AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

INTRODUCTION TO AI AND CYBER SECURITY

Experiences and Lessons Learned

Eric Eifert
Research Engineer
Center for Digital Safety & Security
Eric.Eifert@ait.ac.at

AGENDA

- Introduction
- Artificial Intelligence and Cyber Security
- Current use of AI in Cyber Security
- Future use of AI in Cyber Security
- Potential applications in Nuclear
- Risks and concerns
- Recommendations

INTRODUCTION – ERIC EIFERT

AUSTRIAN INSTITUTE OF TECHNOLOGY

28 YEARS EXPERIENCE IN CYBER SECURITY

- Former US Air Force Special Agent investigated cyber crimes, computer intrusions, cyber espionage, and cyber counterintelligence
- Built and operated computer forensic labs, security operations centers, insider threat programs, incident response teams, and VAPT programs
- Mature critical infrastructures cyber programs

PROFESSOR AND INSTRUCTOR

- Developed and taught graduate cyber investigations course for George Mason University
- Developed and teach Cyber Security course in the Nuclear Industry globally
- Developed and teach Insider Threats, Incident Response, and Forensics

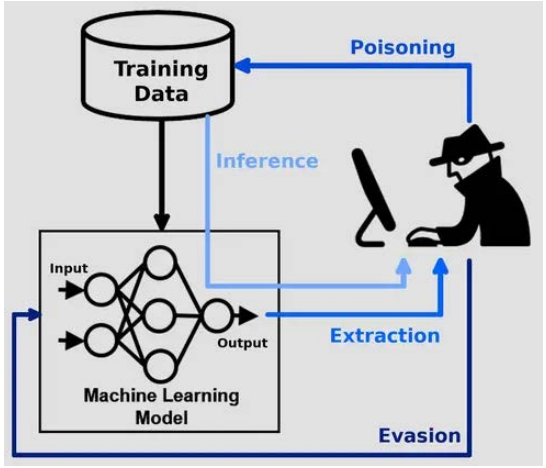
SECURITY RESEARCHER

- Support research projects in digital twins, Industrial Control Systems/Operational technology (ICS/OT) cyber security, next generation security operation centers, and insider threat detection.

ARTIFICIAL INTELLIGENCE AND CYBER SECURITY

Cyber Security for AI

- Security policies and procedures for AI
- Governance framework for AI
- Securing the underlying AI training data
- Secure the model development
- Secure the usage of the AI models



https://research.ibm.com/projects/adversarial-robustness-toolbox

AI for Defensive Cyber

- Threat Detection and Analysis
- Automated Response and Remediation
- Adaptive Defense Systems
- Education and training cyber security staff
- Threat Intelligence
- Report writing

https://www.cyberdb.co/reason-for-using-artificial-intelligence-in-cyber-security

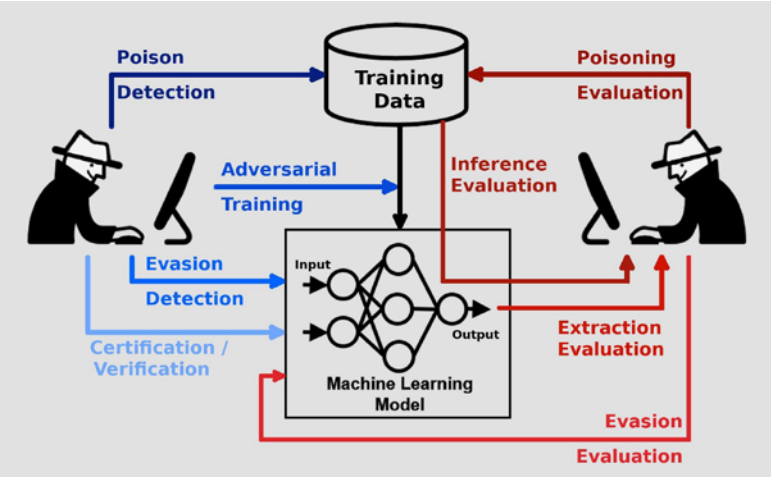
AI for Offensive Cyber

- Automated Attack Tools
- Social Engineering and Deception
- Deep fakes and information operations
- Weaponized AI-powered malware
- Vulnerability research and zero-day vulnerability discovery

https://www.business-reporter.co.uk/ai--automation/the-ai-safety-summit-regulating-the-talking-shop

CYBER SECURITY FOR AI

- Understand the Attack Surface across the AI Life Cycle
- Privacy and sensitive data utilization
  - Where is data being sent for LLMs
- Open Source tool like Adversarial Robustness Toolbox (ART) help to make AI Systems more secure



Generative AI Attack Surfaces Across the AI Life Cycle

AI TRiSM Technology — Content Anomaly Detection — Data Protection — Application Security

Life Cycle	Development and Deployment			Runtime
	Initial Steps (e.g., Collect data for training)	**Development** (e.g., Model training)	**Deployment** (e.g., Prompt services)	**Run** (e.g., model fine-tuning adding plug-ins)

Attack Surfaces[a]	
	Training Data (e.g., Poisoning)
	Prompts (input and output)
	Prompt Integration (RAG, engineering)
	Runtime Data ("plug-ins")
	Orchestration (application code)
	Model Integrations (APIs)
	Model Attacks

IT Supply Chain	
	Data Stores
	Third-party Models
	Code and Libraries

Source: Gartner

[a] Main sample attack surfaces only; others not shown

796422_C

CURRENT USE OF AI IN CYBER SECURITY

AUSTRIAN INSTITUTE OF TECHNOLOGY

Defensive Cyber

ChatGPT Integration

- Summarize and extract contextual meaning from large data sets
- Natural language to computer language – example SQL queries
- Secure coding and insecure code detection
- Explaining technical details and how to perform actions
- Report writing and data presentation

Cyber Threat Intelligence

- Analyze unstructured data feeds for relevant threat intelligence
- Threat Hunting Assistance
- Intelligence sharing and integration

Adversarial Cyber

Traditional Attacks

- Impersonation and spear phishing attacks
- More effective ransomware and cyber extorsion attacks
- Misinformation, deep fakes, and attacks on data integrity
- Distributed Denial of Service Attacks

Enablement

- Automation of reconnaissance and targeting activities
- Malware development to avoid detection
- Speed of monetizing stolen data/information

FUTURE USE OF AI IN CYBER SECURITY

AUSTRIAN INSTITUTE OF TECHNOLOGY

Defensive Cyber

Automation

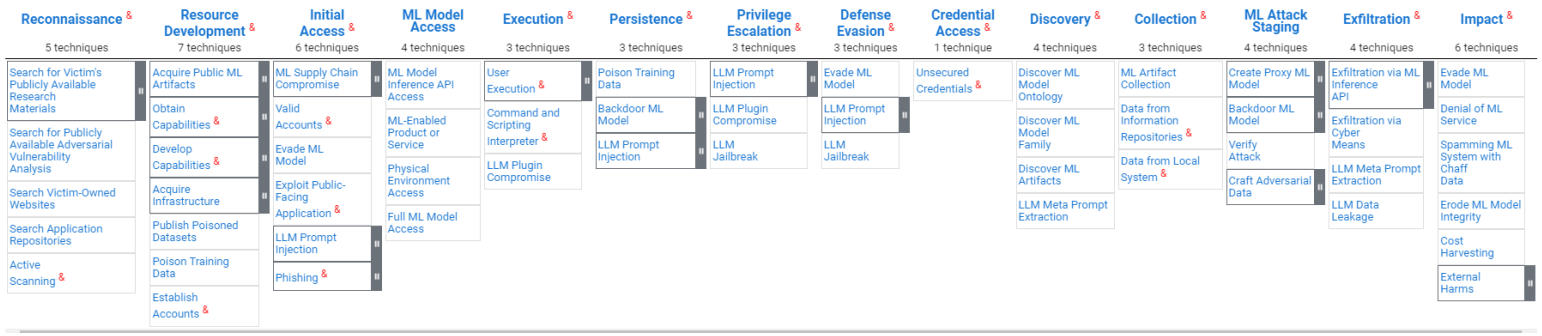
- Assist SOC analysts in detecting cyber events in real time
- Automation of routine manual tasks
- Workflow development for incident response
- Automated response actions
- Automating security patching
- Automated governance, risk, and compliance audits

Analysis

- Behavioral Analytics
- Actionable Threat Intelligence
- Predictive Analytics

Adversarial Cyber

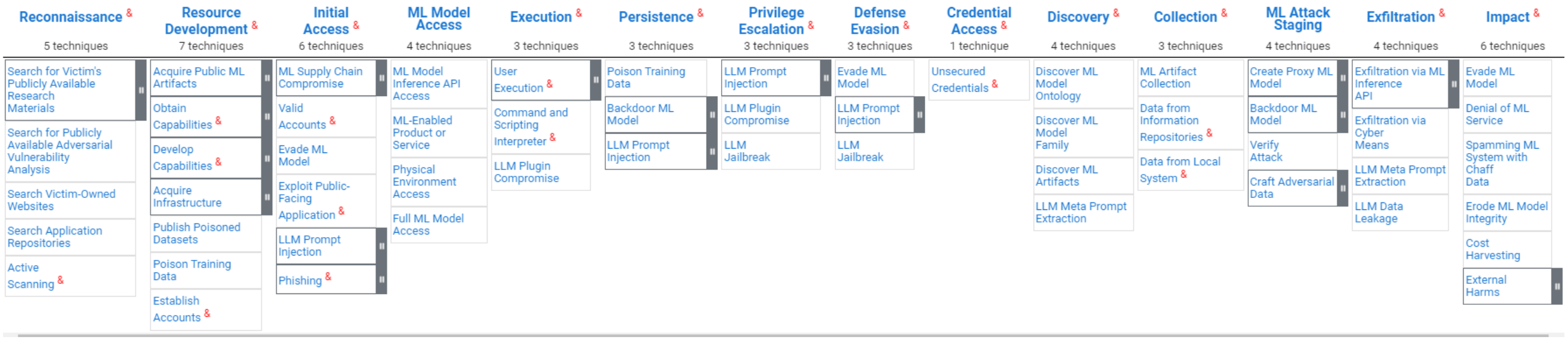
- Vulnerability research allowing the identification of zero day vulnerabilities and subsequent exploits
- Phishing campaigns going global with enhanced language translation features
- Automation of attacks throughout the Cyber Kill Chain
- Cyber attacks against AI Systems



https://atlas.mitre.org/

ADVERSARIAL THREAT LANDSCAPE FOR ARTIFICIAL-INTELLIGENCE SYSTEMS (ATLAS)

AUSTRIAN INSTITUTE OF TECHNOLOGY

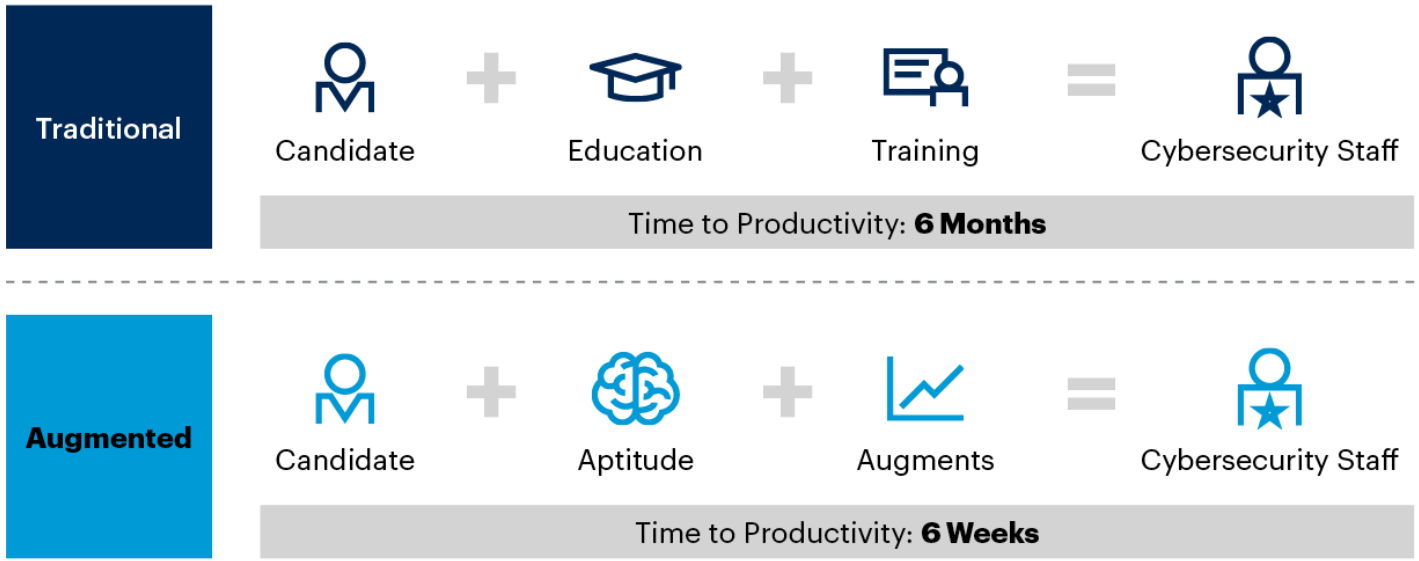


https://atlas.mitre.org/

POTENTIAL USE IN THE NUCLEAR SECTOR

Cyber Security Training and Skills Development

Illustration of a Generative Augment Workflow vs. Traditional
Workflows for Knowledge Worker Onboarding



Source: Gartner
800663_C

Enhancements to Cyber Security Operations Center

- Supplement understaffed security teams
  - Assist in anomaly detection
  - Report writing
- Actionable Intelligence

Secure Code Review

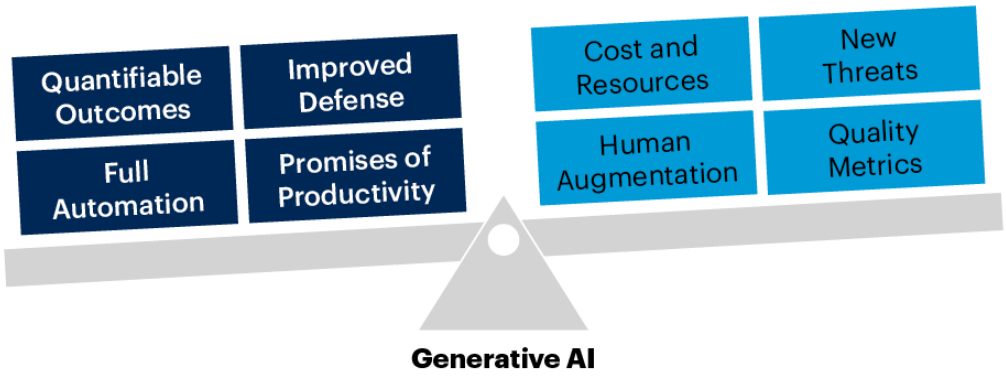
- Analyze vendor patches and updates

https://securityintelligence.com/is-your-soc-overwhelmed-artificial-intelligence-and-mitre-attck-can-help-lighten-the-load/

RISKS AND CONCERNS

Balance the Hype vs Reality

Balancing Cybersecurity Reality With GenAI Hopes



Source: Gartner
800663_C

Potential privacy and data sovereignty concerns

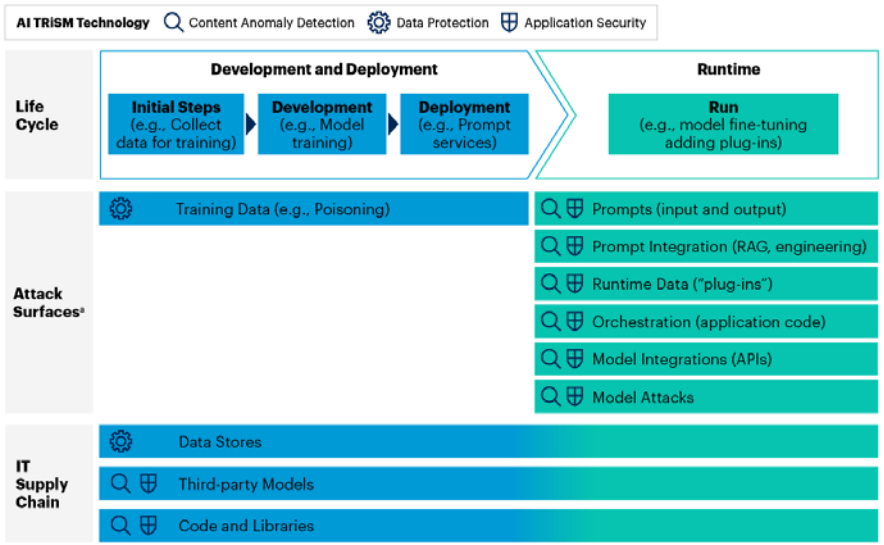
Lack of on-premises (non-SaaS) options

Lack of relevant metrics to measure AI benefits

Challenge of integrating AI into existing workflows or with third party vendors

Limited use of AI TRiSM (trust, risk, and security management) technical controls or policies

Generative AI Attack Surfaces Across the AI Life Cycle



Source: Gartner
[a] Main sample attack surfaces only; others not shown
796422_C

RECOMMENDATIONS

Develop a plan on how your organization will leverage AI

- ISO 42001 was published in Dec 2023

Develop an AI Governance Model that factors in cyber security of the AI technology

Evaluate existing Cyber Security tools in use and determine if AI enhancements can be leveraged to create value for the organization

Develop a multiyear approach to progressively integrate AI features and products when they augment security workflows

Introduce business value-driven AI evaluation frameworks to measure speed, accuracy, and productivity

Run pilots with real use cases and measure false positive reduction, accuracy, and cost savings

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

THANK YOU!

Eric Eifert

Eric.Eifert@ait.ac.at