AI Guardians: The Next Frontier in Cybersecurity

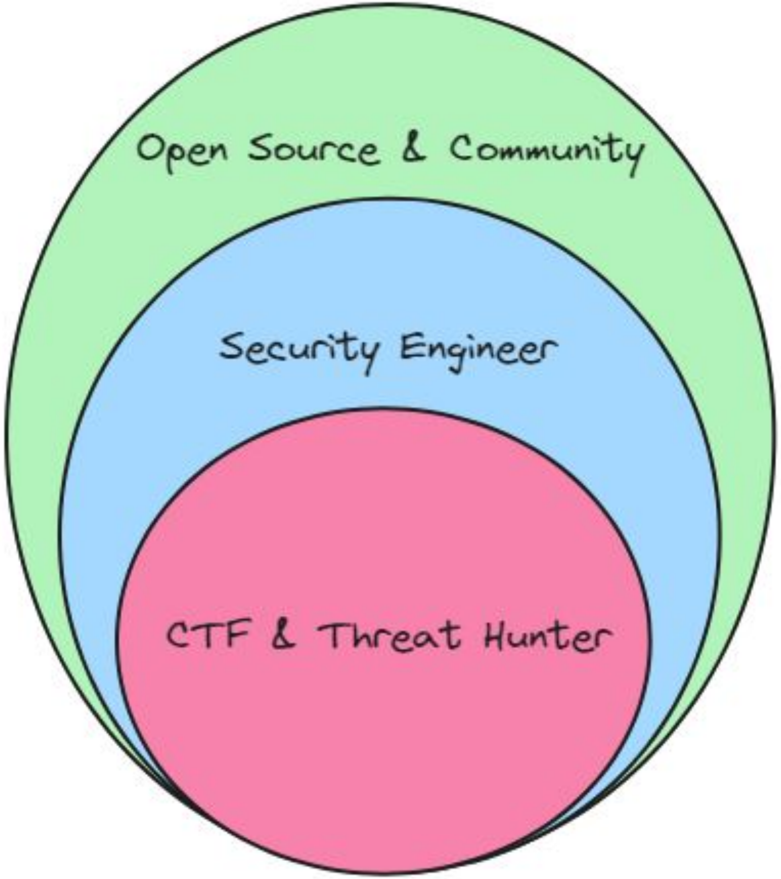
GenAI in Cyber Security

$whoami

Shubhendu Shubham

“sudo rm -rf / problems”

aka Troubleshooter



CTF BADGES

CERTIFICATIONS

SC 100

AZ 305

AZ 104

AZ 500

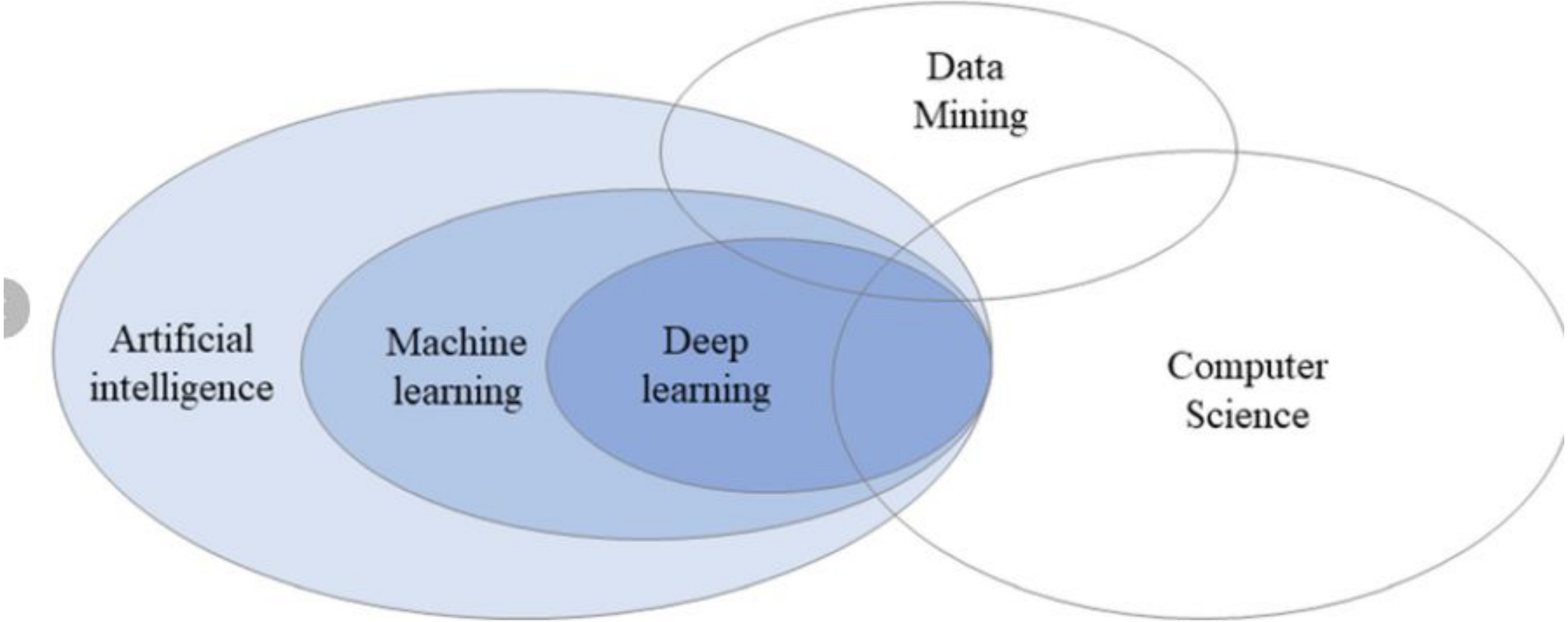
AZ 700

Community

Disclaimer

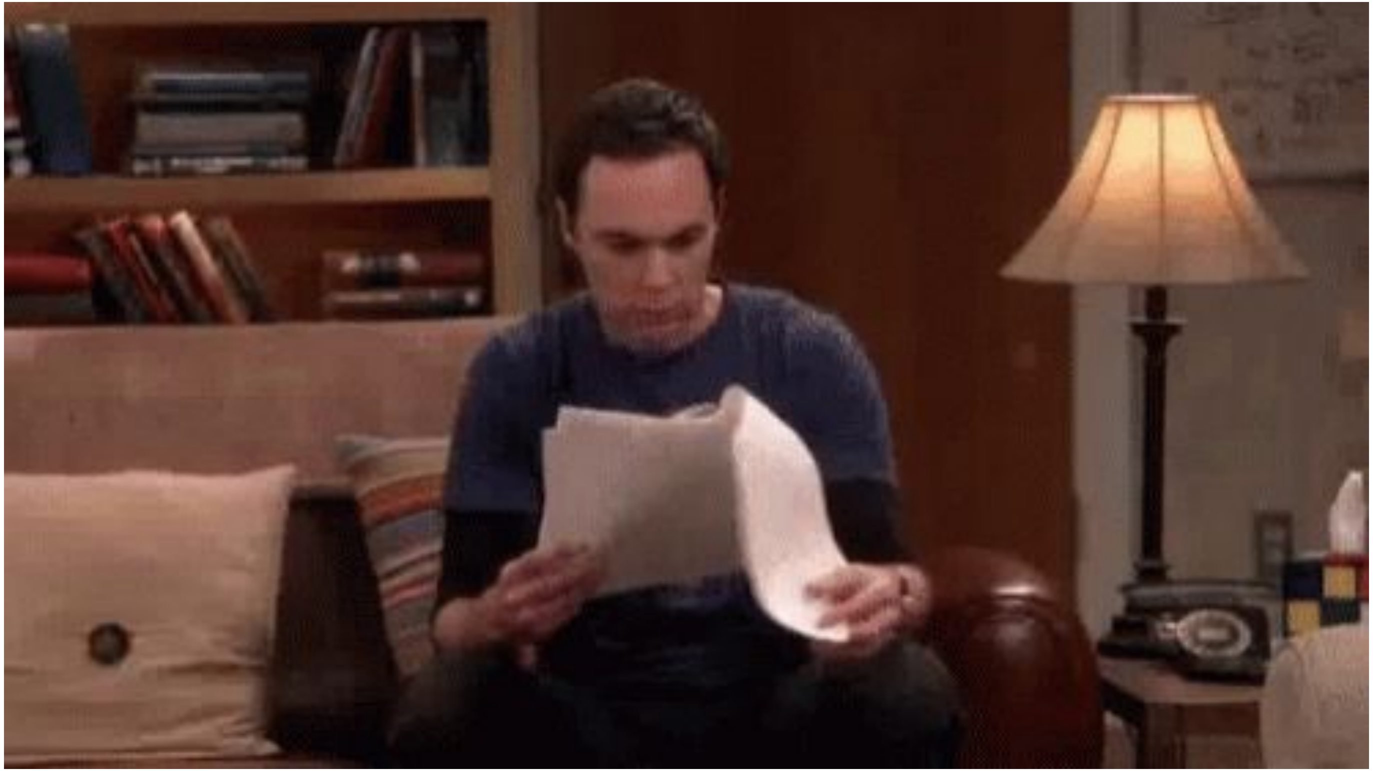
“If you think technology can solve your security problems, then you don’t understand the problems, and you don’t understand the technology.” – Bruce Schneier

Back 2 Basics



Venn diagram representing the relationships between AI, ML and DL. (Adapted from Goodfellow et al. 2016)

Why AI in Cybersecurity?



Statistics

- 4000 Password attacks per second (Year 2023) , 579(Year 2022)
- 72 minutes Median time for an attacker to access your private data if you open a phishing email
- 3.5 million Global shortage of skilled cybersecurity professionals

Source : Microsoft Digital Defence Report 2023

Do we need AI ?

Hints:- With great power comes great responsibility

We need RAI

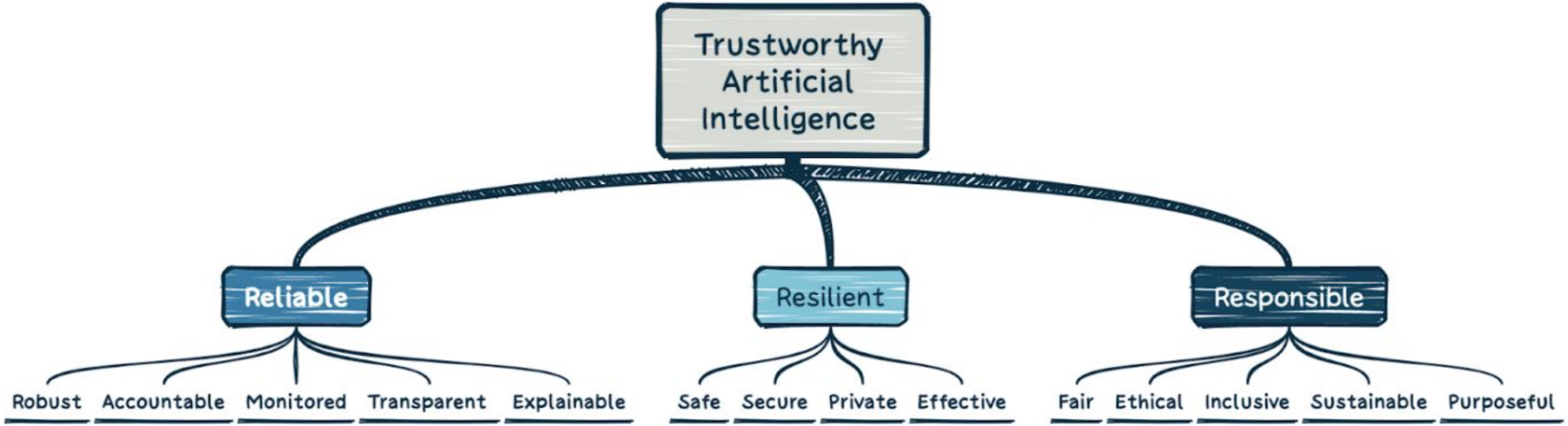


Figure 1.1: Image depicting Pillars of Trustworthy Artificial Intelligence: created from Montreal Ethics Institute Example

LLM Threat Category

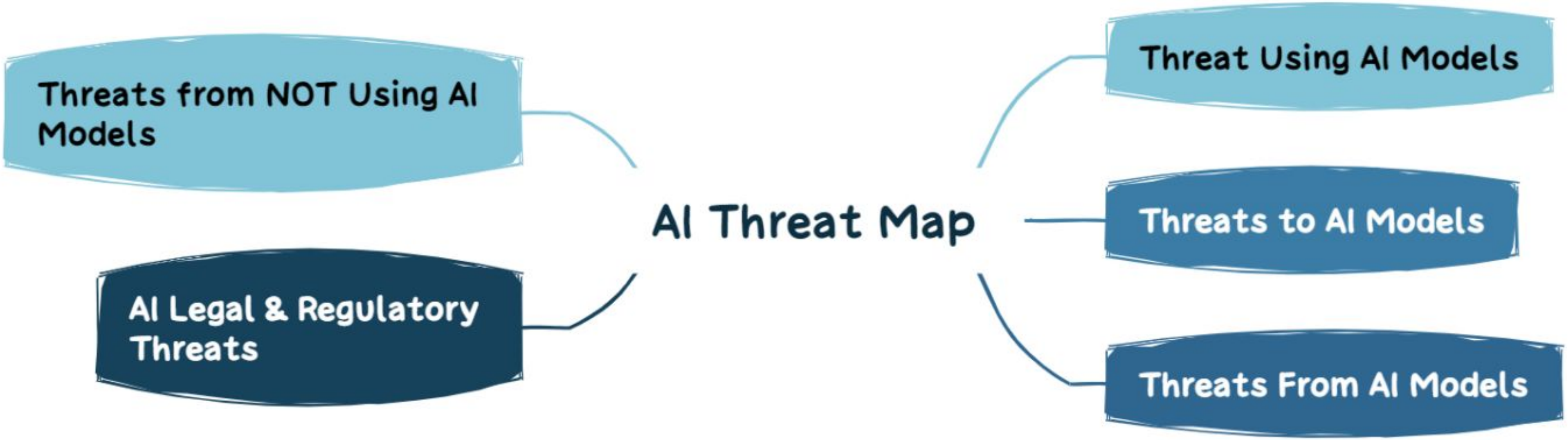
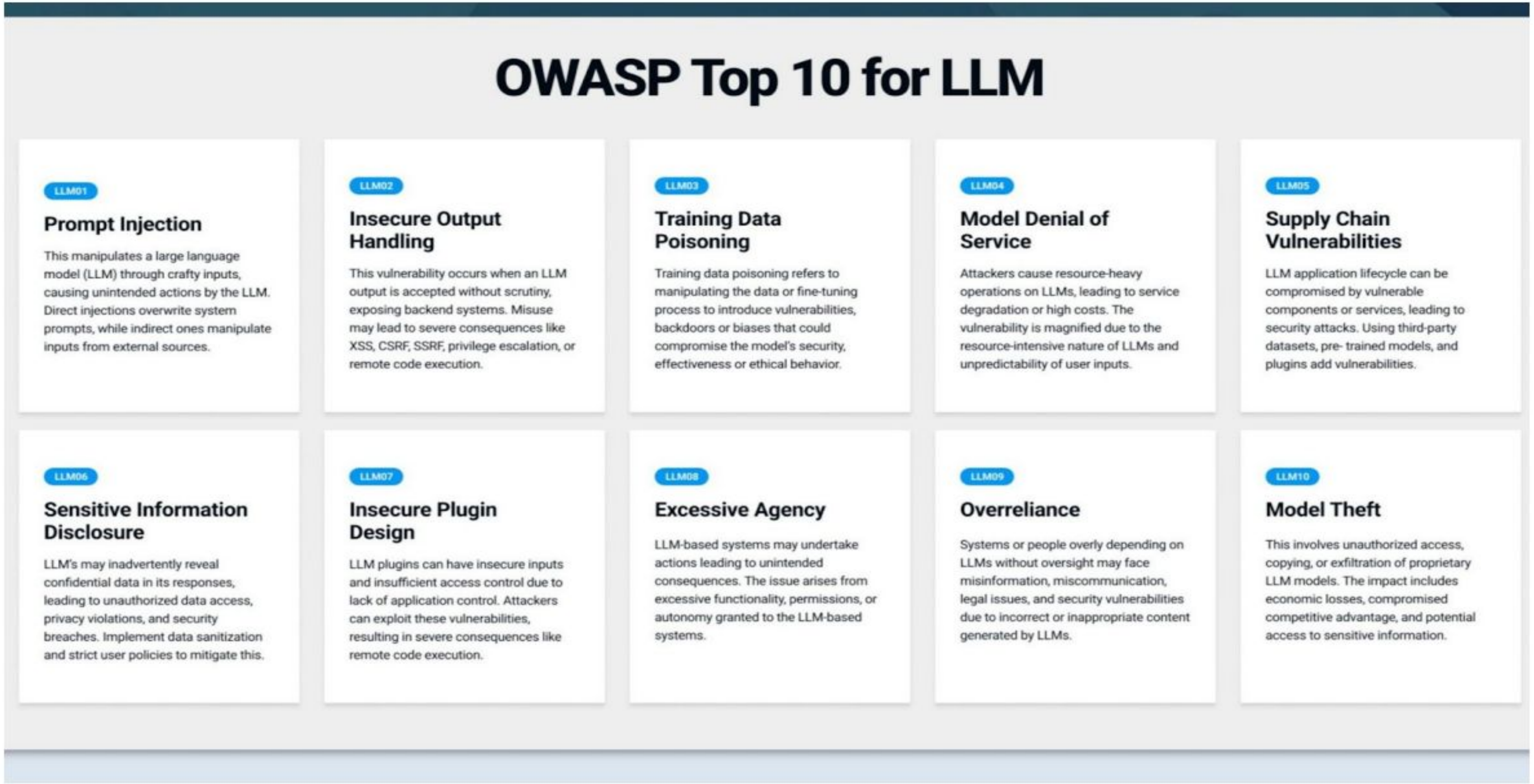


Figure 1.2: Image depicting the types of AI threats: credit sdunn

OWASP Top 10 for Large Language Model Applications



OWASP Top 10 for LLM

LLM01

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

Training Data Poisoning

Training data poisoning refers to manipulating the data or fine-tuning process to introduce vulnerabilities, backdoors or biases that could compromise the model's security, effectiveness or ethical behavior.

LLM04

Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05

Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre- trained models, and plugins add vulnerabilities.

LLM06

Sensitive Information Disclosure

LLM's may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. Implement data sanitization and strict user policies to mitigate this.

LLM07

Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control due to lack of application control. Attackers can exploit these vulnerabilities, resulting in severe consequences like remote code execution.

LLM08

Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09

Overreliance

Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

LLM10

Model Theft

This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

Debunk Cybersecurity AI Myths

Myth 1 : Unauthorised Data Access

Answer: No.

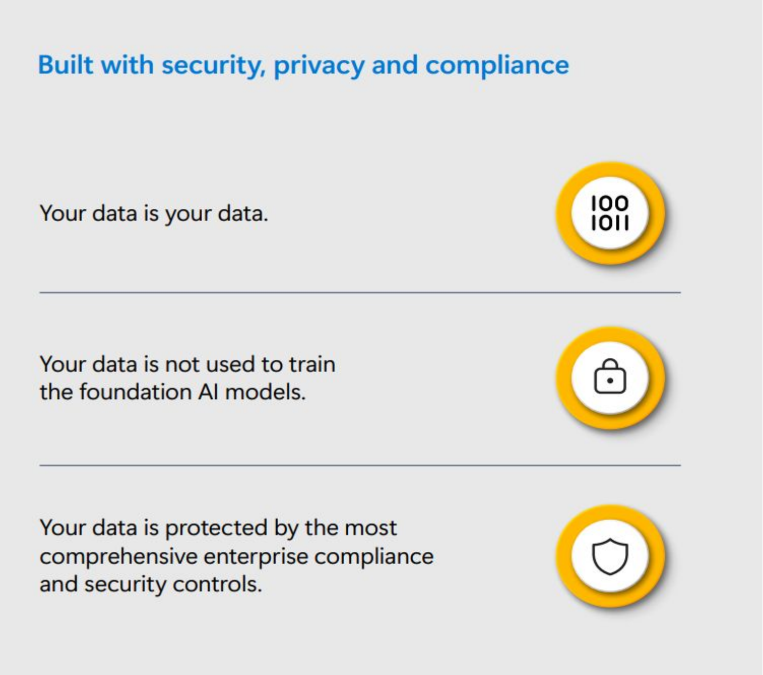
This won’t happen with Copilot because it uses ‘admin on behalf of’ rights for the user logged in. This means the rights are limited to that specific user and that user only. Copilot runs queries as the user, so it never has elevated privileges beyond what the user has.

Myth 2 : Data Privacy and Ownership

Will my customer data be used to train language models in Copilot?

Answer: No

When it comes to data, unlike ChatGPT, Copilot is grounded in the unique context of your organisation. That means when you ask Copilot any question, the answer will be based on what's happening in your organisation at that moment. Your data isn't used to train the foundation AI models. It's a closed learning loop that continuously improves based on your use



Built with security, privacy and compliance

Your data is your data.

Your data is not used to train the foundation AI models.

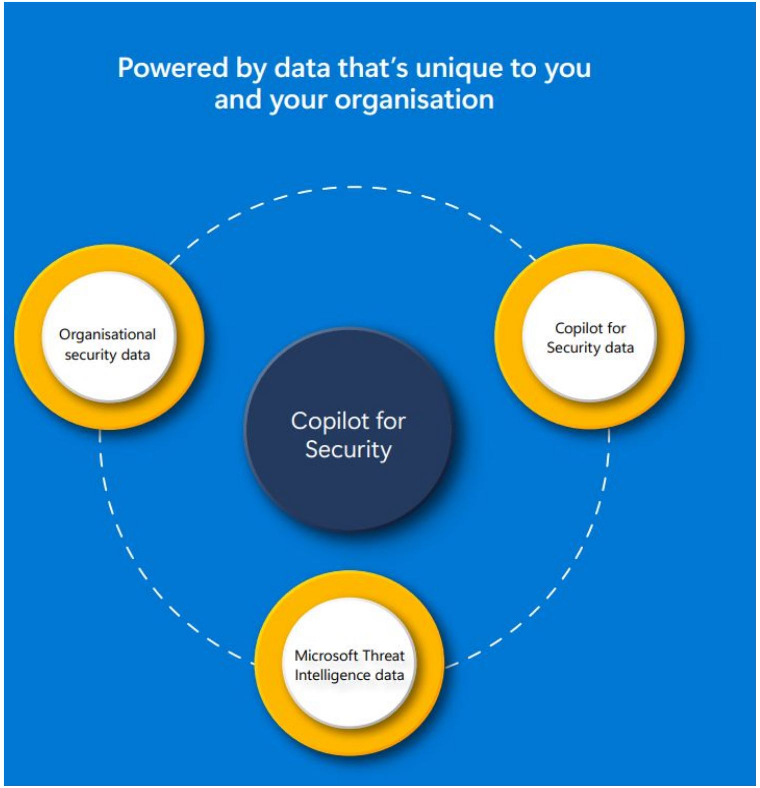
Your data is protected by the most comprehensive enterprise compliance and security controls.

Is transferred data protected from unauthorised access?

Answer: Yes.

When using Copilot for Security, your data:

- Is your data.
- Is stored where you choose and always encrypted at rest.
- Isn't used for sales or shared with third parties.
- Is housed in systems governed by Microsoft SOC and International Organisation for Standardisation-certified processes.
- Isn't used to train foundation AI models.
- Is never shared with OpenAI.
- Is protected by the most comprehensive enterprise compliance and security controls


Powered by data that's unique to you and your organisation
Organisational security data
Copilot for Security
Copilot for Security data
Microsoft Threat Intelligence data

Myth 5 : Hallucinations

Does Copilot for Security help detect hallucinations?

Answer : Yes

Trust is paramount in security. If you can't trust security data and insights, you can't achieve the right outcomes. For humans to confidently work with AI-powered tools such as Copilot, it's critical to build trust in the technology.

Microsoft, is committed to responsible AI, which is why Copilot is designed to:

- Show reasoning, sources, debug and runtime.
- Ensure data is compliant, secure and private.
- Address harms and hallucinations.
- Be transparent and allow for an open dialogue

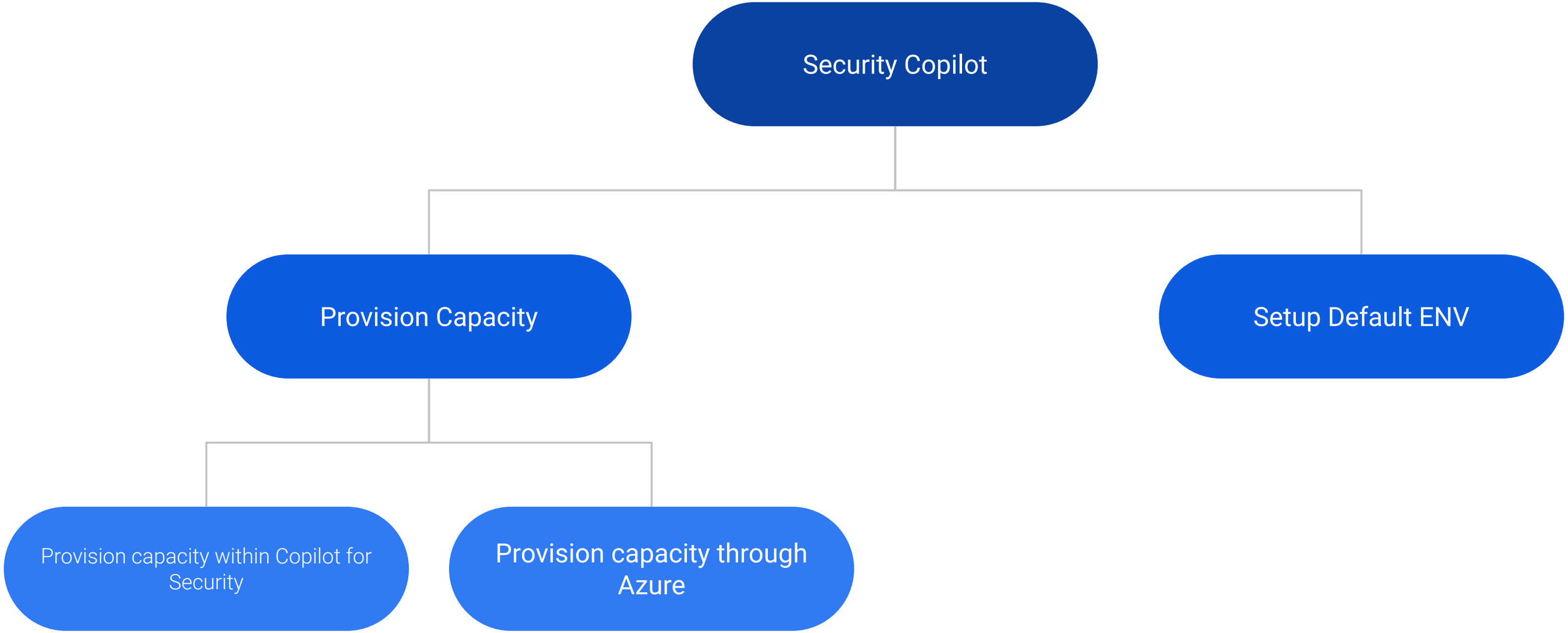
Myth 4 : Compliance Issues

Does Copilot for Security meet industry or regional compliance requirements?

Answer: Yes

Copilot meets General Data Protection Regulation (GDPR) requirements for EU markets by implementing the Azure Public Preview requirements. It stores all EU customer data within the EU Data Boundary and is available in multiple languages. Copilot also provides compliance controls to help you meet business and regulatory requirements

Onboarding

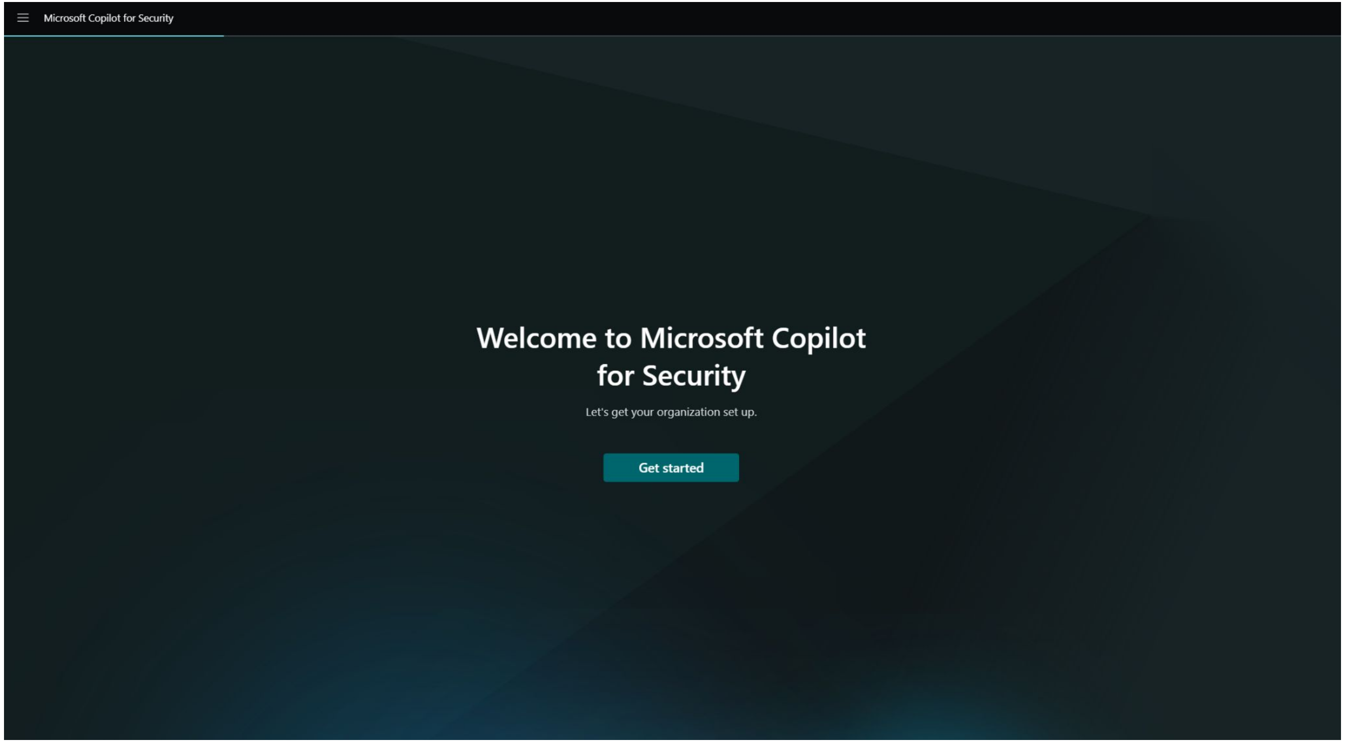

Security Copilot
Provision Capacity
Setup Default ENV
Provision capacity within Copilot for Security
Provision capacity through Azure

Demo

Prerequisite

- Azure Subscription
- Azure subscription owner or contributor to create capacity

1. Sign in to Copilot for Security (https://securitycopilot.microsoft.com).
2. Select Get started.

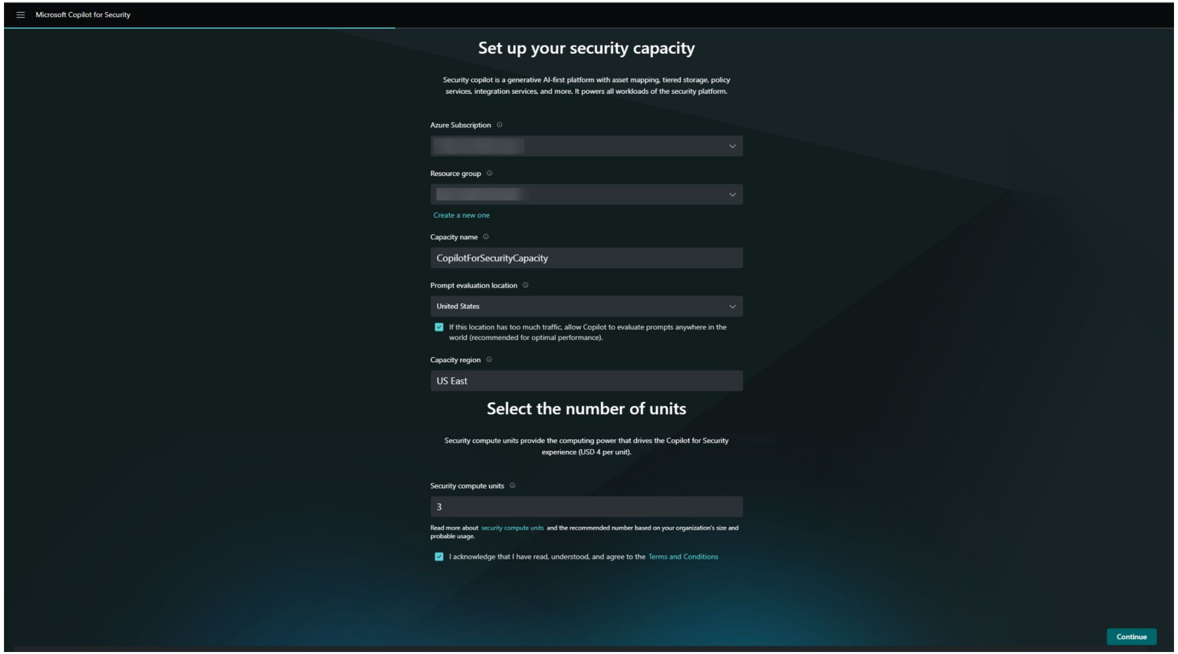


Welcome to Microsoft Copilot for Security

Let's get your organization set up.

3.Set up your security capacity:

Select the Azure subscription, associate capacity to a resource group, add a name to the capacity, select the prompt evaluation location, and specify the number of Security Compute Units (SCUs). Data is always stored in your home tenant geo.



Set up your security capacity

Security copilot is a generative AI-first platform with asset mapping, tiered storage, policy services, integration services, and more. It powers all workloads of the security platform.

Azure Subscription: [redacted]

Resource group: [redacted]

Create a new one

Capacity name: CopilotForSecurityCapacity

Prompt evaluation location: United States

☑ If this location has too much traffic, allow Copilot to evaluate prompts anywhere in the world (recommended for optimal performance).

Capacity region: US East

Select the number of units

Security compute units provide the computing power that drives the Copilot for Security experience (USD 4 per unit).

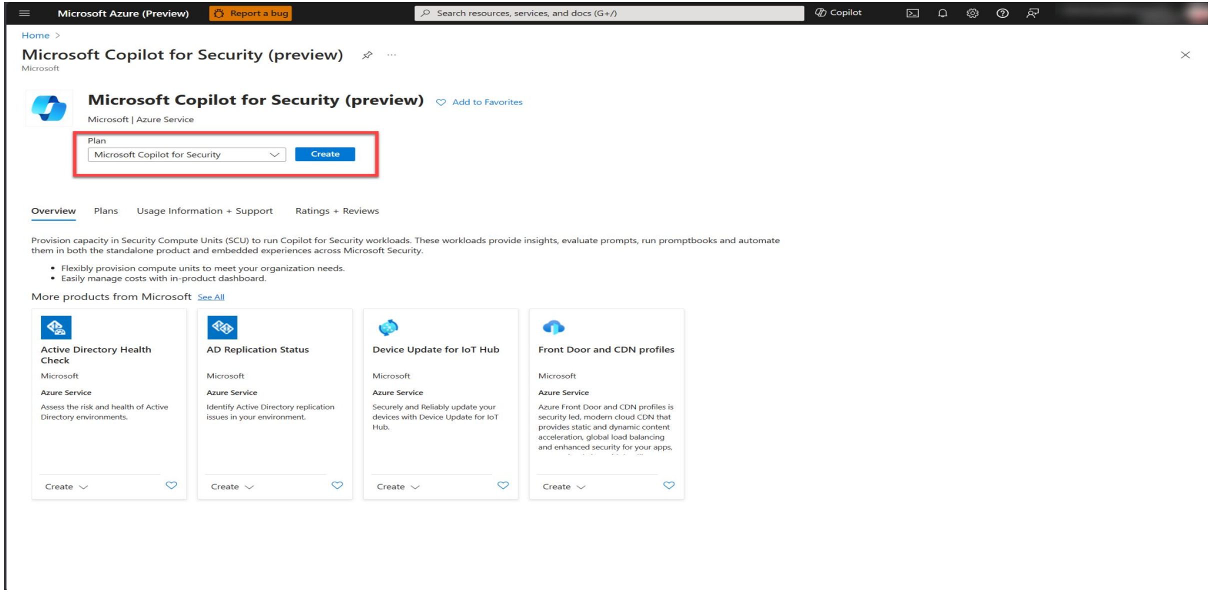
Security compute units: 3

Read more about security compute units and the recommended number based on your organization's size and probable usage.

☑ I acknowledge that I have read, understood, and agree to the Terms and Conditions

Provision capacity in Azure

1. Sign in to the Azure portal.
2. Search for Copilot for Security in the list of services, then select Copilot for Security.
3. Select Resource groups.
4. Under Plan, select Microsoft Copilot for Security. Then select Create.



Microsoft Azure (Preview)
Report a bug
Search resources, services, and docs (G+/)
Copilot
Home >
Microsoft Copilot for Security (preview)
Microsoft
Microsoft Copilot for Security (preview)
Add to Favorites
Microsoft | Azure Service
Create
Overview
Plans
Usage Information + Support
Ratings + Reviews
More products from Microsoft
See All
Active Directory Health Check
Microsoft
Azure Service
AD Replication Status
Microsoft
Azure Service
Device Update for IoT Hub
Microsoft
Azure Service
Front Door and CDN profiles
Microsoft
Azure Service
Create
Create
Create
Create

Plan: Microsoft Copilot for Security

Provision capacity in Security Compute Units (SCU) to run Copilot for Security workloads. These workloads provide insights, evaluate prompts, run promptbooks and automate them in both the standalone product and embedded experiences across Microsoft Security.

- Flexibly provision compute units to meet your organization needs.
- Easily manage costs with in-product dashboard.

Assess the risk and health of Active Directory environments.

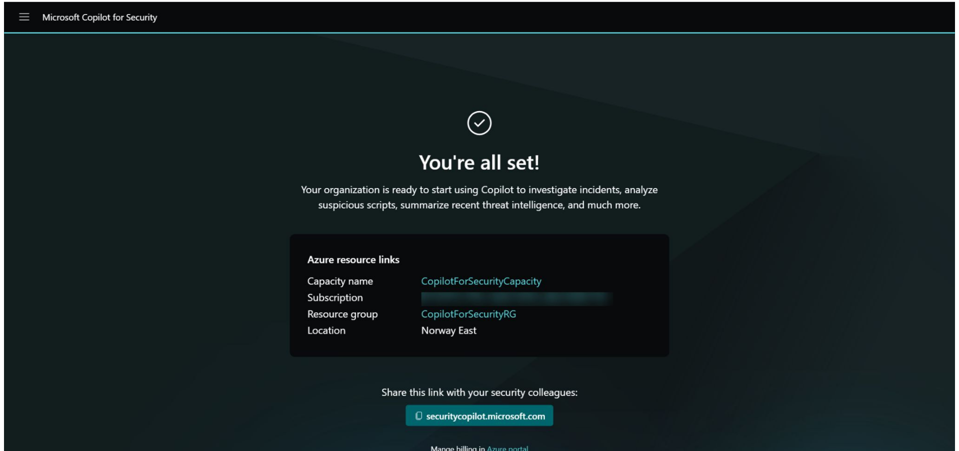
Identify Active Directory replication issues in your environment.

Securely and Reliably update your devices with Device Update for IoT Hub.

Azure Front Door and CDN profiles is security led, modern cloud CDN that provides static and dynamic content acceleration, global load balancing and enhanced security for your apps,

Set up Default environment

1. Associate your capacity to the Copilot for Security environment if the capacity was created in the Azure portal.
2. You're informed where your Customer Data will be stored. Select Continue.
3. Select among the data sharing options. Select Continue. For more information on data sharing, see Privacy and data security.
4. You'll be informed of the default roles that can access Copilot for Security. Select Continue.
5. A confirmation page is displayed. Select Finish



You're all set!

Your organization is ready to start using Copilot to investigate incidents, analyze suspicious scripts, summarize recent threat intelligence, and much more.

Azure resource links

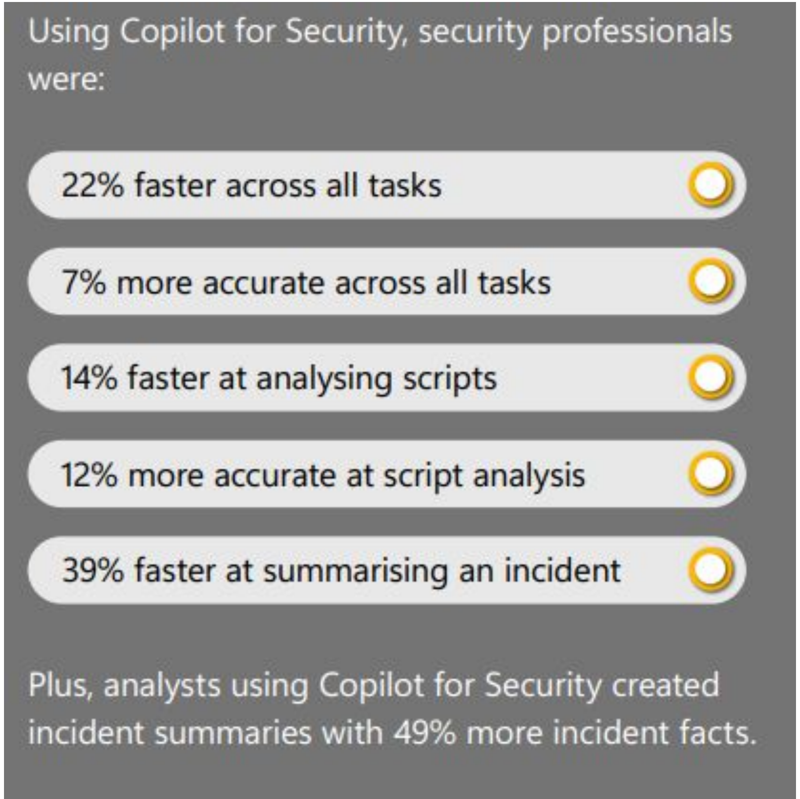
Capacity name	CopilotForSecurityCapacity
Subscription	[redacted]
Resource group	CopilotForSecurityRG
Location	Norway East

Share this link with your security colleagues:

securitycopilot.microsoft.com

Mange billing in Azure portal

Give your security team an edge with Industry-leading Generative AI



Using Copilot for Security, security professionals were:

- 22% faster across all tasks
- 7% more accurate across all tasks
- 14% faster at analysing scripts
- 12% more accurate at script analysis
- 39% faster at summarising an incident

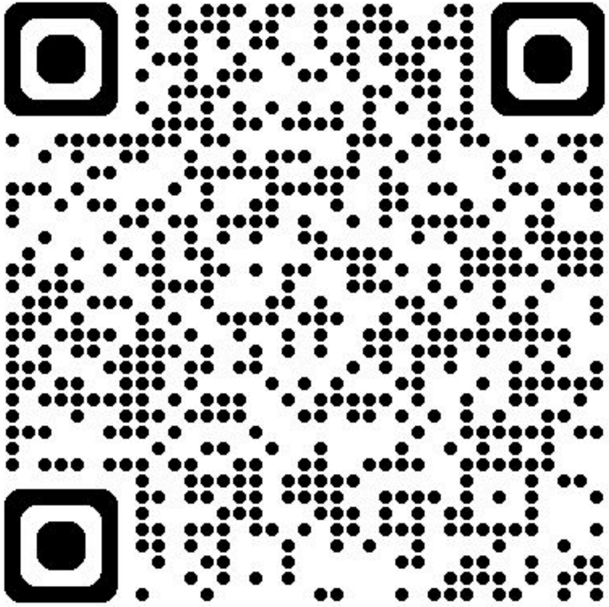
Plus, analysts using Copilot for Security created incident summaries with 49% more incident facts.

Wanna become Security Copilot Ninja

How to Become a Microsoft Copilot for Security Ninja: The Complete Level 400 Training

What’s Next ?

This is not Phishing QR ,It’s my LinkedIn **Don’t Trust** Always verify



References

https://www.microsoft.com/en-us/security/business/ai-machine-learning/microsoft-copilot-security?msockid=1983daf653f86c1736dcc9ce522a6df0

https://www.microsoft.com/en-us/security

https://owasp.org/www-project-top-10-for-large-language-model-applications/

Thank you!