-on-box-in-close-up-view-8850709/

What an AIBOM cannot do

- It does not determine whether an AI system is safe or effective.
- It does not verify every supplier claim.
- It cannot expose information a closed provider does not supply.
- It does not resolve model identity or version ambiguity automatically.
- It does not measure exploitability or business impact by itself.
- It does not replace testing, monitoring, risk assessment, or human accountability.

RSAC

Questions are familiar, but now a little harder

Familiar BOM Questions	AI makes things more complicated
What exactly is the component?	Model name or API endpoint may not identify stable weights or a stable provider snapshot.
Who supplied or created it?	The base-model developer, adapter author, hosting provider, and system integrator may all be different.
What was it derived from?	Lineage may include datasets, base models, fine-tunes, adapters, merges, quantization, and other transformations.
What evidence supports the claim?	Closed or hosted components may support only provider declarations or attestations.
Is the inventory complete and current?	Retrieval sources, routing, tools, and permissions can change independently.
What is unknown, unavailable, or withheld?	Scale and provider opacity can make provenance data particularly hard.

Different consumers need different information

- Developers need dependency and lineage detail.
- Security teams need exposure and incident-response context.
- Procurement teams need supplier and usage information.
- Auditors need evidence and provenance.
- Customers may need a deliberately limited disclosure.

RSAC

How do we get there?

Technical lessons from SBOM

- Naming and identity are harder than they appear.
- Risk context must be joined to inventory data.
- Completeness and accuracy require measurement.
- Operational workflows matter more than static artifacts.

Adoption Lessons from SBOM

Generating BOMs is easier than consuming them well.

More fields do not automatically create more value.

Minimum data and use-case-specific extensions must coexist.

Procurement can drive adoption, but checkbox compliance can hollow it out.

RSAC

Where are we today

Policy Goals and Emerging Expectations

G7 Guidance • Legislation • Procurement Goals

SHARED BASELINE

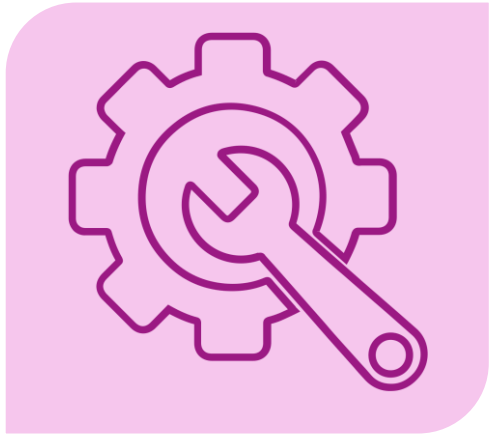
Not Yet Agreed

Technical Building Blocks

CycloneDX • SPDX • OWASP • Generation Tools

RSAC

Minimum is a floor, not a wish list



UTILITY

AVAILABILITY

CONSISTENCY

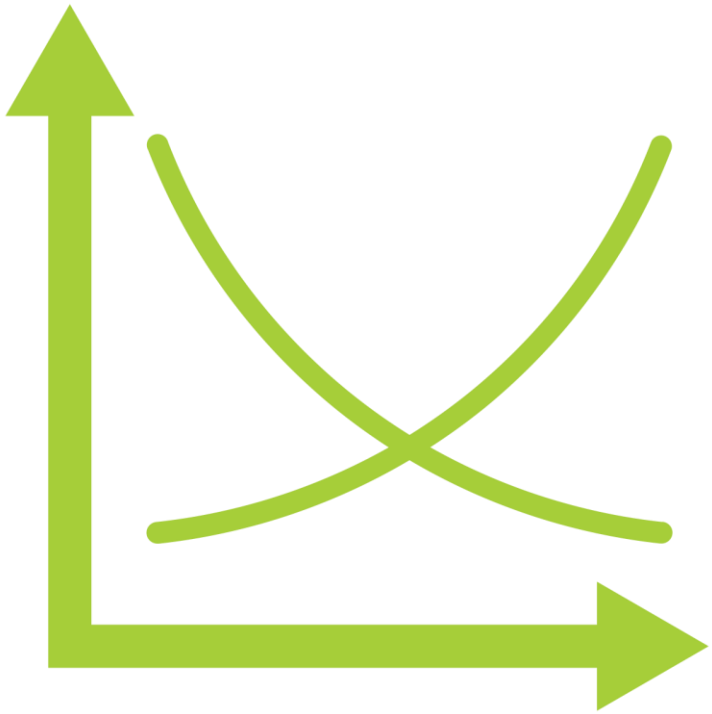
RSAC

Big Picture – What do we, collectively, need to do?

Policy Path Forward

Supply

Demand



Policy Path Forward

Supply

- Agree on a useful, achievable baseline.
- Implement it through existing formats.
- Build repeatable generation and consumption tools.
- Make AIBOM production less expensive and more consistent.

Demand

- Buyers require suppliers to know what is in their systems.
- Security teams ask operational questions of that information.
- Contracts, procurement, regulation, and norms reward better transparency.
- Experience using the data clarifies what the baseline should contain.

Data models *without* use becomes an academic standards exercise.
Policy or procurement *without* consensus creates fragmented, incompatible requirements.

AIBOM resources available today

G7 Cybersecurity Working Group. *Shared Vision* (2025) and *Minimum Elements* (2026)

cisa.gov/resources-tools/resources/software-bill-materials-ai-minimum-elements

Friedman and Leiserson. *Driving AI Transparency (2026)*
Supply- and demand-based paths toward an AIBOM ecosystem
securityandtechnology.org/virtual-library/policy-memo/driving-ai-transparency/

Linux Foundation. *Implementing AI Bill of Materials with SPDX 3.0 (2024)*
AIBOM concepts, use cases, and analysis
linuxfoundation.org/research/ai-bom

Guidance and framing

CycloneDX AI/ML-BOM
cyclonedx.org/capabilities/mlbom

SPDX 3.0 AI Profile
spdx.github.io/spdx-spec/v3.0.1/model/AI/AI

OWASP GenAI Security Project, AIBOM Generator
github.com/GenAI-Security-Project/aibom-generator

Standards, communities, and tools

What you and your organization can do **today**

How organizations can get started today

1. Choose one high-value AI system and the operational questions its inventory should help answer.
2. Map not just its foundation model.
3. Distinguish declared information, observed evidence, and explicit unknowns.
4. Assign responsibility for maintaining the inventory and define when it must be updated.
5. Test one operational workflow, such as responding to a model or dataset incident.
6. Ask suppliers for machine-readable information and experiment with CycloneDX or SPDX.

RSAC

Do you know what's in your AI?

Know the components.
Know the relationships.
Connect them to risk.

This is an evolving conversation. Join it.

allan@securityandtechnology.org

RSAC

Join our RSAC Membership Platform

RSAC