

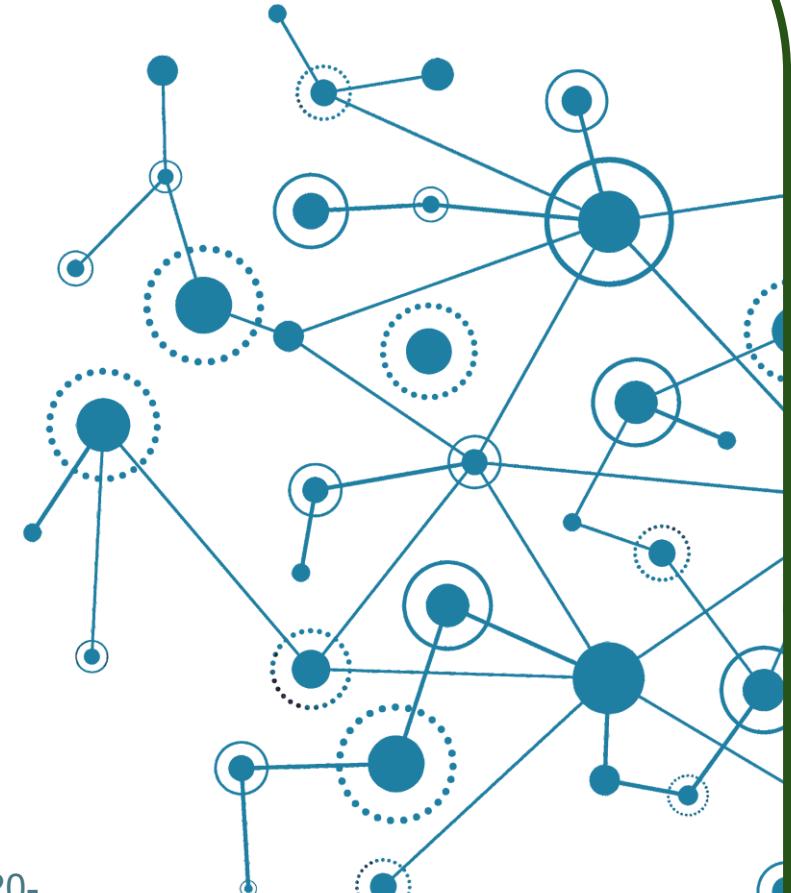
Beyond the Firewall: Evolving Cyber Skills for the Age of AI and Automation

Ron Woerner, CISSP, CISM

vCISO, Consultant, and Educator

<https://www.linkedin.com/in/ronwoerner/>

Slides are available at: <https://github.com/hackerron/AI-Cybersecurity/blob/main/RWoerner%20-%20Evolving%20Cyber%20Skills%20-%20VetSecCon%202025.pdf>



creative
commons



Every October, **Cybersecurity Awareness Month** teaches people simple ways to protect themselves online. This year's theme, **Stay Safe Online**, is all about the simple ways to protect yourself, your family and your business from online threats.

Small actions can make a big difference. That's why we're focusing on the **Core 4**. Four easy steps anyone can take to boost their online safety:

- Use strong passwords and a password manager
- Turn on multifactor authentication
- Recognize and report scams
- Update your software

What are you doing for it?



SHARE!

>whoami

I'd tell you who I am,
but do you really
care?

Ron Woerner, CISSP, CISM
USAF, Intel Officer

LinkedIn:

<https://www.linkedin.com/in/ronwoerner/>



Ron Woerner – Evolving Cyber Skills



Websites & Social Media:
<https://linktr.ee/cyberron>



A black silhouette of Alice from Disney's Alice in Wonderland, standing in her white apron and blue skirt, looking slightly to the right. She is positioned on the left side of the slide, partially cut off by the frame.

Warning!

- These are my thoughts based on my studies and experiences and NOT necessarily those of my employers or anyone else
- I AM NOT AN AI EXPERT!
- Ethical uses only / NOT a presentation on breaking AI
- Use at your own risk / Normal caveats apply
- There is homework!



This is an Interactive Workshop!

Intermediate technical level

Follow along on your computer



Ask questions if you're unsure or need help.



Please limit other work distractions.



Stay focused on the session's outcomes.



Have fun!

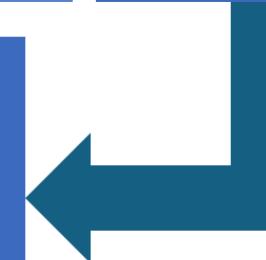


Stay open minded and tell us what you REALLY think

Psychological Safety

This means...

- Postpone / withhold your judgement of IDEAS
- Encourage WILD exaggerated ideas
- Quantity counts at this stage NOT quality
- Build on IDEAS put forward by others
- EVERY person and every idea has worth.





WHY
are we here?



Learning objectives:

1. **Identify the evolving skill sets** required in modern cyber roles, including AI literacy and strategic thinking
2. **Analyzing** how emerging technologies and threats are **reshaping professional competencies**
3. **Develop a personal upskilling strategy**
4. **Integrate** interdisciplinary perspectives into **technology and cyber practices**

Exploring the strategic and practical dimensions of AI in cybersecurity.

Agenda

Exploration into AI

Enhancing Your AI Skillz

AI Risks & Threats

Experiment with Adversarial Techniques

Establishing AI Guardrails

Hacking Cyber Careers

Poll Questions

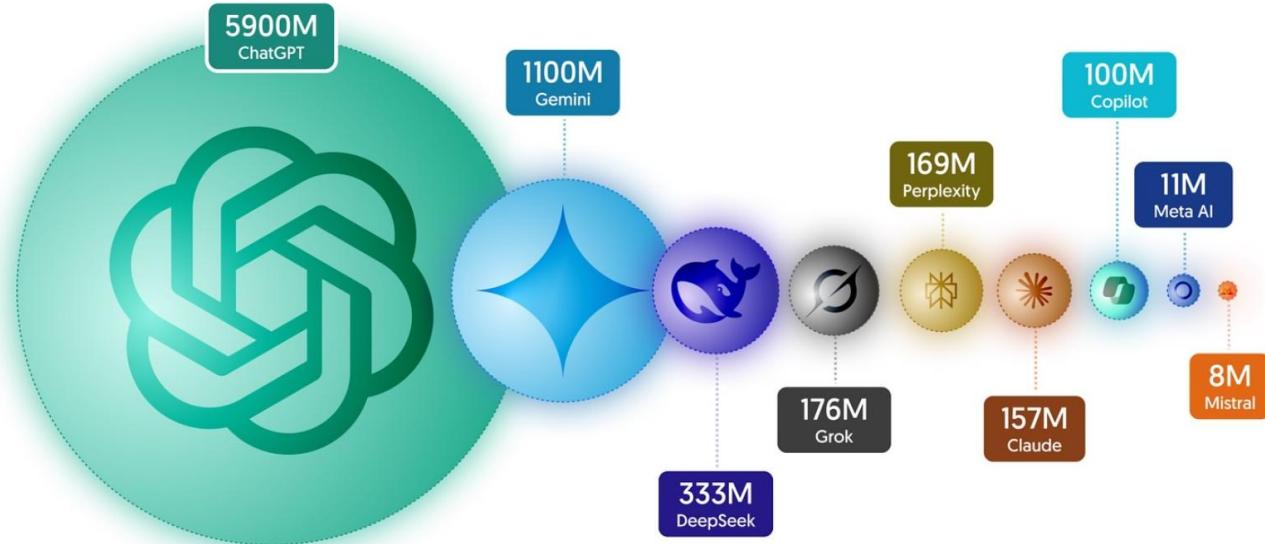
- What skill has become most critical in your role?
- How / Where / When / Why do you use AI?
- What AI Apps / Platforms / Tools do you regularly use?

AI is Now Ubiquitous



Marketing & AI Insights

Most Popular AI Chatbots By Visits



Source: SimilarWeb — Oct/2025.

Windows Blog

Making every Windows 11 PC an AI PC

New Copilot and agentic experiences make powerful AI easy on Windows 11. Today, we're taking an exciting step forward with a new wave of...

2 days ago



CNBC

Microsoft will test a Copilot AI feature that performs work on local files in Windows 11

Microsoft will bring its Copilot Actions feature that can perform certain tasks to its Windows 11 operating system, enabling work with...

2 days ago



Tom's Hardware

Microsoft wants you to talk to Windows 11 PCs again — Copilot gets 'conversational' input to complement your mouse and keyboard

Microsoft is going deep on AI with Windows 11, making voice a third input aside the keyboard and mouse with new wake words and additional...

2 days ago



WIRED

As Windows 10 Support Ends, Microsoft Is 'Rewriting' Windows 11 Around AI

Microsoft saved its most powerful AI tools for paying customers in the first phase of its AI evolution. Now, the company has announced a...

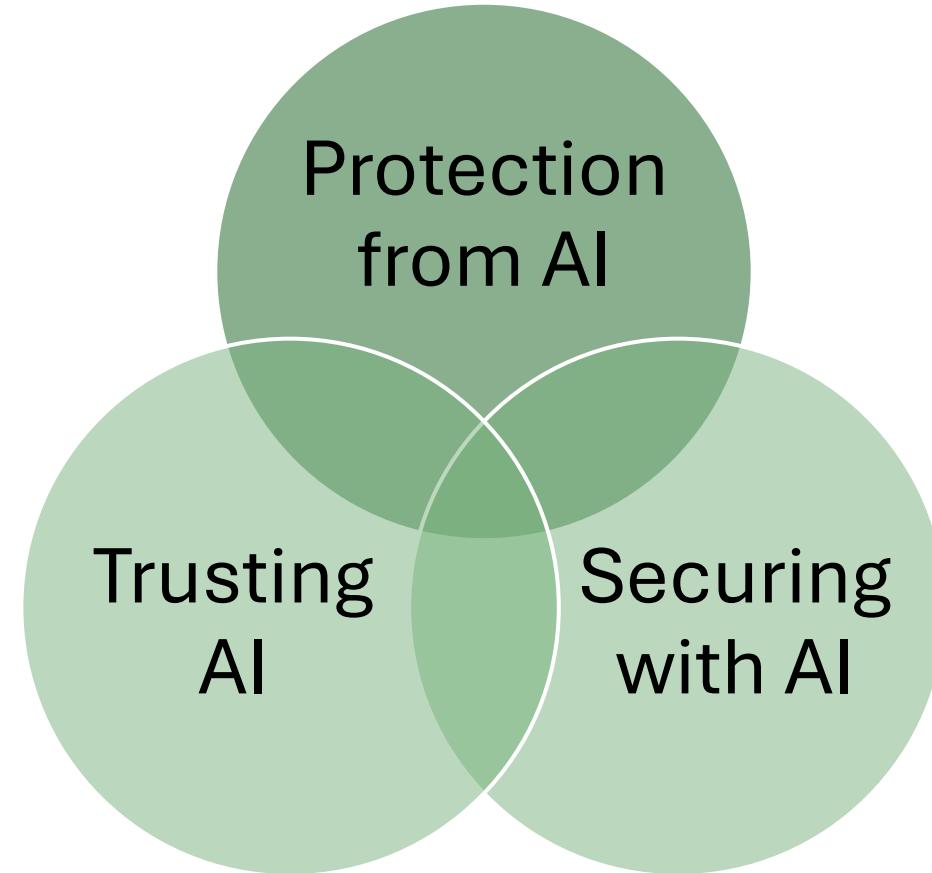
2 days ago



<https://www.similarweb.com/top-websites/ai-chatbots-and-tools/>

Setting the Stage

Why?



All Security is Local

Setting the Stage

What are your top concerns about emerging technologies?

Evolving AI-powered threats

Additional complexity

Expanding attack surfaces

Reduced barrier to entry

Automated, dynamic decision-making

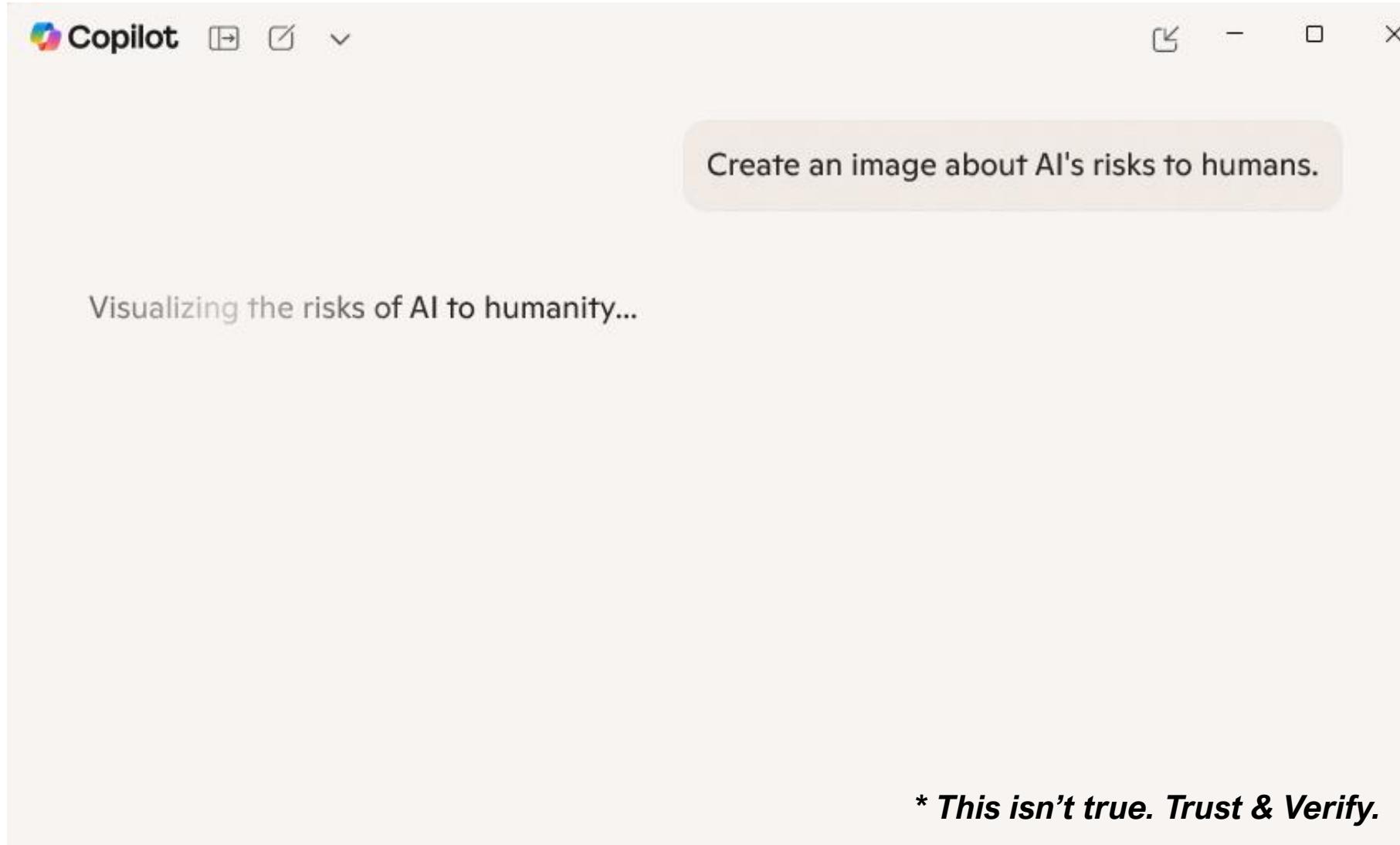
Insufficient and outdated skill sets

Top Questions about AI in Cybersecurity

- **GenAI in Security Tools:** Common use cases include incident summarization, threat research chatbots, and behavior modeling for triage and investigation.
- **AI Agents:** Task-specific agents (e.g., phishing triage) automate discrete security functions with high accuracy through focused training and prompts.
- **Agentic Systems:** Emerging concept where multiple AI agents collaborate to handle complex workflows (e.g., full incident response), but still in early stages.
- **Chatbot Use:** Helpful for documentation and threat research, but underutilized by practitioners due to workflow disruption.
- **Caution on Hype:** Many capabilities are not yet broadly available or reliable; rigorous evaluation is needed before adoption.

Source: <https://www.forrester.com/blogs/your-top-questions-on-generative-ai-ai-agents-and-agentic-systems-for-security-tools-answered/>

AI Human Risks



AI Human Risks



AI: THE PANDORA'S BOX

 FINANCIAL TIMES

Accenture to 'exit' staff who cannot be retrained for age of AI

Group details \$865mn restructuring programme and outlook reflecting sluggish corporate demand for consulting projects



The job cuts allowed Accenture to say it would cut costs by 10% to 12% a year, or at least 10 basis points in the next fiscal year.

FORTUNE

Home News Fortune 500 Tech Finance Leadership

AI-DELOITTE

Deloitte was caught using AI in \$290,000 report to help the Australian government crack down on welfare after a researcher flagged hallucinations

BY NINO PAOLI
NEWS FELLOW

October 7, 2025 at 5:10 PM EDT

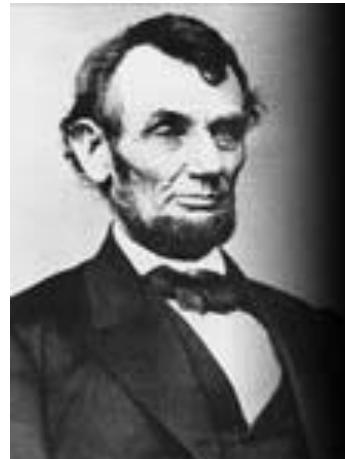


Core Principle



Trust, AND Verify

“Never trust a single data source.”*
from my life as a military intelligence officer



“Just because you see something on the Internet with a quote, a picture and a date, it doesn’t mean it’s going to be true.”
- Abraham Lincoln, 2006

Quick animal fact #38



Less than 5% of ducks have attempted world domination

AI Ethical Use

NIST AI 100-1

AI RMF 1.0



Fig. 1. Examples of potential harms related to AI systems. Trustworthy AI systems and their responsible use can mitigate negative risks and contribute to benefits for people, organizations, and ecosystems.

NIST AI Risk Management Framework 1.0,

<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>, p.5

Ethical Use – Reflection



NIST AI Risk Management Framework

AI Risks & Trustworthiness

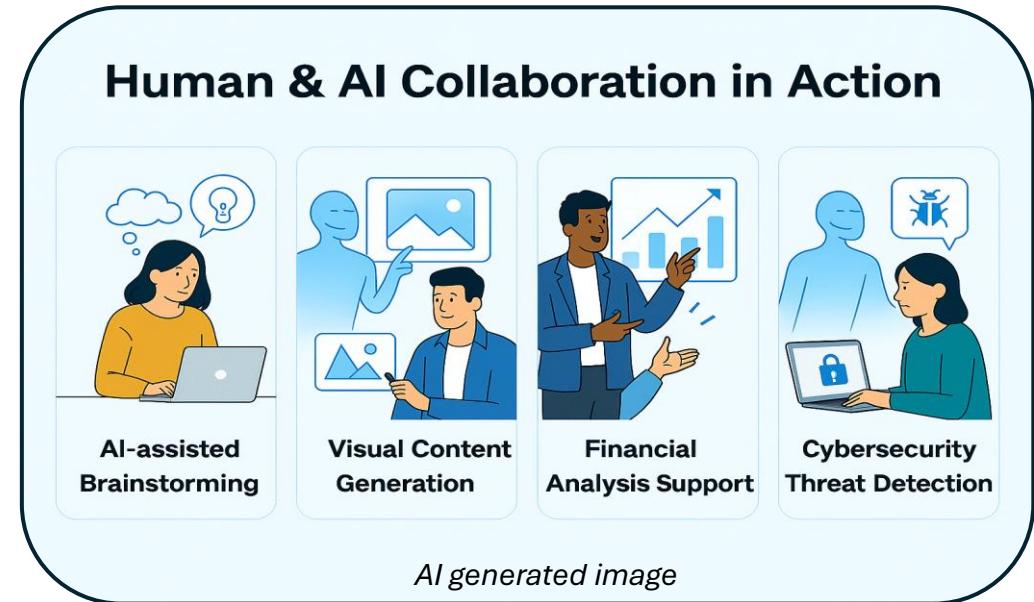
1. Valid and Reliable
2. Safe
3. Secure and Resilient
4. Accountable and Transparent
5. Explainable and Interpretable
6. Privacy-Enhanced
7. Fair – with Harmful Bias Managed

- How do you know what's “true” and verify results?
- Explore ethical dilemmas
 - What bias may be introduced?
 - What does it mean to “trust” a machine?
 - Can AI be “self-aware” of its misuse?
- Balancing automation with human involvement
- Embracing continuous learning and interdisciplinary fluency

AI Replacing Humans

Prompt: *Will AI replace humans in the workplace?*

Response (Perplexity): *AI will not replace humans entirely in the workplace*, but it is transforming many roles and automating routine tasks, leading to the reduction of some jobs while creating new opportunities in others. Recent studies show that while AI is accelerating the automation of certain tasks, it has not led to mass unemployment or total displacement of the workforce so far. Many **roles are simply evolving**, with humans increasingly focusing on oversight, creativity, and judgment rather than repetitive work.



AI Replacing Humans

FORRESTER®

REPORT

Ground Your Workforce AI Strategy In Human Experience

Put Employees At The Center Of Your AI Efforts To Drive Business Success

Mar 6, 2025 • 11 min read



J. P. Gownder
Vice President, Principal Analyst

With contributors:
James McQuivey, PhD, Shynise McElveen and Bill Nagel

Employees Harbor Misgivings About Workforce AI



Too many leaders see their workforce AI deployment as primarily a technology and data exercise. That perception couldn't be further from the truth: **Today's AI remains intimately tied to human users**, whose experience with the technology will be a principal determinant of its success or failure. **Framing AI as a tool that builds new opportunities** for employees in their jobs, lives, and future careers can help. And **designing a human-centric culture of AI** to reinforce positive practices, beliefs, and behaviors will drive both employee experience and business success.

1. Most AI augments humans rather than replacing them.
2. Employee buy-in is crucial to successful adoption.
3. Misuse (and abuse) can lead to quality breakdowns and ethical lapses.

What You Can (Should) Do

Preparation & Resilience

Embrace Continuous Learning

Learn / Review Fundamentals

Enhancing AI Skills

Risks, Threats, &
Adversarial Techniques

Establishing Resilience

Starting
here



Embrace Continuous Learning

A-B-C =
Always Be Curious



Key Skills for Professionals

Regardless of the specific role, certain skills are essential for experienced professionals:

- **Leadership and mentorship:** Ability to guide and develop junior team members.
- **Strategic and business acumen:** The capacity to align AI initiatives with an organization's business objectives.
- **Communication skills:** The ability to explain complex technical concepts to non-technical stakeholders and leadership.
- **Ethical awareness:** Understanding of potential biases and responsible AI practices.
- **Technical depth:** Expertise in specific technical areas like enterprise architecture, Zero Trust, and networking, and AI domains like NLP, computer vision, or reinforcement learning.
- **Programming proficiency:** Mastery of languages such as Python and frameworks like TensorFlow or PyTorch.

Core Competencies for AI & Cybersecurity Professionals

Technical Skills to Cultivate:

- **AI Fundamentals:** Machine learning, deep learning, NLP, automation, and model evaluation
- **AI-Specific Threats:** Deepfakes, automated phishing, model inversion, data poisoning
- **Secure Development Practices:** Threat modeling, secure coding, and adversarial resilience
- **Data Handling & Privacy:** Governance, categorization, encryption, anonymization, and compliance (GDPR, CCPA)
- **Security Architecture:** Identity & access management, container security, and Zero Trust architecture

Mastering Fundamentals

AI:

- SDLC / Coding Best Practices
- Core Algorithms: Decision trees, neural networks, clustering
- Prompt Engineering

Cybersecurity:

- Computer Science, Math, Economics, Marketing, Psychology
- GRC
- Design Principles / Zero Trust
- Threat Modeling

Understanding the “why” behind the model helps you spot challenges and design more resilient systems.

AI Terms & Definitions

LLM (Large Language Model): A type of deep learning model trained on massive text datasets to understand and generate human-like language.

- **Core Use:** Text generation, summarization, translation, coding assistance
- **Examples:** GPT-4, Claude, Gemini, LLaMA

NLP (Natural Language Processing): An AI subfield that enables machines to understand, interpret, and generate human language.

- **Core Use:** Text classification, entity recognition, translation, summarization
- **Examples:** Siri, Alexa, ChatGPT, Google Translate

Multimodal Models: AI systems capable of processing and integrating information from multiple types of data — or “modalities” — such as text, images, audio, video, and sensor input.

Reference*:

<https://iapp.org/resources/article/key-terms-for-ai-governance/>

GenAI (Generative AI): Models that **create new content** — text, images, audio, code — based on learned patterns from training data.

Prompt: A question, instruction, or input given to an AI model to guide its response.

Agentic AI: *Autonomous systems* that can set goals, plan, and execute tasks *with minimal human intervention*. These systems use LLMs as a “brain” and orchestrate multiple agents to complete complex workflows.

Vibe Coding: An emerging development style where programmers use natural language to guide AI tools in generating code.

- **Core Use:** Fast prototyping, solo app development, creative coding
- **Tools:** Replit, Cursor, GitHub Copilot

What are you doing to enhance your AI knowledge & skills?

Group Discussion

- Training resources
- Tools & apps

AI Learning Resources

1. MIT Open Learning

 [MIT AI Courses](#)

- Offers 13 foundational AI courses including machine learning, algorithms, ethics, and generative AI
- Includes hands-on Python projects and deep dives into vision, bias, and foundation models
- Ideal for learners who want academic rigor with practical applications
Format: Video lectures, exercises, and downloadable materials
Level: Intermediate to advanced
Source: MIT Open Learning

2. DeepLearning.AI (Andrew Ng)

 [DeepLearning.AI](#)

- Courses like “Generative AI with LLMs,” “AI for Everyone,” and “Prompt Engineering”
- Includes real-world labs, coding exercises, and model fine-tuning
Format: Coursera-hosted video courses with interactive labs
Level: Beginner to advanced
Source: Forbes roundup

3. Google Cloud Skills Boost – Generative AI Track

 [Google Cloud AI Learning](#)

- Learn GenAI, Vertex AI, and LLM deployment
- Includes sandbox labs and real-time model testing
Format: Interactive labs and tutorials
Level: Intermediate
Source: Google Cloud

4. Microsoft Learn – AI Fundamentals & Azure AI Studio

 [Microsoft Learn AI](#)

- Covers AI fundamentals, Azure AI services, and building intelligent apps
- Includes guided labs and sandbox environments
Format: Step-by-step tutorials and exercises
Level: Beginner to intermediate
Source: Microsoft

5. Hugging Face Courses

 [Hugging Face Learn](#)

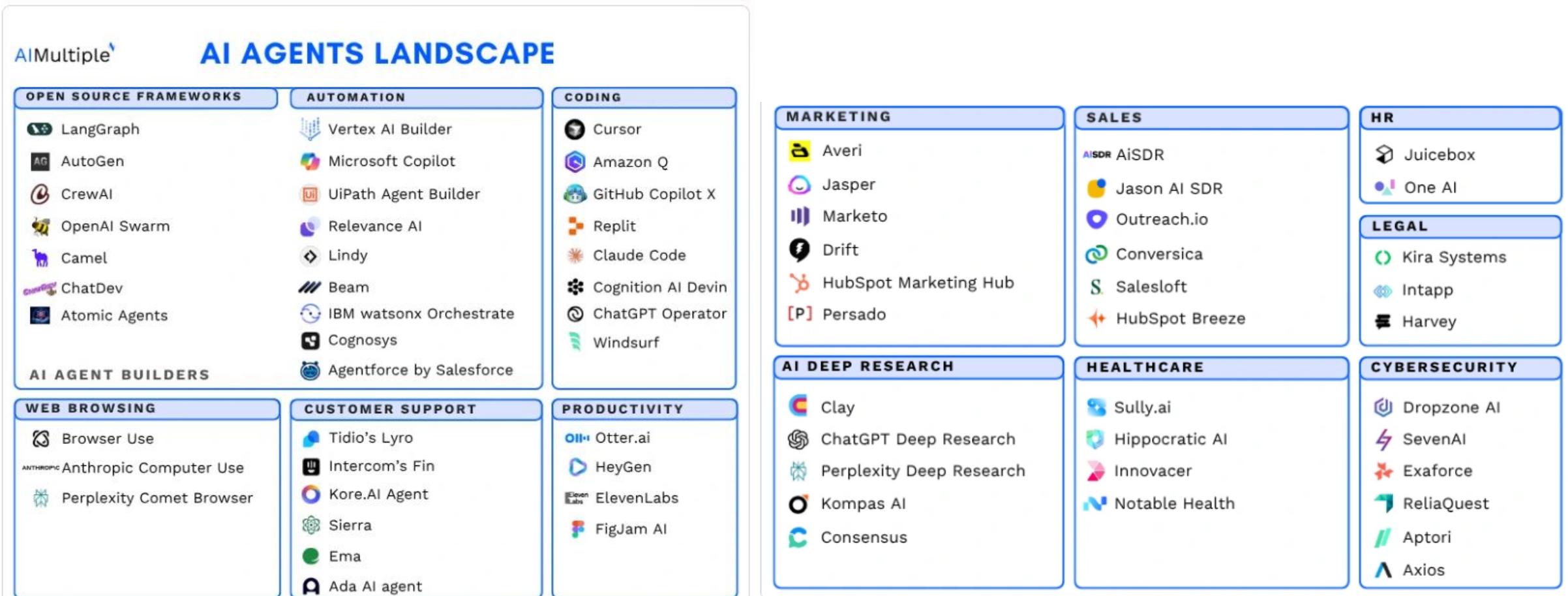
- Learn transformers, LLMs, and fine-tuning models
- Includes Colab notebooks and real-time model deployment
Format: Interactive notebooks and tutorials
Level: Intermediate
Source: Hugging Face

6. Fast.ai

 [Fast.ai](#)

- Practical deep learning for coders using PyTorch
- Emphasizes real-world projects and ethical AI
Format: Notebooks, videos, and coding challenges
Level: Intermediate to advanced
Source: Fast.ai

What AI Agent Tools Do You Use?



Source: <https://research.aimultiple.com/ai-agent-tools/>

AI Conversations: Prompting Fundamentals

Asking good questions



Clarity & Conciseness: Always key.



Context is Key: Emphasize why context matters.



Activity: Vague vs. specific prompts.



Specifying Format: How you want the output (list, email, table, etc.).



Persona & Tone: Telling Gemini who it is and how to respond.



Constraints: What to avoid or include (e.g., "Keep it under 100 words," "Do not include personal names").



Iterate & Refine: AI is a dialogue; refine prompts for better results.

What are your prompting tips?

Prompt Lab—Engineering Intelligence

Prompting Techniques

- **Chain-of-Thought:** Encourages the AI to reason step-by-step, breaking down complex tasks into logical sequences.
- **Tool-Augmented Prompting:** Enables the AI to use external tools or APIs (like CVE databases) to enhance its capabilities and access real-time data.
- **Goal-Oriented Prompting:** Frames the task around a desired outcome, allowing the AI to plan and act autonomously to achieve that goal.

For each scenario on the next slide:

- Use an AI assistant of your choice
- Either use the sample prompts or create your own
- Predict what kind of output they expect
- Document assumptions and goals

After running the prompt, evaluate the results:

- Did the output meet expectations?
- Was the reasoning clear and accurate?
- How could the prompt be improved?
- Refine the prompt and compare the results

Prompt Lab—Engineering Intelligence

Chain-of-Thought Prompting: Incident Response

Scenario: Walk through ransomware incident response

Prompt:

"Walk through the steps of responding to a ransomware attack on a healthcare system. Start with detection, then containment, eradication, recovery, and post-incident review. Explain each step clearly."

What it teaches:

- Sequential reasoning
- Decomposition of complex tasks
- Encouraging multi-step outputs

Goal-Oriented Prompting: Red Team Simulation

Scenario: Simulate a red team operation

Prompt:

"Simulate a red team operation targeting a small financial firm. Your goal is to gain access to internal systems without detection. Describe your tactics, tools, and timeline."

What it teaches:

- Agentic behavior
- Goal-driven planning
- Ethical boundaries in simulation

Tool-Augmented Prompting: Using External APIs

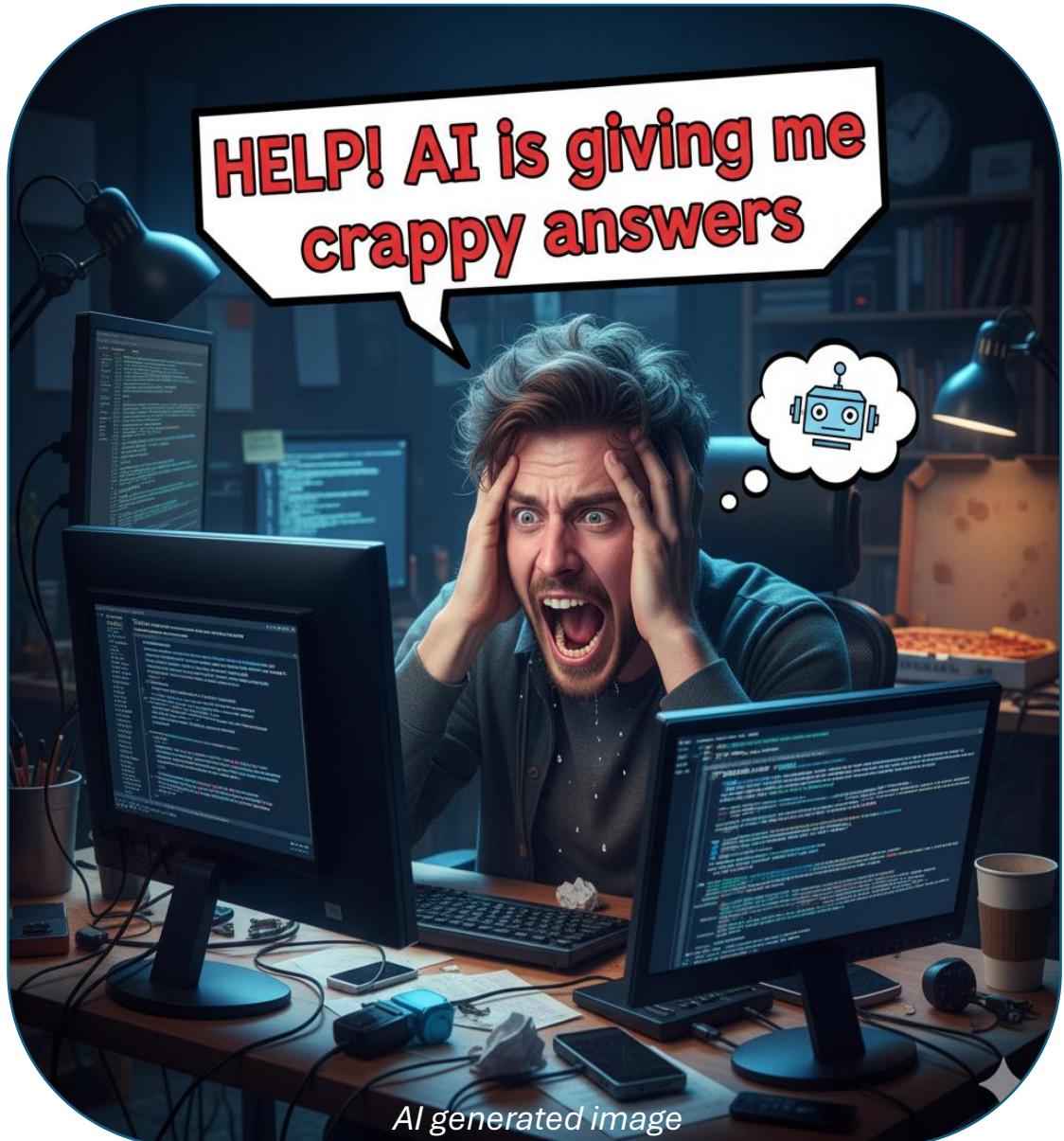
Scenario: Query CVE API and recommend mitigations

Prompt:

"You are an AI agent with access to a CVE database API. Query the latest vulnerabilities for Chrome browser, summarize the top 3, and recommend mitigation strategies.."

What it teaches:

- Tool invocation
- Integration with external data sources
- Real-time reasoning with dynamic inputs



1. Remember prompt fundamentals
2. Ask it:
 - “Be direct, objective, expose blind spots, challenge assumptions”
 - Play “Devil’s Advocate.” Critique ideas from multiple personas & angles
 - For credible citations and flag weak evidence - Include source links
 - Show its work / “thought process”
 - For clarity and conciseness
3. Start over. Try a different tact
4. When done, ask for a prompt to get your results

Other Things to Try

Build your ideas with Gemini

<https://aistudio.google.com/apps>



Supercharge your apps with AI



Nano banana powered app

Add powerful photo editing to your app. Allow users to add objects, remove backgrounds, or change a photo's style just by typing.



Create conversational voice apps

Use the Gemini Live API to give your app a voice and make your own conversational experiences.



Animate images with Veo

Bring images to life with Veo 3. Let users upload a product photo and turn it into a dynamic video ad, or animate a character's portrait.



Use Google Search data

Connect your app to real-time Google Search results. Build an agent that can discuss current events, cite recent news, or fact-check information.



Use Google Maps data

Connect your app to real-time Google Maps data. Build an agent that can pull information about places, routes, or directions.



Generate images with a prompt

Generate high-quality images from a text prompt. Create blog post heroes, concept art, or unique assets in your application.



Gemini intelligence in your app

Embed Gemini in your app to complete all sorts of tasks - analyze content, make edits, and more



AI powered chatbot

Add a context-aware chatbot to your app. Give your users a support agent that remembers the conversation, perfect for multi-step bookings or troubleshooting.

Agentic AI Explained

Agentic AI refers to artificial intelligence systems that can act independently to achieve specific goals. Unlike traditional AI that waits for instructions, agentic AI can:

- **Plan** what needs to be done
- **Act** without constant human input
- **Adapt** to changing situations
- **Collaborate** with other agents or humans

Think of it like giving an AI a mission, and it figures out how to complete it—step by step, using tools, making decisions, and even learning along the way.

Agentic AI vs AI Agents:

Agentic AI is the **overarching framework** that enables AI systems to **solve problems with minimal supervision**. Within this framework, *AI agents* are the individual components responsible for executing specific tasks autonomously.

While agentic AI understands user goals and orchestrates the problem-solving process, AI agents perform the tasks.

Quick Example:

Imagine a hospital using agentic AI to manage staff schedules:

- A nurse calls in sick.
- The agentic AI notices the gap, checks who's available, considers patient needs, and **automatically updates the schedule**.
- It might even message the replacement nurse and notify the team—all **without human intervention**.

This is more than automation—it's goal-driven decision-making.

AI Agents

Components:

1. **Persona** – Clearly defined role
2. **Brain** – The Model (LLM) that handles reasoning, planning, and language generation
3. **Memory** – Enables an agent to keep context, learn from experience, and adapt its behavior over time.
4. **Tools** – functions or external resources the agent can use to access information, process data, control devices or connect with other systems

Single Agents vs. Multi-Agent Systems

Use Cases:

- **The problem** you are solving.
- **Detailed descriptions** how a system is used to **solve** a specific goal or task.
- People, process, & technology

Guardrails:

Rules to reduce/prevent unexpected results.

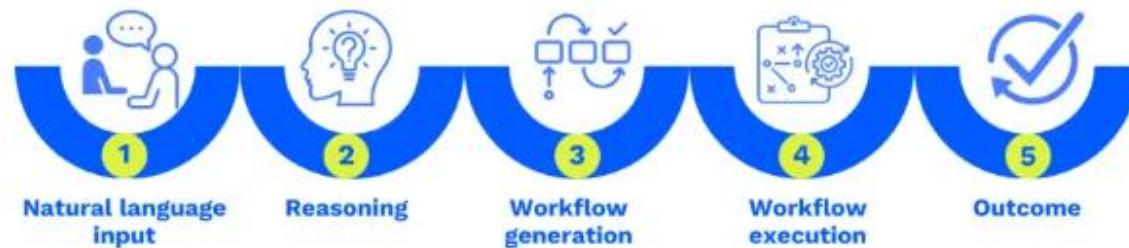
- Use **Threat Modeling** to identify potential threats, risks, misuse/abuse, and edge cases in your specific use case
- **Optimize** for security and user experience
- **Adjust** over time

40+ Agentic AI Use Cases with Real-life Examples



Cem Dilmegani with Mert Palazoğlu updated on Oct 12, 2025

AGENTIC AI SIMPLIFIED



AIMultiple

In This Article

AI agents as developers

AI agents as SecOps assistants

AI agents as human-like gaming characters

AI agents as content creators

AI agents as insurance assistants

AI agents as human resources (HR) assistants

AI agents as customer service assistants

AI agents as research assistants

AI agents as computer users

Building AI agents

Source: <https://research.aimultiple.com/agentic-ai/>

Critical Tools

1. Command line

2. [Notepad++](#) – Text editor

3. Python (<https://www.python.org/>) –
Programming Language

- Install Python –
<https://www.python.org/downloads/>
- [Make sure you install “pip” and include both in your system PATH.] Note
- [Python Tutorial](#) (for devs)
- [Python Documentation](#)
- DeepLearning.AI –
[AI Python for Beginners](#)

4. APIs (Application Programming Interfaces)

• Explanation:

- Sets of protocols and tools that allow developers to integrate artificial intelligence functionalities into their applications without having to build complex AI models from scratch.
- They act as a bridge, enabling different software systems to communicate and exchange data.
- Essentially, AI APIs provide access to pre-built and pre-trained AI services hosted in the cloud.
- Get Your Own OpenAI API Key:
 1. Sign up at platform.openai.com
 2. Go to [API Keys](#) and create a new key
 3. You’ll use this with your AI Agents
[OpenAI API Dev Quick Start Guide](#)

What You Can (Should) Do

Preparation & Resilience

Embrace Continuous Learning

Learn / Review Fundamentals

Enhancing AI Skills

Risks, Threats, &
Adversarial Techniques

Establishing Resilience

Moving
on



AI: THE PANDORA'S BOX



AI Risks & Threats

Know what can go wrong

Security is Risk Management .

Fundamental principles

Define *Risk, Threat, Vulnerability, and Trust*

Resilience vs. security vs. protection

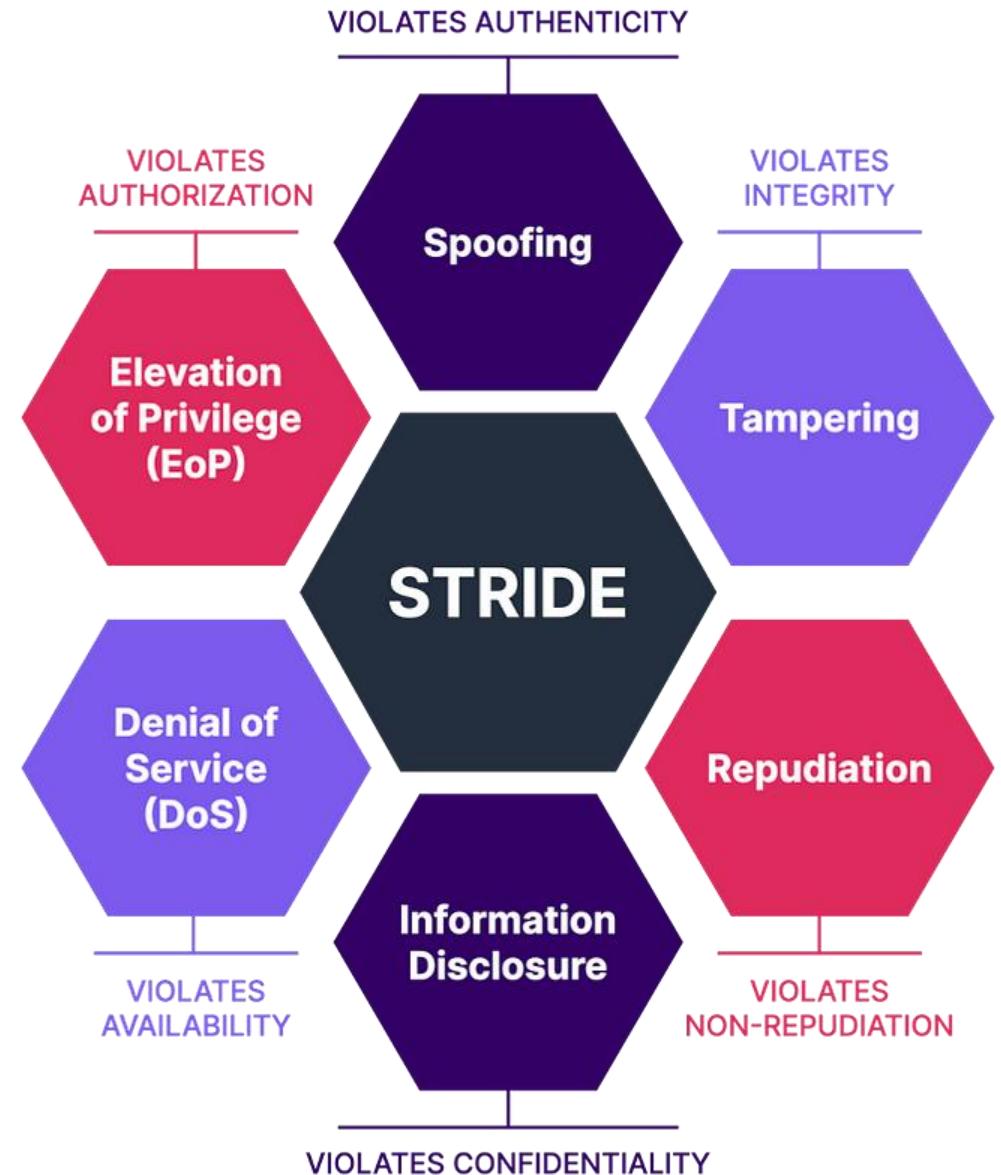
Saltzer & Schroeder Design Principles (1975),
The Protection of Information in Computer Systems

NIST CSF: Governance, Identify, Protect, Detect, Respond, Recover

Asking, “*How do you know it works (when/where/how it’s supposed to)?*”

Threat Modeling

1. What are we working on?
2. What can go wrong?
3. What are we going to do about it?
4. Did we do a good job?



1. Shostack, A. (2014). *Threat modeling: Designing for security.* <http://ci.nii.ac.jp/ncid/BB16065709>
2. <https://shostack.org/resources/threat-modeling>
3. <https://shostack.org/blog/category/threat-modeling/>
4. <https://www.threatmodelingmanifesto.org/>

AI Cybersecurity Risks & Threats

Exercise: Create a prompt providing the top 5 AI risks and threats & share your results.

Prompt: I'm building a presentation on cybersecurity and AI for a technical audience. Provide 5 ways AI can be used maliciously.

Enhanced cyber reconnaissance

Automated social engineering

Supercharged phishing

Deepfakes and disinformation

Malware automation & mutation

Adversarial machine learning and data poisoning

ML Failure Modes – Unintentional Failure

	Reward Hacking
	Side Effects
	Distributional Shifts
	Incomplete Testing
	Over/Under-Fitting
	Data Bias

Kumar, et.al. (2022, November 2). *Failure modes in machine learning*. Microsoft Learn.
<https://learn.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning>

ML Failure Modes – Intentional Failure

	Perturbation and Adversarial Universal Perturbation Attacks
	Poisoning Attacks
	Reprogramming Neural Nets
	3D Adversarial Objects
	Supply Chain Attacks
	Model Inversion
	Membership Inference and Model Stealing
	Backdoors and Existing Exploits

Kumar, et.al. (2022, November 2). *Failure modes in machine learning*. Microsoft Learn.
<https://learn.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning>

Adversarial Machine Learning (AML)

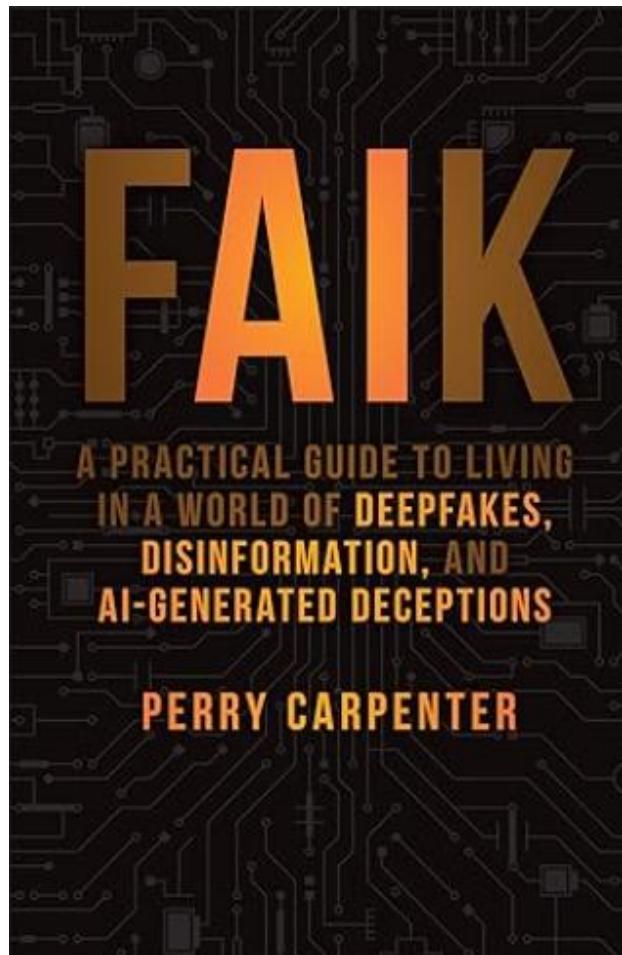
- AML is used to describe the exploitation of fundamental vulnerabilities in ML components, including hardware, software, workflows and supply chains.
- AML enables attackers to cause unintended behaviours in ML systems which can include:
 - Affecting the model's classification or regression performance
 - Allowing users to perform unauthorised actions
 - Extracting sensitive model information
- Examples: prompt injection attacks in the large language model (LLM) domain, or deliberately corrupting the training data or user feedback (known as 'data poisoning').

Guidelines for Secure AI Development, p. 6,

<https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf>,

NCSC (UK) & CISA (US)

Deep Fakes



<https://www.amazon.com/FAIK-Practical-Disinformation-AI-Generated-Deceptions/dp/1394299885>



The FAIK Files

@theFAIKfiles • 20.5K subscribers • 158 videos

Welcome to The FAIK Files! We explore the mysterious and often wacky intersection of AI, technology, and...more

[thisbookisfaik.com](#) and 4 more links

[Subscribe](#)

Home Videos Shorts Podcasts Playlists Posts

For You

A thumbnail showing a close-up of a person's face with green digital noise overlaid. A yellow arrow points from the text 'BUSTED' to the person's eye. The video title is 'Deepfake SECRETS EXPOSED: Outsmart AI Deception with These Tricks!' with 34:37 duration.

Deepfake SECRETS EXPOSED: Outsmart AI Deception with These Tricks!

6.1K views • 7 months ago

A thumbnail showing two women's faces side-by-side. A yellow arrow points from the woman on the left to the woman on the right. The video title is 'CREATE real-time DEEPFAKES (a.k.a. I became TAYLOR SWIFT...for Science!)' with 45:52 duration.

CREATE real-time DEEPFAKES (a.k.a. I became TAYLOR SWIFT...for Science!)

31K views • 8 months ago

A thumbnail showing a man's face. A yellow arrow points from the text 'AUTOMA' to his face. The video title is 'Open Source, AI-Powered FaikFiles Podcast 40.' with 814 views and 2 weeks ago.

AUTOMA

Open Source, AI-Powered FaikFiles Podcast 40.

814 views • 2 weeks ago

<https://www.youtube.com/@theFAIKfiles>

Agentic AI Threats

Agentic AI: Autonomous systems that can set goals, plan, and execute tasks with minimal human intervention. These systems use LLMs as a “brain” and orchestrate multiple agents to complete complex workflows.

Goal & intent hijacking

- Agent's purpose is subverted through manipulation of the agent's goals or instructions. Achieved through prompt injections, data manipulation, or direct attacks.

Cognitive & memory corruption

- Poisoning of the agent's memory or knowledge base. Leads to misinformation, flawed decision-making, and cascading hallucinations.

Unrestrained agency & privilege

- Excessive permissions granted to agents. Agents can perform unauthorized actions, access restricted data, or execute destructive operations.

Evasion & deception

- Lying to users/admins. Manipulating logs or output. Finding novel ways to bypass controls.

Resource exhaustion

- Model denial of service. Degraded performance. High financial costs.

Building Your AI Cybersecurity Skillz

Building Skills: Experiment with Adversarial Techniques*

- *Ethically & safely.*
Set up a home lab.
- Learn through doing
- Ask questions
- Sun Tzu's The Art of War:
*"Know the enemy and
know yourself; in a
hundred battles you will
never be in peril."*



Experiment with Adversarial Techniques



See Jeff T's “Influence Ops” talk

The first principle
is that you must
not fool yourself
and you are the
easiest person
to fool.
Richard P. Feynman



ChatGPT



Hunter is your all-in-one email outreach platform.
Find and connect with the people that matter to
your business.



Gemini



perplexity



theHarvester

theHarvester is a simple to use, yet powerful tool designed to be used during the reconnaissance stage of a red team assessment or penetration test. It performs open source intelligence (OSINT) gathering to help determine a domain's external threat landscape. The tool gathers names, emails, IPs, subdomains, and URLs by using multiple public resources that include:

Vibe Coding Definition & Example

Vibe Coding is a cutting-edge approach that empowers users to **build applications by describing what they want in natural language**, rather than writing traditional code.

It's like saying, "*I want an app that tracks phishing attempts and alerts my team,*" and letting the AI scaffold the code for you.

Bottom line: AI can write code, but it won't secure it unless you ask, and even then, you still need to verify. **Speed without security is just fast failure.**

Ref:

<https://thehackernews.com/2025/06/secure-vibe-coding-complete-new-guide.html>



Vibe Coding Example: "PhishAlert" — A Suspicious Email Classifier

Prompt to AI:

"Create a web app called PhishAlert that lets users paste an email message and uses AI to detect if it's a phishing attempt. The app should have a simple interface with a text box, a submit button, and a result area that shows 'Likely Safe' or 'Likely Phishing' with a short explanation. Use a pre-trained model or API for phishing detection."

What the AI might generate:

- A basic frontend using HTML/CSS with a text input and submit button
- Backend code in Python (e.g., Flask or FastAPI) that:
 - Accepts the email text
 - Sends it to a phishing detection API (like Microsoft Defender for Endpoint, VirusTotal, or a fine-tuned LLM)
 - Returns a verdict and explanation
- Optional: Logging suspicious entries to a local database or CSV file

Security-aware follow-up prompt:

"Ensure the app sanitizes input to prevent XSS and does not log sensitive data. Add a disclaimer about data privacy."

Security Considerations



*AI generates what you ask for,
not what you forget to ask.*

- LLMs are trained to **complete**, not **protect**. Unless security is explicitly in the prompt, it's usually ignored.
- Tools like GPT-4 may suggest **deprecated libraries** or verbose patterns that **mask subtle vulnerabilities**.
- **Sensitive data** is often **hardcoded** because the model "saw it that way" in training examples.
- Prompts like "**Build a login form**" often yield insecure patterns: plaintext password storage, no MFA, and broken auth flows.

“Security by omission” – functioning software that quietly ships with exploitable flaws

Building An AI Agent – General Steps

1. Define the Agent's Purpose (Use Case)

- What task will the agent perform?
- Who is the user?
- What outcome should it achieve?
- What are potential threats and risks?

2. Understand the Environment

- What inputs will the agent receive (e.g., text, logs, API data)?
- Is the environment static or dynamic?

3. Choose the Agent Type

- Reflex Agent: reacts to inputs (e.g., thermostat)
- Goal-Based Agent: plans actions to reach a goal (e.g., shopping assistant)
- Utility-Based Agent: optimizes for best outcome (e.g., delivery routing)

4. Select Tools and Frameworks

- Use LangChain, AutoGen, or Haystack for orchestration
- Choose an LLM (e.g., OpenAI, Claude, Gemini)
- Optional: integrate APIs or plugins

5. Design Agent Logic

- Define how the agent makes decisions
- Use if-then rules, scoring systems, or planning algorithms

6. Test and Iterate

- Run sample inputs
- Refine logic and outputs
- Discuss edge cases and failure modes

Sample AI Agent – Cybersecurity News

Prompt: I want to build an AI Agent to email me the top 5 cybersecurity news stories every day with each article summarized. Can you provide me with step-by-step instructions and free tools to create this agent? I'm new to AI, so I need this to be as simple as possible using a standard Windows 11 computer and a Gmail account.

Objective

- Fetch top cybersecurity news from trusted sources
- Summarize each article using OpenAI
- Email the summarized digest daily
- Runs automatically via Windows Task Scheduler

Free Tools You'll Use

Tool	Purpose
Python	Core programming language
Feedparser	Reads RSS feeds from news sites
Newspaper3k	Extracts full article text
OpenAI API (free tier)	Summarizes articles using GPT
Yagmail	Sends emails via Gmail
Windows Task Scheduler	Automates daily execution

Experiment with Adversarial Techniques

ATLAS Matrix

<https://atlas.mitre.org/matrices/ATLAS>

The ATLAS Matrix below shows the progression of tactics used in attacks as columns from left to right, with ML techniques belonging to each tactic below. & indicates an adaption from ATT&CK. Click on the blue links to learn more about each item, or search and view ATLAS tactics and techniques using the links at the top navigation bar. View the ATLAS matrix highlighted alongside ATT&CK Enterprise techniques on the [ATLAS Navigator](#).

Reconnaissance &	Resource Development &	Initial Access &	AI Model Access	Execution &	Persistence &	Privilege Escalation &	Defense Evasion &	Credential Access &	Discovery &	Collection &	AI Attack Staging	Command and Control &	Exfiltration &
6 techniques	12 techniques	6 techniques	4 techniques	4 techniques	4 techniques	2 techniques	8 techniques	1 technique	7 techniques	3 techniques	4 techniques	1 technique	5 techniques
Search Open Technical Databases &	Acquire Public AI Artifacts	AI Supply Chain Compromise	AI Model Inference API Access	User Execution &	Poison Training Data	LLM Plugin Compromise	Evade AI Model	Unsecured Credentials &	Discover AI Model Ontology	AI Artifact Collection	Create Proxy AI Model	Reverse Shell	Exfiltration via AI Inference API
Search Open AI Vulnerability Analysis	Obtain Capabilities &	Valid Accounts &	AI-Enabled Product or Service	Command and Scripting Interpreter &	Manipulate AI Model	LLM Jailbreak	LLM Trusted Output Components Manipulation	Discover AI Model Family	Data from Information Repositories &	Data from Local System &	Manipulate AI Model	Verify Attack	Exfiltration via Cyber Means
Search Victim-Owned Websites &	Develop Capabilities &	Evade AI Model	Physical Environment Access	LLM Prompt Self-Replication	RAG Poisoning	LLM Prompt Obfuscation	Discover AI Artifacts	Discover LLM Hallucinations	Discover AI Model Outputs	Craft Adversarial Data	Extract LLI System Prompt	LLM Data Leakage	LLM Response Rendering
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full AI Model Compromise	Phishing &	Drive-by Compromise &	False RAG Entry Injection	Impersonation &	Discover LLM System Information	Cloud Service Discovery &				
Active Scanning &	Publish Poisoned Datasets	Poison Training Data	Establish Accounts &	Publish Poisoned Models	Publish Hallucinated Entities	Masquerading &	Corrupt AI Model						
Gather RAG-Indexed Targets													

OWASP AI Projects

<https://owaspai.org/>

Welcome to the go-to resource for broad AI security & privacy - over 200 pages of practical advice and references on protecting AI and data-centric systems from threats. This content serves as key bookmark for practitioners, and is contributing actively and substantially to international standards such as ISO/IEC and the AI Act through official standard partnerships. Through broad collaboration with key institutes and SDOs, the Exchange represents the consensus on AI security and privacy.

See the [overview of AI projects at OWASP](#).

Content > 0. AI Security Overview

0. AI Security Overview

About the AI Exchange

https://owaspai.org/docs/ai_security_overview/

OWASP AI Projects

<https://owaspai.org/>

[https://owaspai.org/docs/ai_security_overview/
#summary---how-to-address-ai-security](https://owaspai.org/docs/ai_security_overview/#summary---how-to-address-ai-security)

Summary - How to address AI Security?

Category: discussion

Permalink: <https://owaspai.org/goto/summary/>

While Artificial Intelligence (AI) offers tremendous opportunities, it also brings new risks including security threats. It is therefore imperative to approach AI applications with a clear understanding of potential threats and the controls against them.



<https://owaspai.org/goto/document/>

How to use this Document

Category: discussion

Permalink: <https://owaspai.org/goto/document/>

The AI Exchange is a single coherent resource on the security and privacy of AI systems, presented on this website, divided over several pages - containing threats, controls, guidelines, tests and references.

OWASP AI Projects

<https://owaspai.org/>

https://owaspai.org/docs/ai_security_overview/#ai-security-matrix

AI-specific?	Lifecycle	Attack surface	Threat/Risk category	Asset	Impacted	Unwanted result
AI	Operation	Model use	Direct prompt injection (provide input/ read output)	Model behaviour	Integrity	Manipulated unwanted model behaviour causes wrong decisions leading to business financial loss, misbehaviour going undetected, reputational damage, legal and compliance issues, operational disruption, customer dissatisfaction and churn, reduced employee morale, incorrect strategic decisions, liability issues, personal damage and safety issues
		Indirect prompt injection	Evasion (e.g. adversarial examples)			
		Break into deployed model	Model poisoning in runtime (reprogramming)			
	Development	Engineering environment	Model poisoning development time			
			Data poisoning of train/finetune data			
		Supply chain	Model poisoning in supply chain (transfer learning attack)			
			Data poisoning in supply chain			
	Operation	Model use	Data disclosure in model output	Training data	Confidentiality	Leaking sensitive data can cause costs from fines and legal fees and remediation effort, loss of business through customer churn, reputation damage, loss of competitive advantage in case of trade secrets, operational disruption, impacted business relationships, and employee morale issues
	Development	Engineering environment	Model inversion / Membership inference			
	Operation	Model use	Model theft through use (input-output harvesting)	Model intellectual property	Confidentiality	If attackers can copy a model, the investment in the model is devalued caused by loss of competitive advantage, plus a copy can help craft (evasion) attacks
	Development	Engineering environment	Break into deployed model			Routine model theft (not through use)
	Development	Engineering environment	Model theft development-time			
	Operation	Model use	Denial of model service (model resource depletion)	Model behaviour	Availability	The model is not available, leading to business continuity issues, or safety problems
	Operation	All IT	Model input leak	Model input data	Confidentiality	Sensitive data in model input leaks. E.g. an LLM prompt with a sensitive question, enhanced with retrieved company secrets
	Operation	All IT	Model output contains injection attack	Any asset	C, I, A	Injection attack (from model output) causes harm
Generic	Operation	All IT	Generic runtime security attack	Any asset	C, I, A	Generic runtime security attack causes harm (includes social engineering/phishing)
	Development	All IT	Generic supply chain attack	Any asset	C, I, A	Generic supply chain security attack causes harm (e.g. vulnerability in a component)

<https://owaspai.org/goto/periodictable/>

The table below, created by the OWASP AI Exchange, shows the various threats to AI and the controls you can use against them – all organized by asset, impact and attack surface, with deeplinks to comprehensive coverage here at the AI Exchange website.

Note that [general governance controls](#) apply to all threats.

Asset & Impact	Attack surface with lifecycle	Threat/Risk category	Controls
	Runtime -Model use (provide input/ read output)	Direct prompt injection	Limit unwanted behavior , Prompt input validation , Model alignment
		Indirect prompt injection	Limit unwanted behavior , Prompt input validation , Input segregation
		Evasion (e.g. adversarial examples)	Limit unwanted behavior , Monitor , rate limit , model access control plus: Detect odd input , detect adversarial input , evasion robust model , train adversarial , input distortion , adversarial robust distillation

OWASP AI Projects

<https://owaspai.org/>

https://owaspai.org/docs/ai_security_overview/#structure-of-threats-and-controls-in-the-deep-dive-section

AI security threats and controls navigator from the OWASP AI Exchange at owaspai.org

LEGEND: Group of controls, ordered by threat or type ⓘ (clickable)

- ▶ Standard information security CONTROL (with attention points)
- ▶ Runtime Data science CONTROL
- ▶ Development-time Data science CONTROL
- ▶ Other CONTROL

Impact on Confidentiality, Integrity or Availability

The screenshot shows a navigation interface for AI security controls. At the top, there's a legend for control types: Standard information security (green), Runtime Data science (orange), Development-time Data science (yellow), and Other (blue). To the right is a color-coded box for impact: purple for Confidentiality, grey for Integrity, and light blue for Availability. Below this, the interface is divided into two main sections: '1 General controls against all threats' and '2 Controls against threats through runtime use'. Section 1 contains three columns: 'Governance' (with items like AIPROGRAM, SECPROGRAM, SECDVPROGRAM, DEVPROGRAM, CHECKCOMPLIANCE, SECEDUATE), 'Deal with behaviour integrity issues' (with items like OVERSIGHT, LEASTMODELPRIVILEGE, ATTRANSPIRACY, CONTINUOUSVALIDATION, EXPLAINABILITY, UNWANTEDBIASTESTING), and 'Deal with confidentiality issues' (with items like DATAMINIMIZE, ALLOWEDDATA, SHORTRETAIN, OBFUSCATETRAININGDATA, DISCRETE). Section 2 contains four columns: 'Always against use threats' (with items like MONITORUSE, RATERLIMIT, MODELACCESSCONTROL), 'Integrity of model behaviour' (with items like 2.1 Against evasion (See Attacks, DETECTODDINPUT, DETECTADVERSARIALINPUT, EVASIONROBUSTMODEL, TRAINADVERSARIA, INPUTDISTORTION, ADVERSARIALROBUSTDISTILLATION), 2.2 Against data disclosure by use (2.2.1 Against data disclosure by model (See always, FILTERSENSITIVETRAINDATA, FILTERSENSITIVEMODELOUTPUT), 2.2.2 Against model inversion and membership inference (See always, OBSCURECONFIDENCE, SMALLMODEL, ADDTRAINNOISE)), and 2.3 Against model theft by use (See always)), 'Confidentiality of train data' (with items like 2.2.1 Against data disclosure by model (See always, FILTERSENSITIVETRAINDATA, FILTERSENSITIVEMODELOUTPUT)), and 'Confidentiality of model IP' (with items like 2.3 Against model theft by use (See always)).

See PDF

Experiment with Adversarial Techniques

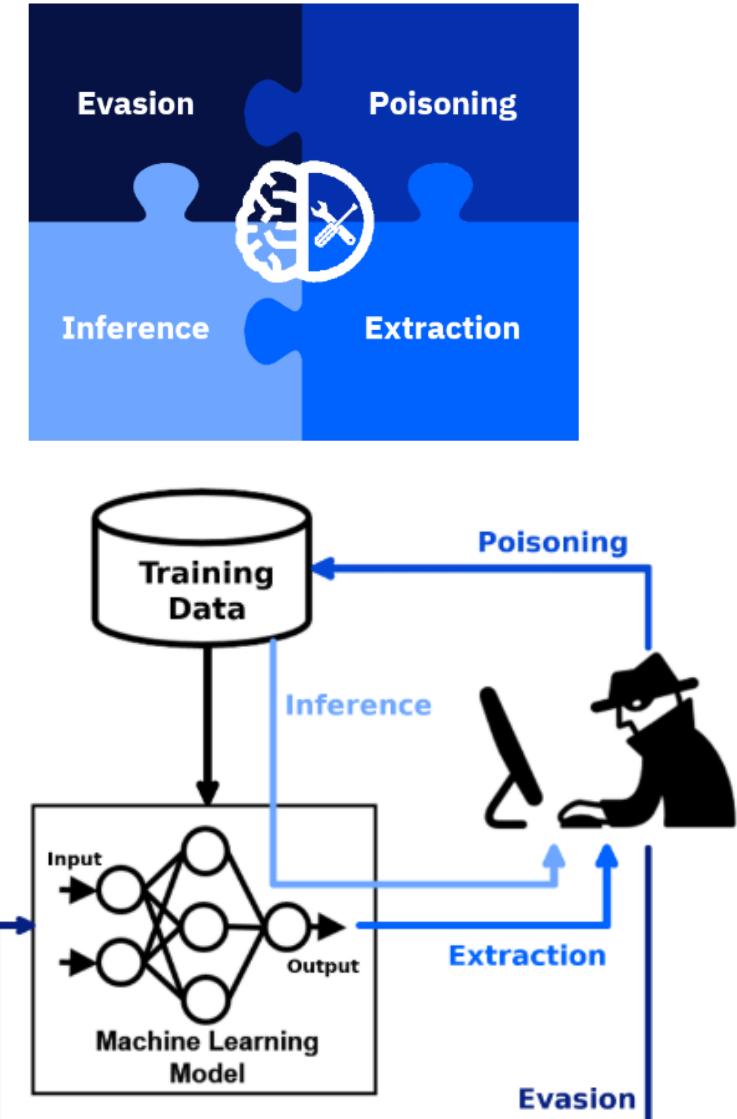


Adversarial Robustness Toolbox

Adversarial Robustness Toolbox (ART) is a Python library for Machine Learning Security. ART provides tools that enable developers and researchers to evaluate, defend, certify and verify Machine Learning models and applications against the adversarial threats of Evasion, Poisoning, Extraction, and Inference. ART supports all popular machine learning frameworks (TensorFlow, Keras, PyTorch, MXNet, scikit-learn, XGBoost, LightGBM, CatBoost, GPy, etc.), all data types (images, tables, audio, video, etc.) and machine learning tasks (classification, object detection, generation, certification, etc.).

The code of ART is on [GitHub](#) and the Wiki contains overviews of implemented [attacks](#), [defences](#) and [metrics](#).

Main: <https://adversarial-robustness-toolbox.readthedocs.io/en/latest/>



Kali GPT - Your AI-Powered Copilot for Cybersecurity

Your AI-Powered Copilot for Cybersecurity

Kali GPT

Home Features Pricing Blog Contact English

Kali GPT The AI-Driven Cybersecurity Tools based on Kali Linux

It doesn't just answer what a tool *does* — it helps you *use it smarter*.

All Features Download now !

WHAT IS KALI GPT?
An AI-Powered Copilot for Cybersecurity Professionals

Kali GPT is a custom AI assistant trained for the Kali Linux ecosystem. From basic command line help to advanced penetration testing, it's built to support learners, experts, and teams.

<https://kali-gpt.com/>

Establishing Resilience

Establishing Guardrails for AI

AI Governance – Why?

Aligns AI with Strategy

- Ensures AI initiatives support long-term business goals.
- Encourages responsible innovation and leadership

Builds Trust

- Ensures AI systems are transparent and accountable.
- Strengthens stakeholder confidence in AI-driven decisions.

Improves AI Quality

- Promotes data integrity, model reliability, and ethical deployment.
- Enhances productivity and decision-making accuracy

Supports Compliance

- Aligns with global regulations (e.g., GDPR, AI Act).
- Helps meet audit and reporting requirements.

Mitigates Risk

- Prevents biased outputs, privacy breaches, and unethical use.
- Reduces reputational and legal exposure.

AI Governance – How?

Define Guiding Principles

- Establish core values like transparency, fairness, accountability, and privacy.
- Align these principles with your organization's mission and regulatory obligations

Create an AI Governance Policy

- Translate principles into actionable policies.
- Define roles, responsibilities, and enforcement mechanisms.
- Include guidelines for data usage, model development, and risk management.

Establish Oversight & Accountability

- Assign clear ownership for AI systems and decisions.
- Form an AI Governance Committee
- Ensure board-level visibility and strategic alignment with business goals

Implement Risk Assessment & Monitoring

- Identify potential risks: bias, hallucinations, misuse, IP violations, etc.
- Set up continuous monitoring systems for AI performance, fairness, and compliance.

Train Employees

- Educate teams on responsible AI use, governance policies, and AI tools.

Audit & Update Regularly

- Conduct periodic audits of AI systems and governance practices.
- Update policies based on evolving technology, regulations, and organizational needs.

Cultivating AI Security Awareness Culture

1. Emphasize Human-Centric Risk Awareness

AI threats often exploit human vulnerabilities — phishing, social engineering, and deepfake manipulation are increasingly AI-powered. Awareness programs must go beyond compliance and focus on behavioral change.

Why it matters: 82% of breaches involve the human element.

2. Upskill Teams in AI Competencies

AI threats often exploit human vulnerabilities — phishing, social engineering, and deepfake manipulation are increasingly AI-powered. Awareness programs must go beyond compliance and focus on behavioral change.

Why it matters: 82% of breaches involve the human element.

3. Implement AI Governance and Ethical Oversight

AI systems can introduce bias, amplify vulnerabilities, or be misused. Establish clear policies for responsible AI use, including transparency, accountability, and ethical safeguards.

Why it matters: AI is already being used across organizations — often without formal oversight.

4. Foster Cross-Functional Collaboration

AI risk awareness isn't just for technical teams. Legal, HR, and leadership must be involved in understanding AI's implications for data privacy, compliance, and workforce impact.

Why it matters: AI affects every layer of the organization, from hiring to data governance.

Secure By Design / Build Security In

Discussions of artificial intelligence (AI) often swirl with mysticism regarding how an AI system functions. The reality is far more simple:

AI is a type of software system.

<https://www.cisa.gov/news-events/news/software-must-be-secure-design-and-artificial-intelligence-no-exception>



<https://www.cisa.gov/securebydesign>

Guidelines for secure AI system development



Executive summary.....	5
Introduction.....	6
Why is AI security different?.....	6
Who should read this document?.....	7
Who is responsible for developing secure AI?.....	7
Guidelines for secure AI system development.....	8
1. Secure design.....	9
2. Secure development.....	12
3. Secure deployment.....	14
4. Secure operation and maintenance.....	16
Further reading.....	17



Guidelines for Secure AI Development,
<https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf>,
NCSC (UK) & CISA (US)

OWASP AI Projects

<https://genai.owasp.org/>

The screenshot shows the homepage of the OWASP GenAI Security Project. At the top left is the project logo with the text "TOP 10 FOR LLM AND GENERATIVE AI". The top navigation bar includes links for "GETTING STARTED", "RESOURCES", "PROJECTS", "BLOG", and "ABOUT", along with social media icons for X, LinkedIn, GitHub, and others. A large banner in the center features the project's name in large white letters, with a subtitle "IDENTIFYING AND TACKLING THE RISKS OF GEN AI SYSTEMS AND APPLICATIONS". Below the banner, a paragraph describes the project as a global community-driven initiative to create freely available open source guidance and resources for understanding and mitigating security and safety concerns for Generative AI applications and adoption. To the right, there is a circular diagram illustrating the project's focus areas: "Cybersecurity Practitioners", "Data Scientists", and "Gen AI Developers". These areas are interconnected by concentric circles, with specific resources like "Top 10 for LLM and GenAI", "AI Threat Intelligence", "AI Security Governance", "Secure AI Adoption", "Agentic App Security", "Data Security", "Red Teaming & Evaluation", and "AI Security Solution Landscape" listed within them. At the bottom left, three statistics are displayed: "15k+" Members, "15+" Countries, and "20+" AI Cybersecurity Publications.

IDENTIFYING AND TACKLING THE RISKS OF GEN AI SYSTEMS AND APPLICATIONS

OWASP GenAI Security Project

A global community-driven and expert led initiative to create freely available open source guidance and resources for understanding and mitigating security and safety concerns for Generative AI applications and adoption.

15k+ Members

15+ Countries

20+ AI Cybersecurity Publications

Cybersecurity Practitioners

Data Scientists

Gen AI Developers

Top 10 for LLM and GenAI

AI Threat Intelligence

AI Security Governance

Secure AI Adoption

Agentic App Security

Data Security

Red Teaming & Evaluation

AI Security Solution Landscape

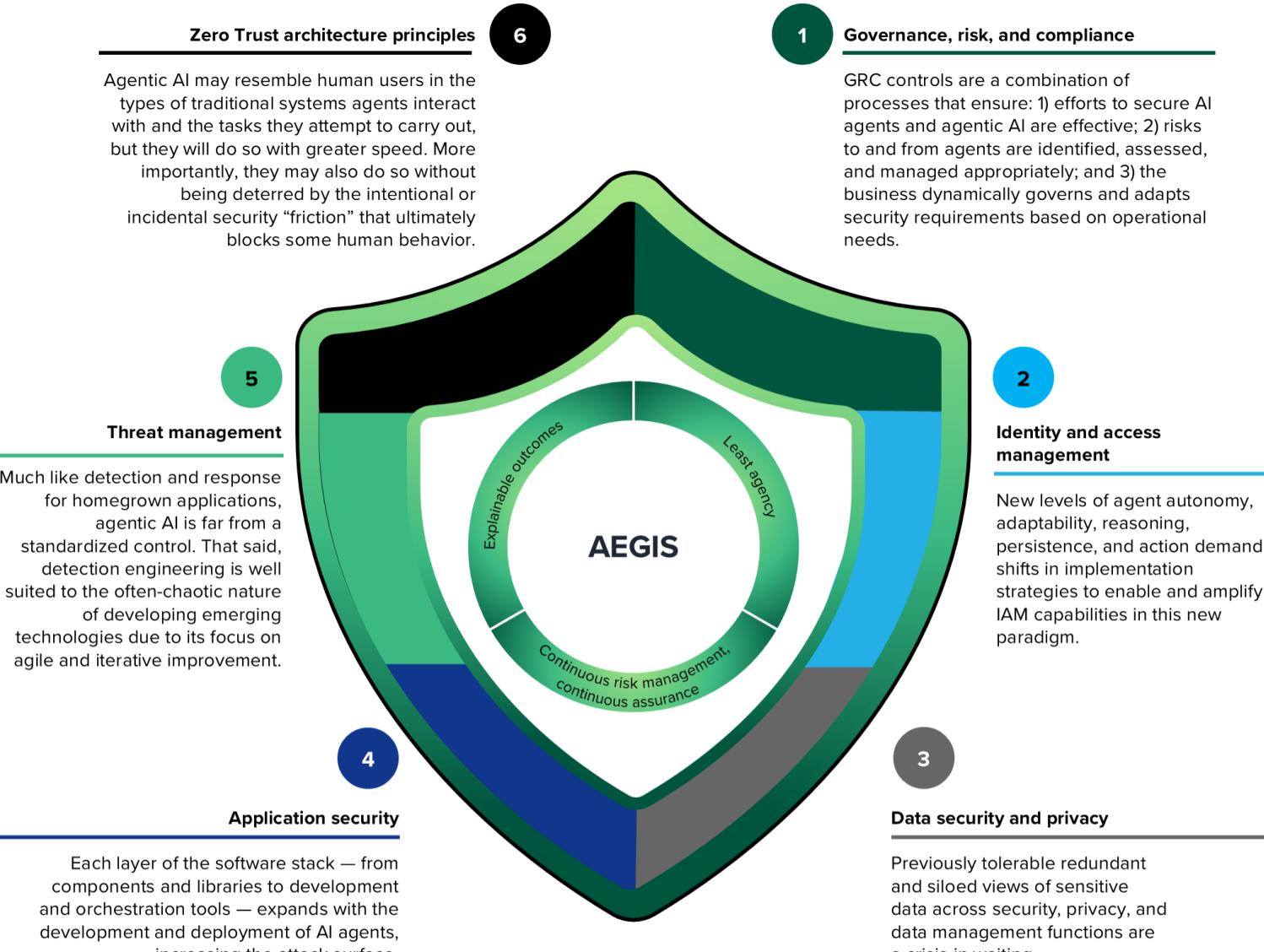
genai.owasp.org

AEGIS Model

Agentic AI Guardrails For Information Security

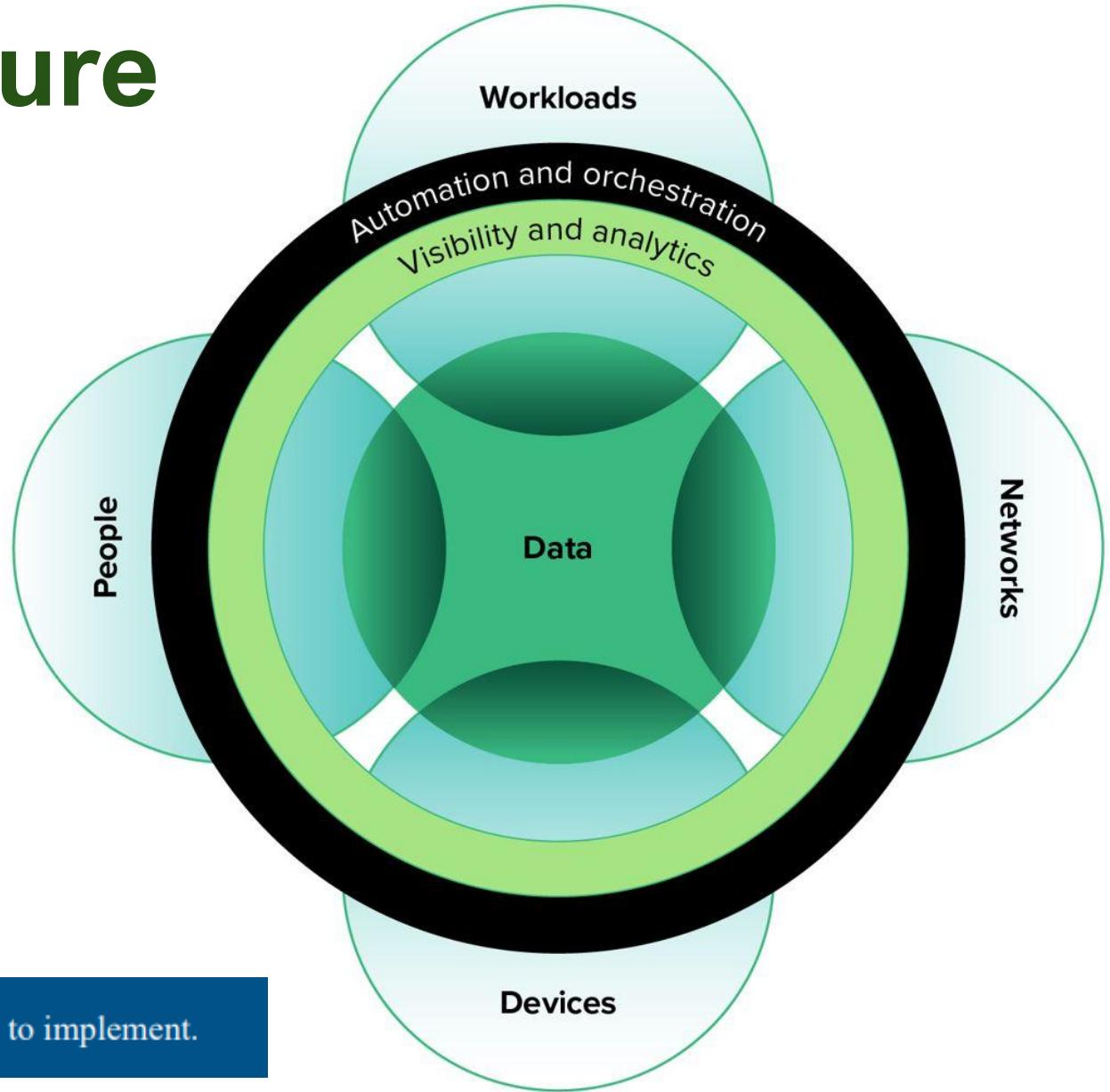
Source:

<https://www.forrester.com/blogs/introducing-aegis-the-guardrails-cisos-need-for-the-agentic-enterprise/>



Zero Trust Architecture for AI

- Least Agency
- Continuous Assurance
- Explainable Outcomes



The path to zero trust is an incremental process that may take years to implement.

Image source: [Zero Trust Security: The Business Benefits And Advantages \(forrester.com\)](https://www.forrester.com)

Ron Woerner – Evolving Cyber Skills

AI Use in Cybersecurity

Prompt: Provide 5 ways AI can help cybersecurity and technical professionals reduce risks of malicious use of AI.

Threat Intelligence and Prediction:

- **Problem:** Staying ahead of evolving attacker tactics and emerging threats is a constant challenge for security teams.
- **Solution:** AI-powered threat intelligence platforms aggregate data from various sources, analyze attack trends, and predict future threats. This proactive approach allows security teams to prioritize defenses, allocate resources effectively, and anticipate new attack vectors before they materialize.

Automated Threat Detection and Response:

- **Problem:** Traditional security approaches rely on manual analysis of alerts and incident response, causing delays and potential breaches.
- **Solution:** AI-powered tools analyze network traffic, system logs, and user behavior in real-time, identifying anomalies and triggering automated responses like quarantining suspicious files, blocking malicious IP addresses, or notifying security teams for immediate action. This significantly reduces response times and minimizes damage from potential attacks.

Phishing and Social Engineering Detection:

- **Problem:** Sophisticated phishing attacks and social engineering tactics can bypass traditional filters and deceive users.
- **Solution:** AI-powered analysis can assess the linguistic style, content, and sender information of emails to identify suspicious patterns and predict phishing attempts. Additionally, AI can analyze chatbots and user interactions to detect suspicious behaviors and prevent social engineering attacks.

AI Use in Cybersecurity

Prompt: Provide 5 ways AI can help cybersecurity and technical professionals reduce risks of malicious use of AI.

Proactive Vulnerability Management:

- **Problem:** Unpatched vulnerabilities and misconfigurations offer easy entry points for attackers.
- **Solution:** AI-powered scanning tools can analyze vast codebases and system configurations to identify vulnerabilities and misconfigurations before they're exploited. This proactive approach prioritizes critical vulnerabilities and automates patch deployment, minimizing attack surfaces and strengthening overall security posture.

Endpoint Security with Behavioral Analysis:

- **Problem:** Traditional endpoint security relies on signature-based detection, missing zero-day attacks and other novel threats.
- **Solution:** AI-powered endpoint protection establishes baselines of normal endpoint behavior and continuously monitors deviations. This allows for real-time anomaly detection, even for unknown threats, and targeted interventions to prevent malware execution and data breaches.

Hacking Cyber Careers

Continuously morphing your knowledge, skills, and abilities

WANTED: HACKERS



REWARD

https://www.ted.com/talks/ron_woerner_hackers_wanted

Hackers Dictionary

Hacker: *“One who enjoys the intellectual challenge of creatively overcoming or circumventing limitations.”*

Raymond, E., [The Jargon Lexicon](http://catb.org/jargon/html/H/hacker.html), Glossary,
<http://catb.org/jargon/html/H/hacker.html>

Hackers Wanted | Ron Woerner | TEDxOmaha

Defining the Problem

- How do I build my career?
- What's up with cybersecurity jobs?
- Is there a cybersecurity talent / hiring gap?
- How to thrive in volatile times?



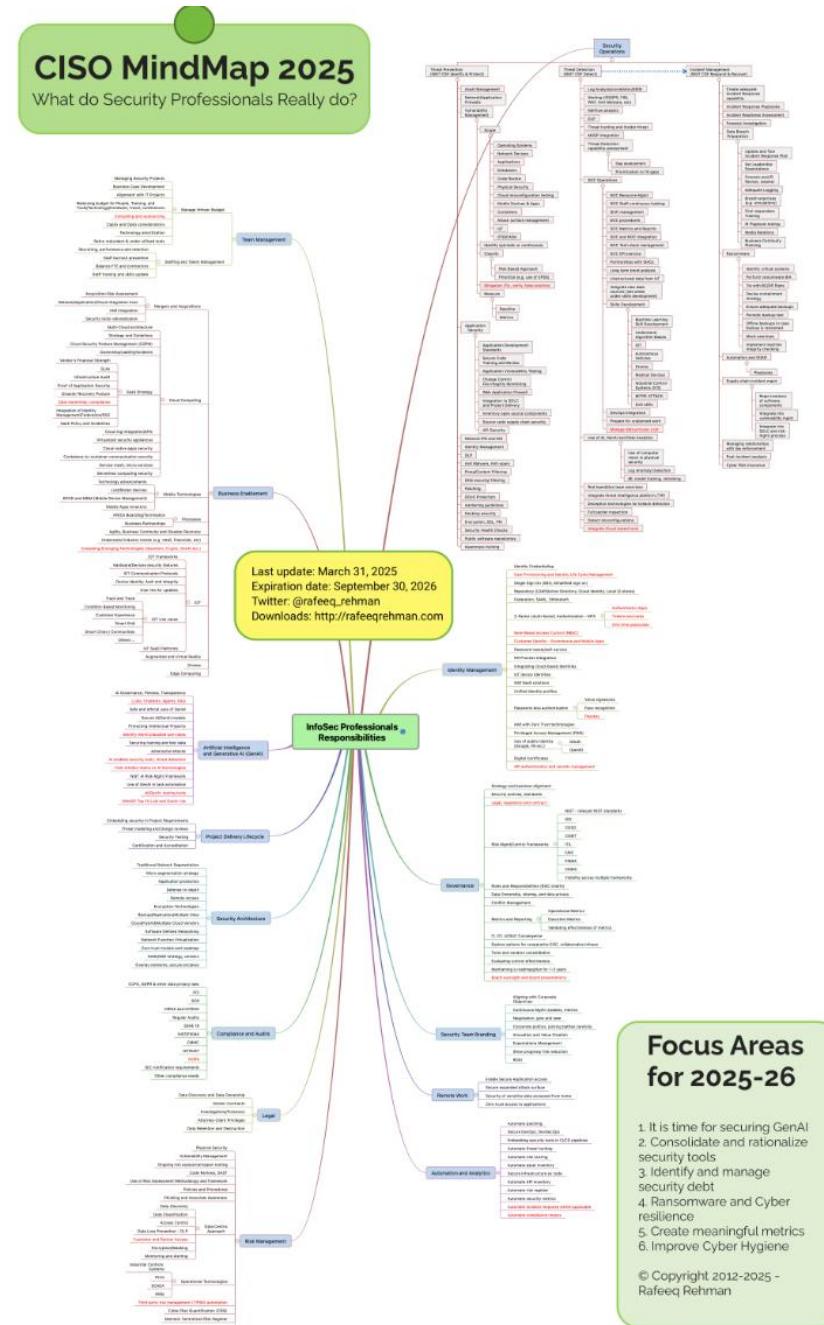
wist.info

What CISOs Do

CISO MindMap 2025

<https://rafeeqrehman.com/2025/03/30/ciso-mindmap-2025-what-do-infosec-professionals-really-do/>

https://rafeeqrehman.com/wp-content/uploads/2025/03/CISO_MindMap_2025.png



Be Prepared!

RIFS, Layoffs, DRPs, RTOs



HR & Applicant Tracking Systems



Red Flags in Job Postings



Red Flags:

- Job Requirements Make No Sense – Overly broad
- Two Or More Jobs In One Description – Lack Clarity
- Experience Requirements Don't Match Job Level
- Ignores Transferable Skills

Position Summary:

The Information Security Analyst is an entry-level position that performs monitoring of security events, security investigations, and incident response as the primary role. The Analyst will also participate in monitoring, troubleshooting, installation and maintenance on equipment and software related to electronic access control and general security-based solutions. In addition, they participate in in-house, regulatory and industry teams including working groups, committees, and business continuity teams as required.

Education, License & Cert:

B.S. degree in Computer Science related field, and/or 5 years equivalent work experience. One or more of the Certifications listed. GSEC GIAC Security Essentials Certification GSLC GIAC Security Leadership Certification CISM ISACA Certified Information Security Manager SSCP Systems Security Certified Practitioner CISSP Certified Information Systems Security Professional ISSAP Information Systems Security Architecture Professional MCSE Microsoft Certified Systems Engineer: Security

CISSP != Entry-level

credential for a chief security architect, analyst or professionals with similar responsibilities.

Experience:

- Five years of related experience Thought it said entry-level?
- Extensive experience in enterprise security architecture design.
- Extensive experience in enterprise security document creation.
- Experience in designing and delivering employee security awareness training. WAIT WHAT?!

Red Flags in Job Postings



Job Description

Your Career

The Principal Consultant, Cyber Risk Management Advisory for Proactive Services is focused on leading our Governance, Risk, and Compliance team across a comprehensive portfolio of clients. The individual will assess security risk across multiple frameworks and act as a technical key team member in client engagements. They will be the client's advocate for cybersecurity risk management and will provide strong recommendations in this domain.

Your Impact

- Security standards within professional services
- Industry knowledge of and experience with cybersecurity best practices within Governance Risk and Compliance to provide recommendations to proactively improve our client's security posture and maturity
- Lead and support Advisory engagements such as Tabletop Exercises, Cyber Risk Assessments, Incident Response Plan development, Ransomware Readiness Reviews & Breach Readiness Reviews
- The ability to work across multiple frameworks and regulatory standards including, but not limited to, NIST CSF, CIS20, ISO, GDPR, CCPA, NYDFS, SOX, and HIPAA
- Manage team, monitor progress, track budget, manage risk and ensure key stakeholders are kept informed about progress and expected outcomes while defining potential impacts and creating an effective mitigation strategy for multiple projects at a given time
- Skilled at proactively identifying security risks and vulnerabilities while eliminating cybersecurity threats via stakeholder interviews, documentation review, and deep-dive testing and control validation
- Ensure client controls meet legal, regulatory, privacy, policy, standards and security requirements

Red Flags in Job Postings



Red Flags:

- Job Requirements Make No Sense – Overly broad
- Two Or More Jobs In One Description – Lack Clarity
- Experience Requirements Don't Match Job Level
- Ignores Transferable Skills

A **bad job posting** can signal any number of *bad omens for good candidates*:

- This is a firm that doesn't know what they need from this security position
- This is a firm that will overwork me and my security colleagues
- This is a firm that is compliance driven and not committed to security and risk

2024 ISC2 Cybersecurity Workforce Study

Key Findings – 1:

- **Skills shortages*** make it difficult to secure the organization.
- **Lack of budget** is the main cause for talent shortages and skills gaps.
- **Participants' pathways** to enter the cybersecurity workforce are changing, as are their priorities.
- **Diverse backgrounds** can help solve the talent gap.

*Debate “Cybersecurity Skills Shortage”

2024 ISC2 Cybersecurity Workforce Study

Key Findings – 2:

- The expected advancements of **AI** will change the way cyber respondents view their skills shortage.
- Cyber professionals confident **Gen AI** will not replace their role.
- **Gen AI** presents benefits and challenges for cybersecurity.
- Organizations need a **Gen AI** strategy to responsibly implement the technology.

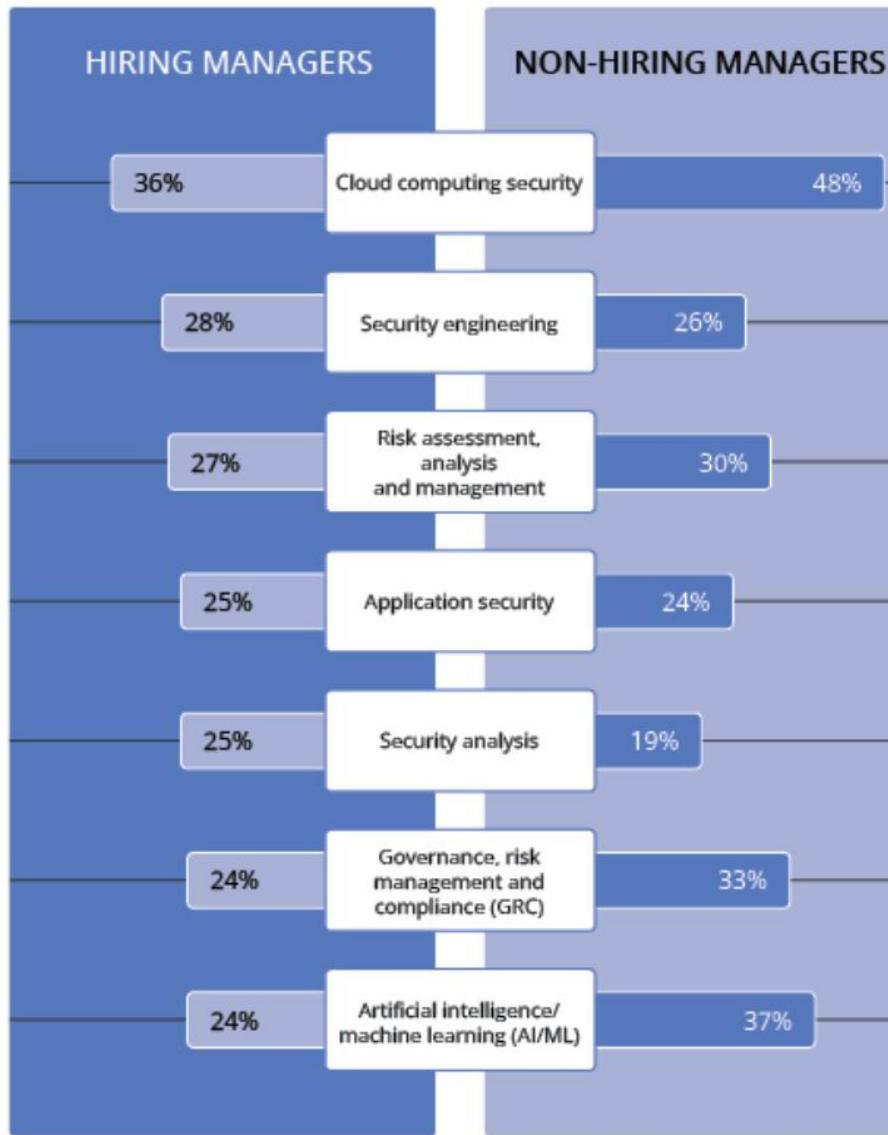
Workforce Insights

Technical Skills*

* Technical is only a part of needed skills

What technical skills are you most looking for right now when hiring?

What technical skills do you think are most in demand for security professionals looking to advance their careers?



Cloud

Security Engineering

Risk Management

Application Security

Security analysis

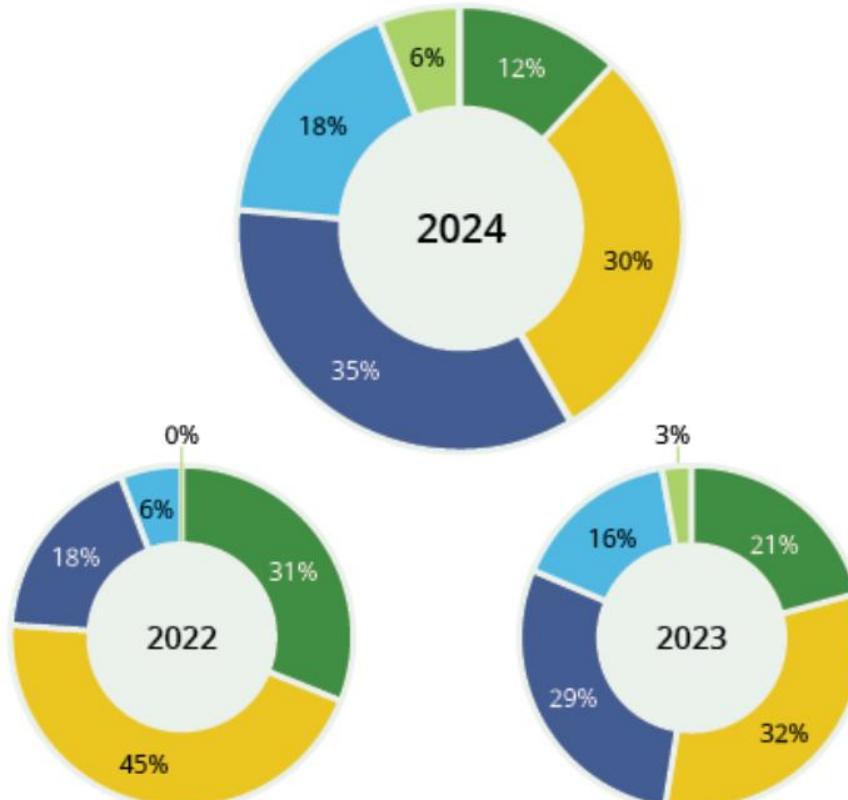
GRC

AI/ML

Workforce Insights

Ages of new entrants into the cybersecurity profession
(i.e., started in cybersecurity within the last year).

Under 30 30-38 39-49 50-59 60 or older



Base: 488 global cybersecurity professionals who started in the past 12 months
695 surveyed in 2023; 356 surveyed in 2022

Note: Individual percentage values may not sum to totals due to rounding.

How valuable have you found each of the following in your career growth in cybersecurity?

Somewhat valuable Valuable Very valuable

Worked in an IT position



Got a cybersecurity certification in cybersecurity



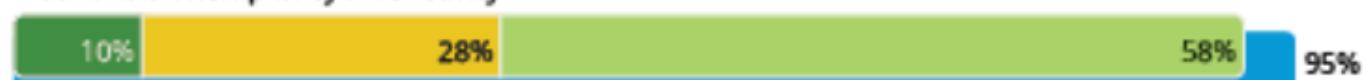
Got a bachelor's or postbaccalaureate degree in cybersecurity or other related field



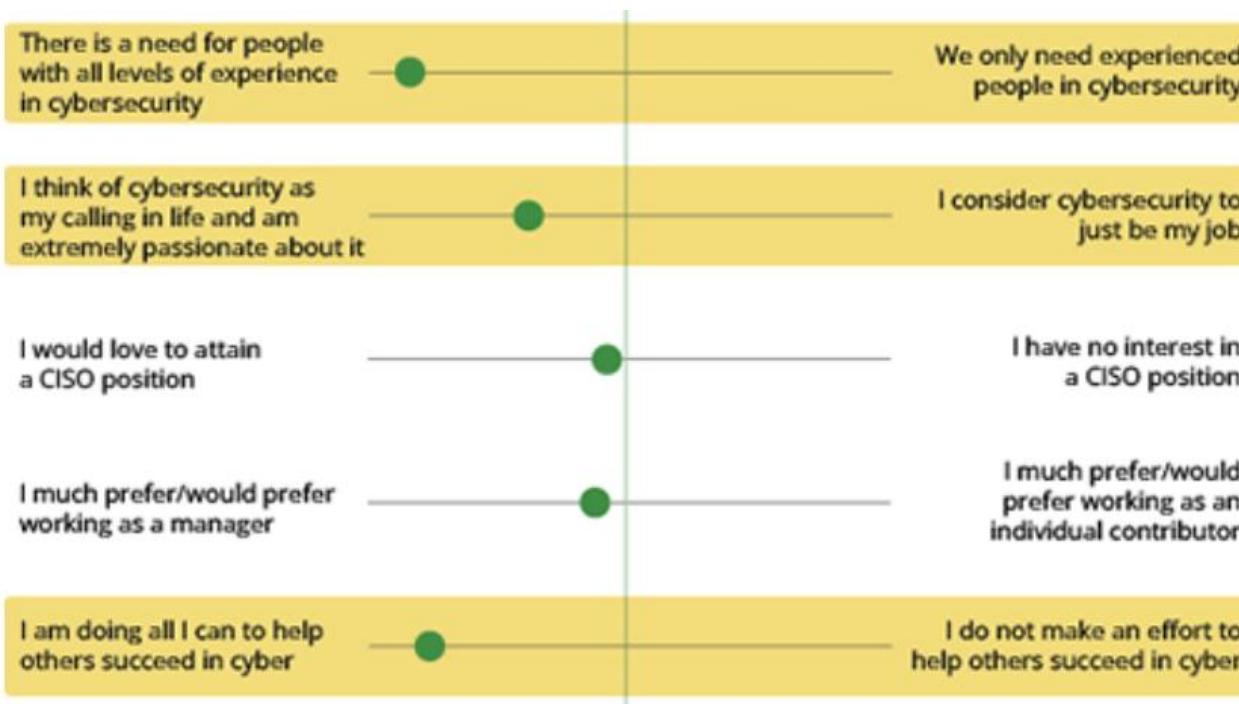
Got an advanced degree in cybersecurity or other related field



Got an internship in cybersecurity



Cybersecurity Career Growth Aspirations and Pathways



Rank the following in terms of how meaningful they would be (or already are) for you to achieve in your cybersecurity career.
(Showing top three responses)



Base: 15,852 global cybersecurity professionals

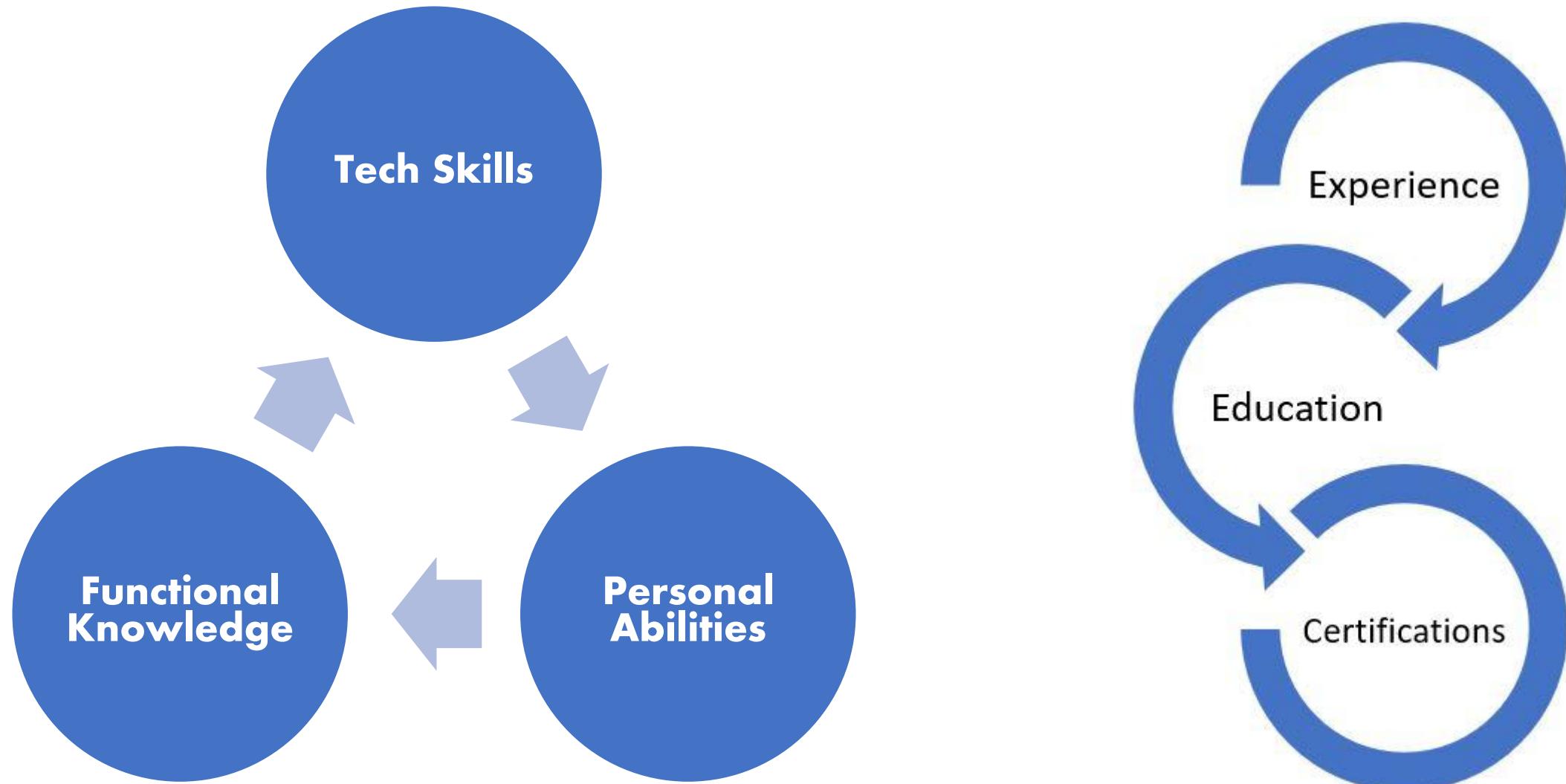
What can we (personally) do?

Hint: Lots

See your career as a continual learning journey

Build a Resilient Mindset

Cyber Career Requirements



Hacking Your Career Tips

- Visualize Your Goals – what do you want to be when you grow up & why?
- Social Engineer Your Next Boss / OSINT
- ABC = Always Be Curious | Actively Learn
- Practice, Practice, Practice
- Be Tenacious
- Be a Hacker



When all else fails

Fake It ‘Til You Make It



Photo by [Sander Sammy](#) on [Unsplash](#)

Ignore ‘Em



Photo by [I have no idea](#)

Building Your Brand

Differentiating yourself

Social Media (primarily LinkedIn)



Ron Woerner

Cybersecurity Trusted Advisor / Consultant || CISO/vCISO/ISSO ||
Keynote Speaker || Author || Educator || Veteran

Omaha Metropolitan Area · [Contact info](#)

[Ron's Social Media links](#)

13,614 followers · 500+ connections

F Forrester



Michigan State University

Personal Website



AARON GROTTHE

OMAHA, NEBRASKA 68106 · (402)305-7265 · AJGROTHE@GMAIL.COM

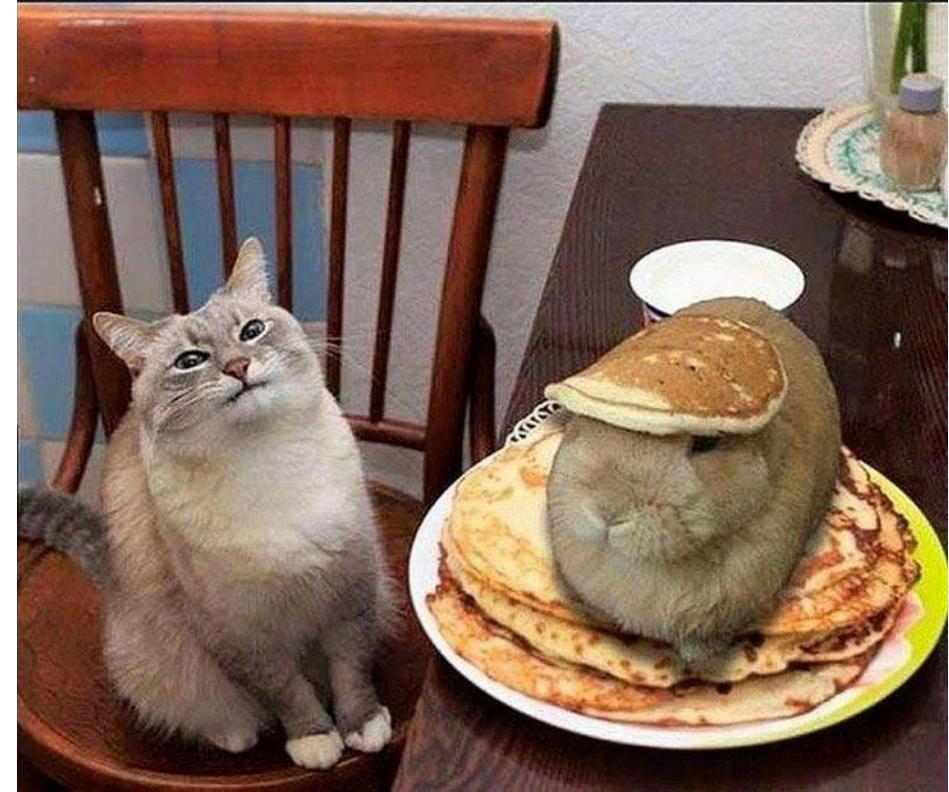
WELCOME TO GROTTHE.US!

Your LinkedIn Profile and Website should match your resume

Building Interpersonal (soft) Skills

- Attitude (preferably good)
- Communications
(Writing & Speaking)
- Teamwork / Collaboration
- Critical Thinking
- Tenacity
- Humor

Excuse me sir. There appears to be a hare in my pancakes.



I'm allowed 1 "Dad" Joke

Improving Communications & Leadership



Tech Talkers Toastmasters Club 6424

Main Menu	Welcome to "Tech Talkers"
<p>Home</p> <p>Meeting Information / Directions</p> <p>Contact Us</p> <p>Club Calendar</p> <p>Public Downloads</p> <p>Free Resources</p> <p>Toastmasters Video</p> <p>About & FAQs</p> <p>Ice Breaker Project</p> <p>Member Meeting Roles</p> <p>Pathways</p> <p><i>For more information on Toastmasters International, visit www.toastmasters.org</i></p>	<p>Welcome to "Tech Talkers"</p> <p>Geeks (and aspiring geeks) welcome! 😊</p> <ul style="list-style-type: none">• We love techies, programmers, sys admins, testers, UX professionals, security professionals, students, and all types of geeks.• We encourage our members to convey their passion.• We help our members present information in a memorable way.• Our members gain the confidence to speak up at work or present at a conference or meetup.• Our members learn from each other as they improve their technical knowledge, communication, and leadership skills. <p>Tech Talkers Toastmasters can help you! Tech Talkers Toastmasters is a club that enables those in technology careers to improve communication and leadership skills in a safe, encouraging, and supportive environment. We're one of the few <u>Toastmasters International</u> clubs focused on technology topics.</p> <p>We meet online via Zoom on the 1st and 3rd Fridays at noon CT. While we're based around Omaha, Nebraska, USA, anyone in tech may join us. No experience required; just a drive to improve yourself and an Internet connection.</p>

Be part of a community - Contribute



<https://cybersecurityomaha.com/>



Omaha - Lincoln

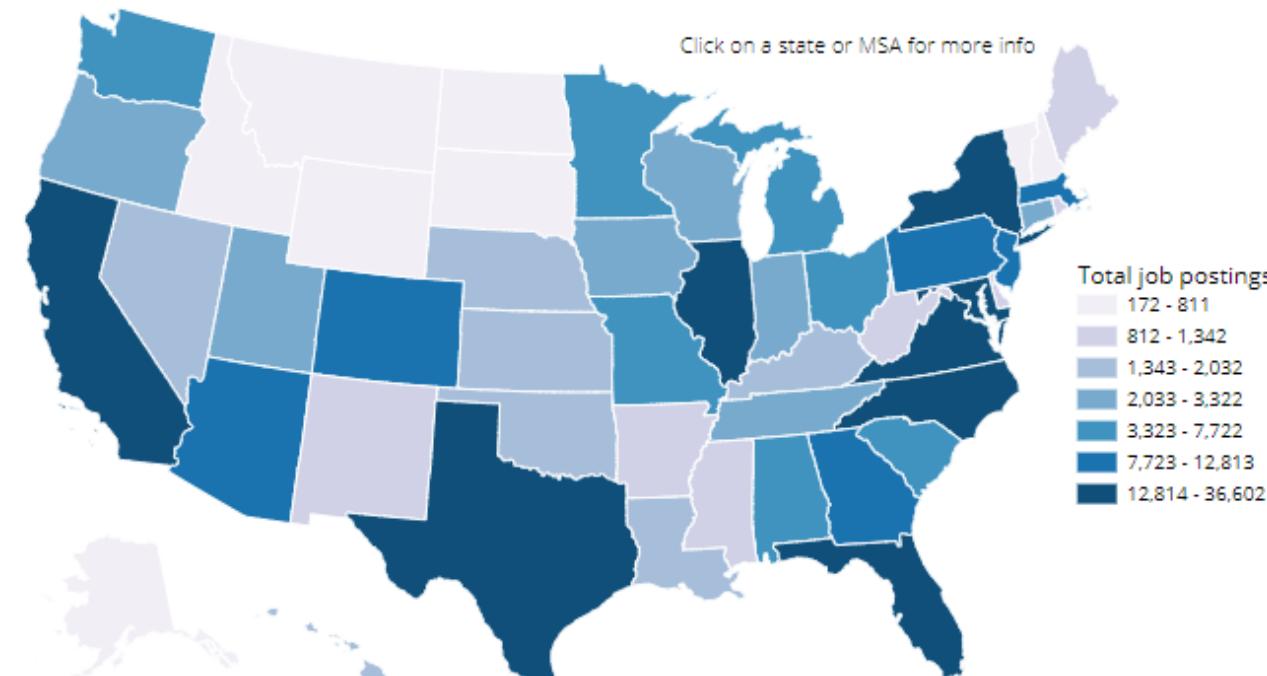


Cybersecurity Supply/Demand Heat Map

Cybersecurity talent gaps exist across the country. Closing these gaps requires detailed knowledge of the cybersecurity workforce in your region. This interactive heat map provides a granular snapshot of demand and supply data for cybersecurity jobs at the state and metro area levels, and can be used to grasp the challenges and opportunities facing your local cybersecurity workforce.

[States](#)[Metro Areas](#)[All Data](#)[Public Sector Data](#)[Private Sector Data](#)[Total job openings](#)

Search State

 [Share](#) [Embed](#)

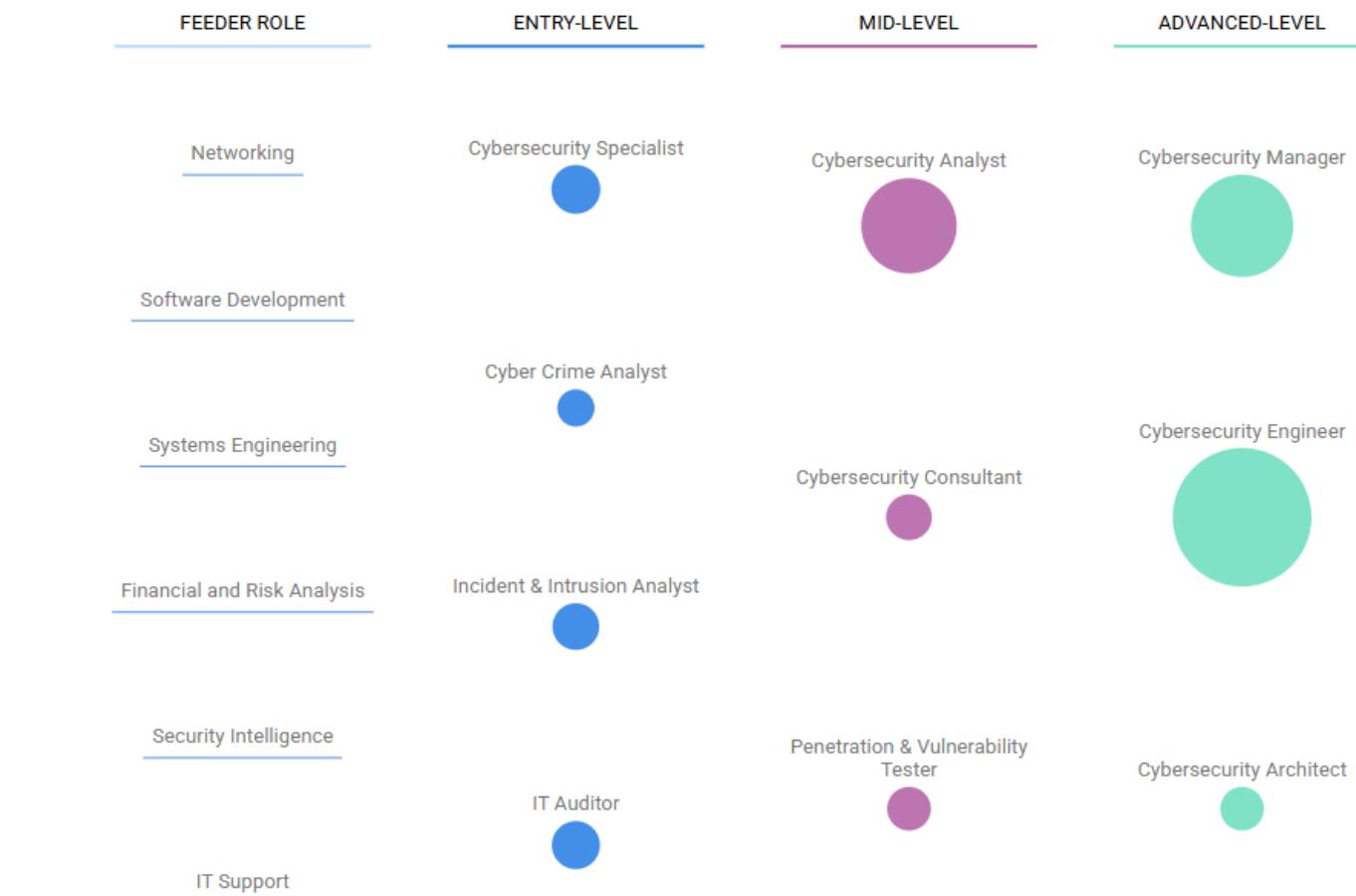
<https://www.cyberseek.org/heatmap.html>



Cybersecurity Career Pathway

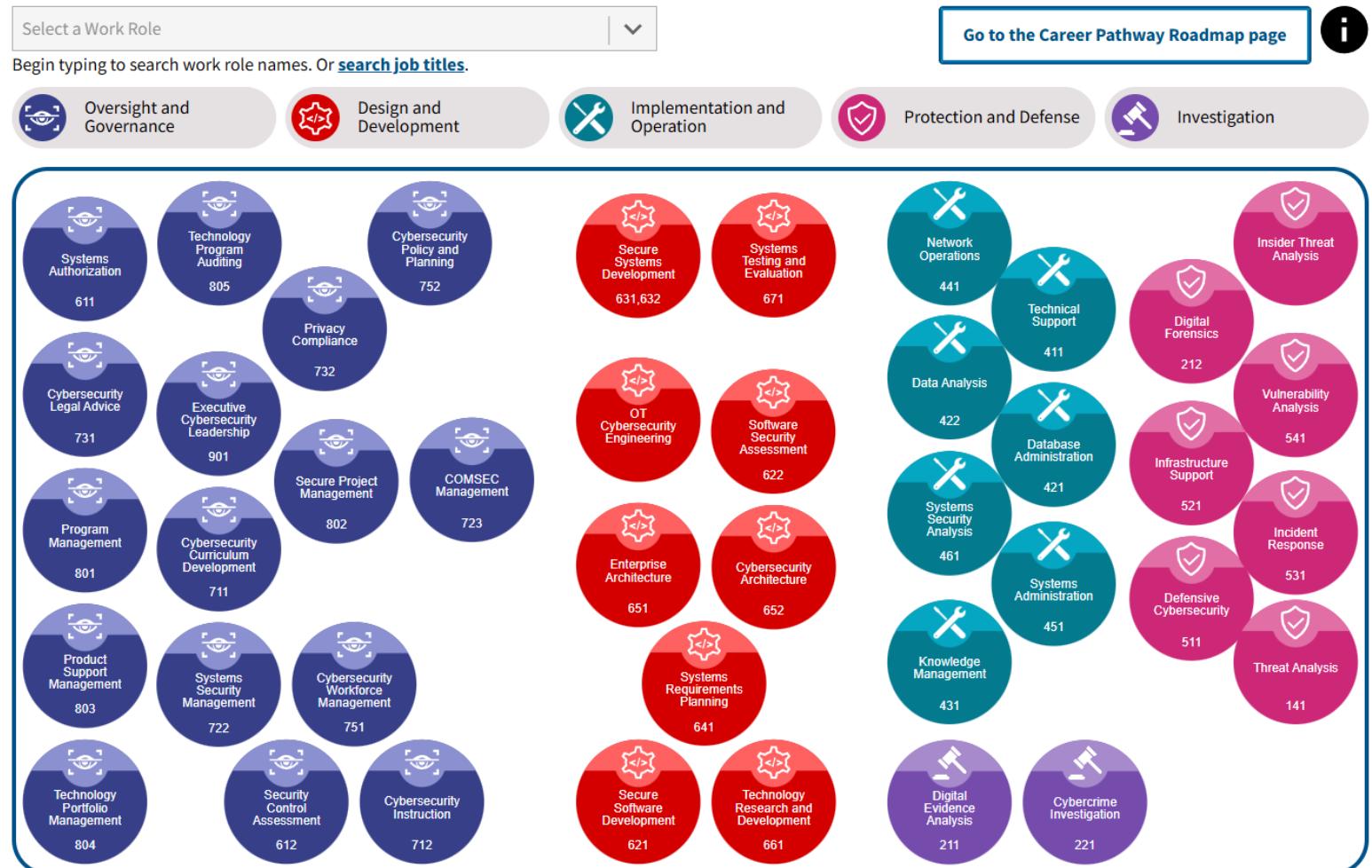
Roles

Skills and Certifications

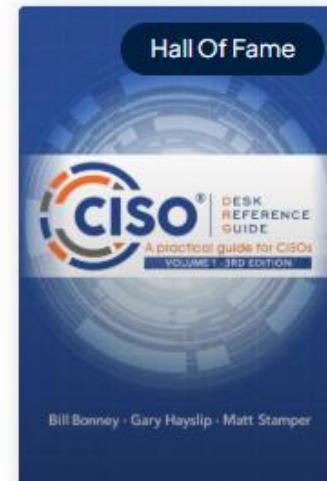
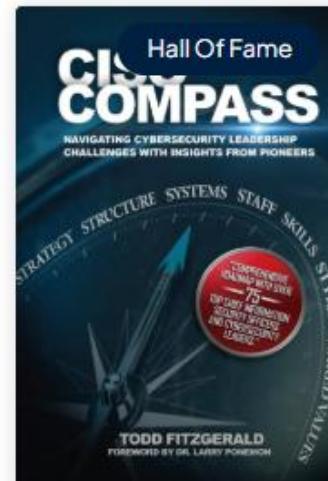
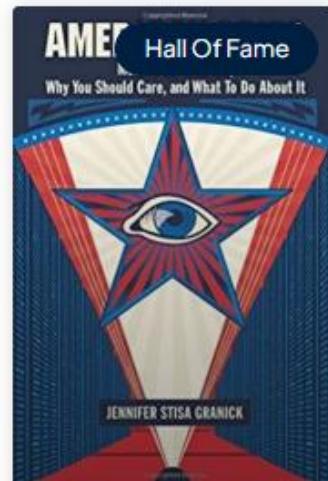
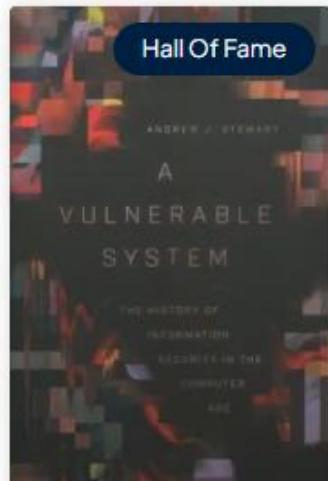
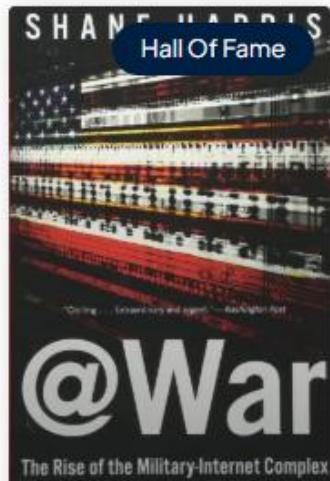
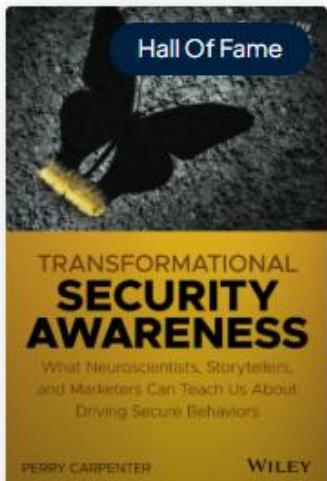


Cyber Careers Pathways Tool

This tool presents a new and interactive way to explore the updated work roles within the [Workforce Framework for Cybersecurity \(NICE Framework\)](#). It also highlights core attributes among each of the Work Roles and offers actionable insights for employers, professionals, and those considering a career in Cyber. To start, select a Work Role below, or enter keywords in the search bar.



For Book Nerds



Transformation
Security
Awareness:...

@War: The
Rise of the
Military-...

A Vulnerable
System: The
History of...

American
Spies: Modern
Surveillance,...

CISO
Compass

CISO Desk
Reference
Guide: A...

Your community for InfoSec's essential reads

In the chaotic flood of cybersecurity content,
CyberCanon is your signal in the noise—the InfoSec
professional's first stop for must-consume wisdom.

RESOURCE CENTER

[Centers for Academic Excellence](#)

[Job Guide](#)

[Veteran's Guide](#)

[Women's Guide](#)

[Internship Guide](#)

[Security Clearance Guide](#)

[Ethical Hacker Guide](#)

[Coding for Cybersecurity Guide](#)

[Cybersecurity 101](#)

[Student Guide to Internet Safety](#)

[Scholarship Guide](#)

[Cybersecurity Math Guide](#)

[Small Business Guide](#)

[Cybersecurity for K-12 Students](#)

[Career Networking Guide](#)

[What is a Cyber Range?](#)

[Code Like a Hacker](#)

[Reacting to a Cyber Incident](#)

[Introduction to Cyber Defense](#)

[Cybersecurity Courses Online](#)

[Recommended Reading](#)

[Phishing Attacks](#)

[Cybersecurity Responsibility](#)

[How to Get Into Cybersecurity](#)

One Ring To Rule Them All...



What is Cybersecurity Guide?

Over the past few years, Cybersecurity Guide has grown into a massive resource referenced by leading cybersecurity professionals in academia, industry, and beyond.

Our global team of professional cybersecurity writers — leaning on their first-hand experience in the field — conduct interviews and deep-dive research, all in an effort to create the most up-to-date and comprehensive cybersecurity education resource available.

<https://cybersecurityguide.org/>



The Obligatory Questions Slide

1 - aka Let's Play “Stump the Professor”

2 - Obligatory AI art

Next Steps / “Apply” Slide – Summary

Immediate (the next week):

- Review and try 2-3 resources from this presentation
- Practice prompting and AI Agent development
- Set a calendar reminder for 1 month...

Long-term (the next 1-2 months):

- Review the deck and Experiment with 2-3 new resources
- Share with others
- Lead with curiosity, defend with wisdom





Thank You!

For More Information:

ronw@cyber-aaa.com

<https://www.linkedin.com/in/ronwoerner/>

<https://linktr.ee/cyberron>



MANIFEST YOUR INNER CYBER SUPERHERO!

