“A (kiber)biztonság gyakorlatAI” konferencia, 2025. április 10, NKE

Kiberbiztonság és AI: kockázatok és megoldások

Varga Pál
tanszékvezető
BME-TMIT

MŰEGYETEM 1782

BME VILLAMOSMÉRNÖKI ÉS INFORMATIKAI KAR

TMiT

BME

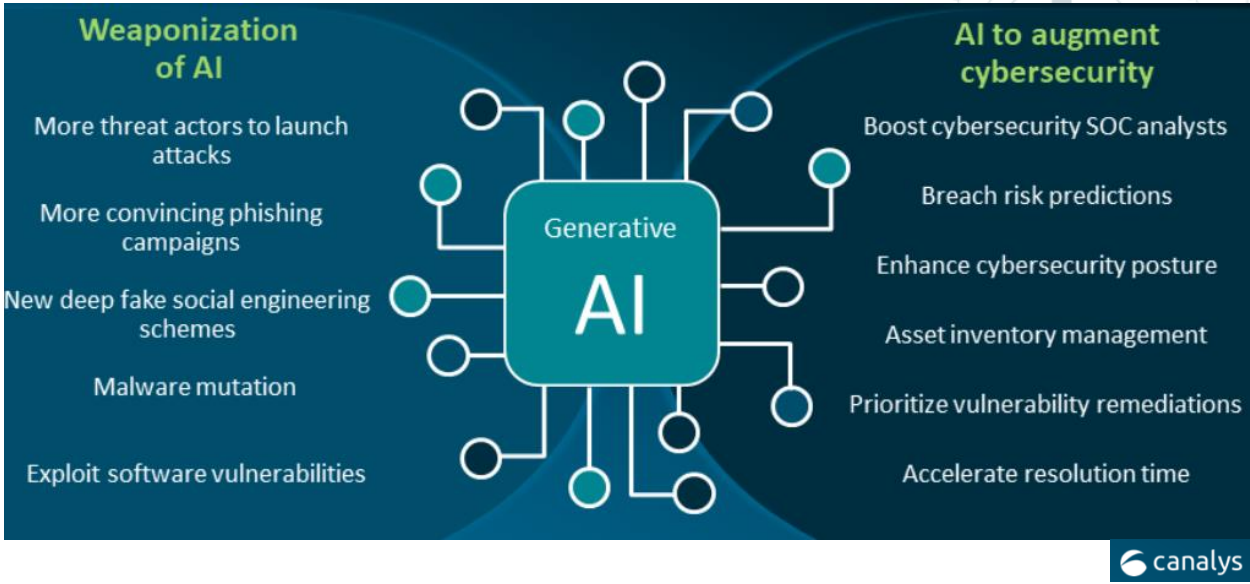
TÁVKÖZLÉSI
ÉS
MESTERSÉGES INTELLIGENCIA
TANSZÉK

BME-TMIT Bemutatkozás

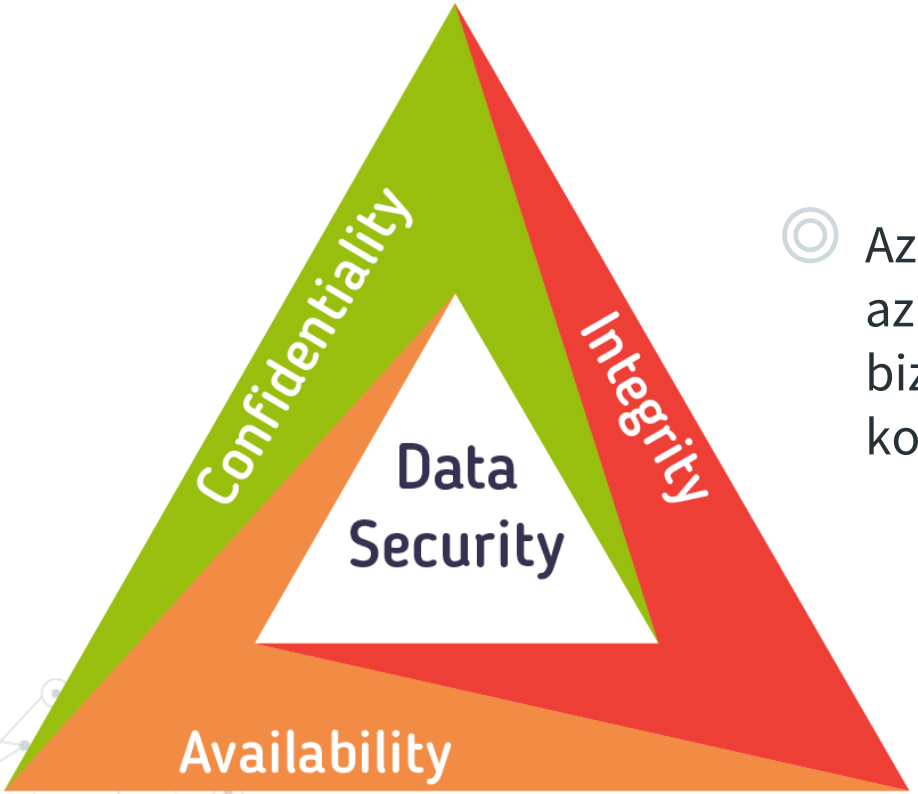
https://app.colossyan.com/videos/4039a741-9f18-4473-8d28-590c75ac20d0

Kiberbiztonság: trendek

- Generatív MI
- Ransomware
  - (~1000Mrd USD in 2025)
- Jármű hackelés
- Insider threats
- Deepfake
- “S” means security in IoT
- Zero Trust cybersecurity



Az Információbiztonság alapelvei



- Az információbiztonság valójában az adatok és információk biztonságához kapcsolódó kockázatok folyamatos kezelése

Műszaki kihívások

Az MI rendszerek összetettsége

- Többszintű modellek, bonyolult döntési folyamatok
- Nehéz azonosítani, hogy „miért” hozott egy adott döntést

Az adatok minősége és elérhetősége

- A tanításhoz használt adatok gyakran hiányosak, elavultak vagy torzítottak
- A rossz adat rossz modellt eredményez

Érthetőség és magyarázhatóság

- Sok MI-modell „fekete dobozként” működik
- Az átláthatóság hiánya biztonsági kockázatot jelenthet

Hibás riasztások (false positive/negative)

- Túl sok téves riasztás csökkenti a rendszer hitelességét
- A kihagyott támadások (false negative) pedig veszélyesek

Etikai és jogi kihívások

⚖️ Torzítások a tanító adatokban

- A torz adathalmazok **biztonsági réseket** hagynak
- Például alulreprezentált csoportok nem megfelelő védelmet kapnak

🔐 Adatvédelmi aggályok és szabályozások

- MI rendszerek gyakran kezelnek érzékeny adatokat
- GDPR, HIPAA és egyéb adatvédelmi rendeletek betartása kritikus

❓ Felelősség kérdése az autonóm döntéseknél

- Ki a felelős egy MI-rendszer hibás vagy káros döntéséért?
- Szükség van jogi és etikai elszámoltathatóságra

📄 Szerzői jog és MI által generált tartalmak

Kié a jog, ha az MI hoz létre szöveget, képet, kódot?

Új típusú szellemi tulajdonvédelmi kérdések

Generatív MI a kiberbűnözők kezében

Kép: Szuchy Roland

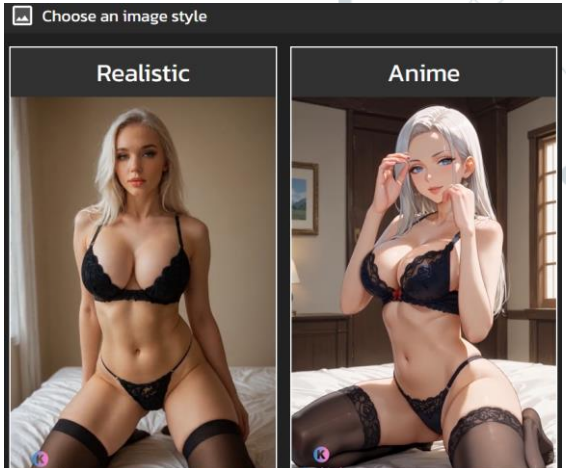
A generatív MI fokozza a kibertámadások hatékonyságát…

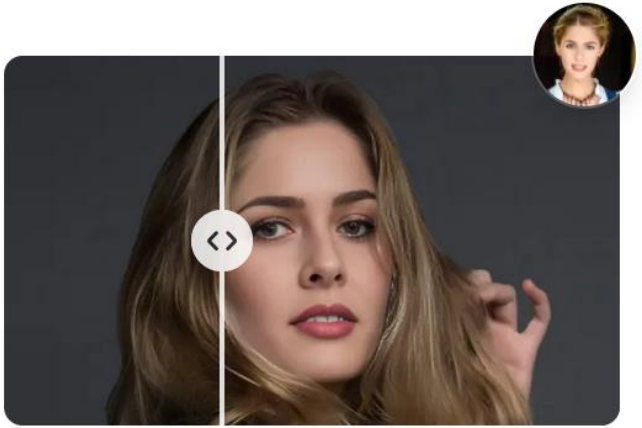
- Nemcsak eszköz, hanem erőforrás is
- Mai valóság
  - az MI által generált tartalmak (szöveg, kép, hang) “szinte” megkülönböztethetetlenek a valóstól.
- Testreszabás:
  - célzott támadásokhoz könnyen adaptálható, személyre szabható
- Automatizálás: phishing, malware kreálás, DDoS, chatbot,…
- Skálázódás
  - egyetlen támadó százezres++ csoportot elér
  - többnyelvű támadás – javuló minőség
  - tömeges, személyre szabott megtévesztés

Megtévesztés: Deepfake

- **Best DeepFake AI Tools for Face Swapping on Video & Photo (2025.04.08)**
  - Herahaven – Best AI Girlfriend App
  - BasedLabs Face Swap Tool – Face Swap Changer
  - Ghibli Filter & AI Face Swap – by MaxStudio

https://gdpratlas.com/best-deepfake-ai-tools-for-face-swapping-on-video-photo-updated-2025



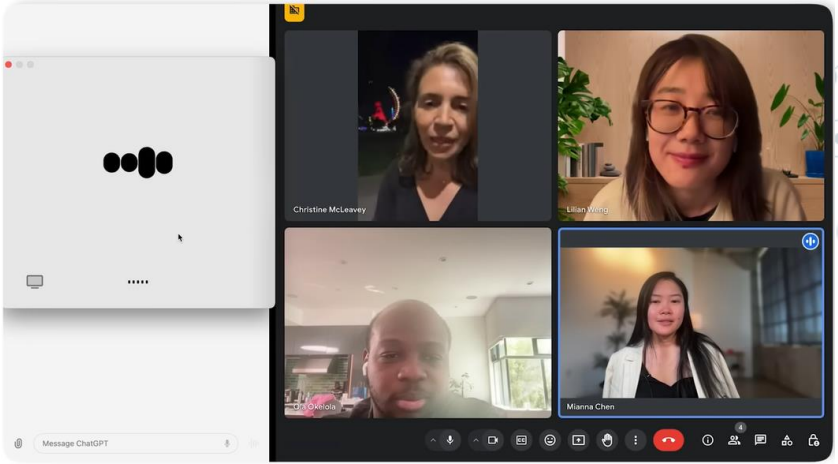


Deepfake – dokumentumok, kommunikáció

- Hitelesnek tűnő weboldalak, dokumentumok, üzenetek
  - generálása egyszerű és gyors
- A generált tartalmak megbízhatóságának illúziója

  hamis biztonságérzetet kelt a célpontban.
- Deepfake AI Agent

  – online meetingeken



További fenyegetések a Generatív MI térnyerésével...

- **Automatizált adathalászat és manipuláció**
  - Hiperrealisztikus e-mailek, chatbotok támadó célokra
  - Célzott támadások automatizálva
- **Generatív Malware**
  - MI által generált új, ismeretlen kártevők
  - Változó, nehezen észlelhető viselkedésminták
- **Modell-támadások (Adversarial Attacks)**
  - MI rendszerek félrevezetése szándékosan manipulált bemenetekkel
  - Adatmérgezés, manipulált tanulási folyamatok

Generatív MI – a jó csapat kezében

Kép: Szuchy Roland

Kibervédelmi technikák – az MI használatával

🔍 MI-alapú fenyegetésészlelés és prediktív elemzés

- Az MI képes valós időben észlelni szokatlan viselkedési mintákat
- Előrejelzi a potenciális támadásokat a múltbéli adatok alapján

⚙️ Automatizált incidenskezelés és válasz

- Az MI képes gyorsan reagálni a támadásokra emberi beavatkozás nélkül
- Automatikus izoláció, zárolás, naplózás, helyreállítás

🔁 Folyamatos tanulás és adaptív védelem

- A rendszer folyamatosan frissíti saját szabályait új támadások alapján
- Alkalmazkodik az új fenyegetésekhez és változó környezethez

Robosztus és biztonságos MI-rendszerek építése

☒ Az MI védelme: robusztus tanítási módszerek

- Adversarial training: támadások szimulálása a modell tanítása során
- Ellenállás a manipulált inputokra: robusztus gépi tanulás

🔐 Adatbiztonság és adatvédelem

- **Federated learning**: az adatok helyben maradnak, csak a modellek tanulnak
- Minimalizálja az érzékeny adatok központi kezelését

☒ Magyarázhatóság és átláthatóság

- Az MI döntéseinek érthetővé tétele (pl. SHAP, LIME, XAI módszerek)
- Transzparens modellek növelik a bizalom szintjét, támogatják a megfelelést

CIS (Center of Internet Security) Top 18 Security Controls

Generatív MI alkalmazása a kiberbiztonságban

- **Védekezés új lehetőségei**
  - Szintetikus tanítóadatok adatok használata
  - Anomáliák generálása
- **Proaktív védelem Generatív MI-vel**
  - Előrejelzés és megelőző válaszok
  - Szimulációs környezetek valósághű fenyegetések modellezésére
- **Kutatás és Innováció támogatása**
  - Eszközpark fejlesztése
  - Új védelmi eszközök gyors fejlesztése
  - Változatos szcenáriók tesztelése generatív technológiákkal

Megoldások és jó gyakorlatok – Generatív MI

- **Műszaki megoldások**
  - deepfake detektálás
  - szintetikus tartalom azonosítása
  - adatvédelem, MI-biztonságos architektúrák
  - Komplex DDoS detekció
- **Etikai és szabályozási Keretek**
  - Etikus MI-használat elvei
  - Aktuális és jövőbeli szabályozások (EU AI Act, NIS2, stb.)

Szervezeti ajánlások

- Tudatosságnövelő tréningek, belső protokollok frissítése
  - Rendszeres képzések a deepfake, adathalászat és automatizált támadások felismeréséről
  - Gyakorlati példák, szimulált támadási forgatókönyvek
  - Frissített belső szabályzatok
- Készenléti tervek generatív MI incidensekre
  - Gyanús tartalmak (pl. hamis videók, e-mailek) hitelesítése
  - Kommunikációs stratégia belső és külső érintettek felé
  - Gyors reagálás és helyreállítási lépések
- Együttműködés különböző szakterületek között
  - IT, biztonsági, jogi, etikai, kommunikációs, HR: „AI Response Team”
  - Rendszeres belső workshopok
  - Egységes szervezeti álláspont és eljárásrend kialakítása

A Természetes és Mesterséges Intelligencia együttműködése

Együttműködés az M.I-vel – a módszer

- Az új technológiák a hasznunkra válnak, ha jól alkalmazzuk
- Tanuljuk meg – és kövessük a fejlődést – ha lehet nyújtsunk újat

- Hallgatók és oktatók: **csináljuk közösen – jól**.

A saját életünk ez – a jelent és a jövőt alakítjuk

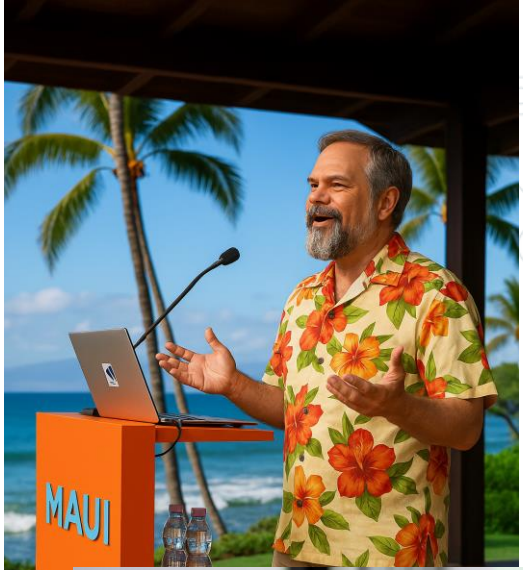
Pal Varga • You
BME - Head of Department, Telecommunications and Artificial Intelligence
1w •

I'm happy to share that I'm starting a new position as Vice Chancellor for Budapest-Maui Internships at the University of Hawaii Maui College!

Starting a New Position

Gergely Biczok and 204 others 34 comments

Reactions



AWARENESS

- 🚀 Panel Discussion on
- **“Cybersecurity Challenges in the 6G Cloud-Edge Continuum”**
- at IEEE/IFIP NOMS 2025 in Honolulu, Hawaii, USA
- ✨ Expert Panelists:
  Dr. Pawani Porambage, VTT Technical Research Centre of Finland
  Prof. Dr Shahzad Sarfraz, FAST National University of Computer and Emerging Sciences, Pakistan
  Dr. Engin Zeydan, CTTC, Spain
  Dr. Pal Varga, Budapest University of Technology and Economics, Hungary

A jövő
az MI kezében
van
Ne engedjük el ☺

TMiT

BME

TÁVKÖZLÉSI
ÉS
MESTERSÉGES INTELLIGENCIA
TANSZÉK

Köszönöm

a figyelmet

Varga Pál

TMiT