Bundesministerium
Klimaschutz, Umwelt,
Energie, Mobilität,
Innovation und Technologie

Bundesministerium
Finanzen

1. Vernetzungstreffen im Rahmen des AI Policy Forums

KI in der öffentlichen Verwaltung

Mittwoch, 19. Oktober 2022
Festsaal Technisches Museum Wien

AIM AT 2030
Artificial Intelligence Mission Austria

Themensession 4: **Cybersicherheit und KI**

- **Mario Drobics**, AIT - *Sicherheit mittels KI – Was kann KI zu einer sicheren Verwaltung beitragen?*
- **Martin Pirker**, FH St. Pölten - *IT-Security und AI vor dem Hintergrund des AI Acts*
- **Camillo Nemec**, BMLV - *Cybersicherheit und KI aus verteidigungspolitischer Sicht*

SICHERHEIT MITTELS KI

1. Vernetzungstreffen “KI in der öffentlichen Verwaltung”

Wien, 19.10.2022

Mario Drobics
AIT Austrian Institute of Technology

https://fu

Cyberangriffe auf das österreichische Außenministerium angeblich erfolglos

"Schadsoftware konnte keine Auswirkungen entfalten." Angreifer hatten versucht, sich mithilfe von Phishing-Mails in interne Systeme

Hackerangriff auf Kärnten: 80.000 Stammdatenblätter ausgelesen

Im Rahmen des Leaks sind Datenblätter mit Namen, Geburtsdaten, Adress
und Telefonnummern aufgetaucht

10. Juni 2022, 15:04, 300 Postings

https://www.derstandard.at/story/20001364
stammdatenblaetter-ausgelesen

IT-SICHERHEIT

Cyberangriff auf I
Uni: Erste Daten i
aufgetaucht

Unter anderem dürften Reisepässe,
worden sein. Den Angriff beanspruch
für sich

Mickey Manakas, Andreas Proschofsky
27. Juni 2022, 17:02, 83 Postings

https://www.derstandard.de/story/2000136948190/cyb
uni-erste-daten-im-darknet-aufgetaucht

ORFplus WIRTSCHAFT

Cyberangriff auf Gunskirchner Motorenhersteller BRP-Rotax

steiermark ORF.at 6.9.2022

Steiermark-News | Steiermark-Magazin | Der ORF Steiermark | Volksgruppen | Ganz Österrei

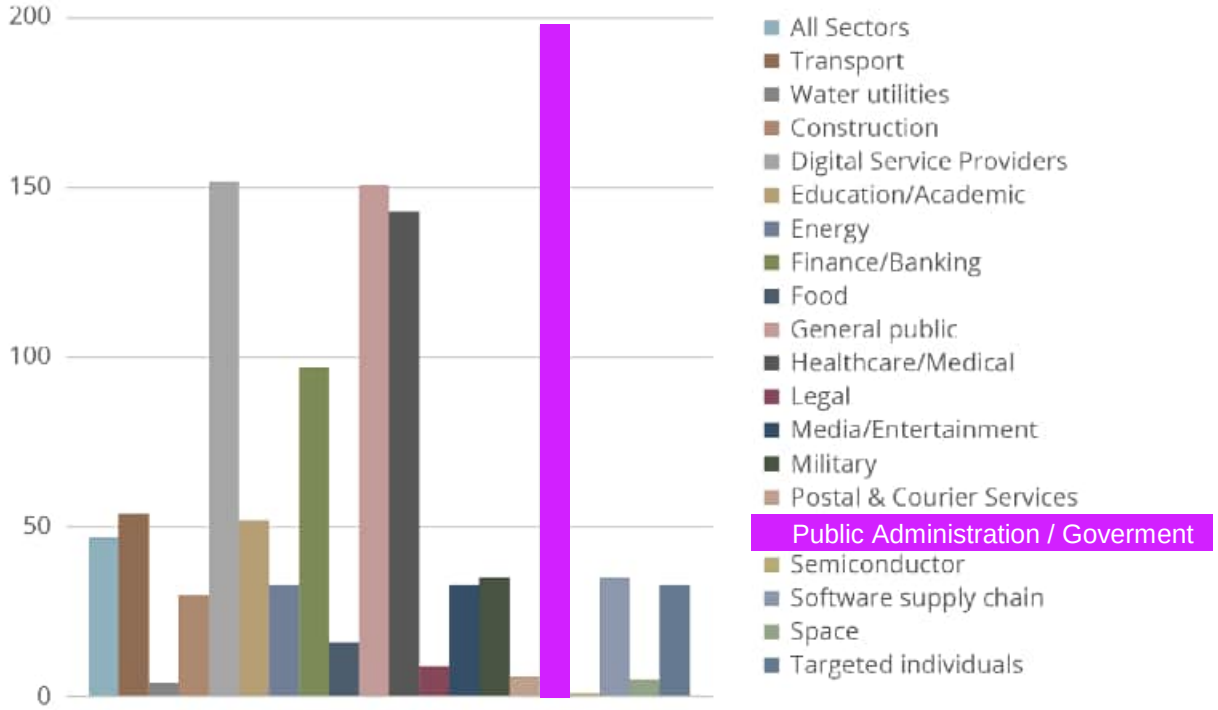
CHRONIK

https://steiermark.orf.at/stories/3172339/

Hackerangriff legt Feldbacher EDV lahm

Am Wochenende ist die Stadt Feldbach Opfer eines Hackerangriffs geworden: Das EDV-System wurde übernommen; sollte die Stadt ihre Daten wiederhaben wollen, müsse Lösegeld bezahlt werden.

CYBER-SECURITY VORFÄLLE



Targeted sectors per number of incidents (April 2020-July 2021)

Source: ENISA THREAT LANDSCAPE 2021

CYBER-SECURITY BEDROHUNGEN

ENISA Threat Landscape 2021 - Prime threats

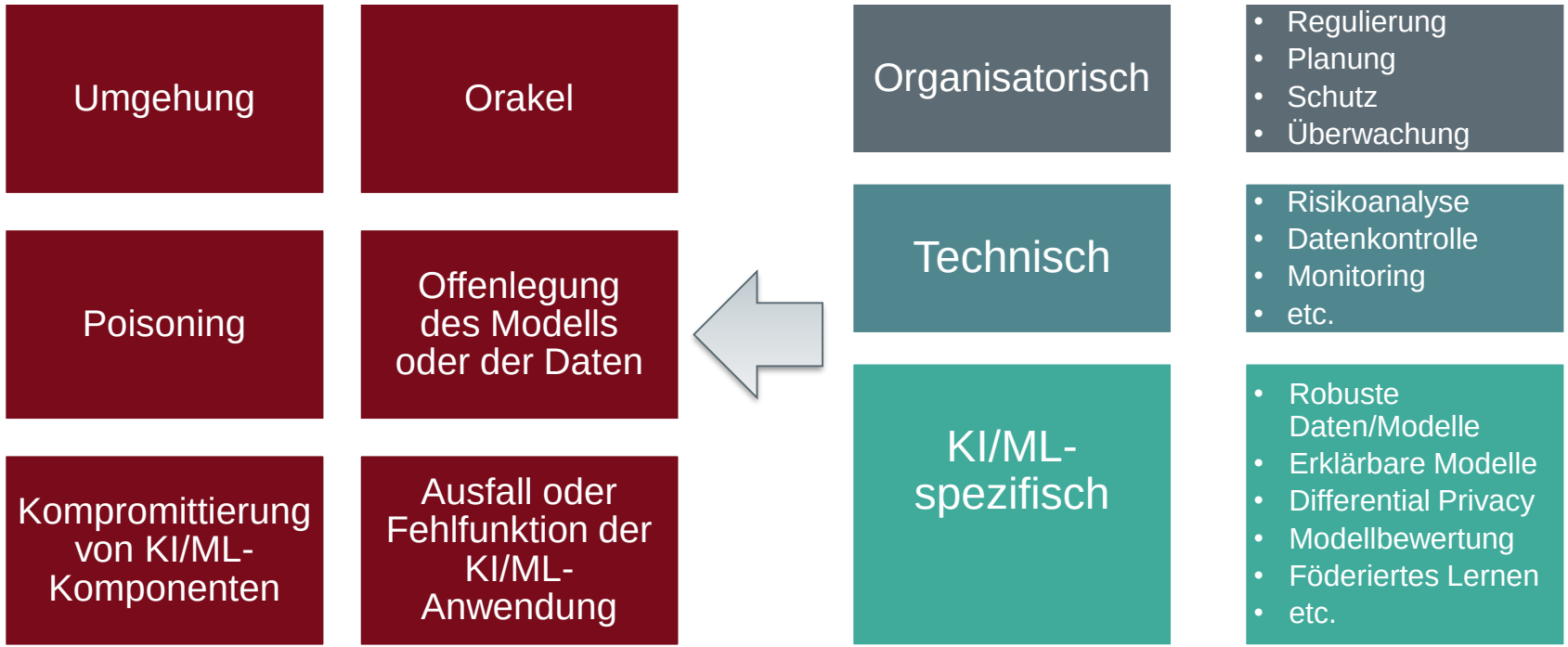
Source: ENISA THREAT LANDSCAPE 2021

BEDROHUNGEN KI-BASIERTER SYSTEME

ENISA Threat Landscape 2021 - Prime threats

Source: ENISA THREAT LANDSCAPE 2021

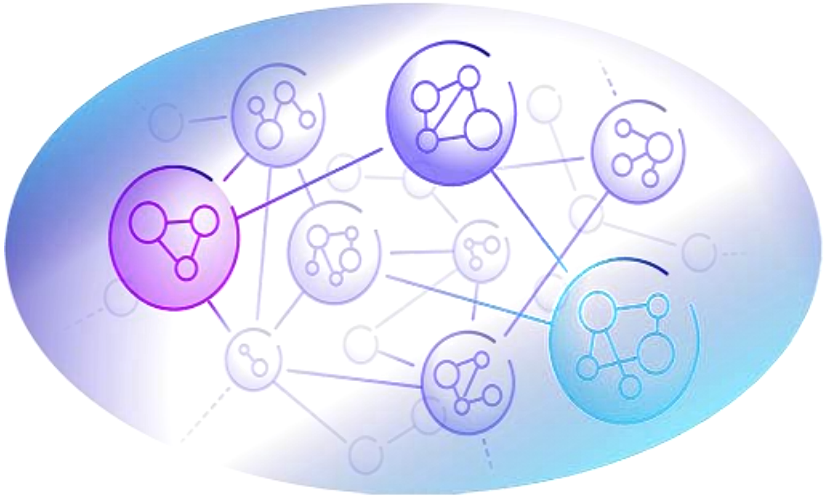
BEDROHUNGEN FÜR KI-BASIERTE SYSTEME



SCHUTZ DER DATEN

- Vernetzte Datennutzung erfordert verstärkte Maßnahmen zum Schutz der Daten
  - **Identifikation** der Systeme & Anwender:innen
  - **Autorisierung** der Zugriffe
  - **Sichere** Datenübermittlung
  - **Vertrauensvolle & nach-vollziehbare** Datenspeicherung / -nutzung

➔ **Schutz des Datenökosystems**



KI ZUM SCHUTZ KOMPLEXER SYSTEME

ENISA Threat Landscape 2021 - Prime threats

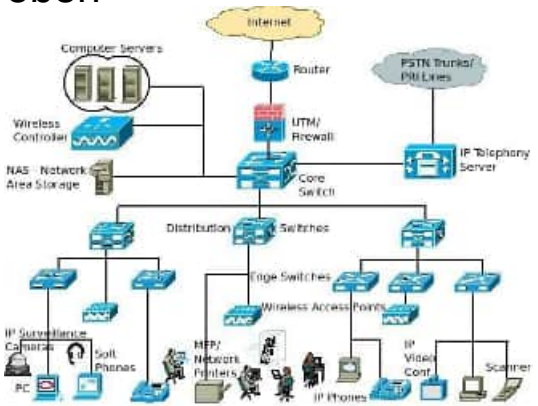
Source: ENISA THREAT LANDSCAPE 2021

CYBER-SECURITY
IN KOMPLEXEN SYSTEMEN

- **Komplexe Systeme wachsen organisch** von unten nach oben
  - Implementierungs- und Konfigurationsfehler führen zu Schwachstellen
  - Konstruktionsfehler verursachen Schwachstellen

  ➔ **Prävention scheitert letztlich**
- Monitoring konzentriert sich auf die frühzeitige Erkennung von feindlichen Aktionen
- Der heutige Stand der Technik ist immer noch:
  - Hauptsächlich Untersuchung des Netzwerkverkehrs (gängige Tools)
  - Signaturbasierte Suche nach bekannten fehlerhaften Elementen
  - Protokolldatenuntersuchung mit SIEMs – meist nur vordefinierte Regeln
  - Begrenzte Anomalieerkennung

  → **Wie erkennt man unbekannte Angriffe?**

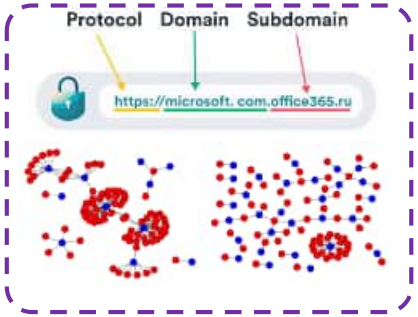


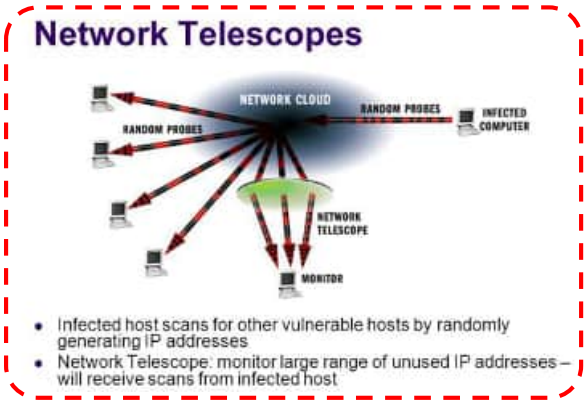
AI-BASIERTE ANALYSEN AUF NETZWERKEBENE

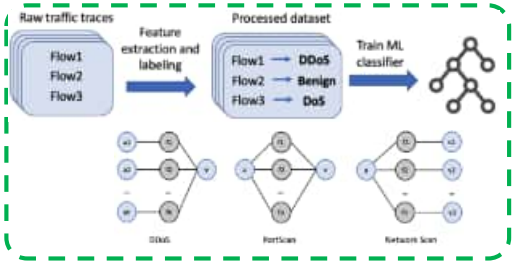
AI4SEC

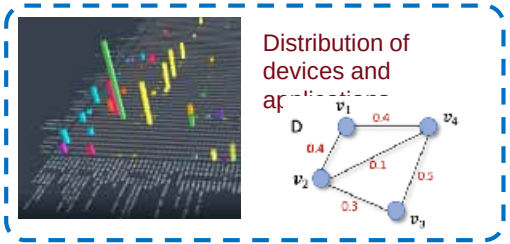
AUSTRIAN INSTITUTE OF TECHNOLOGY

- Erkennung von Phishing-Websites durch **lexikographische Analyse**
- Identifizierung von Schadgeräten durch **Darkspace UnSolicited Traffic Analyse**
- **Deep Learning zur Malware-Erkennung** über verschlüsselten Netzwerkverkehr
- **Generative KI zur Anomalie-erkennung**
- **Explainable AI** zur Interpretation von Deep-Learning-Modell-entscheidungen







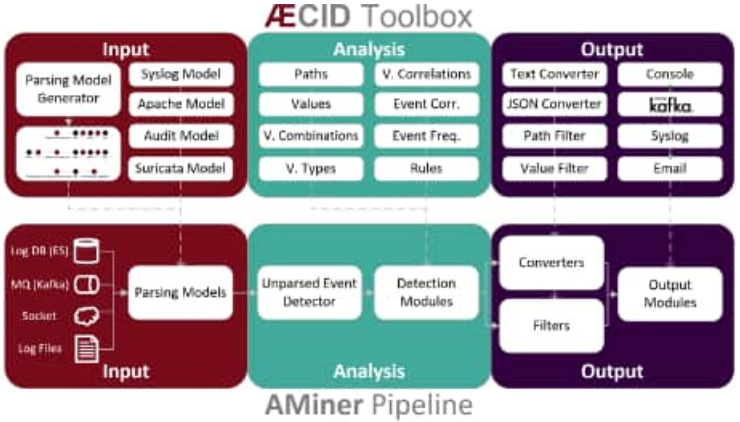


AI-BASIERTE ANALYSE AUF SYSTEMEBENE

- **Angreifer nutzen Systeme jedoch anders** als legitime Benutzer...
  - Zugriff auf andere DMZ-Server von einem kompromittierten Webserver
  - Verwendung der SSH-Wartungsschnittstelle anstelle der Weboberfläche
  - Anmeldung mit Backup-System-SSH-Schlüssel, der für die SFTP-Dateiübertragung vorgesehen ist
- **Neuartige Ansätze des maschinellen Lernens**
  - **Beobachten** eines Systems und seiner "normalen" Auslastung
  - **Dynamischer Aufbau** eines Modells, das eine Baseline darstellt
  - **Warnung bei signifikanten Abweichungen** von diesem Ausgangswert
- **Sichtbarkeit** von gegnerischen Aktionen ist der Schlüssel!
  - **Ausführliche Protokolldaten** von Diensten, Anwendungen, Betriebssystemen

ÆCID UND AMINER

- **AECID** ist ein ausgereiftes Einbruchmeldesystem, das Protokolldaten verwendet
- Erfasst **Protokolldaten von beliebigen Systemen**
  - Funktioniert mit domänenspezifischen und bisher unbekannten Systemen, ist also nicht auf vordefinierte Parser angewiesen
    ➔ **selbstlernend!**
  - **Light-weight**, verteilte Anomalydetektion
  - Clients werden mit **geringem Speicherbedarf** und **minimaler CPU-Auslastung** ausgeführt
  - Nicht in Konkurrenz zu etablierten Systemen, sondern als zusätzlicher Detektionsmechanismus



https://github.com/ait-aecid/logdata-anomaly-miner

KI ZUR ERKENNUNG VON DESINFORMATION

ENISA Threat Landscape 2021 - Prime threats

Source: ENISA THREAT LANDSCAPE 2021

BEKÄMPFUNG VON DESINFORMATION

Fake News und Desinformation in sozialen Medien	Audio/Video/Text Analytics			Verbraucherschutz
	Identification of **Fake Shops**	**Price discrimination in the Internet**	**Fact checker**	

EU-HYBNET
Empowering a Pan-European Network to Counter Hybrid Threats
2020 2025

https://euhybnet.eu/

DEVELOPED BY AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

Federal Chancellery

Federal Ministry Republic of Austria Europe, Integration and Foreign Affairs

Federal Ministry Republic of Austria Defence

Federal Ministry Interior

Bundesministerium Soziales, Gesundheit, Pflege und Konsumentenschutz

ORF APA

AIT's Fact Checking Platform

FFG netidee MAL ZWEI primming SINBAD preis.wert KIRAS defalsif-ai HORIZON 2020 VICTORIA VIDEO ANALYSIS FOR INVESTIGATION OF CRIMINAL AND TERRORIST ACTIVITIES

NATIONALE LEITINITIATIVE – KI-BASIERTE FACT-CHECKING PLATTFORM

Bundeskanzleramt

Bundesministerium Landesverteidigung

Federal Ministry Republic of Austria Europe, Integration and Foreign Affairs

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

defalsif-ai

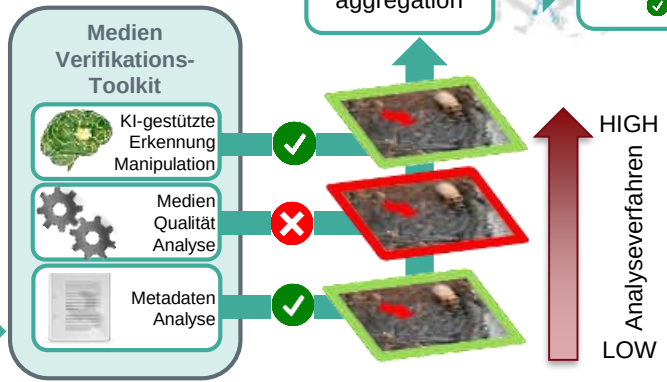
APA

ORF

Einfaches Benutzerinterface

Multimedia Analyse

Multimodale KI



Hybrid CoE

VICTORIA
VIDEO ANALYSIS FOR INVESTIGATION OF CRIMINAL AND TERRORIST ACTIVITIES

UNITED NATIONS OFFICE OF COUNTER-TERRORISM
UN Counter-Terrorism Centre (UNCCT)

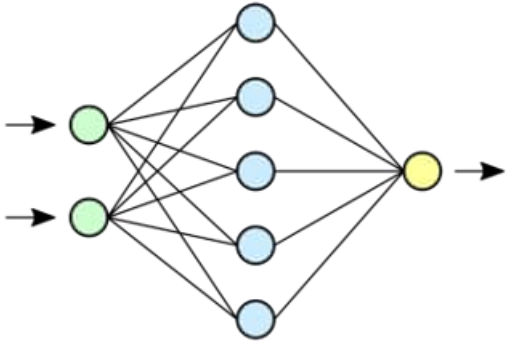
AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

THANK YOU!

Mario Drobics, 19th October 2022

IT-Security und AI vor dem Hintergrund des AI Acts

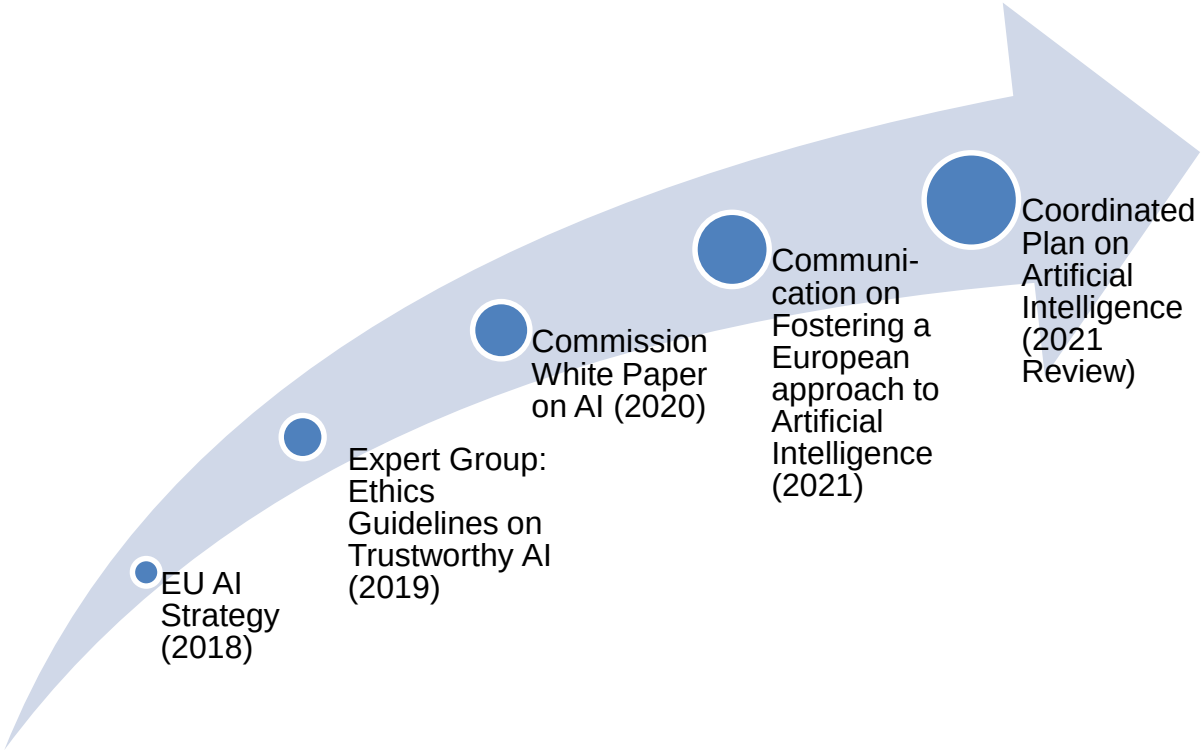
AI everywhere



AI …und Security?

- AI wird allgegenwärtig
- AI wird „commodity“
- AI trifft / beeinflusst Entscheidungen
- AI schafft neue Möglichkeiten
- AI schafft ein Sicherheitsproblem?

Der Artificial Intelligence Act



Schlüsselinhalte

- „AI-Systeme"
  - Definition von AI
  - Unterkategorie: "Hochriskantes KI-System"
  - Definitionen in den Anhängen I-III
- Verpflichtungen insbesondere für Anbieter und Nutzer von "Hochrisiko"-KI-Systemen
  - verbotene KI-Praktiken (Art 5); Transparenzpflichten (Art 52); „Sandboxing" (Art 53)
- Hohe Strafen / Sanktionen
- Risikoabschätzung und Security bekommen einen hohen Stellenwert

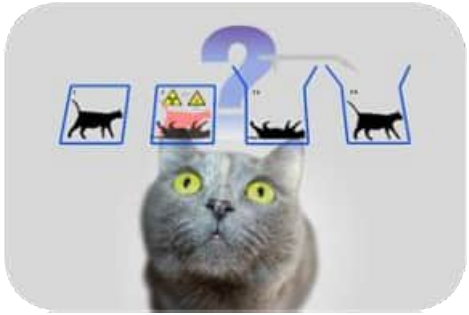
Scope

AI Act

- Definiert in Annex I
- Lehnt sich an OSZE-Definition an
  - Nicht für Gesetzestexte geplant
- Problem
  - Taxativ
  - Anfänglich sehr technologiespezifisch
  - „Catch-All“
    - Rule-Based Systems
    - Statistische Methoden
    - Predictive Algorithms

Generell

- AI extrem schwierig zu definieren
  - Speziell die „Randbereiche“ wie Entscheidungsbäume
  - Emergenz-Problem / Gruppendynamik



Security Testing

„Klassische“ Probleme

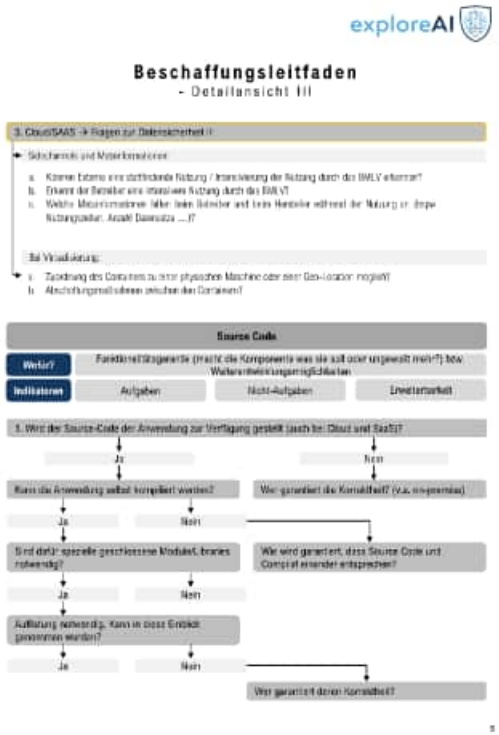
- Größe / LoCs
- Abgekapselter Code
  - Externe Module
  - Stabilität
  - Review?
- Komplexe Algorithmen
  - Systemische Fehler?
- Erwartete vs. Realer Nutzung
- Time & Money

Neue Herausforderungen

- Explainability-Problem
  - Halting-Problem squared
- Modell
  - Aus welchen Daten? Datenkontrolle? Data Cleansing
  - Fehlende Stabilität
  - Model as a Service?
- Beeinflussung durch Testing
  - Fuzzy-Tests versus Echtdaten
- Risikoabschätzung
  - „Klassische“ Security-Intelligence

Beschaffung

- Neue Kriterien
  - Transparenz
  - „Einfachheit“ der Methodik
  - Richtige Ansprechpartner
  - „Ist das überhaupt AI?“
- Neue Angriffsvektoren
  - Modelle
  - Daten (Training & Processing)
  - Aktivitäts(meta)monitoring
- Komplexe Thematik
- Download: www.secureAI.info



Kieseberg, P., Buttinger, C., Kaltenbrunner, L., Temper, M. and Tjoa, S., 2022, July. Security considerations for the procurement and acquisition of Artificial Intelligence (AI) systems. In *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-7). IEEE.

Panik?

Kontakt

Peter Kieseberg
Institut für IT Sicherheitsforschung
Fachhochschule St. Pölten
peter.kieseberg@fhstp.ac.at

Cybersicherheit und Künstliche Intelligenz aus verteidigungspolitischer Sicht

ObstdhmfD Mag. Camillo NEMEC
Generaldirektion Verteidigungspolitik

Bedrohungen Risiken Herausforderungen

Ukraine Russland Krieg

Regionale und globale Krisen und Konflikte

Geopolitische Auswirkungen

Hybride Bedrohungen (**Cyber**, Desinformation, Deepfakes, Bedrohung der kritischen Infrastruktur)

Internationaler Terrorismus und organisierte Kriminalität

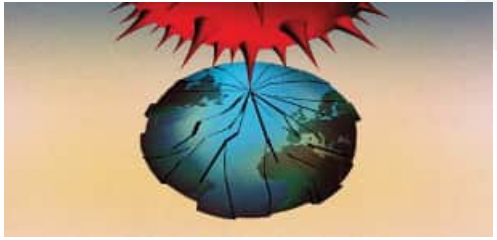
Digitalisierung (**Künstliche Intelligenz**)

Weltraum

Blackout

Klimawandel (Naturkatastrophen, technische oder humanitäre Katastrophen, Kampf um Ressourcen)

Anhaltende Flüchtlingsdebatte und Migrationsströme



Pandemie und globale Wirtschaftliche Folgen

Cybersicherheit

Das Internet als globaler öffentlicher Raum

fortschreitende Digitalisierung

Geopolitische Rivalität

Sicherheits- und verteidigungspolitische politische Relevanz (politisch, wirtschaftlich, militärisch, gesellschaftlich)

Nexus innere und äußere Sicherheit

Akteursvielfalt

Cyber als operativer Bereich von Konflikten (Warfighting Domain)

Rolle der Streitkräfte

Schwellenwert Cyber Angriff

Attribuierung

Künstliche Intelligenz

Domänenübergreifende Querschnittsthematik (multidisziplinär und transformativ)

Streitkräfte und Bedrohungsdispositiv

Instrument der Geopolitik

Militärische Anwendungsgebiete (Chancen und Risiken)

Dual-Use-Charakteristik

Kontroversen um Definitionen und Grad menschlicher Kontrolle von letalen autonomen Waffensystemen

Danke für die Aufmerksamkeit

camillo.nemec@bmlv.gv.at