

CYBER NUTMEG

AI as Weapon and Shield

The new threat landscape for municipal and education networks

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



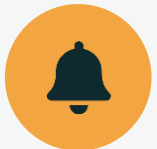
Could you spot a cloned voice of your superintendent or first selectman?



Does a phone call alone ever change vendor banking details?



Is a video call proof that you are talking to who you think?



Has your awareness training changed since AI arrived?

What AI changes for attackers



Cheaper

Tailored attacks at near-zero cost



Faster

Research and execution at machine speed



More convincing
No typos, familiar voices, familiar faces

Today

1 AI as weapon: five real patterns

2 AI as shield: what actually helps

3 The checklist: mostly free



Take-home checklist

Bring it to your IT and leadership teams

PART 1

AI as weapon

Five patterns, all documented

Phishing just got expert-level

12%

clicked generic phishing

54%

clicked fully AI-written spear
phishing

54%

clicked emails by human
experts

AI now matches skilled human attackers, at a fraction of the cost.

Source: Heiding, Schneier, et al., human-subjects study (N=101), 2024 preprint; published 2026

Pattern 1: The deepfake meeting



\$25.6M

Lost in one day through 15 transfers

- A finance employee joins a video call with the "CFO" and colleagues
- Every other person on the call is a deepfake
- Built from publicly available video of real staff

Lesson: A video call is no longer proof of identity.

Source: Hong Kong Police; Arup confirmation to CNN, May 2024

Pattern 2: Cloned voices of officials



Since 2023

Officials impersonated with AI voice and text, per FBI

- AI voice messages posing as senior federal and state officials
- Targets: officials and their contacts
- Goal: move to Signal, Telegram, or WhatsApp, then ask for codes or money

Lesson: A request to switch channels is a red flag.

Source: FBI IC3 public service announcements, May and December 2025

Pattern 3: Deepfakes against school leaders



4 months

Jail sentence for the staff member
who made it

- Fake audio of a Maryland high school principal making racist remarks
- Spread across the school community and social media
- Created by the school's own athletic director

Lesson: Plan your response to a deepfake before you need one.

Source: AP reporting; Baltimore County Circuit Court, April 2025

Pattern 4: Fake people, real jobs



4

Video interviews passed by a fake remote hire

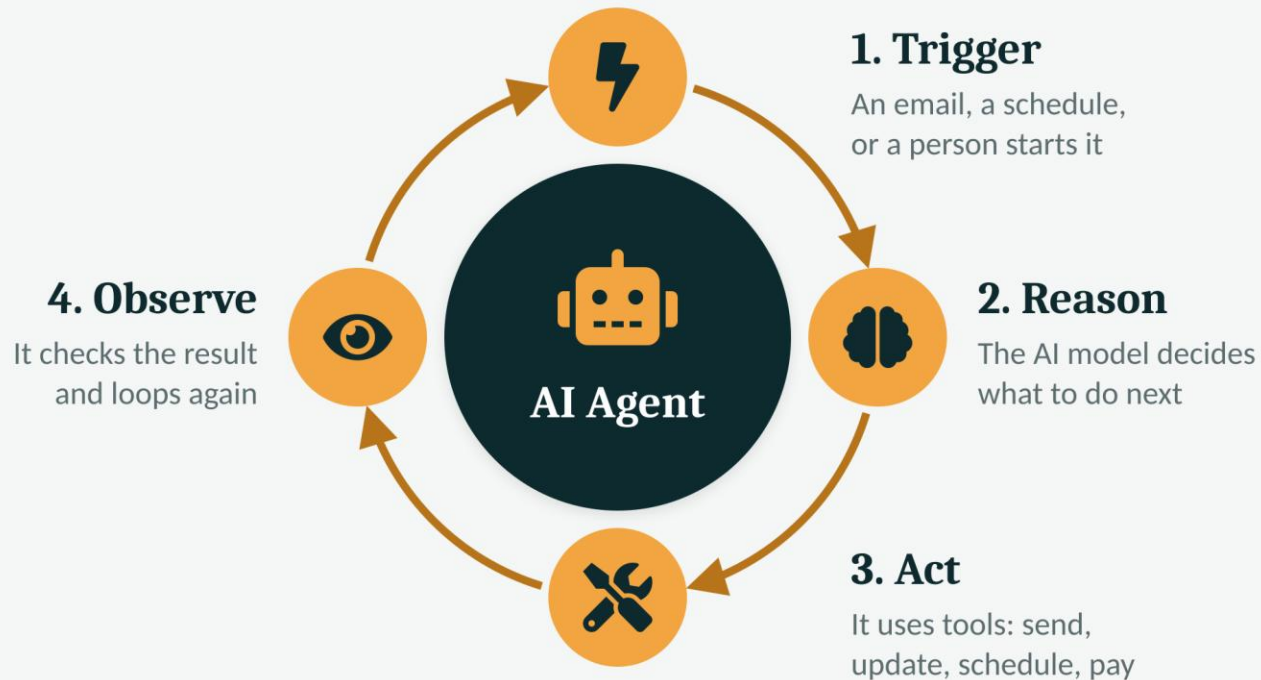
- North Korean operative hired by a U.S. security company
- Stolen U.S. identity plus an AI-enhanced stock photo
- Tried to load malware on day one; caught by endpoint monitoring

Lesson: Verify identity in hiring, especially for remote IT roles.

Source: KnowBe4 disclosure, July 2024

What is an AI agent?

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



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

Pattern 5: AI inside the attack itself



80 to 90%

Of one espionage campaign run by AI, per Anthropic

- Google: experimental malware that asks an AI to rewrite itself
- Anthropic: state actor used an AI agent against about 30 targets
- Real, but early; the AI also made mistakes

Lesson: The basics (patching, MFA, monitoring) still stop most of it.

Source: Google Threat Intelligence Group, Nov 2025; Anthropic, Nov 2025

An AI agent, unchecked

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



Where the losses actually happen

\$3.05B

Business email compromise losses reported
to the FBI in 2025

Attackers target your
processes, not just your
technology.

Source: FBI IC3 2025 Internet Crime Report

Your training is aimed at the wrong cues

Old cues

Typos, bad grammar, strange voice, unfamiliar face

New cues

Urgency, secrecy, a channel switch, a payment change

Verify the request, not the messenger.

PART 2

AI as shield

What actually helps a small team

Where AI genuinely helps defenders



Email analysis

Flags phishing and impersonation



Alert triage

Sorts and summarizes the noise



Identity anomalies

Unusual logins and access



Plain-language search

Ask your logs questions

Hype versus reality

THE PITCH

"Autonomous SOC"

"Detects every deepfake"

"Set and forget"

"Replaces staff"

THE REALITY

Assists analysts; people still decide

Detection is unreliable; use process

Needs tuning and a named owner

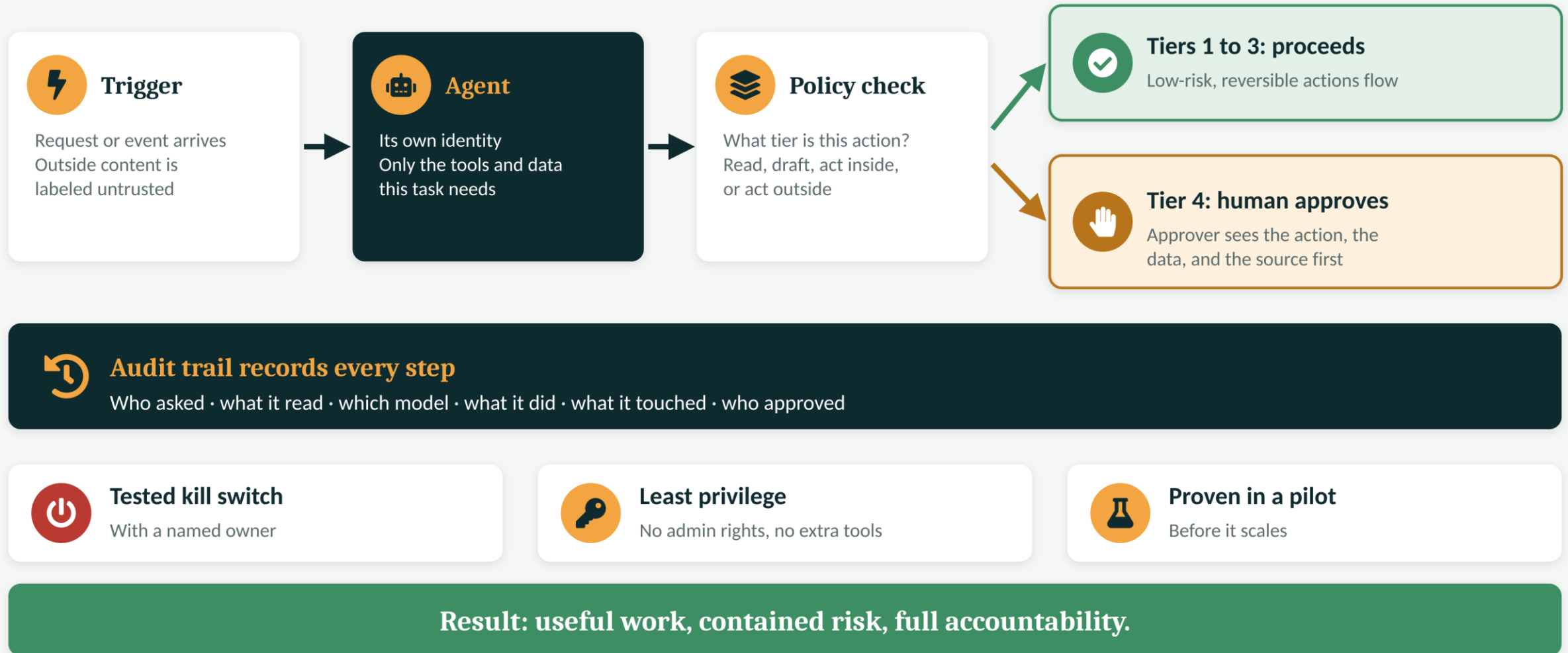
Amplifies the staff you have

Five questions before you buy

- 1 What will it catch that we miss today?
- 2 What is its false-positive rate in places like ours?
- 3 What data leaves our network?
- 4 Who acts on its alerts at 2 a.m.?
- 5 Can we trial it on our own data first?

A well-governed AI agent

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



Start with what you already own



Turn on the security features in your email and identity platforms



Use free government services for public entities



Share threat information with peers



Review settings quarterly; defaults change

PART 3

The checklist

Mostly process, mostly free

Stop the money



Call back on a known number before any banking change



Two people approve every wire and new payee



Executives use a verbal code for urgent requests



Never act on a request that switches channels

Protect identity and reputation



Phishing-resistant MFA for admins and finance



Email authentication (DMARC) at enforcement



Live identity checks when hiring remote staff



A deepfake response plan for your leaders

If money moves, minutes matter

\$679M

Frozen by the FBI IC3 Recovery Asset Team
in 2025

1

Call your bank's fraud line

2

File at ic3.gov

3

Call your FBI field office

Source: FBI IC3 2025 Internet Crime Report

Five things to remember

- 1 AI made old attacks cheap, fast, and convincing.
- 2 Seeing and hearing is no longer believing.
- 3 Verify the request through a channel you trust.
- 4 Use AI defense for volume; keep people for judgment.
- 5 Most of the fix is process, and it is free.



Questions

1

Which leader in your organization is easiest to clone?

2

Would your finance office call back before paying?

3

Who would declare a deepfake fake, and how fast?

Thank you

Isidore "Izzy" Sobkowski
Izzy@Sobkowski.com

Get the take-home checklist
www.TheInstituteForResponsibleAI.com

