Financované Európskou úniou
NextGenerationEU

PLÁN OBNOVY

MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI
ŽILINSKEJ UNIVERZITY V ŽILINE

Umelá inteligencia v KB

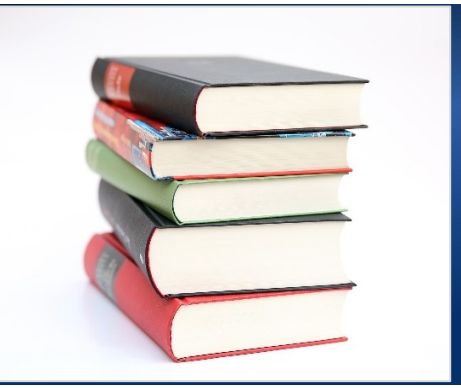
Moderné technológie, bezpečnosť cloudu a IoT (Blok VII)

Kurz: Špecialista kybernetickej bezpečnosti vo verejnej správe

Ondrej Škvarek

KC KYB UNIZA, https://kc.uniza.sk/

skvarek@uniza.sk



Obsah

- Prečo je potrebná AI v kybernetickej bezpečnosti (KB)
- Zverejnené zraniteľnosti a aktuálne hrozby
- Umelá inteligencia, strojové učenie, neurónové siete, veľké jazykové modely
- Autoenkodéry a GAN siete – pre detegovanie anomálií v sieťovej prevádzke
- Nasadenie AI v systémoch kybernetickej bezpečnosti
- Firmy - používajúce AI v KB
- Praktické príklady použitia AI
- Trendy AI v KB na najbližšie roky

Úvod

Umelá inteligencia v KB

Úvod

- rozvoj informačných systémov - nové zraniteľnosti a kybernetické hrozby
- kybernetická bezpečnosť – kľúčová pre digitálny ekosystém
- tradičné prístupy ochrany systémov nestačia
  - založené na signatúrach, manuálnych reakciách a statických politikách
  - nedokážu adekvátne reagovať na dynamicky sa meniace a stále sofistikovanejšie formy útokov
  - útočníci využívajú automatizované nástroje, distribuované kampane a pokročilé techniky sociálneho inžinierstva, ktoré dokážu obísť konvenčné obranné mechanizmy
- umelá inteligencia (AI)
  - nástroj kybernetickej bezpečnosti (KB) - proti kyberzločinu
  - analyzovať obrovské objemy dát v reálnom čase
  - identifikovať anomálie
  - adaptovať sa na nové typy hrozieb
  - predikovať potenciálne útoky ešte pred ich realizáciou, čím sa zásadne mení paradigma ochrany – z reaktívneho na **proaktívny model**
  - na druhej strane, ju využívajú kyberzločinci na vývoj sofistikovanejších a ťažšie odhaliteľných útokov - „závod v zbrojení“, obe strany zdokonaľujú svoje metódy

Umelá inteligencia

Umelá inteligencia (artificial inteligence - AI) – napodobniť človeka

- súbor technológií umožňujúcich počítačom napodobňovať ľudské kognitívne funkcie, ako je **učenie**, **rozhodovanie** a **rozpoznávanie vzorcov**
- oblasť informatiky, ktorá sa zaoberá tvorbou systémov schopných vykonávať úlohy vyžadujúce si ľudskú inteligenciu, napríklad **spracovanie jazyka**, **analýzu dát** či **riadenie**.
- systémy AI sa **trénujú na obrovskom množstve dát**, aby sa z nich naučili rozpoznávať súvislosti a na základe toho robiť rozhodnutia

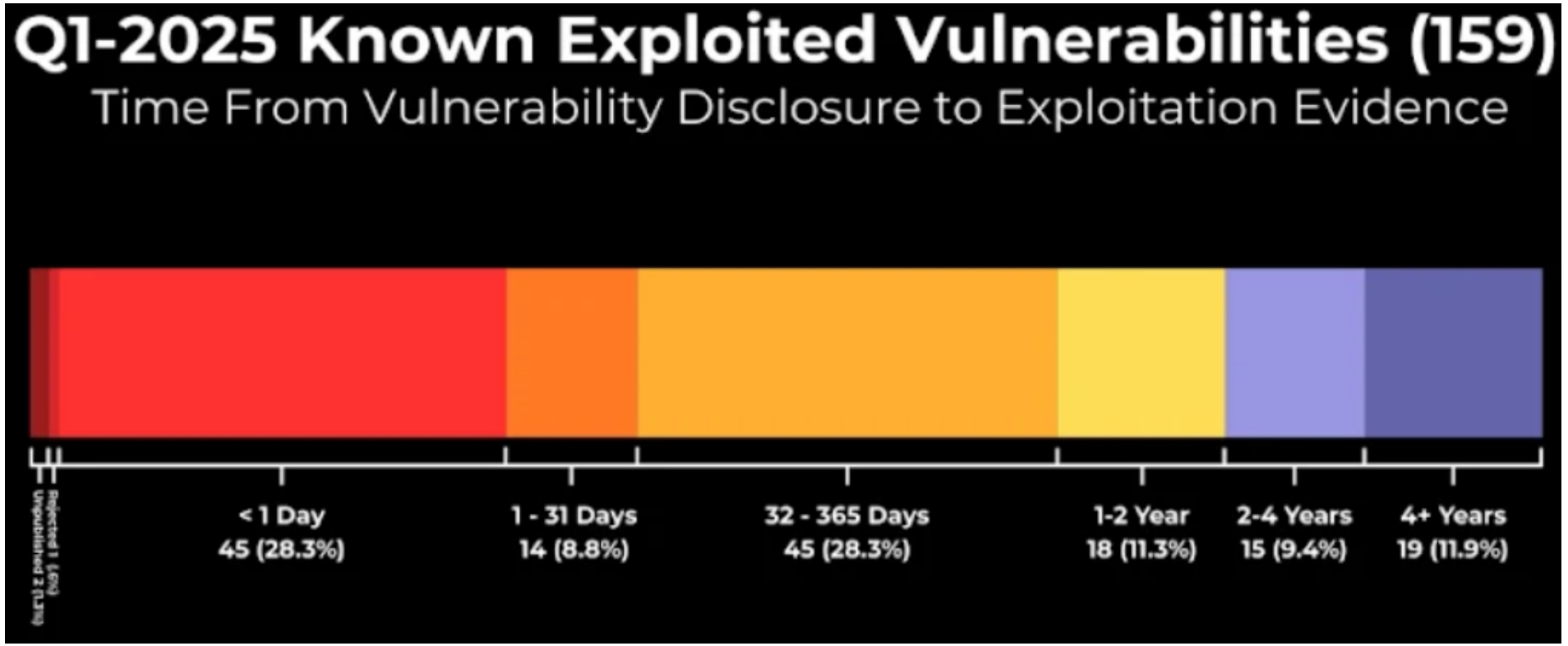
Prínosy umelej inteligencie pre kybernetickú bezpečnosť

1. Zlepšená schopnosť **detekcie hrozieb a anomálií** analýzou veľkého množstva údajov a učením sa z minulých útokov
2. **Automatizácia reakcií** na incidenty, vrátane automatického blokovania škodlivých aktivít a izolácie infikovaných systémov
3. **Prediktívna analýza** potenciálnych hrozieb na základe analýzy historických údajov a aktuálnych trendov
4. **Zneužívanie AI kyberzločincami** na vytváranie sofistikovanejších útokov (phishingové kampane, deepfake videá, oklamanie obetí a šírenie dezinformácií)

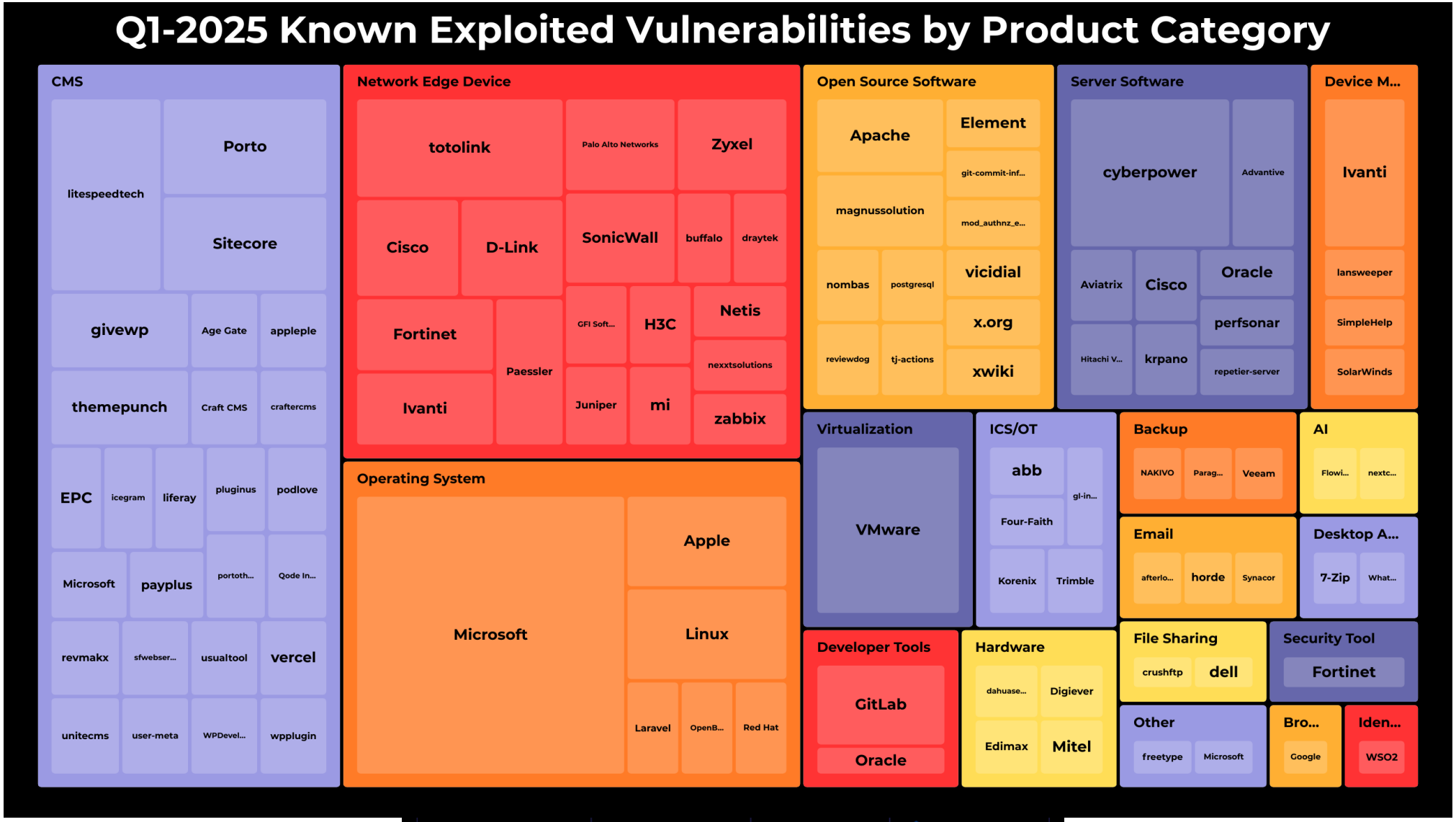
Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Rýchlo reagovať - automaticky aktualizovať obranu

- 23,8% zverejnených bežných zraniteľností (CVE v Q1 2025) – bolo zneužitých do 24 hod [159 CVEs Exploited in Q1 2025 — 28.3% Within 24 Hours of Disclosure]



Aké rôzne zneužité zraniteľnosti

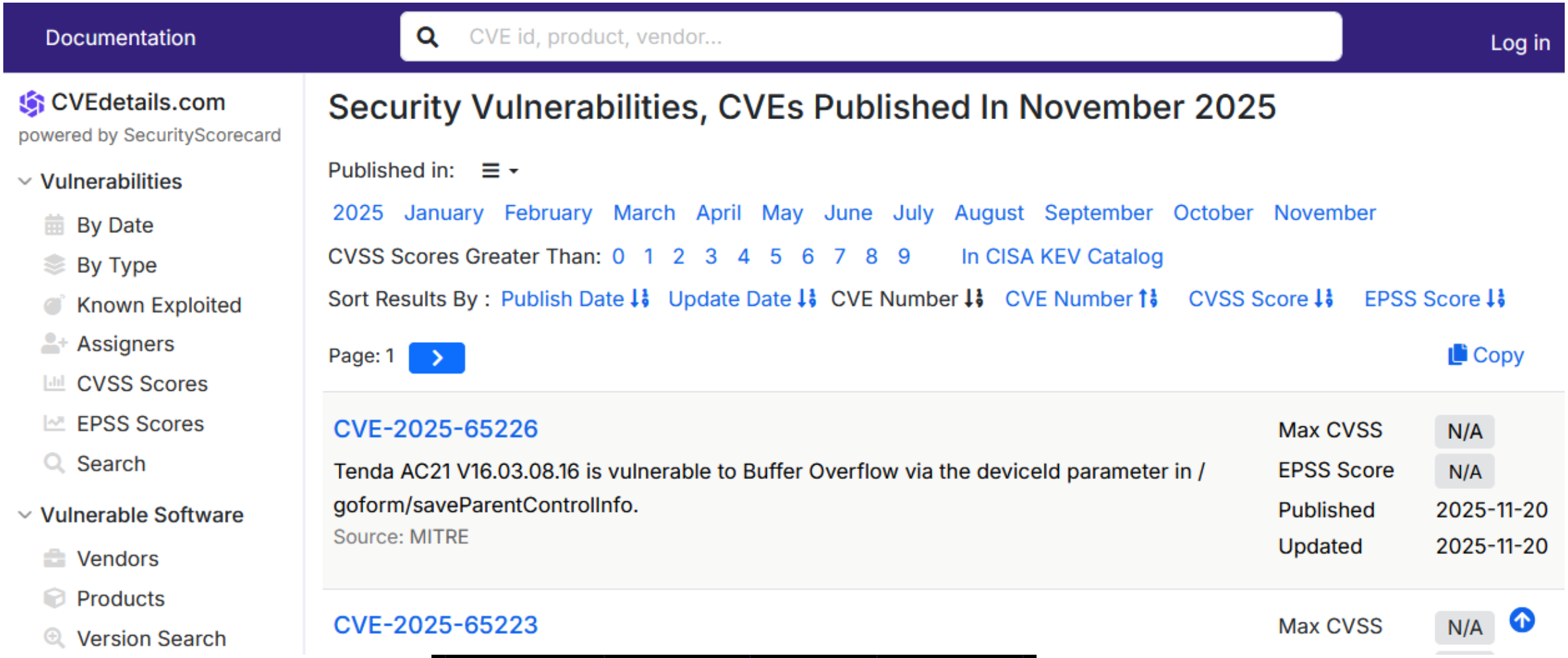


Hlavné hrozby (cybersecurity threats) – podľa fi. Verizon

- Správa telekom. spoločnosti Verizon - o vyšetrovaní únikov údajov za rok 2025 je alarmujúci nárast kybernetických útokov prostredníctvom tretích strán verizon.com/about/news/2025-data-breach-investigations-report
  - **Exploitation of Vulnerabilities** (zneužitie zraniteľností):
    zaznamenal 34 % nárast, s výrazným zameraním na zero-day exploity,
    útočil hlavne na perimetrické zariadenia a VPN
  - **Ransomware:** narástol od predošlého roku o 37% a reprezentuje 44% únikov dát (breaches), aj napriek citeľnému poklesu mediánu výšky zaplateného výkupného (medián – hodnota – polovica paltieb je menšia ako medián)
  - **Third-Party Involvement** (zapojenie tretích strán):
    Percento únikov zahŕňajúcich tretie strany sa **zdvojnásobilo** - riziká spojené s dodávateľským reťazcom a partnerskými ekosystémami
  - **Human Element** (ľudský prvok): účasť človeka na únikoch zostáva vysoká, s výrazným prekrývaním medzi sociálnym inžinierstvom a zneužitím prístupových práv (credential abuse)

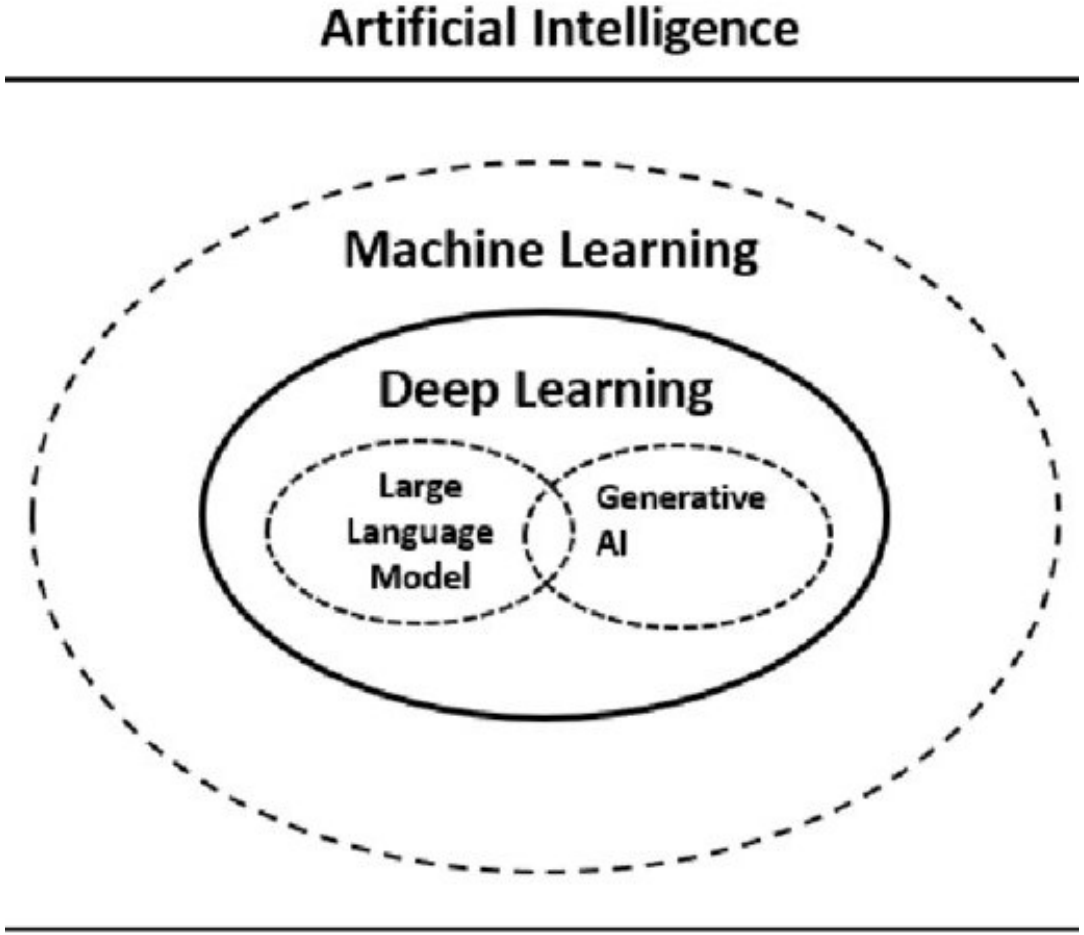
Zraniteľnosti - zverejnené

cvedetails.com/vulnerability-list/year-2025/month-11/November.html



Umelá inteligencia zahŕňa - strojové učenie, neurónové siete, hlboké učenie, veľké jazykové modely (LLM)

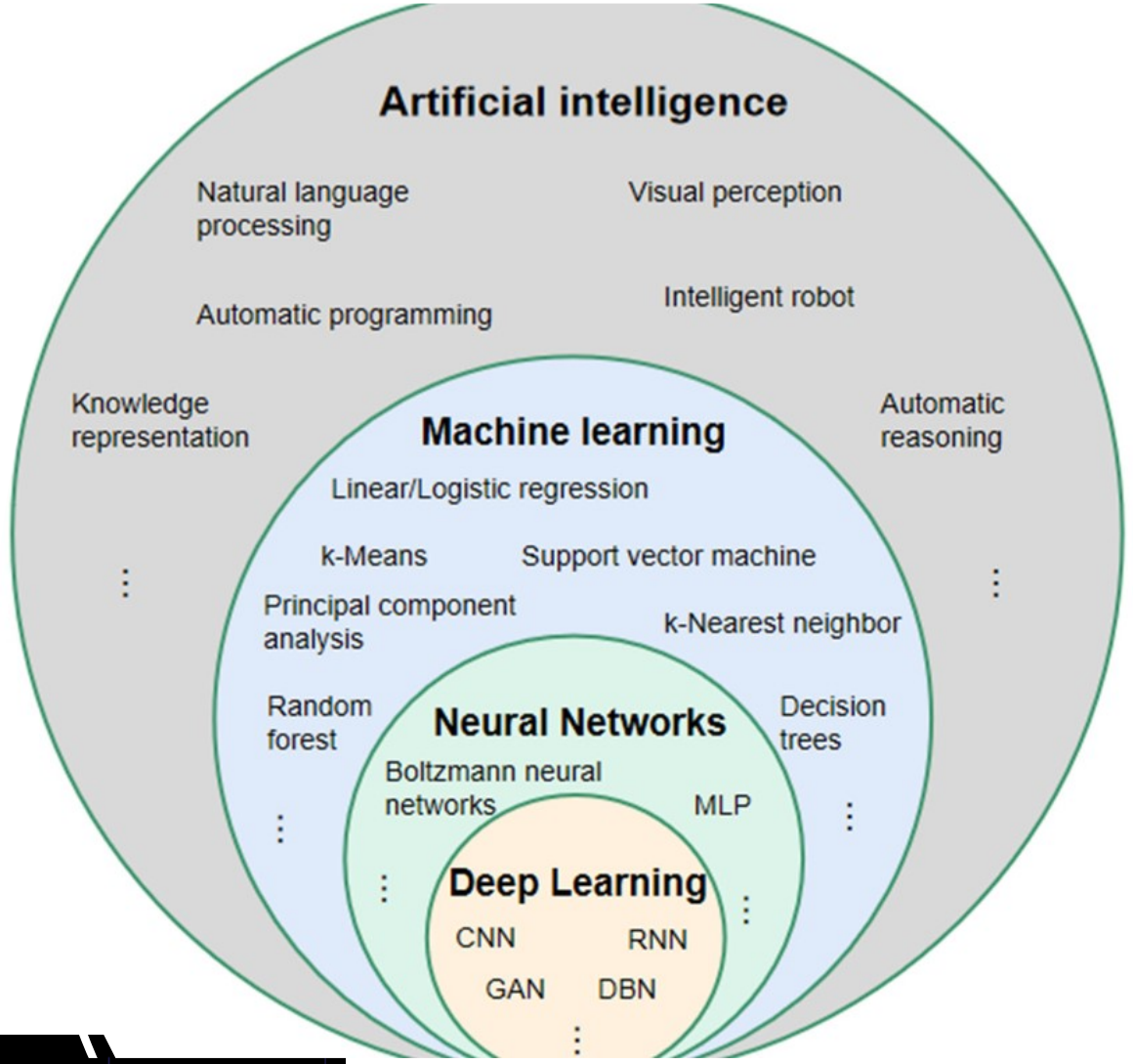
- Umelá inteligencia – napodobniť ľudskú inteligenciu (učiť sa, chápať, uvažovať)
- Strojové učenie - učí sa vzory z dát
- Neurónové siete – napodobňujú prepojenia neurónov
- Hlboké učenie – viac vrstiev neurónov, každá sa učí vzory – komplexné „porozumenie“
- Veľké jazykové modely (LLM) – typ hlbokého učenia – vedia generovať odpovede



Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Každá oblasť AI má veľa metód a algoritmov

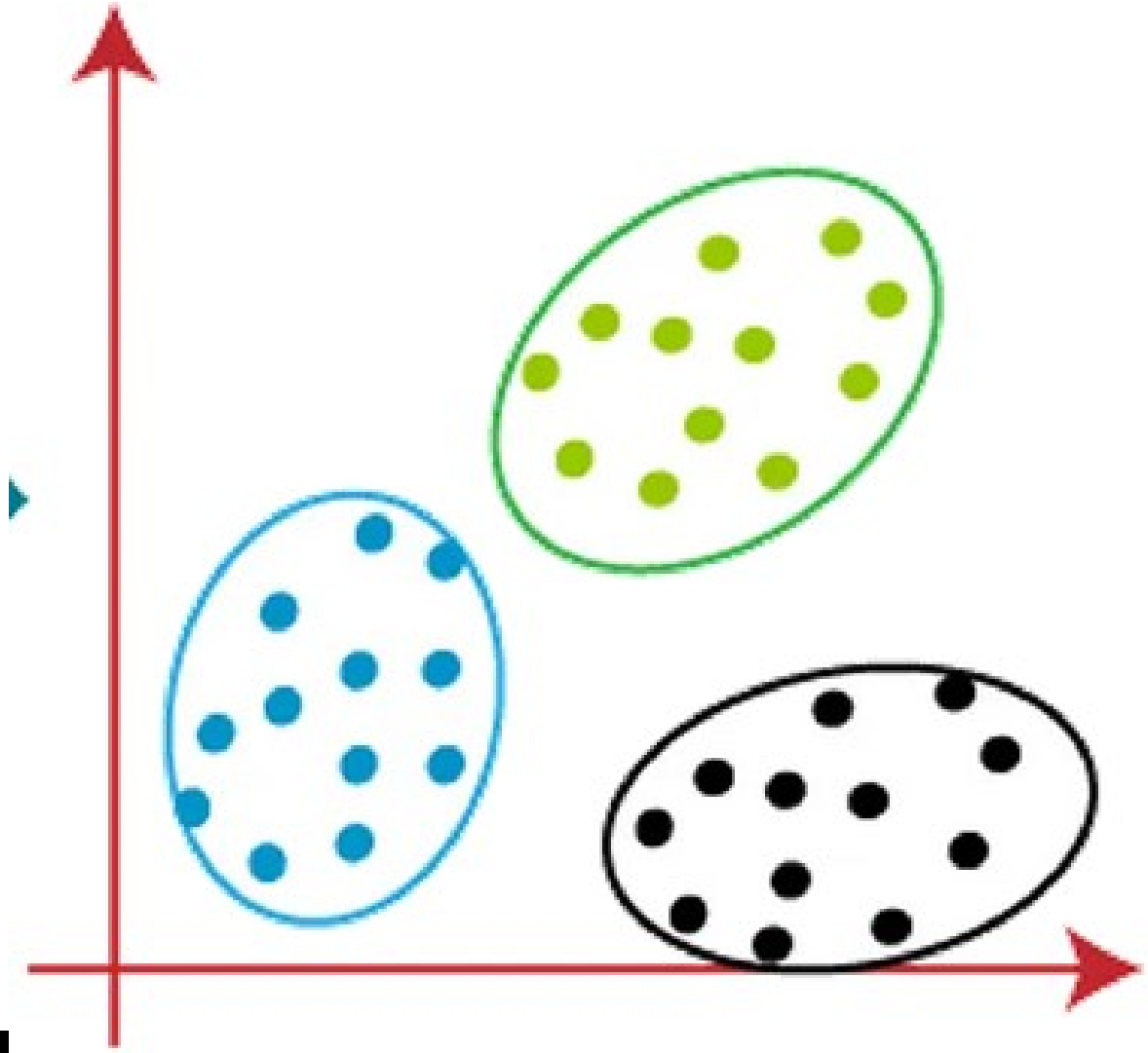
- AI (Artificial intelligence) - umelá inteligencia
- NLP spracovanie prirodzeného jazyka
- ML (machine learning) - strojové učenie
- a) **ML bez neurónových sietí**
  - lin. regeesia, zhlukovanie (k-Means, ...), rozhodovacie stromy, ...
- b) **ML s neurónovými sieťaqmi**
  - Deep Learning - hlboké učenie
    - CNN (convolutional neural network) - konvolučné neurónové siete
    - RNN (recurrent neural network) - rekurentné neurónové siete
    - DBN (deep belief network)
    - GAN (generative adversative network)
    - LLM (large language model) - **chatGPT**, ...



ML bez neurónových sietí

- s učiteľom
  - lin. regeesia
  - rozhodovacie stromy
  - LDA
  - ...
- bez učiteľa
  - zhlukovanie (k-Means, ...)
  - ...
- zobrazenie procesov z viacrozmerných priestorov do 2D a 3D
  - PCA
  - t-SNE
  - UMAP
  - ...

Unsupervised Machine Learning in Cybersecurity: A Comprehensive Analysis
medium.com/@leev574/unsupervised-machine-learning-in-cybersecurity-a-comprehensive-analysis-aa4d854e65d2



Učenie s učiteľom a učenie bez učiteľa

Učenie s učiteľom (supervised learning)

- Data + označené **labels**, napr. normálna prevádzka, útok

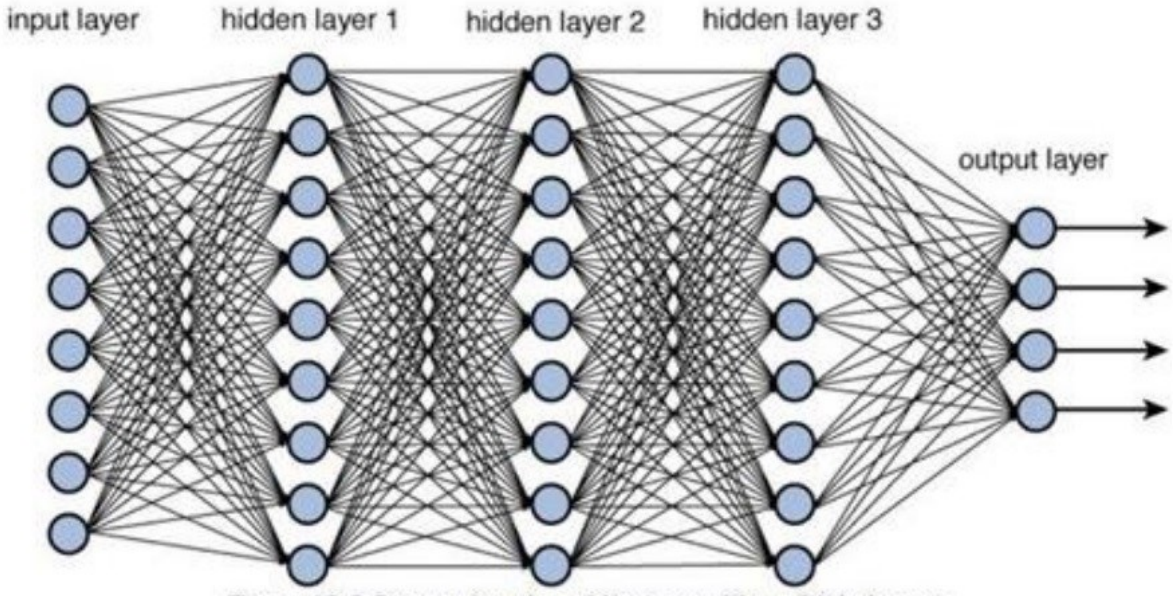
Učenie bez učiteľa (unsupervised learning)

- bez značiek, len dáta (no labels, data only)
- cieľ – naučiť sa skryté vzory v skúmanej štruktúre dát
  (to learn the hidden or underlying structure of the data)

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Hlboká neurónová sieť (viac vrstiev neurónov)

- Výstupy každého neurónu sú privedené na vstupy neurónov nasledujúcej vrstvy
- neurón vynásobí vstupy váhami, spočíta a súčet váži funkciou
- výstup sa privedie na vstupy neurónov nasledujúcej vrstvy
- učí sa tak, že chyba výstupu sa šíri spätne sieťou a váhové koeficienty sa upravujú, aby sa chyba zmašila

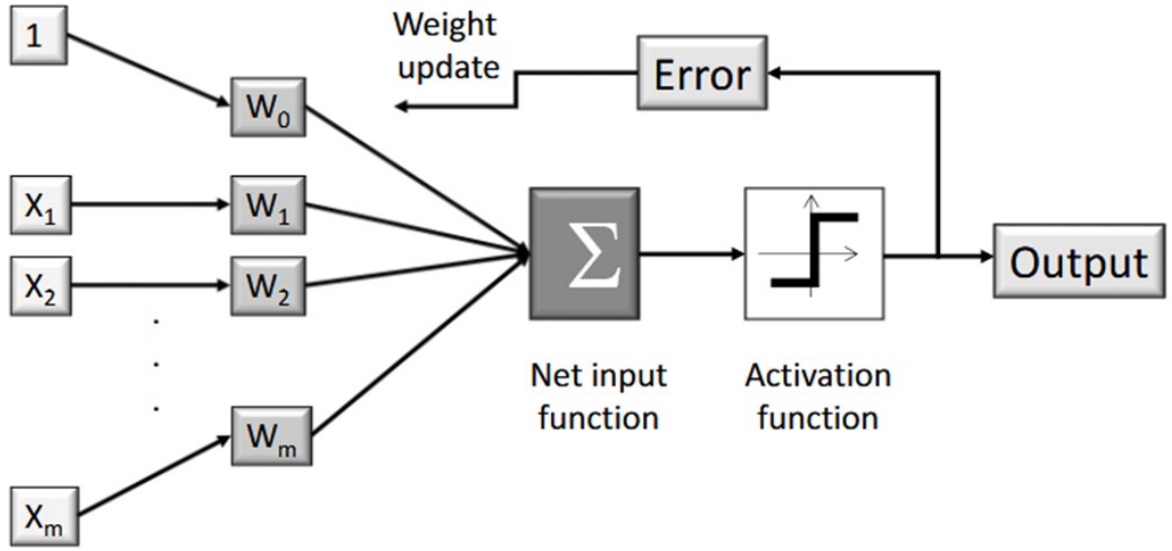


Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

ML s NN - Neurónová sieť sa učí z dát

- Trénovanie neurónovej siete (NN – neural network)
  - na vstup vkladáme postupne informácie o objektoch (vektory X) – napr. štratistiky paketov pre časové úseky
  - výstup nú hodnotu odpočítame od očakávanej – dostaneme chybu
  - chybu spätne šírime cez sieť a upravujeme váhy „wi“ tak, aby sa v budúcnosti chyba zmenšila – sieť sa učí
  - keď je sieť natrénovaná, ukončíme učenie
- Používanie NN na detekciu nezvyčajných udalostí (anomálií – útokov v sieti)
  - na stup privádzame vektor X pre neznámy objekt, napr. štatistiky pre úsek paketovej prevádzky
  - podľa výstupnej hodnoty rozhodneme, že je to napr. anomália alebo známa prevádzka

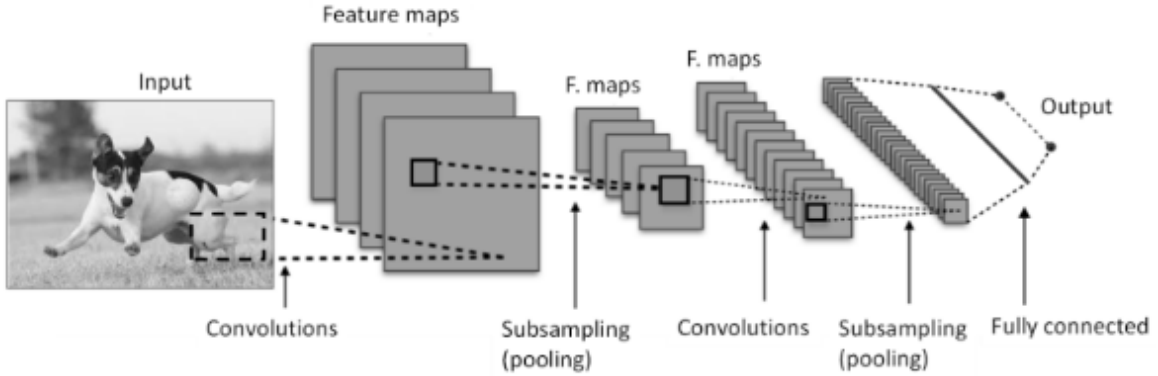
obrázky z researchgate.net/profile/Paolo-Dellaversana/publication/337678793_ARTIFICIAL_NEURAL_NETWORKS_AND_DEEP_LEARNING_A_SIMPLE_OVERVIEW/links/6818745dded4331557424d4d/ARTIFICIAL-NEURAL-NETWORKS-AND-DEEP-LEARNING-A-SIMPLE-OVERVIEW.pdf



Konvolučná neurónová sieť (typ hlbokej NN)

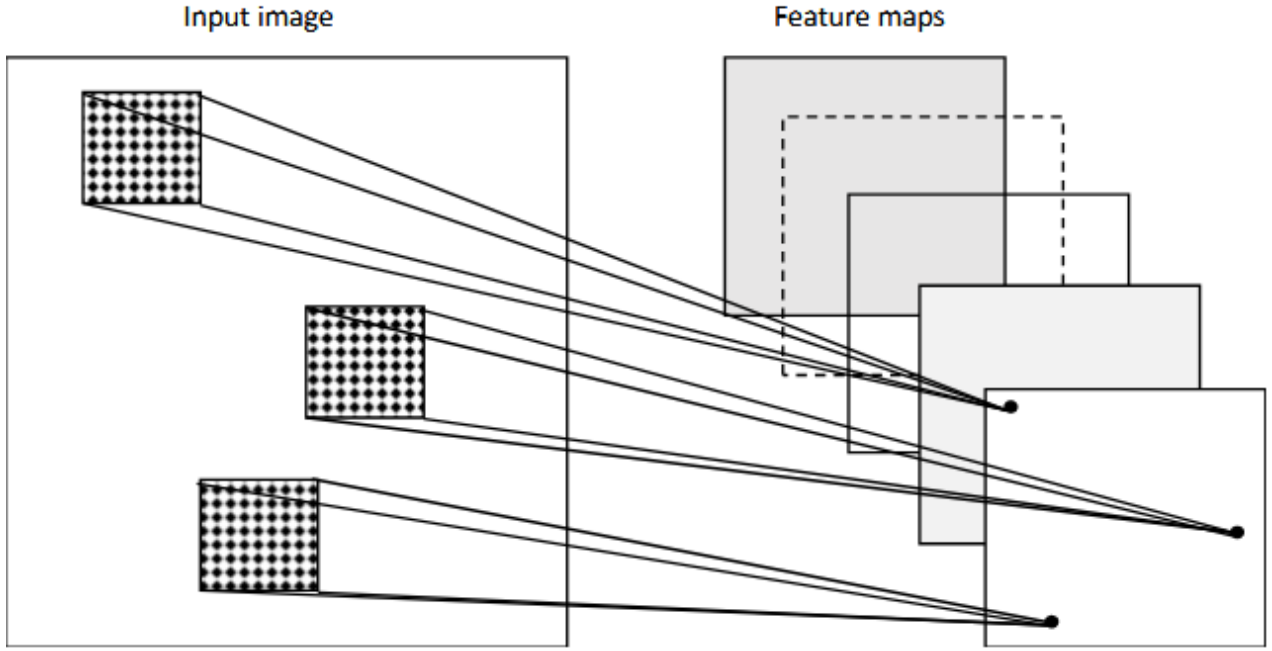
Trénovaie

- Každá vrstva sa učí - koeficienty štvorcových **receptívnych polí**
- pole sa posúva po „obraze“, s obrazom sa vyásobí, súčiny sa spočítajú – získa sa bod príznakovej mapy **feature map** - tým sa „vyťahujú“ z obrazu dôležité vlastnosti **hlavné príznaky**
- príznakové mapy sa ďalej spracujú
- nakoniec na výstupe siete môžu byť pravdepodpbnosti s ktorými detegujeme „psa, mačku, ... (v sieťových aplikáciách anomáliu v prevádzke alebo na zariadeniach)
- Počas trénovanie spätne šírime chybu predpovede a upravujeme koeficinety „receptívnych polí“.



Používanie na detekciu objaktov, anomálií, ...

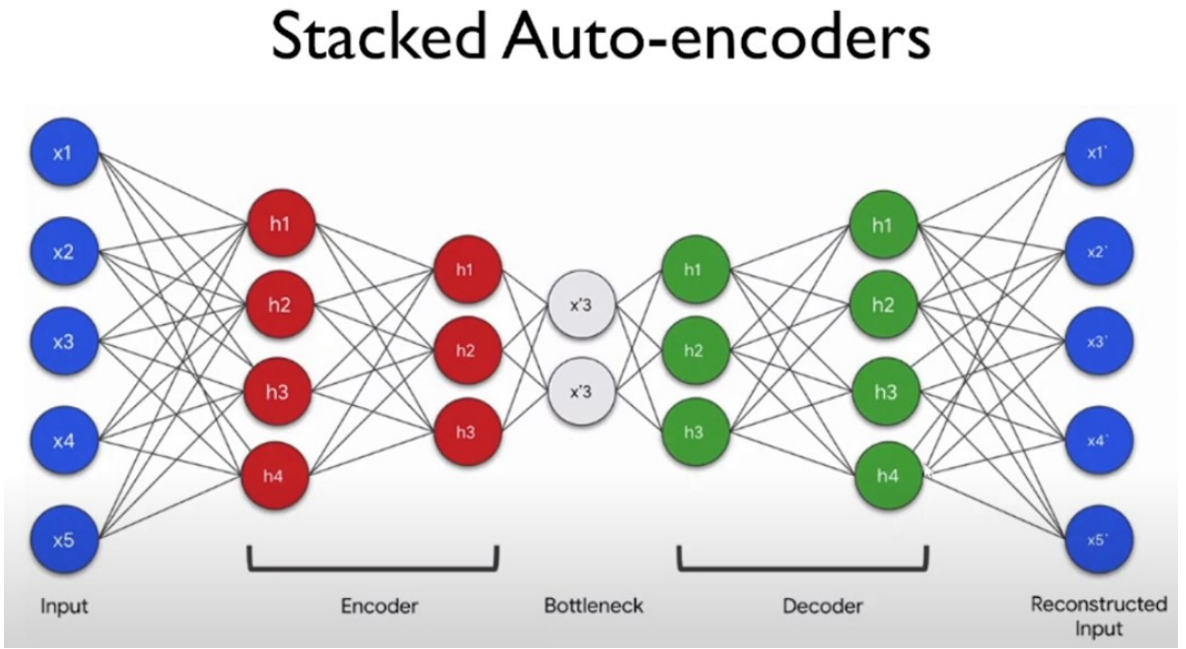
- Na vstup sa privedú informácie o neznámom objekte, sieť – na výstup dá pravddepodobnosti že je to napr. normálna prevádzka, alebo anomália



Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

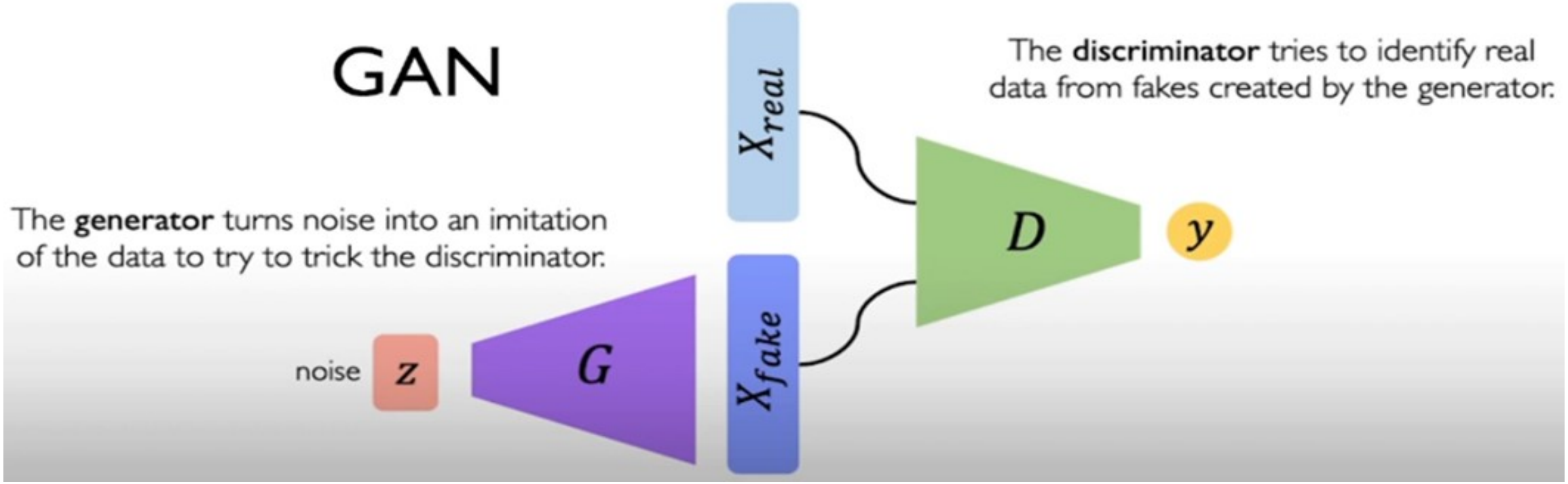
AE

- učí sa zároveň E – enkóder aj D – dekóder, tak aby sa zrekonštruovaný obraz podobal na vstup
- po natrénovaní sa AE použije na odhalenie nezvyčajných dát (anomálií, môžu byť spôsobené útokmi) – pre nezvyčajné vstupy sa výstup bude významne líšiť (na ne nie je AE natrénovaný)



GAN (Generative Adversarial Networks)

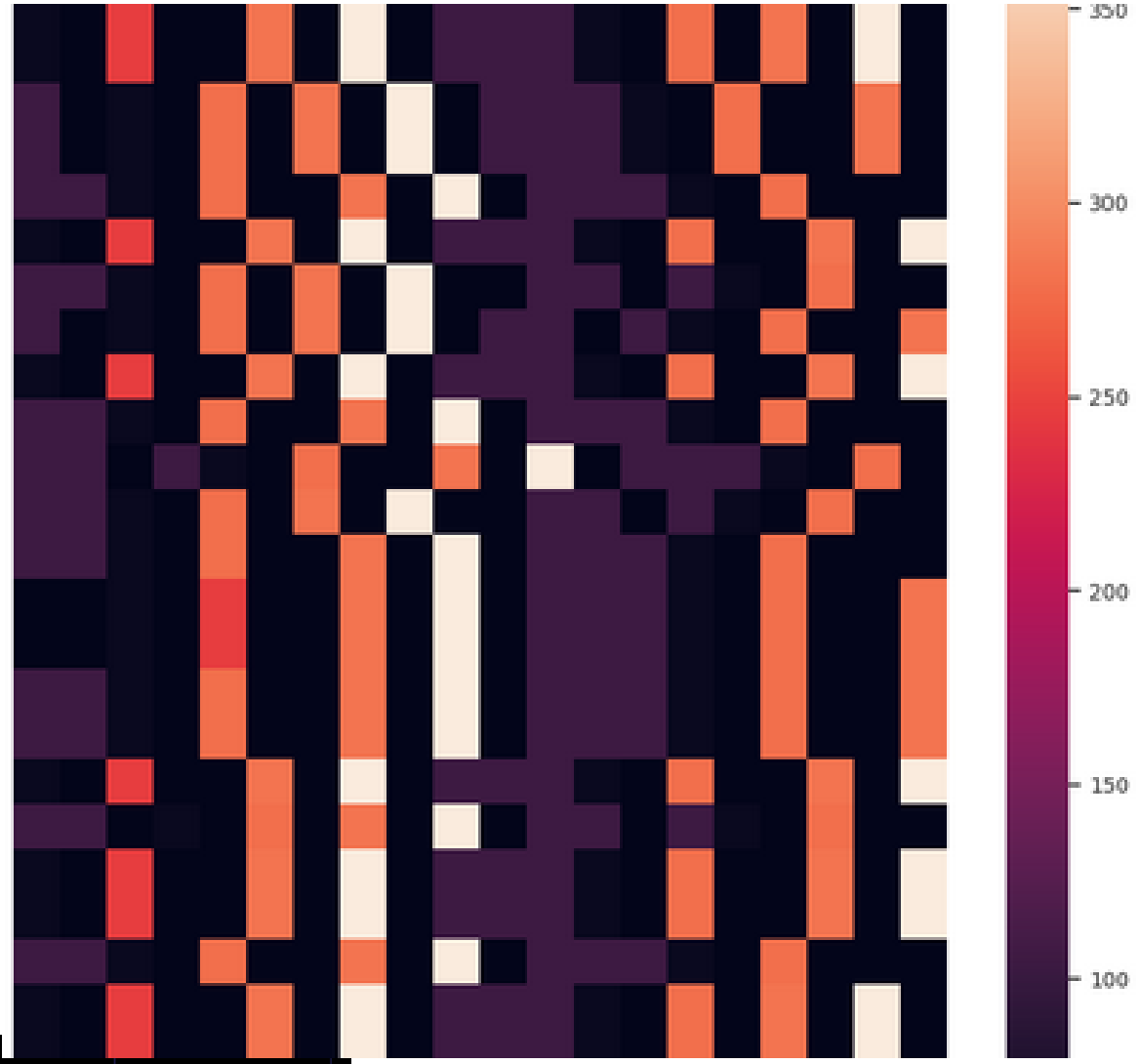
- striedavo sa učí G – generátor a D – diskriminátor
- po natrénovaní sa dá D – použiť na odhalenie nezvyčajnej prevádzky



Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Údaje o počítačovej sieti pre CNN - pre detekciu anomálií

- chceme použiť konvolučnú neurónovú sieť
- pripraviť vstupné údaje v čase
  - o prevádke v počítačovej sieti
    - údaje o paketoch (WireShark, ...)
      - polia z hlavičiek, veľkosti paketov, časy príchodov, intenzita príchodov
      - podľa typu protokolu, IP adereis, počtov, príznakových bitov, ...
    - údaje o tokoch
      - priemern veľkosť paketu, ...
  - logovacie údaje zo zariadení (firewall, smerovače, počítače, ...)
  - údaje o stave zariadení
- údaje môžeme
  - zobaiť napr. teplotnou mapou
  - použiť na trénovanie modelov ML

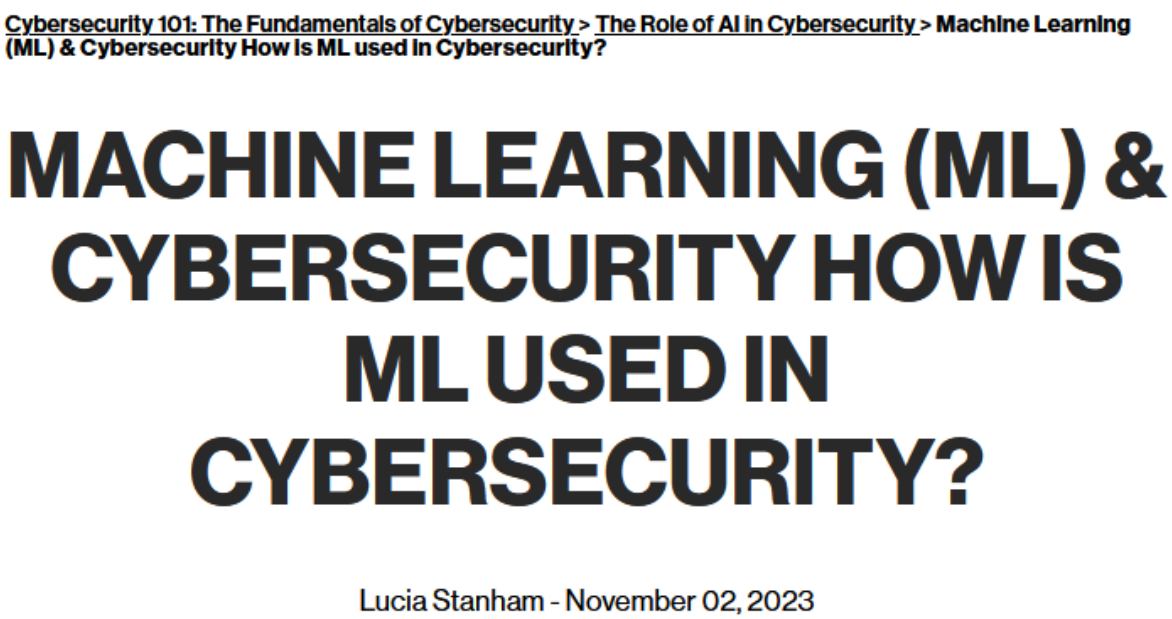


Prečo AI v kybernetickej bezpečnosti?

- Explózia objemu dát (logy, sieťová prevádzka, endpointy)
  → človek to sám nezvládne
- Útočníci používajú automatizáciu a AI → obrancovia musia držať krok
- AI - zrýchlenie detekcie, presnejšie korelácie, menej falošných poplachov cloudsecurityalliance.org+1
- Presun od reaktívneho k proaktívnemu modelu obrany
- Cieľ: znížiť „dwell time“ (čas pobytu útočníka v systéme)
  - zrýchliť incident response

Základné pojmy: AI, ML, DL v bezpečnosti

- AI: všeobecný pojem pre systémy, ktoré napodobňujú ľudské rozhodovanie
- Machine Learning: modely učiace sa z dát (supervised / unsupervised / reinforcement) crowdstrike.com
- Deep Learning: neurónové siete pre komplexné vzory (napr. sieťová prevádzka, obrázky, audio)
- V bezpečnosti sa najviac používajú ML a štatistické metódy
- GenAI/LLM: modely pracujúce s prirodzeným jazykom (chat, sumarizácia, generovanie detekcií)



Prečo tradičné nástroje už nestačia

- Podpisové AV a IDS nedokážu zachytiť nové, neznáme hrozby (zero-day – využije neznámu zraniteľnosť v softvéri, living-off-the-land – použijú legitímny softvér na škodlivé aktivity)
- Masívny počet alertov → SOC analytici sú preťažení
- Ručné korelácie naprieč SIEM, EDR, NDR, IAM sú pomalé
- Útoky sú distribuované (cloud, OT, IoT, SaaS)
- AI pomáha automatizovať triáž (rozdeľovanie podľa závažnosti), koreláciu, odporúčania ďalších krokov cloudsecurityalliance.org+1
- (AI in Cybersecurity: 5 Practical Use Cases for Stronger Defense, Published 07/01/2025)

3. Automation of Routine Tasks

Security teams often deal with an overwhelming number of alerts, many of which turn out to be false positives. AI-driven automation reduces this burden by filtering, categorizing, and responding to threats in real time. This not only improves response time but also allows security analysts to focus on complex threats that require human expertise.

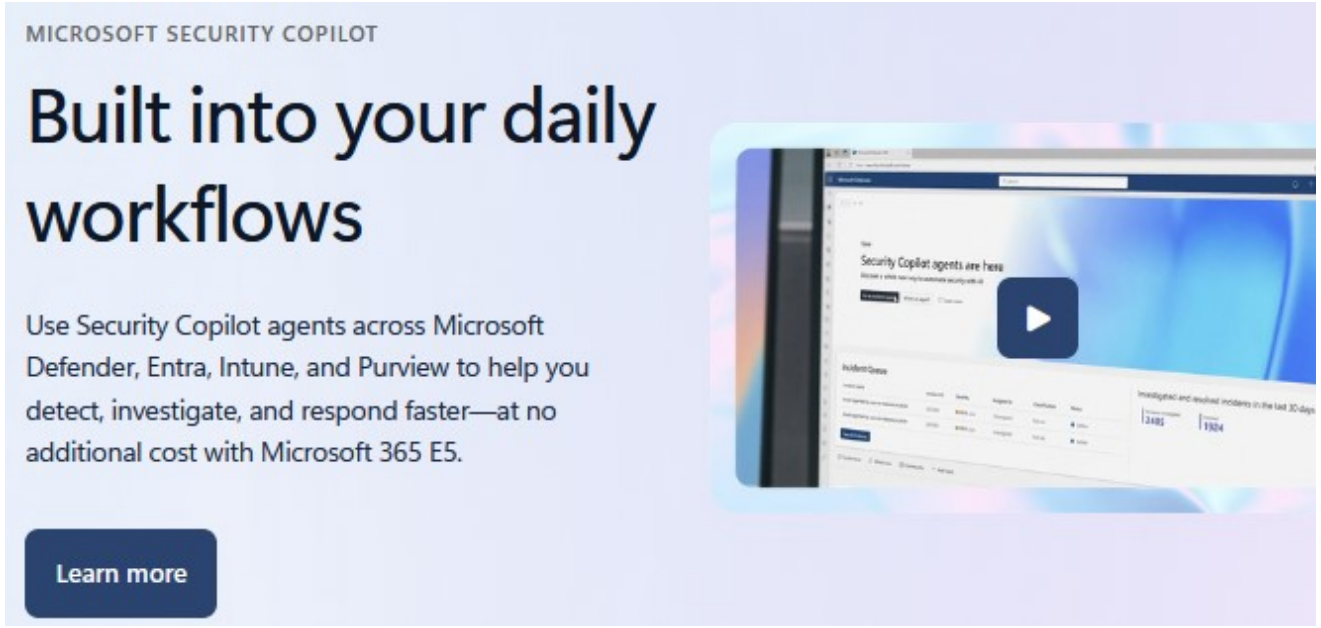
Common automation applications include:

- **Threat Analysis and Correlation:** AI correlates data from multiple sources to detect attack campaigns.
- **Incident Response Workflows:** AI automates security playbooks to contain and remediate threats faster.
- **Phishing Detection and Response:** AI scans emails for suspicious content, flagging or quarantining potential threats before they reach users.

By leveraging AI for routine security tasks, organizations can reduce response times and ensure critical threats receive immediate attention.

Typy AI modelov v kyberbezpečnosti

- Detekčné modely (classification, anomaly detection)
- Korelačné a grafové modely (user/entity behavior, útokové grafy)
- Generatívne modely (LLM, generovanie pravidiel, playbookov)
- Optimalizačné modely (prioritizácia zraniteľností, patch management)
- Agentické modely – samostatne konať (autonómne AI agenti v SOC) Microsoft+1



Použitie AI v SOC: Vysoká úroveň

- AI-pomocník pre SOC analýzu (chat s logmi, incidentmi)
- Automatizovaná triáž (triedenie podľa priority) alertov a priorizácia rizika
- Generovanie timeline (udaostí v čase) útoku a korelácia udalostí
- Návrh playbookov pre responzné akcie
  - postupy ako by ma tím reagovať na útok
  - jasné roly a zodpovednosti, ...
  - napríklad pre incidenty
    - Phishingové útoky
    - Malvér alebo ransomware útoky
    - Úniky dát
    - Neoprávnený prístup
    - exabeam.com
- Učenie sa zo spätnej väzby analytikov (reinforcement / active learning)

Examples of Automated Security Playbooks

Here are a few examples of security playbooks that can be used to automate incident response for specific threats.

Phishing Attack Response

Here are the possible steps of an automated playbook:

1. **Detection:** The automated system monitors incoming emails for signs of phishing, such as suspicious senders, links, and attachments. This step uses email filtering technologies and threat intelligence feeds to identify potential phishing emails.
2. **Alert:** Once a potential phishing email is detected, the system automatically alerts the security operations center (SOC) team via email and dashboard notifications. The alert includes details of the suspicious email and its characteristics.
3. **Isolation:** The email is automatically quarantined to prevent the recipient from accessing potentially harmful content. If the email has already been opened, the system checks if any links were clicked or attachments opened and isolates affected endpoints from the network.
4. **Investigation:** An automated script gathers information about the email, such as the sender's domain, IP address, and the nature of any linked content. This data is cross-referenced with known threat databases for further analysis.
5. **Remediation:** If the email is confirmed as phishing, the system automatically removes the email from all inboxes across the organization. For opened emails, the system initiates a malware scan on affected endpoints and applies necessary patches or isolation measures.
6. **User notification:** Affected users are automatically notified about the phishing attempt, with information on how to recognize such emails in the future and the importance of reporting suspicious messages.
7. **Update defenses:** The system updates its email filtering criteria and threat intelligence database based on the characteristics of the phishing attempt to prevent similar attacks.
8. **Report:** Generate a detailed incident report, including the timeline, actions taken, and lessons learned. This report is automatically shared with relevant stakeholders and used to refine future response strategies.

AI pre detekciu anomálií v sieti (NDR)

- Modely sa učia **normálne správanie** sieťovej prevádzky
- Odchýlky
  - laterálne pohyby
  - neštandardné protokoly
  - exfiltrácia dát (úmyselné odcudzenie dát – malvér, pishing, zamestnanci, slabiny v softéri či v sieti, na USB kľúč)
- Príklad:
  - Darktrace Self-Learning AI sa učí „patterns of life“ zariadení a používateľov darktrace.com+2
- Výhoda:
  detekcia aj úplne nových typov útokov
- Výzva: vysvetliteľnosť
  a redukcia falošných pozitív

Darktrace is the only vendor named the Customers' Choice in the 2025 Gartner Peer Insights Voice of the Customer for Network Detection and Response

AI pre endpoint ochranu (EDR/XDR)

- Modely analyzujú správanie procesov, súborov a pamäte
- ML modely nahrádzajú tradičné AV podpisy
- CrowdStrike Falcon: využíva „signatureless AI/ML“ a Threat Graph na koreláciu miliárd udalostí denne crowdstrike.com+1
- Detekcia ransomvéru, fileless útokov, exploit chainov
- Lokálne modely na endpointoch + cloudové modely pre pokročilú analýzu

CrowdStrike Introduces Enhanced Endpoint Machine Learning Capabilities and Advanced Endpoint Protection Modules

— Company continues to accelerate pace of replacement of legacy AV solutions in both enterprise and SMB markets —

AI v SIEM a XDR platformách

- Automatická normalizácia, obohacovanie a korelácia logov
- Dynamické prahové hodnoty pre detekčné pravidlá
- Prioritizácia incidentov na základe rizika a kontextu
- Príklad:
  Google Security Operations (Chronicle) - používa Gemini na vyhľadávanie cez prirodzený jazyk, gebneruje pravidká, odpovedá na bezpečnostné otázky, vytvára playbook-y a sumarizuje gemini-chronicle
- Integrácia so SOAR pre automatizované playbooky

Key features of Gemini in Google SecOps

Generate search queries

Generate a YARA-L rule using Gemini

Assistance with threat intelligence and security questions

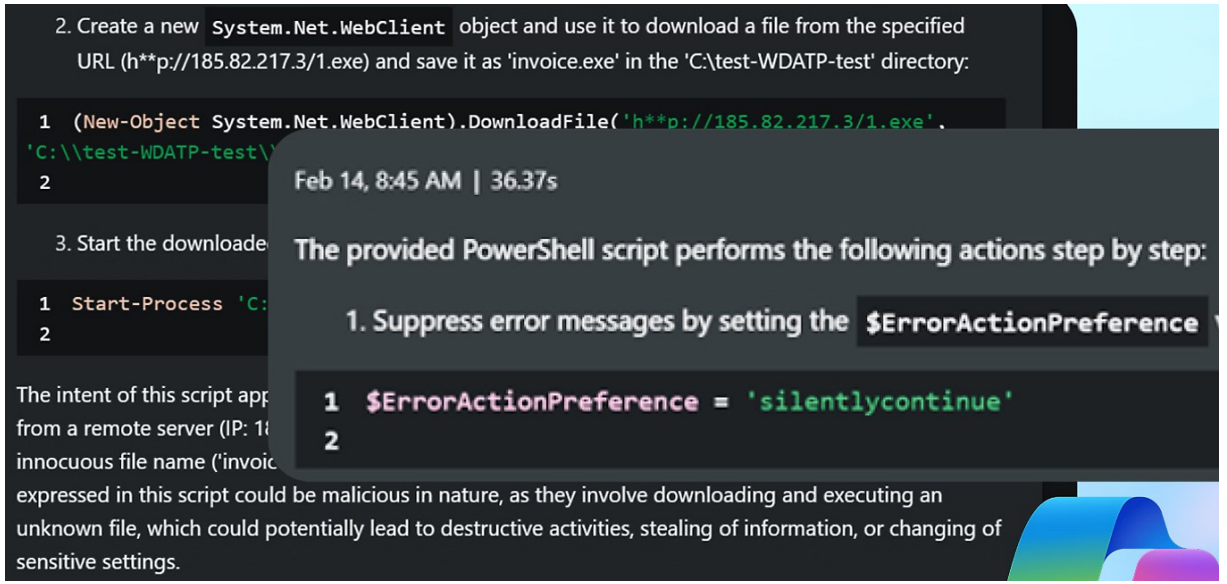
Get documentation summaries

Create and edit a playbook

Use the Gemini case summary widget

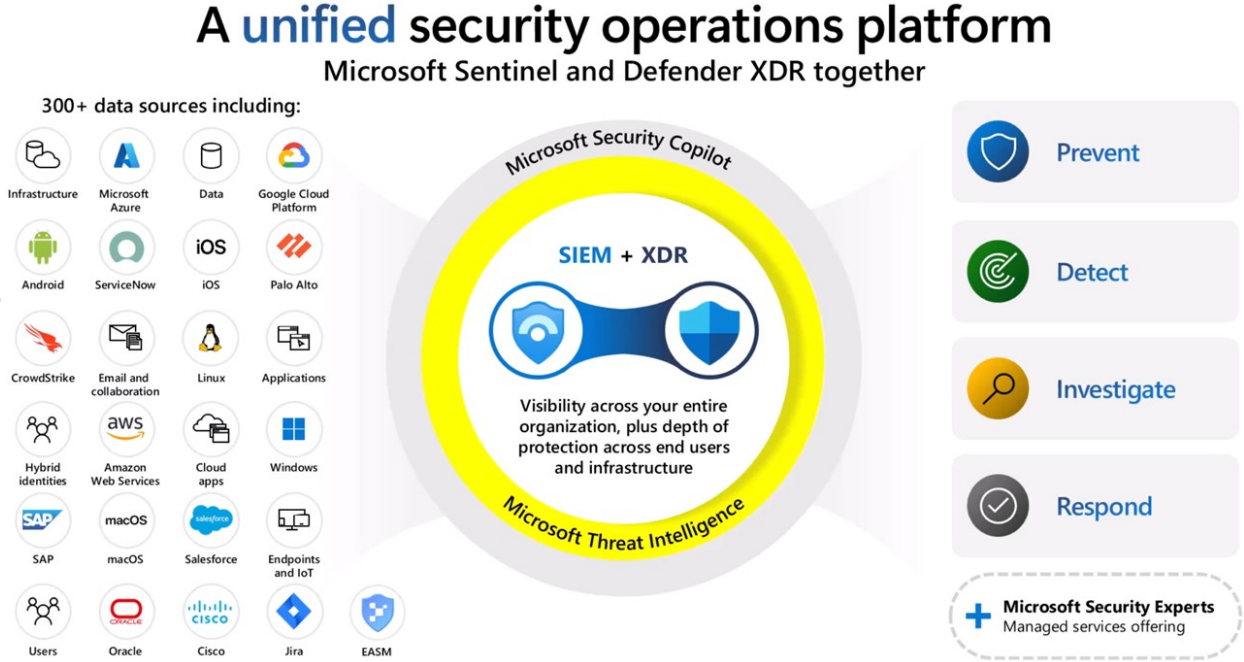
Generatívna AI (LLM) ako „copilot“ v SOC

- Chat rozhranie nad SIEM/EDR: „Vysvetli mi tento incident“
- Sumarizácia alertov, logov, e-mailovej komunikácie a ticketov
- Návrh dotazov (napr. KQL, YARA-L, Sigma) na základe prirodzeného jazyka
- Učiaci sa asistent pre junior analytikov
- Microsoft Security Copilot: LLM + 100+ triliónov bezpečnostných signálov denne Microsoft+1



AI agenti pre automatizáciu bezpečnosti

- AI agent = komponent, ktorý autonomne vykonáva kroky (triáž – treidenie (z fr.) podľa závažnosti – na ktorý reagovať najskôr, izolácia hostu, blokovanie URL)
- Microsoft: Security Copilot AI agents integrovaní do Defender, Entra, Intune atď. Microsoft+2 1: Overview of Microsoft Defender Threat Intelligence with demo
- Príklady úloh: spracovanie phishing e-mailov, monitorovanie zraniteľností, enforcement politiky
- Dôležité: „human in the loop“ pri kritických zásahoch
- Potreba auditovateľnosti a schvaľovacích postupov



AI pre detekciu phishingu a podvodov

- Analýza obsahu e-mailu (text, DOM, URL, prílohy) pomocou strojového učenia (ML/DL)
- Detekcia phishingových šablón, brand spoofingu, BEC kampaní
- LLM-based filtre na identifikáciu sociálneho inžinierstva
- Analýza reputácie domén a IP + behaviorálne signály používateľa
- Prepojenie s anti-fraud systémami v bankách a e-commerce

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

AI v analýze malware

- Automatická klasifikácia vzoriek (static + dynamic analysis)
- Modely predikujú rodinu malware a TTPs na základe príznakov (features - API calls, strings, PE hlavička)
- Clustering podobných kampaní a generovanie YARA pravidiel
- Sandboxy využívajú ML na zníženie času analýzy (prioritizácia zaujímavých vzoriek)
- Prepojenie na threat intel feedy a IOC databázy

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

User & Entity Behavior Analytics (UEBA)

- Modelovanie **správania** používateľov, účtov, hostov, servisných účtov
- Detekcia insider threat, kompromitovaných účtov, laterálneho pohybu
- Skórovanie rizikovosti subjektov (risk score)
- Kombinácia štatistických modelov a grafových algoritmov
- Príklad: mnohé XDR/SIEM (Microsoft, Google, **CrowdStrike**) majú UEBA ako embedded komponent crowdstrike.com+1

AI pre Threat Intelligence

- Automatická extrakcia IOC (Indicator of Compromise) / TTP (Tactics, Techniques, and Procedures) z článkov, blogov, reportov a logov
- Klasifikácia a tagging kampaní podľa MITRE ATT&CK
- LLM na sumarizáciu spravodajských reportov pre manažment
- Gemini v Google SecOps: odpovede na TI otázky a sumarizácie actorov a IOC Google Cloud Documentation+1
- Prepojenie s CTI platformami a SIEM pre auto-obohacovanie alertov

AI v vulnerability managemente

- **Predikcia** exploitability zraniteľností nad rámec CVSS
- **Prioritizácia** patchovania podľa kontextu (expozícia, aktívne zneužívanie, business impact)
- Korelácia so skutočnou aktivitou v logoch a na sieťovej vrstve
- Doporučenie mitigácií (konfigurácie, segmentácia, hardening)
- Integrácia s ticketingom a CMDB

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

AI v oblasti Identity & Access Management (IAM)

- Detekcia anomálnych prihlasovaní (lokalita, device, čas, risk signály)
- Adaptívna MFA (viacfaktorová autentifikácia) na základe rizikového skóre
- Anomálne správanie privilegovaných účtov (PAM)
- Automatizované odporúčania pre revízie prístupov
- Kontextové politiky: „just-in-time“ prístupy a least privilege

AI a OT/ICS bezpečnosť

- Modely sledujú špecifické „patterns of life“ OT zariadení (Operational Technology - prevádzkové technológie), PLC, SCADA Website Files, a ICS - Industrial Control Systems (priemyselné riadiace systémy)
- Detekcia odchýlok, ktoré môžu znamenať sabotáž alebo chybné nastavenia
- Nutnosť modelov odolných voči šumu a neúplným dátam
- Integrácia s fyzickou bezpečnosťou (video, senzory)
- Príklad: Darktrace AI aj v OT segmentoch (elektrárne, výroba, doprava)

AI v cloude a v prostredí SaaS

- Analýza cloudových logov (CloudTrail, Azure Activity, GCP Audit Logs) pomocou ML
- Detekcia misconfigurations (public buckets, otvorené porty, neštandardné identity)
- CASB/SSPM s ML na sledovanie SaaS použití a dátových tokov
- AI-poháňané politiky DLP (obsah + kontext)
- Integrácia s CSPM nástrojmi

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Firmy: Darktrace

- „The Essential / ActiveAI Security Platform“ so Self-Learning AI darktrace.com+2
- Učí sa z reálneho správania organizácie, nie len zo signatúr
- Pokrýva e-mail, sieť, endpoint, cloud, OT
- Autonómna odpoveď (AI-driven response) – selektívne blokovanie aktivít
- Cielené na stredné a veľké organizácie, vrátane kritickej infraštruktúry Financial Times

Firmy: CrowdStrike

- Falcon platforma: endpoint, cloud, identity, log scale XDR
- Threat Graph analyzuje 30+ miliárd udalostí denne zo 176+ krajín crowdstrike.com+1
- Používa behaviorálny ML na detekciu známych aj neznámych útokov
- Kontextové IOC, TTP a hunting vďaka AI
- Silná integrácia s incident response službami

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Firmy: Microsoft (Defender + Security Copilot)

- Microsoft Defender (XDR) využíva ML na endpoint, identitu, cloud a e-mail
- Security Copilot: špeciálny LLM pre bezpečnosť, napojený na signály z Microsoft ekosystému Microsoft+1
- AI agenti priamo v nástrojoch (Defender, Entra, Intune, Purview) Microsoft+1
- Scenáre: triáž phishingu, investigácia incidentov, tvorba detekcií
- Vhodné pre organizácie s Microsoft 365 E5

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Firmy: Google SecOps (Chronicle + Gemini)

- Google Security Operations = Chronicle SIEM+SOAR + SecOps konzola Google Cloud
- Gemini v SecOps: Q&A nad bezpečnostnými dátami, TI, generovanie detekcií Google Cloud Documentation+1
- Podpora jazyka YARA-L a prirodzeného jazyka na detekčné pravidlá
- Experimenty s AI bez narušenia produkcie (SecOps labs)
- Zameranie na vysokú škálovateľnosť a low-cost uloženie logov

Ďalší hráči: SentinelOne, Palo Alto, Sophos, atď.

- SentinelOne: AI-poháňaný EDR/XDR so silným autonómnym response
- Palo Alto Networks: Cortex XDR, XSIAM – AI pre koreláciu signálov z viacerých domén
- Sophos: integrácia Sophos Intelix (TI) do Microsoft Security Copilot a M365 Copilot AI Bawaba
- Mnoho menších vendorov využíva AI aspoň v čiastkových moduliach (UEBA, sandbox, phishing)

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

AI ako dvojsečná zbraň: útočníci

- AI používaná na generovanie phishingových e-mailov a deepfake hlasu/videa
- Automatizované vyhľadávanie zraniteľností a exploitov
- LLM-as-a-service pre útočníkov (skriptovanie, obchádzanie EDR, social engineering)
- Príklad: kampaň využívajúca AI na riadenie rozsiahlych útokov (Anthropic incident) AP News
- Potreba detekcie AI-generovaných kampaní

Hlavné prínosy AI pre kyberbezpečnosť

- Rýchlejšia detekcia a skrátenie „mean time to detect/respond“ cloudsecurityalliance.org+1
- Schopnosť spracovať obrovské množstvá dát v reálnom čase
- Znižovanie falošných pozitív a lepšia priorizácia
- Podpora menej skúsených analytikov
- Automatizácia rutinných úloh → sústredenie na komplexné incidenty

Limity a riziká AI riešení

- Závislosť na kvalite a reprezentatívnosti tréningových dát
- Model drift pri zmene prostredia a vzorov správania
- „Black box“ modely s nízkou vysvetliteľnosťou
- Možnosť zneužitia (prompt injection, data poisoning)
- Falošný pocit bezpečia pri slepej dôvere v AI

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

. Explainable AI (XAI) v bezpečnosti

- Potreba vysvetliť, prečo bol incident označený ako hrozba
- Techniky
  - feature importance
  - rule extraction
  - vizualizácia anomálií (2D teplotné mapy, ...)
- Dôležité pre forenznú analýzu a audit
- Podpora compliance (NIS2, DORA, ISO/IEC 27001)
- Trend: vendorov tlačia zákazníci na lepšiu transparentnosť modelov

Riadenie dát pre AI (data governance)

- Kvalita logov a telemetrie
  = kvalitné AI výstupy
- Normalizácia, deduplikácia, obohacovanie (TI, CMDB, IAM)
- Ochrana citlivých dát pri použití LLM (PII, tajné informácie) cynet.com
- On-prem vs. cloud vs. hybridné spracovanie
- Data retention a právne/regulačné požiadavky

Bezpečné používanie LLM v bezpečnostnom stacku

- Riziko úniku citlivých informácií pri vkladaní logov, konfigurácií, kódu
- Potreba „private LLM“ alebo tenant-isolated služieb
- Politiky pre používanie LLM (čo je a nie je dovolené vkladať)
- Monitorovanie promptov a odpovedí (logging)
- Hardening
  - filtrácia promptov
  - detekcia prompt injection
  - výstupné filtre

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

AI v incident response a forenznej analýze

- Automatické generovanie incident timeline z logov a artefaktov
- Sumarizácia diskových a pamäťových nálezov
- Návrh ďalších forenzných krokov (čo ešte skontrolovať)
- Prepojenie s SOAR
  - automatická izolácia hostu
  - reset hesiel
  - blokovanie IOC
- Post-incident reporting pre manažment a regulátora

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

AI v SOCaaS a MSSP službách

- Poskytovatelia SOCaaS používajú AI na škálovanie tímu analytikov
- Automatizovaná triáž pre stovky klientov súčasne
- Štandardizované playbooky + adaptívne AI rozhodovanie podľa klienta
- Zákazník často dostáva AI funkcie „v balíku“ v rámci služby
- Dôležité: jasný popis, kedy koná AI a kedy ľudia

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Ekonomický pohľad: AI vs. ľudské kapacity

- Nedostatok kvalifikovaných odborníkov → AI pomáha zaplniť medzeru
- AI neznamená zrušenie SOC, ale zvýšenie produktivity
- ROI: menej incidentov, nižšie škody, rýchlejšia reakcia
- Nutné investície: licencie, infraštruktúra, tréning tímu
- Benchmarky vendorov: často demonštrujú zníženie MTTR o desiatky percent

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Regulačné a etické aspekty

- EU AI Act: rizikové kategórie AI systémov (bezpečnostné AI môže spadať do „high-risk“ oblasti)
- NIS2/DORA: požiadavky na riadenie rizík, logging, incident reporting
- Etika: minimalizácia biasu v modeloch (nespravodlivé skórovanie používateľov)
- Transparentnosť voči používateľom o používaní AI
- Audity a nezávislé hodnotenie AI riešení

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Budovanie AI-ready SOC

- Základ: kvalitný zber logov (SIEM), štandardizované playbooky, CMDB
- Identifikácia use-casov, kde AI prinesie rýchly prínos (triáž, UEBA, TI)
- Postupná integrácia: POC → pilot → produkcia
- Nastavenie metrik (MTTD, MTTR, false positives, analyst productivity)
- Priebežné ladenie modelov a playbookov

Skillset nového SOC analytika

- Znalosť AI/ML **konceptov** a ich **limitov**
- Schopnosť pracovať s detekčnými jazykmi + LLM (**prompt engineering**)
- **Kritické** myslenie: nebrať AI výstupy ako „absolute truth“
- Znalosť cloudových **platforiem a API**
- Komunikačné zručnosti – **vysvetliť AI-založené zistenia biznisu**

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Praktický príklad: AI-podporovaná analýza incidentu

- Alert z EDR: podozrivé správanie procesu → AI doplní kontext (história, podobné incidenty)
- SIEM/LLM: automatická sumarizácia udalostí a návrh hypotéz
- UEBA: zvýšené rizikové skóre používateľa
- SOAR: navrhne kroky – izolácia hostu, reset hesla, blokovanie IP
- Analytik len potvrdzuje/koriguje kroky a dopĺňa ľudský kontext

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Praktický príklad: Phishing s AI podporou

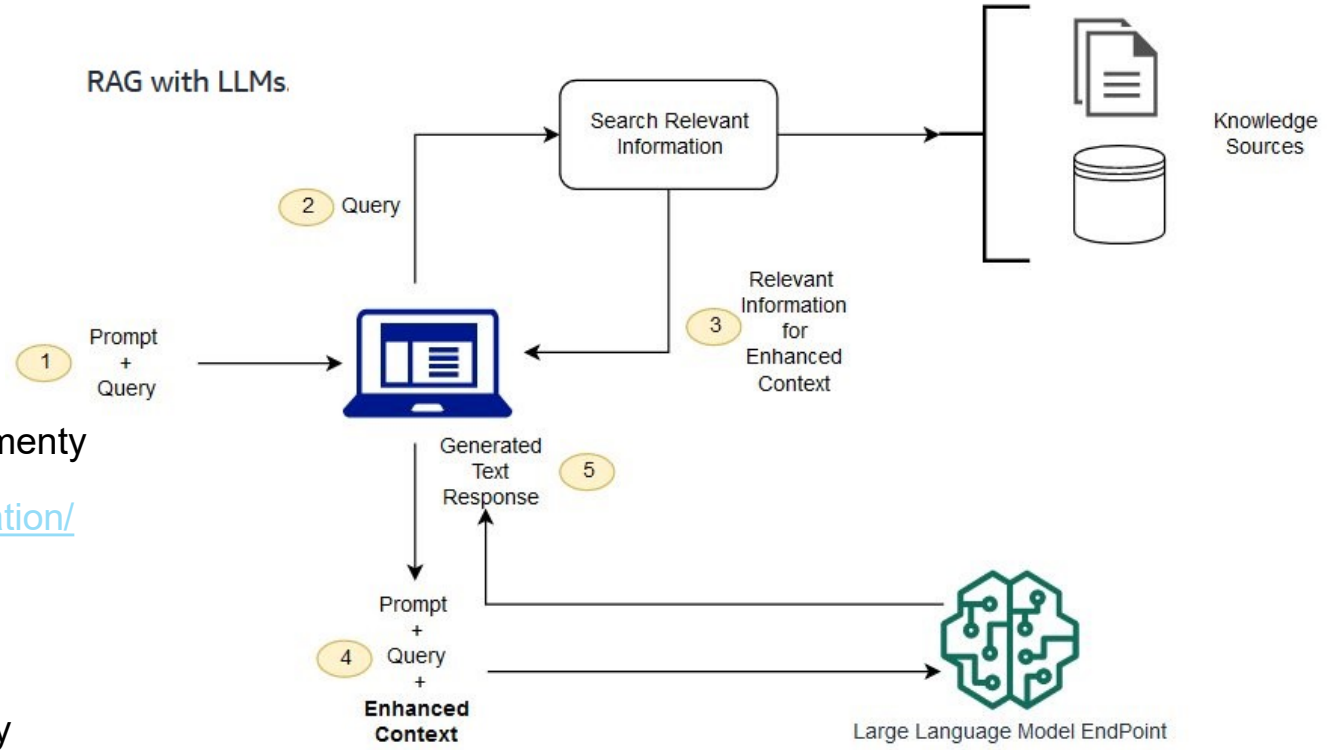
- E-mail detegovaný ML klasifikátorom ako phishing
- LLM vysvetlí, ktoré časti textu sú podozrivé (ton, jazyk, linky)
- SOAR + AI agent: izoluje e-mail, skenuje mailboxy, blokuje doménu
- Generovanie notifikácie pre používateľa v „ľudskej reči“
- Vytvorenie znalostného článku pre budúce incidenty

Ako vyberať AI riešenia pre organizáciu

- Use-case driven prístup (čo konkrétne chceme riešiť)
- Integrácie s existujúcim stackom (SIEM, EDR, ticketing)
- Kvalita a objem tréningových dát vendora (globálny vs. lokálny footprint) crowdstrike.com+1
- Transparentnosť modelov a možnosti ladenia
- Licenčný model, TCO, lokálna podpora

Trendy na najbližšie roky

- Viac agentických AI systémov v SOC (autonómnejšie rozhodovanie) Microsoft+1
- Úzke prepojenie AI v IT a OT bezpečnosti
- Nástup špecializovaných bezpečnostných LLM (domain-specific)
  - **RAG (Retrieval-Augmented Generation) s LLM**
    - LLM sa netrénuje
    - pridáme vektorovú databázu (embeddings)
    - pre otázku sa z db vyberú relevanté informácie
    - tie sa s otázkou pošlú do LLM
    - rozšírenie znalostí o špecifickú oblast, aktuálne dokumenty
    - ľahko aktualizovať vedomosti aws.amazon.com/what-is/retrieval-augmented-generation/
  - **finetooning LLM**
    - dotrénuj LLM
    - dlhšie trvá
    - vhodné na špecifické veci, čo sa často nemenia, napr. aby generoval údaje v zaužívanom formáte fimry
- AI-poháňané red teaming, simulácie útokov
- Prísnejšia regulácia a štandardy pre bezpečnostné AI



Záver: AI ako multiplikátor obrany, nie náhrada ľudí

- AI dramaticky zvyšuje schopnosť brániť sa voči moderným hrozbám
- Ľudský faktor ostáva kľúčový (rozhodovanie, kontext, etika)
- Úspech závisí od správneho dizajnu procesov, dát a školení
- Odporúčanie: začať malými use-casmi a iteratívne škálovať
- AI v kyberbezpečnosti je evolúcia SOC, nie „magická krabička“

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Reflexia (otázky + vyberme z odpovedí)

- Umelá inteligencia v kybernetickej bezpečnosti
  1) vie zlepšiť schopnosť detekcie hrozieb a anomálií analýzou veľkého množstva údajov, ale nevie sa učiť z minulých útokov
  2) nevie automaticky reagovať na incidenty, blokovať škodlivé aktivity a izolovať infikované systémy
  3) vie pomocou analýzy predikovať potenciálne hrozby - na základe analýzy historických údajov a aktuálnych trendov
  4) nepoužívajú ju kyberzločinci na vytváranie sofistikovanejších útokov (phishingové kampane, deepfake videá, oklamanie obetí a šírenie dezinformácií)

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY
KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Financované Európskou úniou
NextGenerationEU

PLÁN OBNOVY

MINISTERSTVO INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA A INFORMATIZÁCIE SLOVENSKEJ REPUBLIKY

KOMPETENČNÉ CENTRUM KYBERNETICKEJ BEZPEČNOSTI ŽILINSKEJ UNIVERZITY V ŽILINE

Ďakujem za pozornosť

Umelá inteligencia v KB

Moderné technológie, bezpečnosť cloudu a IoT (Blok VII)

Kurz: Špecialista kybernetickej bezpečnosti vo verejnej správe

Ondrej Škvarek

KC KYB UNIZA, https://kc.uniza.sk/

skvarek@uniza.sk