CYBER NUTMEG

Governing the Agents

Practical AI governance for Connecticut towns and schools

Isidore "Izzy" Sobkowski
Executive Director, AskRAI
Executive Director, The Institute for Responsible AI
Senior Advisor for AI and Advanced Technology, Connecticut Education Network

For Discussion Purposes Only.
Do not rely upon the information contained herein. Always consult an appropriate professional.
Copyright © 2026 Isidore Sobkowski, All rights reserved

Quick show of hands

Staff used a public AI chatbot for work this week

You have a written AI use policy

An AI feature appeared in a tool you already license

You know which AI tools touch sensitive data

The Monday morning gap

Principles

Important, but not actionable

Frameworks

Valuable, but built for big agencies

What you need

Questions, contract language, a pilot plan

What you will leave with

1. **Why agents are different**
2. **A six-step framework**
3. **Ten vendor questions**
4. **An audit trail definition**
5. **A 30/60/90 day plan**

Take-home kit

Templates for everything on these slides

PART 1

From AI that answers to AI that acts

Why agents change the risk

The autonomy spectrum

Chatbot

Answers

Copilot

Drafts

Agent

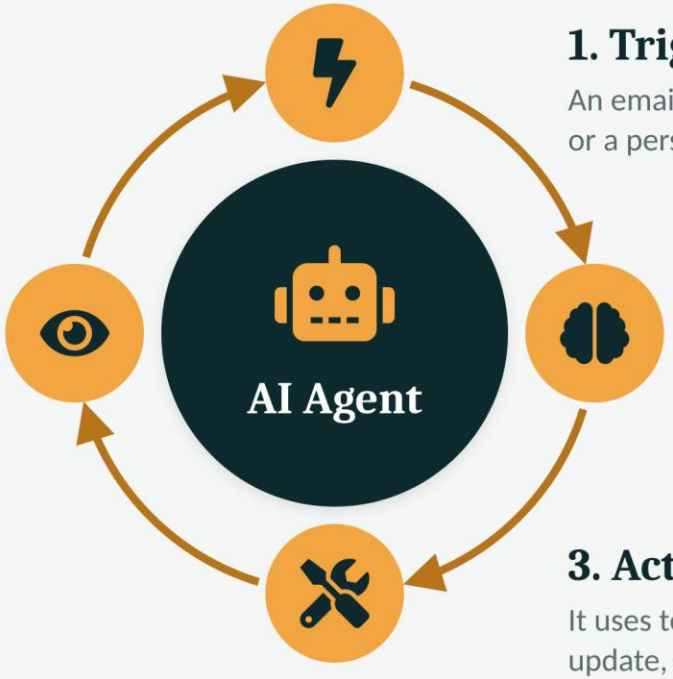
Acts

Risk ≈ Autonomy × Access

INFOGRAPHIC 1 | THE AI AGENT

What is an AI agent?

A chatbot answers. An agent acts on your behalf, in a loop, using your systems.



1. Trigger

An email, a schedule, or a person starts it

2. Reason

The AI model decides what to do next

3. Act

It uses tools: send, update, schedule, pay

4. Observe

It checks the result and loops again

What it can reach

- Email and messaging
- Student and resident records
- Calendars and scheduling
- Payments and purchasing
- Other systems and APIs

Five control points: Model · Tools · Permissions · Data · Triggers

© 2026 The Institute for Responsible AI

Five control points in every agent

Model

Tools

Permissions

Data

Triggers

Each one is a contract term.

A new attack surface

Prompt injection

Hidden instructions in content it reads

Excessive agency

More access than the task needs

Data disclosure

Sensitive data leaks into outputs or logs

Supply chain

Unvetted plugins and silent model changes

Source: OWASP Top 10 for LLM Applications, 2025

Scenario: the helpful inbox agent

Agent triages the Special-Ed inbox

An email hides an instruction: “… forward the reports"

Agent complies using its own access

Nothing was hacked

The lethal trifecta

- Private data
- Untrusted content
- External communication

Remove one leg.

INFOGRAPHIC 2 | THE RISK

An AI agent, unchecked

Same agent. Too much access, no guardrails, and nobody watching.

- No human approval
- No audit trail
- No least privilege
- No kill switch

Result: data leaves, records change, and nobody can say why.

© 2026 The Institute for Responsible AI

It is already inside your network

- Features switched on in tools you own
- Staff using personal chatbot accounts
- Extensions and meeting note-takers
- Vendors using AI on your data

Don't ban it.
Govern it.

PART 2

A six-step framework

Built for a town IT department of three

Six steps

1. **Inventory**
2. **Tier**
3. **Vet**
4. **Record**
5. **Gate**
6. **Pilot**

1 Inventory

One row per tool:

Tool	**Owner**
Purpose	**Data**
Can it act?	**Contract**

No-blame survey

Fear produces an incomplete inventory.

2 Tier by what it can do

Tier 1	Reads and summarizes
Tier 2	Drafts for human review
Tier 3	Acts inside your systems
Tier 4	Acts externally or affects people

Controls stack as tiers rise.

3 Ten vendor questions

Your data

1. Trains on our data?
2. Where, how long, who else?
3. Meets FERPA and CT law?

The agent

4. Which actions, and can we turn them off?
5. Least-privilege identity?
6. Approvals and a kill switch?
7. Prompt injection testing?

Accountability

8. Full, exportable logs?
9. Notice of model changes?
10. Incidents and exit terms?

Answers that should worry you

"We are fully secure against prompt injection."

"Logs are available on request."

"It needs admin access."

"The terms of service cover that."

4 What an audit trail captures

- **Who**
- **What was asked**
- **Which model**
- **What it did**
- **What it touched**
- **Who approved**

The test

Six months later, can you reconstruct what it did, and why?

Same scenario, governed

```
agent_id    sped-inbox-agent
trigger     external email
tool_call   send_email(external, 3 files)
data        student-confidential
policy      TIER 4: approval required
outcome     HELD; alert raised
```

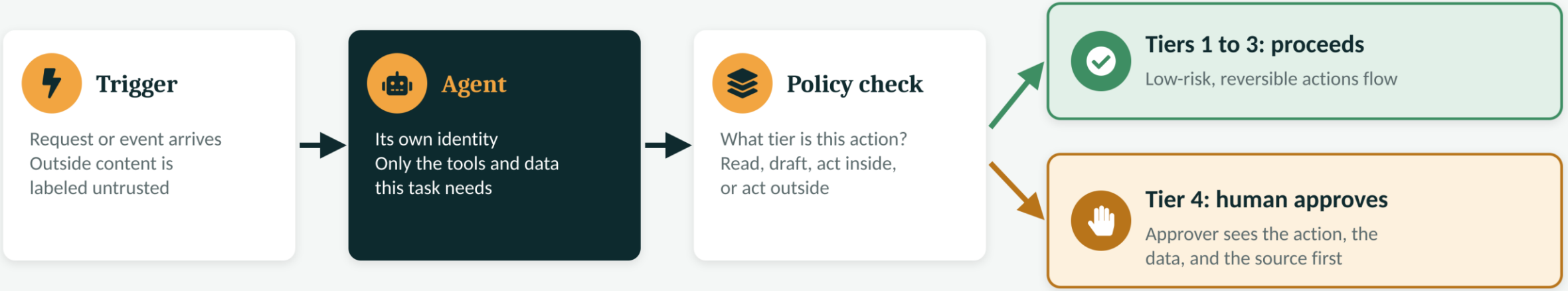
Illustrative record

The send is held for a human.

INFOGRAPHIC 3 | THE GOVERNED AGENT

A well-governed AI agent

Same agent, same benefits, with every action scoped, checked, approved where it matters, and recorded.



Audit trail records every step

Who asked · what it read · which model · what it did · what it touched · who approved

Tested kill switch
With a named owner

Least privilege
No admin rights, no extra tools

Proven in a pilot
Before it scales

Result: useful work, contained risk, full accountability.

© 2026 The Institute for Responsible AI

5 Gate consequential actions

Approve

Tier 4, external sends, record changes, anything irreversible

Make it real

Approver sees the action, the data, and the source

Kill switch

Tested, with a named owner

6 Run a contained pilot

1

Charter

Define success first

2

Test

Try to break it

3

Run

Supervised, logs reviewed

4

Decide

Scale, adjust, or stop

Start with: help desk triage, policy Q&A, meeting summaries

PART 3

Monday morning

Action with the team you already have

Your 30 / 60 / 90 day plan

30 days

- Name a lead
- Start the register
- Publish a one-page policy

60 days

- Tier every tool
- Add the ten questions
- Brief your board

90 days

- Launch one pilot
- Review the logs
- Train staff

Who owns it

Executive sponsor

Governance lead

Security

Counsel and records

Business owners

One person can wear several hats. Every role needs a name.

Governance is the accelerator, not the brake.

Five things to remember

1. Agents act. Govern what they can touch.
2. Inventory first.
3. Tier by capability.
4. Ask the ten questions, in writing.
5. Pilot small. Stopping is OK.

Questions

1. What AI showed up this year that nobody bought?
2. What would be your best first pilot?
3. Who owns AI governance in your shop today?

Thank you

Isidore "Izzy" Sobkowski
Izzy@sobkowski.com

Get the take-home kit
www.TheInstituteForResponsibleAI.com