„Az MI gondolatAI?”

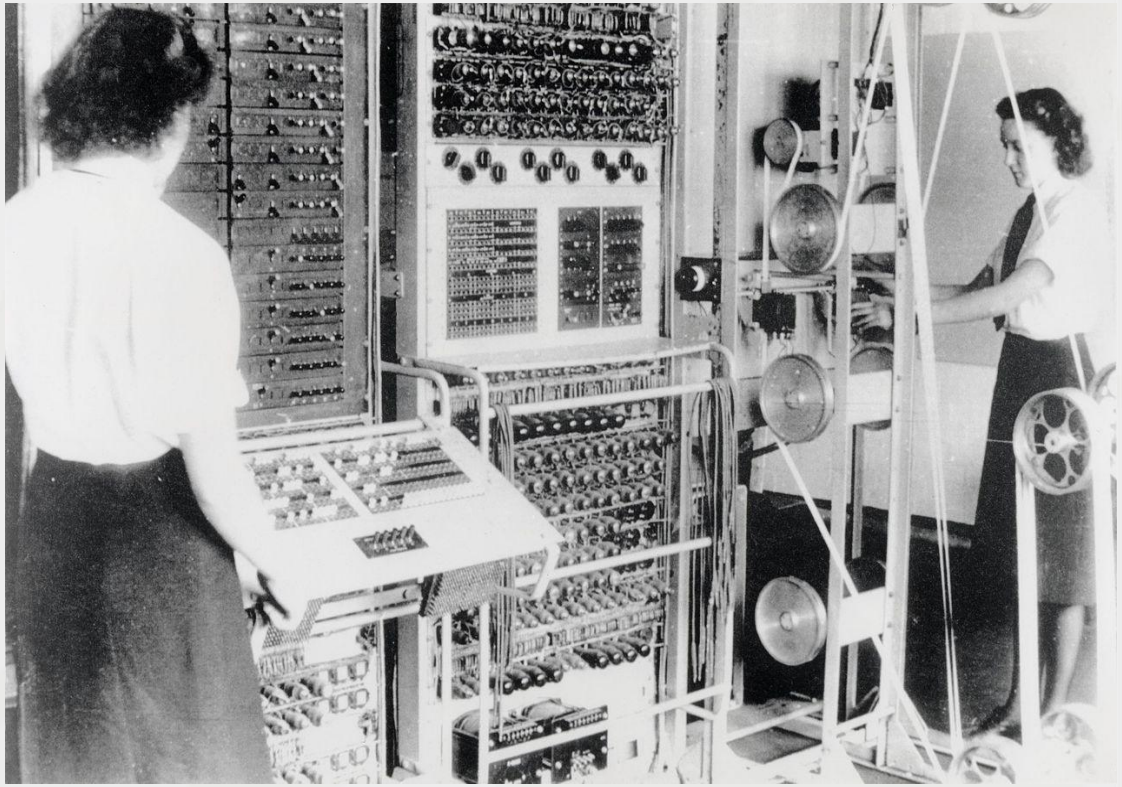
Oláh István

EIVOK alelnök
OTP Bank Nyrt.
istvan.olah@hte.hu

c. egyetemi docens
Nemzeti Közszolgálati Egyetem
olah.istvan.gyorgy@uni-nke.hu

A szoftverek (AI) uralják a fizikai eszközöket (embereket?) is !?

- 30 000 éves az első számoló eszköz
- 5.000 éves az abakusz
- 1620 logarléc
- 1623 első számológép
- 1786 TPV=regiszter
- 1812 programozható számlógép, memóriával
- 1820 TPV alapú szővőgép
- 1853 integrált „áramkör” elmélet
- 1887 statisztikai gép USA-népszámlálás
- 1936 elektromechamikus számológép
- 1939 Z1 szabadon programozható gép, No.
- **1943 Colussus Anglia, nem decimális (az AI kezdete)**
- 1944 ENIAC USA
- 1945- Neumann-elvek szerinti gépek USA (előtte Mao?)



AI történelem

- Már az első számítógép is AI machine volt
- 1960 LISP első AI „nyelv” (család)
- **1962 Hazai egyetemeken oktatott tárgyakban benne van az AI**
- 1965 ELIZA-MIT
- 1969 Szakértői rendszerek, MYCIN
- 1985 az első AI appom
- 1990 ML algaritmusok elérhetőek a civil szférában is, elsőre a fordító appok jöttek
- **1997 Garri Kaszparov-IBM Depp Blue**
- 2006 Deep learning új modellek
- 2011 IBM Watson, nyelvi modellek LLM
- 2014 GAN: Generatív Adeversial Networks
- sok az elérhető adat (de tiszta-e?), olcsó=skálázódó feldolgozási technológiák

Alapfogalmak

- **Artifical Intelligence:** a számítástechnika azon területe, amely olyan intelligens gépek létrehozására törekszik, amelyek képesek az emberi intelligenciát megismételni vagy meghaladni.

- **Machine Learning:** a mesterséges intelligencia olyan részhalmaza, amely lehetővé teszi a gépek számára, hogy a meglévő adatokból tanuljanak, és az adatok alapján döntéseket vagy előrejelzéseket hozzanak.

- **Deep Learning:** olyan gépi tanulási technika, amelyben neurális hálózatok rétegeit használják az adatok feldolgozására és a döntések meghozatalára.

- **Generative AI:** új írott, vizuális és auditív tartalmakat hoz létre kérések vagy meglévő adatok alapján

Forrás: Tóth Márton EIVOK-40

Az AI szelleme „is” kiszabadul a palackból....

Forrás: Microsoft Copilot

Amit az AI tud?

- **Természetes nyelvi chatbotok**: az ügyfélszolgálattól kezdve a személyre szabott korrepetáláson át az öngyilkosság-megelőzési tanácsadók képzéséig minden
- **Fejlett keresési képességek**: nem csak linkek, hanem válaszok keresése
- **Szöveggenerálás**: e-mailek, jogszabálytervezetek, szerződések, forgatókönyvek
- **Tartalom** (hang, videó, szöveg) valós idejű fordítása, átírása, elemzése és összefoglalása
- **Képfelismerés és -generálás**. (Smink felismerés, stb.)

Forrás: Tóth Márton EIVOK-40

Amit az AI tud-II?

Hack-el, és ... védekezik!

https://nki.gov.hu/it-biztonsag-kategoria/mesterseges-intelligencia/
https://nki.gov.hu/wp-content/uploads/2023/03/A-ChatGPT-kiberbiztonsagi-kockazatai_CTI_jelentes.pdf

AI veszélyre felhívó nyilatkozatok

„A mesterséges intelligencia miatti kipusztulás veszélyét az olyan emberiségre leselkedő fenyegetésekkel egyenértékűen kell kezelni, mint a világjárványok vagy az atomháború.”

NIST AI 100-1


ACCOUNNTIALITY ACT
al Intelligence Risk Manag
Framework (AI RM

National Institute for Standards in Technology (NIST) mesterséges intelligencia-kockázati keretrendszere

- 2021-ben a Kongresszus arra utasította a NIST-t, hogy dolgozzon ki egy önkéntes keretet a megbízható mesterséges intelligencia számára
- 2023: Artificial Intelligence Risk Management Framework (AI RMF 1.0)

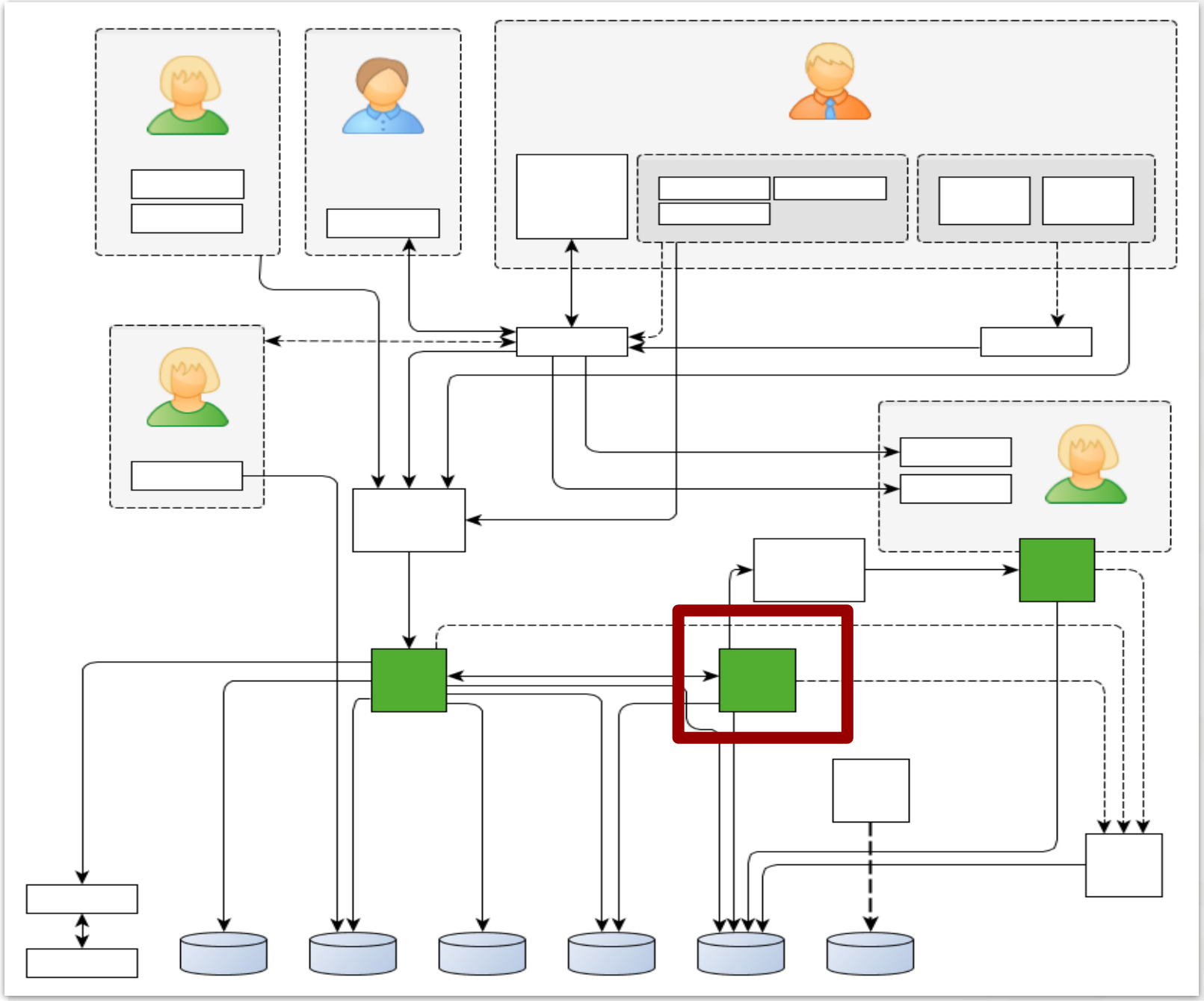
Célja: olyan iránymutatások és bevált gyakorlatok gyűjteménye, amelyek célja, hogy segítse a szervezeteket a mesterséges intelligencia technológiák bevezetésével és használatával kapcsolatos kockázatok azonosításában, értékelésében és kezelésében

Forrás: Tóth Márton EIVOK-40

Bővebben:

- **Érvényesség és megbízhatóság** – Az AI téves információkat is szolgáltathat, amit a GenAI-ban “hallucinációnak” is neveznek. Fontos, hogy a vállalatok validálni tudják, hogy az általuk alkalmazott mesterséges intelligencia pontos-, és megbízható-e
- **Biztonságosság** – Annak biztosítása, hogy az információkat ne osszák meg más felhasználókkal, mint a hírhedt Samsung ügyben
- **Ellenállóság és biztonság** – A szervezeteknek biztosítaniuk kell, hogy az AI rendszer védett és biztonságos legyen a támadásokkal szemben és sikeresen meg tudja hiúsítani a kihasználására vagy a támadások segítésére irányuló kísérleteket
- **Átláthatóság** – Fontos, szempont az AI működésének a megértése, hisz nincs benne semmiféle fekete mágia
- **Adatvédelem** – Biztosítani kell, hogy a kért információk védettek és anonimizáltak legyenek a felhasználás során
- **Tisztességesség** – A káros előítéletek kezelése (például az AI arcfelismerésben gyakran előfordul elfogultság, a világos bőrű férfiakat pontosabban azonosítják, mint a nőket és a sötétebb bőrszínűeket). Ha például a mesterséges intelligenciát a bűnüldözésben használják, ennek súlyos következményei lehetnek

Forrás: Tóth Márton EIVOK-40



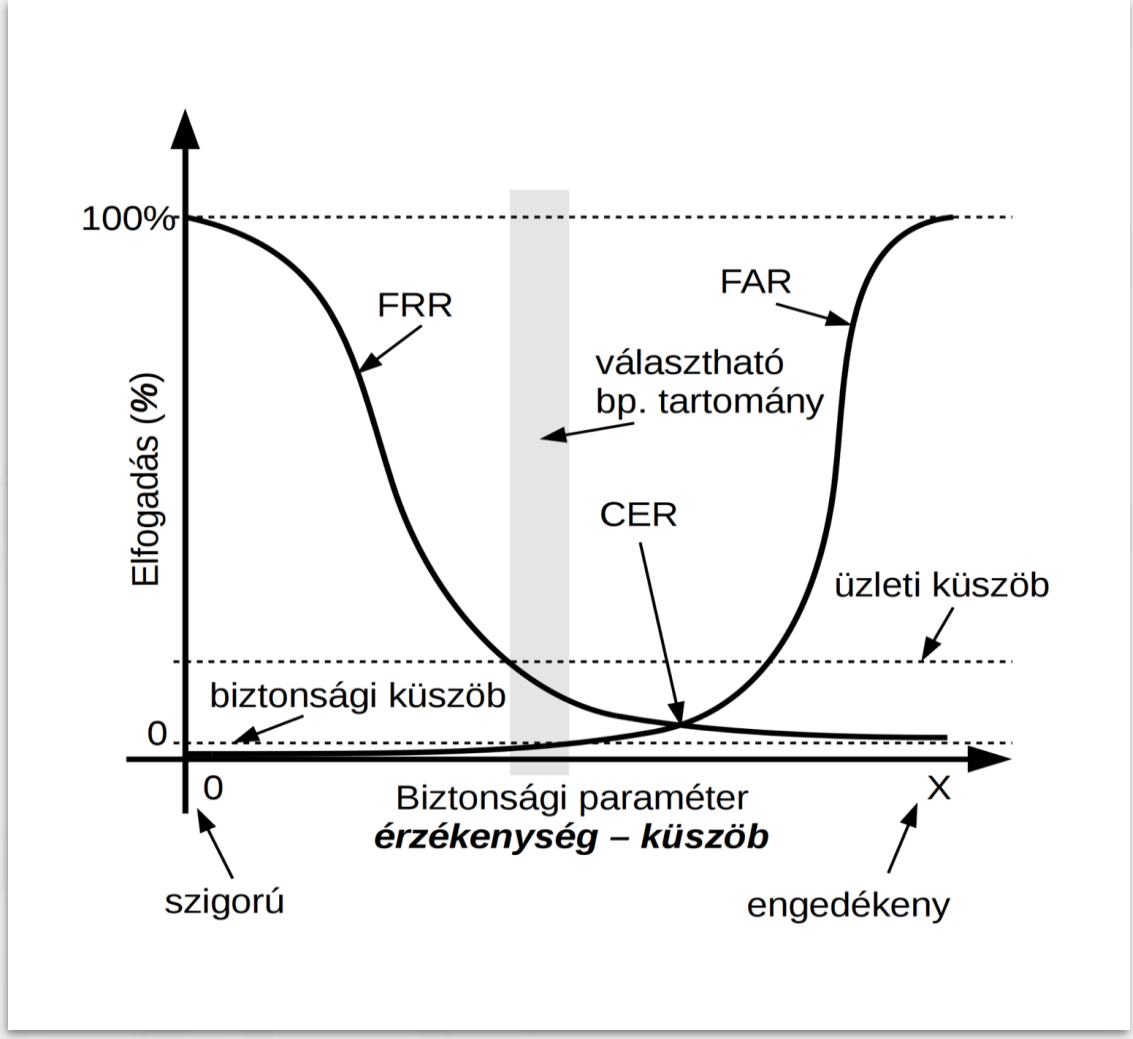
Integrált AI

Forrás: Golda Bence EIVOK-59

Audit

- Külső, független szereplő
- Magyarázható, mérhető, megismerhető az eredmény
  - “Érzésre” mégis más – egyedi esetek, egyén befolyása
- A számítások irreverzibilisek
  - Az eredeti aláírás nem állítható vissza!
- Negatív hipotézis + konfidencia intervallum

  →Elvetésre került → ✅



Forrás: Golda Bence EIVOK-59

Mi, és az MI

Ez az MI magyarázható működésű, mert MI fejlesztettük, ezért lehet auditáltatni, és verifikálni is lehet.

A modell úgy tanul, olyan adaton tanul, és azt tanulja amit MI szeretnénk, és nem azt amit mások!

Forrás: Golda Bence / op EIVOK-59

AI Large Language Model (LLM) támadási vektorok és technológiai sebezhetőségek

- A klasszikus ITsec megoldásokat AI képessé kell tenni
- Az AI új kihívásaira új ITsec megoldások szükségesek

ENISA 2025. októberben publikálta a Threat Landsacpe 2025.összefoglalóját. Az Európai Unió 2022/2555 számú irányelve (NIS2) 6 szerinti ágazati besorolás szerint a pénzintézetek szolgáltatásai a hatodik legjobban támadott szolgálatatások a kibertérben jelenleg. A leggyakoribb támadások adathalászattal kezdődnek. Az ENISA szerint ezek

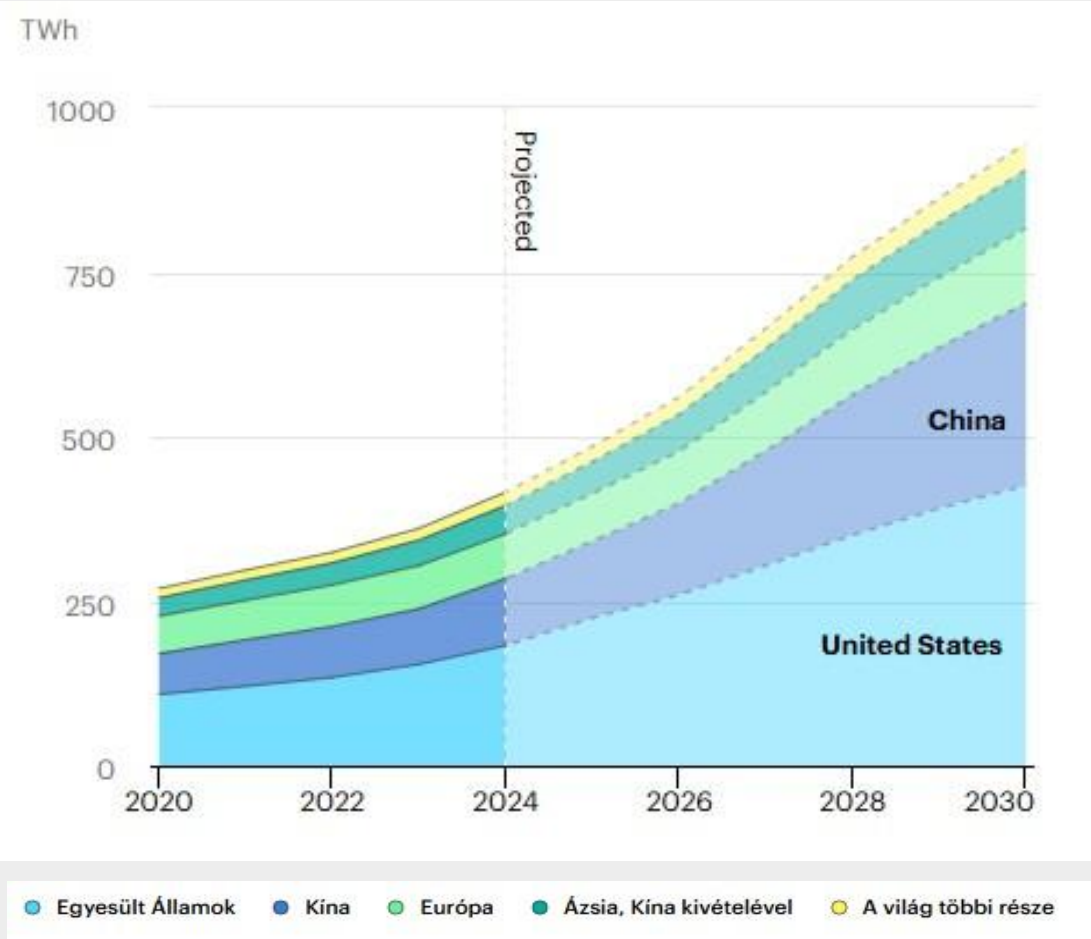
80% már AI alapú.

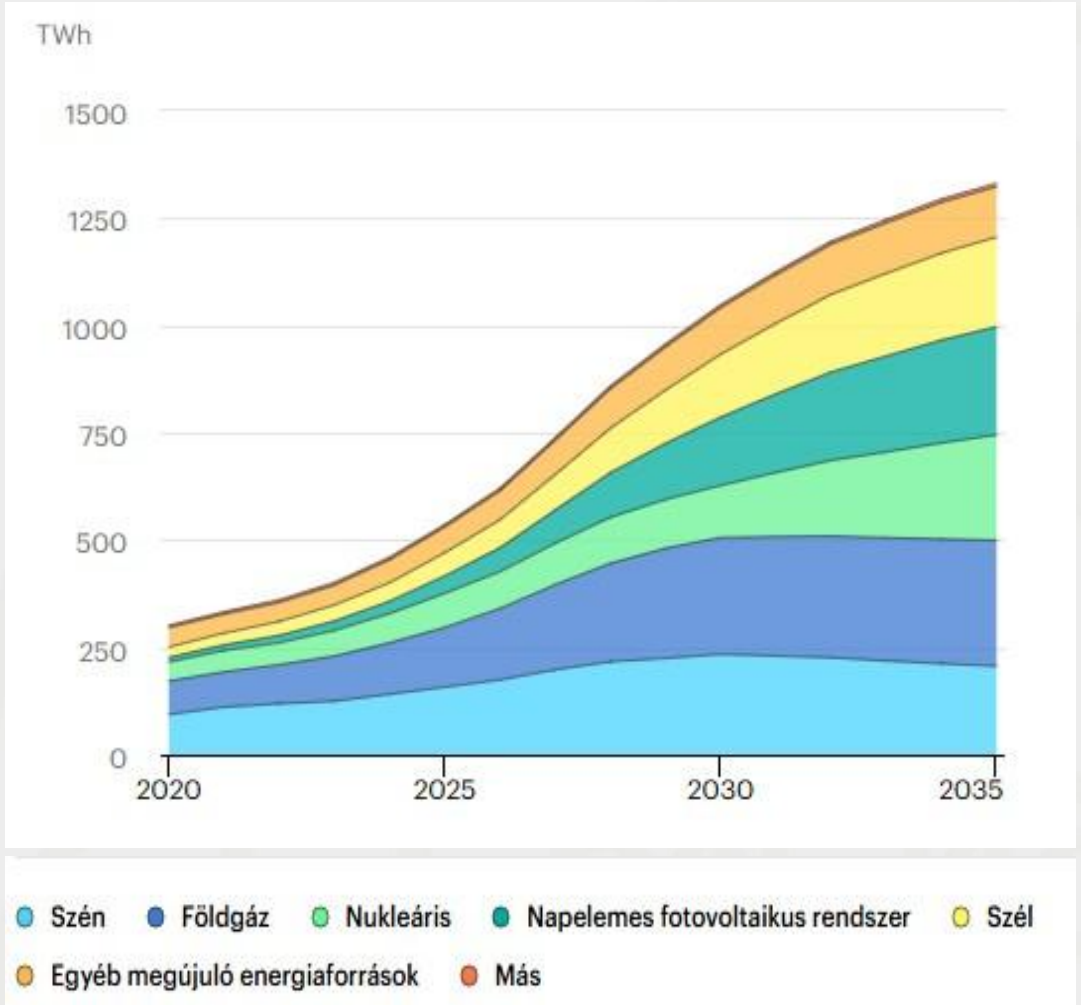
AZ ÚJ TOP

Támadás típusa	Rövid leírás	Főbb példák/hatás	Védekezési stratégia
Prompt Injection	Speciális bemenetekkel manipulált válasz	AI chat szakmai/etikai guardrail megkerülése	Szintaktikus és szemantikus szűrés, input validáció, guardrails, ember-a-hurokban jóváhagyás
Training Data Poisoning	Mérgezett, elfogult vagy rosszindulatú adatokkal tanított modell	Elfogult vagy hibás output, adatvédelmi sérülés	Adatforrás audit, anomaly detection, provenance tracking, adversarial tesztelés
Model Inversion / Membership Inference	Tanítóadatok visszafejtése, PII leak	Személyes adatok kiszivárgása	Differential privacy, likvid lekérdezésszám, titkosítás
Evasion Attack	Finoman módosított inputok, tévesztés céljából	Képfelismerési hibák, azonosítás megkerülése	Adversarial training, input transformation, ensemble modellek
Model Stealing (Extraction)	Modell „lemásolása”, funkcionalitás kisajátítása	IP, tulajdon ellopása	API rate limiting, query log, watermarking
Supply chain attack	Harmadik féltől származó kód, modell vagy dependency támadás	Backdoor/rosszindulatú könyvtár	SBOM, code audit, integrity check, rendszeres függőségfrissítés
Insecure Output Handling	Nem szűrt output, amely downstream rendszerben sérülékenységet okoz	XSS, SQL injection	Output validáció, sandbox, CSP
Hallucination/hamis tartalom	Hihető, de téves vagy veszélyes AI-válasz	Félretájékoztatás, reputációs kár	RAG, fact-check, audit, emberi kontroll
Excessive Agency	AI agent, nem kontrollált autonóm viselkedés	Nem kívánt művelet végrehajtása	Minimális jogosultság, naplózás, felülvizsgálat
Unbounded consumption/denial-of-wallet	Erőforrás (pl. API) túlhasználata/visszaélés	Szolgáltatáskimaradás	Limitálás, throttling, költség-alapú routing

Környezetvédelem-AI

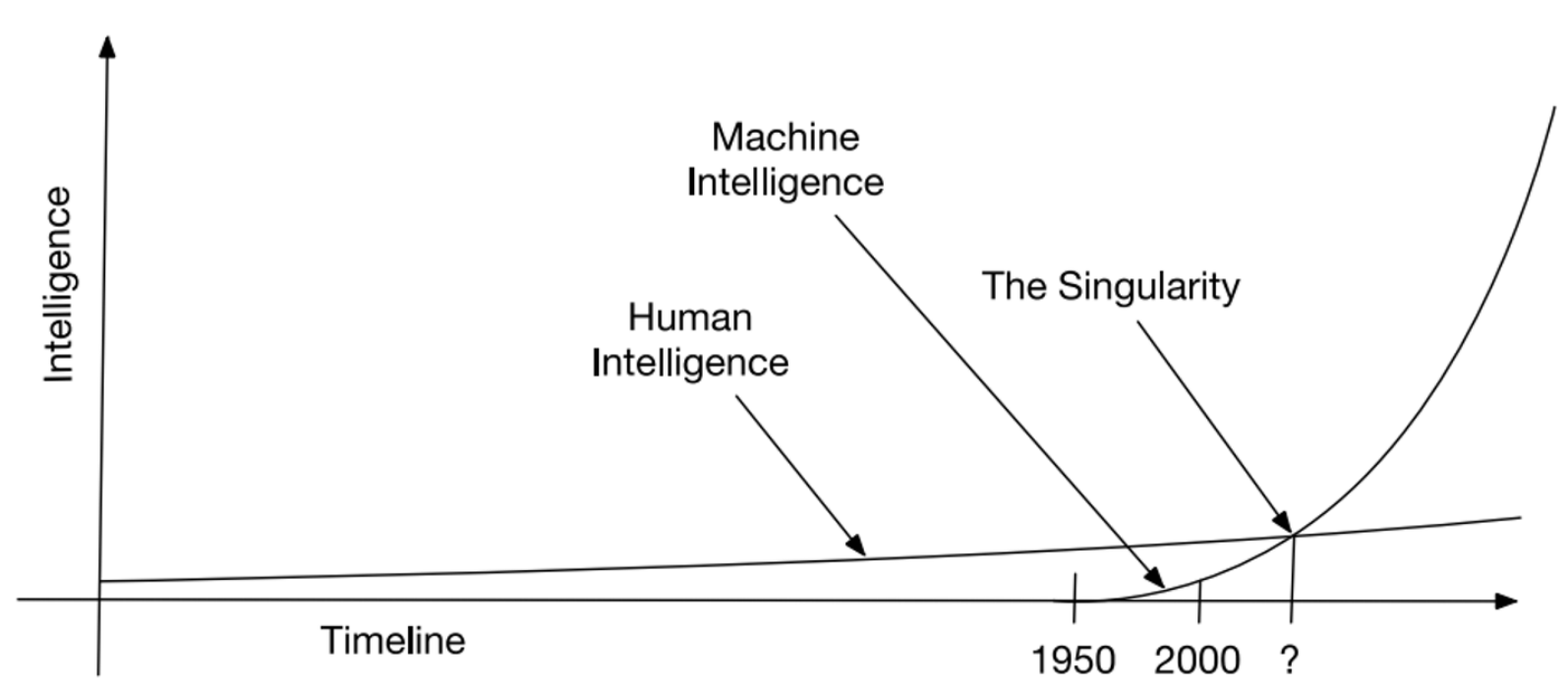
Adatközpontok energia igénye, és forrás mix:





https://www.iea.org/reports/energy-and-ai

Szingularitás?



Forrás: Tóth Márton EIVOK-40

Forrás: Microsoft Copilot

Fókusz a jövőben?

Kvantumbiztonság!

Kvantum+Felhő+AI?

Biológia + Matek processzorok

KÖSZÖNÖM A FIGYELMET!

Oláh István
c. egyetemi docens, CDPSE, EIV, MBA

olah.istvan.gyorgy@uni-nke.hu

olah.istvan@uni-obuda.hu

uni-nke.hu