NEMZETI KÖZSZOLGÁLATI EGYETEM
LUDOVIKA

MI MIóta van AI?

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

olah.istvan.gyorgy@uni-nke.hu
olah.istvan@uni-obuda.hu

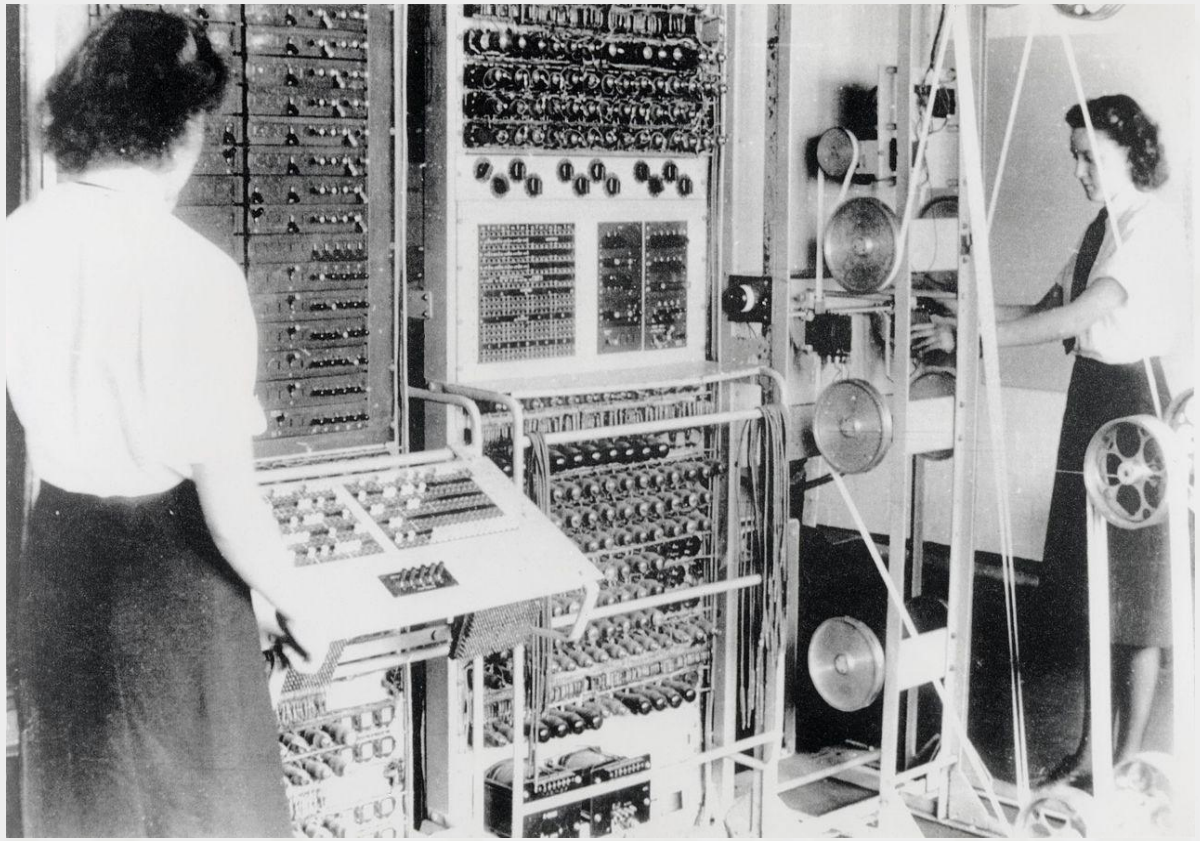
MI óta van MI?

- ie. előtt 300?
- 370 éve?
- 70 éve ?
- 15 éve ?
- pár éve?

Forrás: Microsoft Copilot

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

MI az MI ??????????????????

Jogi fogalom

- **AZ EURÓPAI PARLAMENT ÉS A TANÁCS 2024. június 13-i (EU) 2024/1689 RENDELETE a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról (a mesterséges intelligenciáról szóló rendelet)**

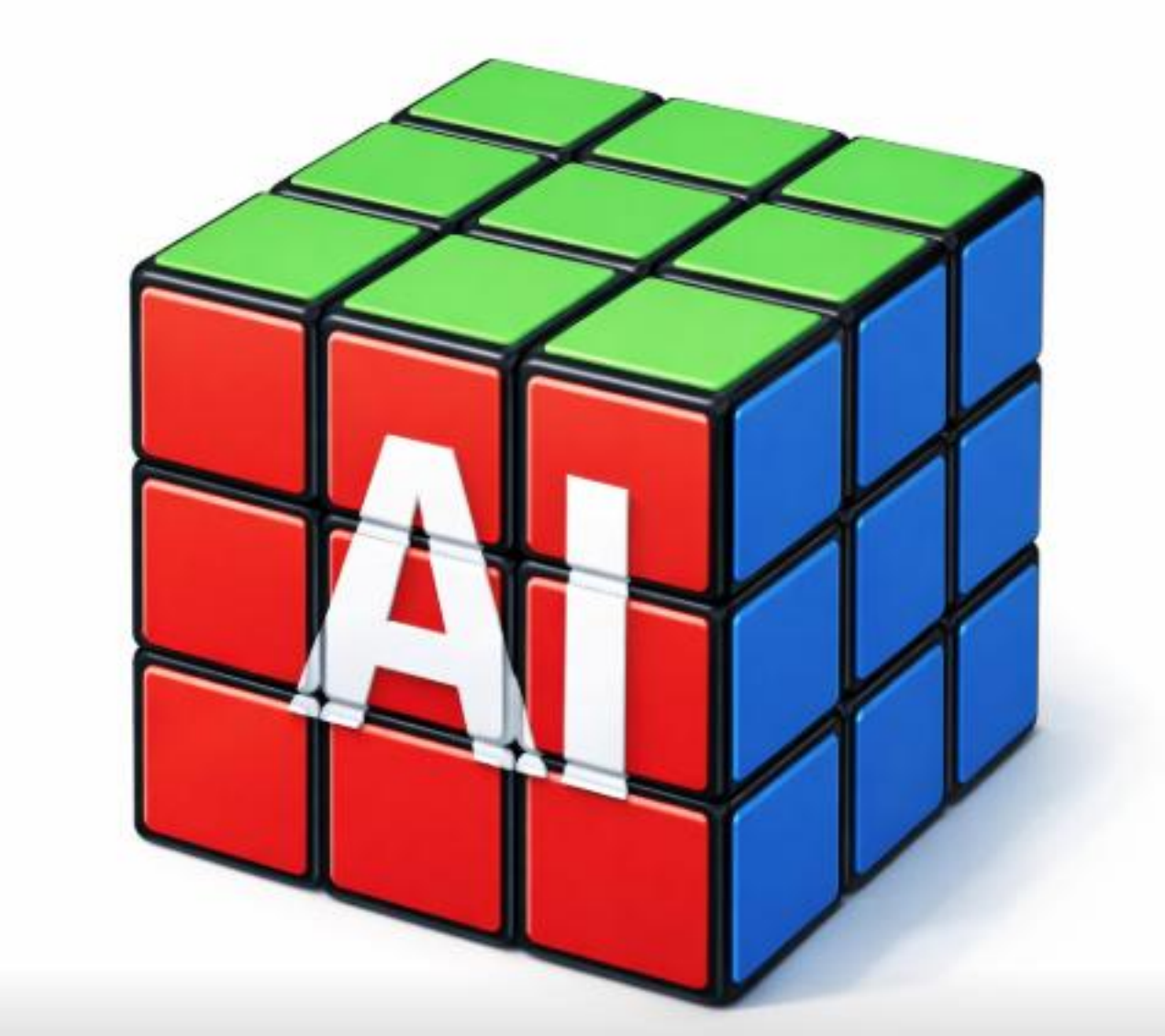
3. cikk

Fogalommeghatározások

- 1. *„MI-rendszer”: gépi alapú rendszer, amelyet különböző autonómiaszinteken történő működésre terveztek, és amely a bevezetését követően alkalmazkodóképességet tanúsíthat, és amely a kapott bemenetből - explicit vagy implicit célok érdekében - kikövetkezteti, miként generáljon olyan kimeneteket, mint például előrejelzéseket, tartalmakat, ajánlásokat vagy döntéseket, amelyek befolyásolhatják a fizikai vagy a virtuális környezetet;*

- https://net.jogtar.hu/jogszabaly?docid=a2401689.eup

Egy kocka (AI by op)



Az adat- és technológiai időszakok megértése

Az adatmenedzsment fejlődésének üteme elmarad az adatmennyiség növekedésének ütemétől

Moore Törvény ideje

1971

Metcalfe Törvény ideje

Hálózatok és kapcsolatok

E korszak győztesei (mint a Google, az Amazon, a Netflix, a Facebook) azzal nyertek, hogy erős digitális súlypontokat hoztak létre felhasználóik és ügyfeleik számára. Kapcsolódtak ügyfeleik életének számos aspektusához.

1995

1995
Metcalfe törvénye

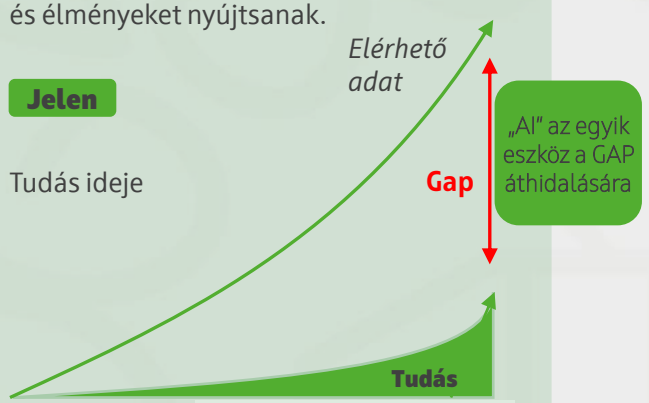
Tudás, ismeret ideje

Az adatok és az MI hasznosítása

Azok a vállalatok lesznek a győztesek, amelyek kihasználják adataikat és információikat, hogy agilisak és előrelátóak legyenek, és kiváló termékeket és élményeket nyújtsanak.

Jelen

Tudás ideje



Gondolkodik-e gép, Turing teszt?

1950 **Alan Turing**: Computing Machinery and Intelligence

A Turing-teszt e szereplős imitációs játék:

- Ember (A) – az egyik válaszadó
- Gép (B) – a másik válaszadó
- Kérdező (C) – aki ember, és nem tudja, melyik válaszadó gép, illetve ember

https://ipon.hu/magazin/cikk/amikor-a-turing-teszt-minket-mer

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

Forrás: Microsoft Copilot

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

AI veszélyre felhívó nyilatkozatok

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

Forrás: Tóth Márton EIVOK-40

AI Kontrollok

Az adat szempontjából érdektelen, hogy:

- User hoston
- Server hoston
- Storageon
- Fentiek virtuális verzión
- Szalagon
- Lemezen
- Hordozható adathordozón
- Mobil eszközön
- …….
- A Felhőben

Van.

- **AI technológia által használtan**

A védelemnek egyenszilárdnak, és kockázatokkal arányosnak kell lennie!!!

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

NIST AI RMF elemei

- **Érvényesség és megbízhatóság** – nincs „hallucináció”
- **Biztonságosság** – nem osztja meg másokkal az adatokat
- **Ellenállóság és biztonság** –az AI rendszer védett és biztonságos a támadásokkal szemben
- **Átláthatóság** –AI működésének a megértése
- **Adatvédelem** – az információk védettek és anonimizáltak legyenek
- **Tisztességesség** – káros előítéletek kezelése

MATEK..

ARR, FRR, xRR, nem determinisztikus esetekre is kell!

Hogyan elemzünk?

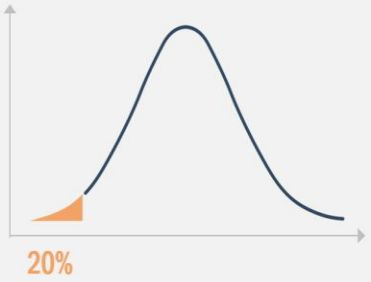
{#define SPACE_METRICS}

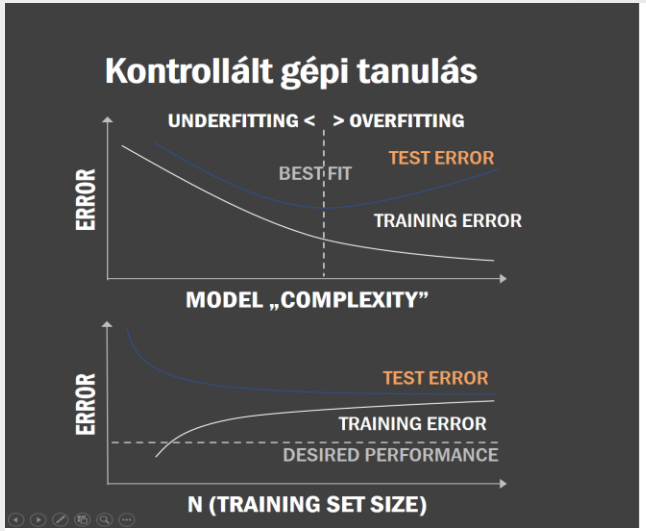
$$(1)\quad d_j(X_k, X_l) = \begin{cases} 0 & \text{if } X_k^j, X_l^j \text{ belong to the same class.} \\ 1 & \text{else} \end{cases}$$

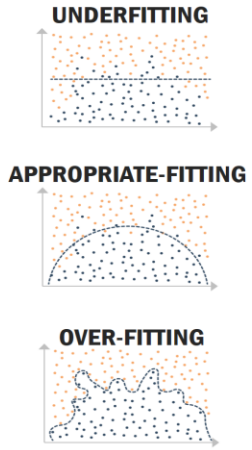
$$(2)\quad d_j(X_k, X_l) = \frac{|X_k^j - X_l^j|^{w_j}}{r_j^{w_j}}$$

$$(3)\quad d(X_k, X_l) = \sum_j \alpha_j d_j(X_k, X_l)$$

$$(4)\quad A_{kl} = d(X_k, X_l)$$

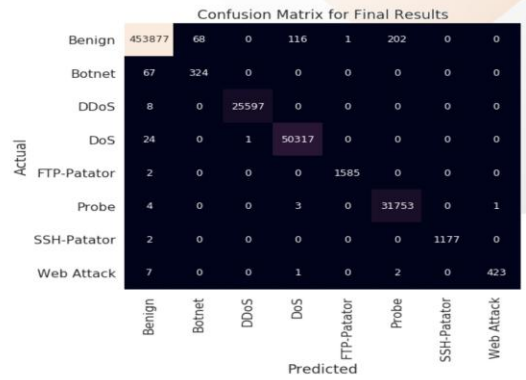






Eredmény scikit-learn használatával

Label	Precision (%)	Recall (%)	F1 Score (%)	% of Training Data
Benign	99.97	99.91	99.94	80.32
Botnet	82.65	82.86	82.76	0.07
DDos	100.00	99.97	99.98	4.53
DoS	99.76	99.95	99.86	8.90
FTP-Patator	99.94	99.87	99.90	0.28
Probe	99.36	99.97	99.67	5.62
SSH-Patator	100.00	99.83	99.91	0.21
Web Attack	99.76	97.69	98.71	0.08
Average	97.68	97.51	97.59	12.50



Forrás: Vágujhelyi Ferenc EIVOK-60

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

AI kontrollok

https://academy.evalcommunity.com/ai-governance-frameworks/

MNB kontrollok

A Magyar Nemzeti Bank 13/2025. (XII.3.) számú ajánlása a hitelintézetek digitális transzformációjáról

XIV. Az MI biztonságos és felelős alkalmazása

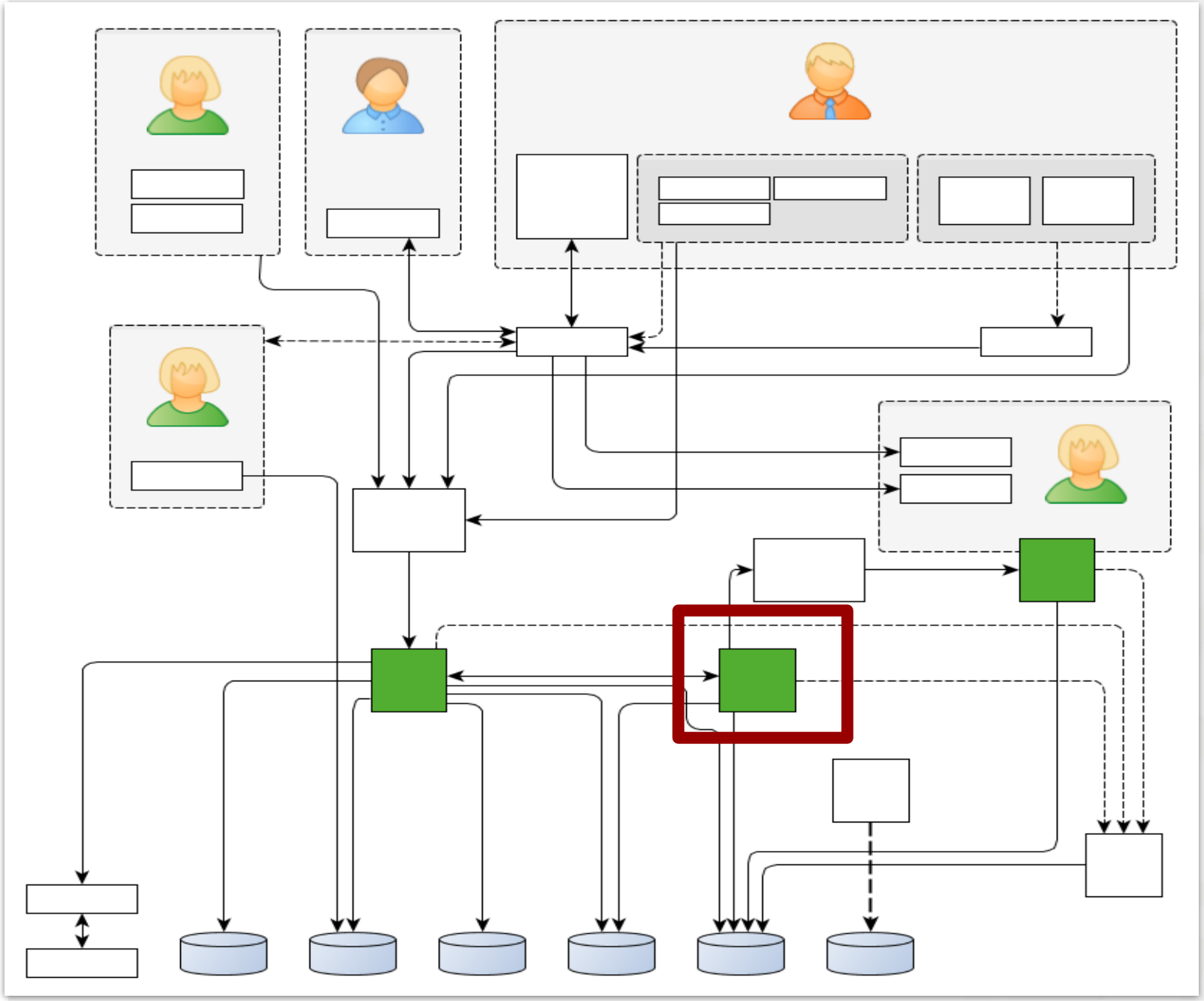
https://www.mnb.hu/letoltes/mnb-ajanlas-13-2015.pdf

Mi a jó kérdés?

- **Kérdés:** Mi is az „AI”? Mit akarunk megoldani?

Válasz:

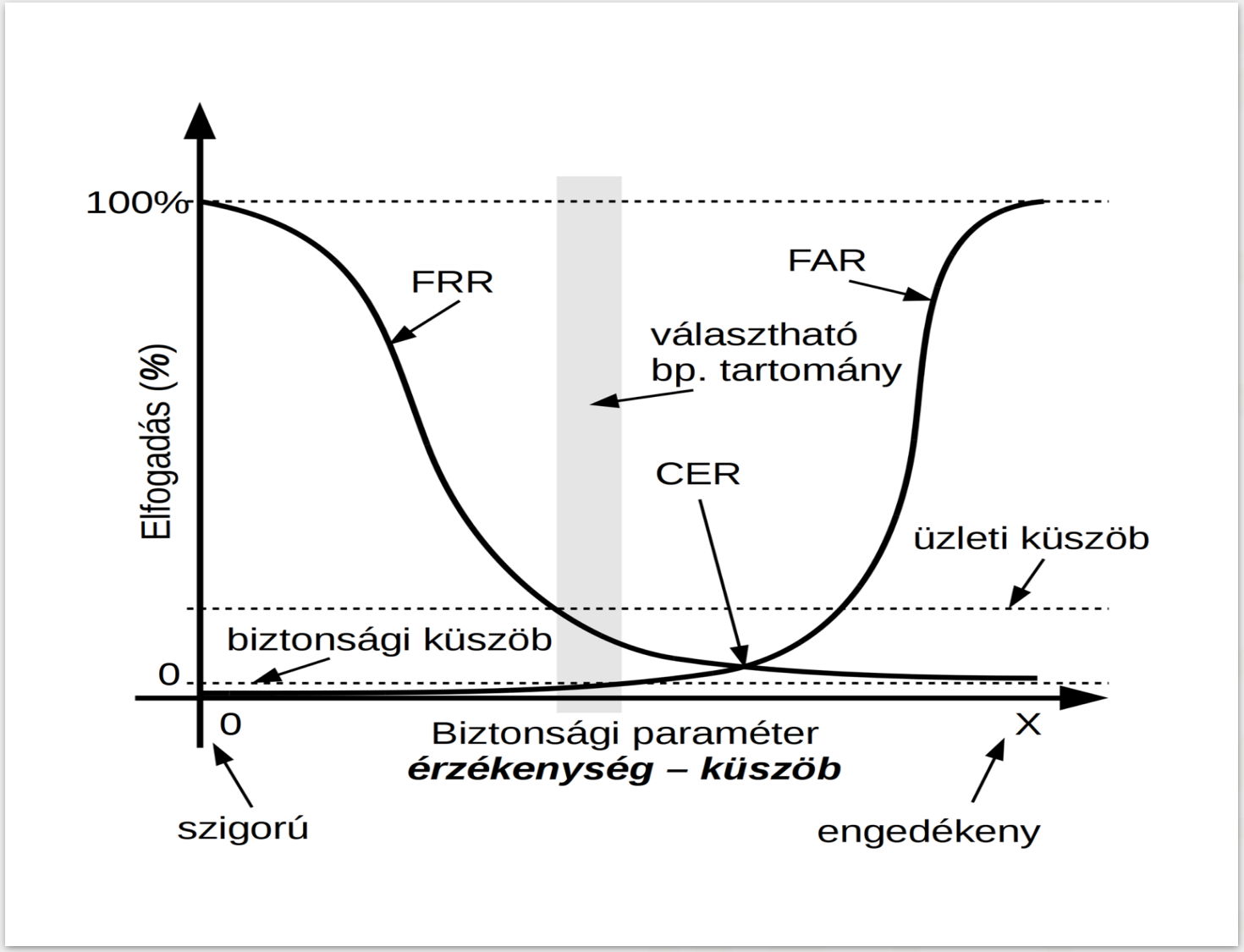
Sok féle működésű „AI”-nak definiálható megoldás létezik.
Nem az AI a cél, csak egy eszköz.
A számoslehetséges megoldásból…!!



Integrált AI

Forrás: Golda Bence EIVOK-59

Audit


100%
FRR
FAR
választható
bp. tartomány
Elfogadás (%)
CER
üzleti küszöb
biztonsági küszöb
0
0
Biztonsági paraméter
érzékenység – küszöb
X
szigorú
engedékeny

Fogalom	Jelentés
α (alfa)	Elsőfajú hiba: „tévesen elutasítod H_0-t”
β (béta)	Másodfajú hiba: „tévesen elfogadod H_0-t”
1–β	Teszt ereje: „helyesen kimutatod a hatást”

📢 Az AI másik oldalAI !!

- 💡 Ha nem tudod a "MI a Matek" benne, akkor nem tudod mire?, mit válaszol, hogy működik!
- 💡 A bevitt tanító adatod, a bevitt promt adatod, a kimeneti adatod örökké benne maradhatnak ‼ azaz kihackelhetőek az adataid egy AI-ból ‼
- 💡 A jó AI-tól kapsz egy „elnézést, de nem tudom választ”! Sok ingyenestől nem, ezzel hitetve el veled….mert a pozitív ügyfélélmény, a képernyőjéhez kötés fontosabb mint a AI „tudása”, vagy a VÉDELMED‼
- 💡 Egy jó LLM-genAI is "eltöri" a számítási láncot, azaz elnézést kapcsolok egy embert is válaszol!
- 💡 Egy jó AI-t ugyanúgy tudsz securizálni, mint bármilyen más szoftvert/IT rendszerszert ‼ ‼

- 📢 Az „AI rabszolgaságod” (by op) ellen tudsz tenni ehhez ‼
- 💡 Olyan az AI-t használj, amely magyarázható működésű!
- 💡 Ha lehet olyat, amelyet Magad fejlesztettél!
- 💡 Ezért lehet auditáltatni, és verifikálni ezt az AI-t!
- 💡 A modell úgy tanul, olyan adaton tanul, és azt tanulja amit TE szeretnél, és nem azt amit Mások!
- 💡 Az AI ott, és úgy fut ahogy TE szeretnéd, így az adataid nem lehet kihackelni belőle!

💡 Ha nem ilyen AI-t használsz, vagy amelyről nem tudod a matekot, akkor gondolj erre:

MI és a „sötét” oldalAI?

Volt e már ország amely gazdasága megszenvedte/megszenvedi az AI hatását?

Volt e már ország amely választását jogerősen kimondva AI eszközökkel csalták el?

Az MI szerepepe kiberbűnözésben mekkora százalékkal van jelen?

Az igazságügyi szakértők által igazolt mód hány ember halt bele a MI használatába?

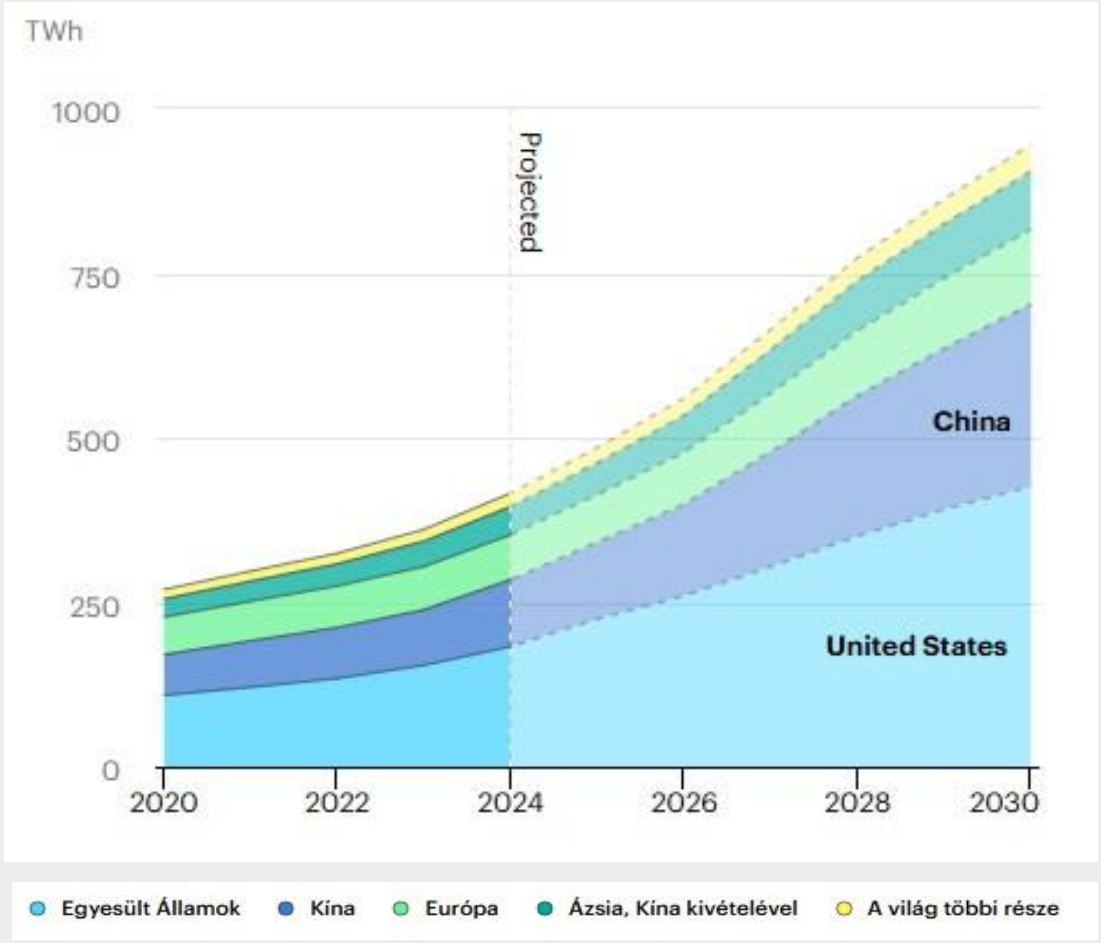
Hol AI lehet a kontrollokat kialakítani?

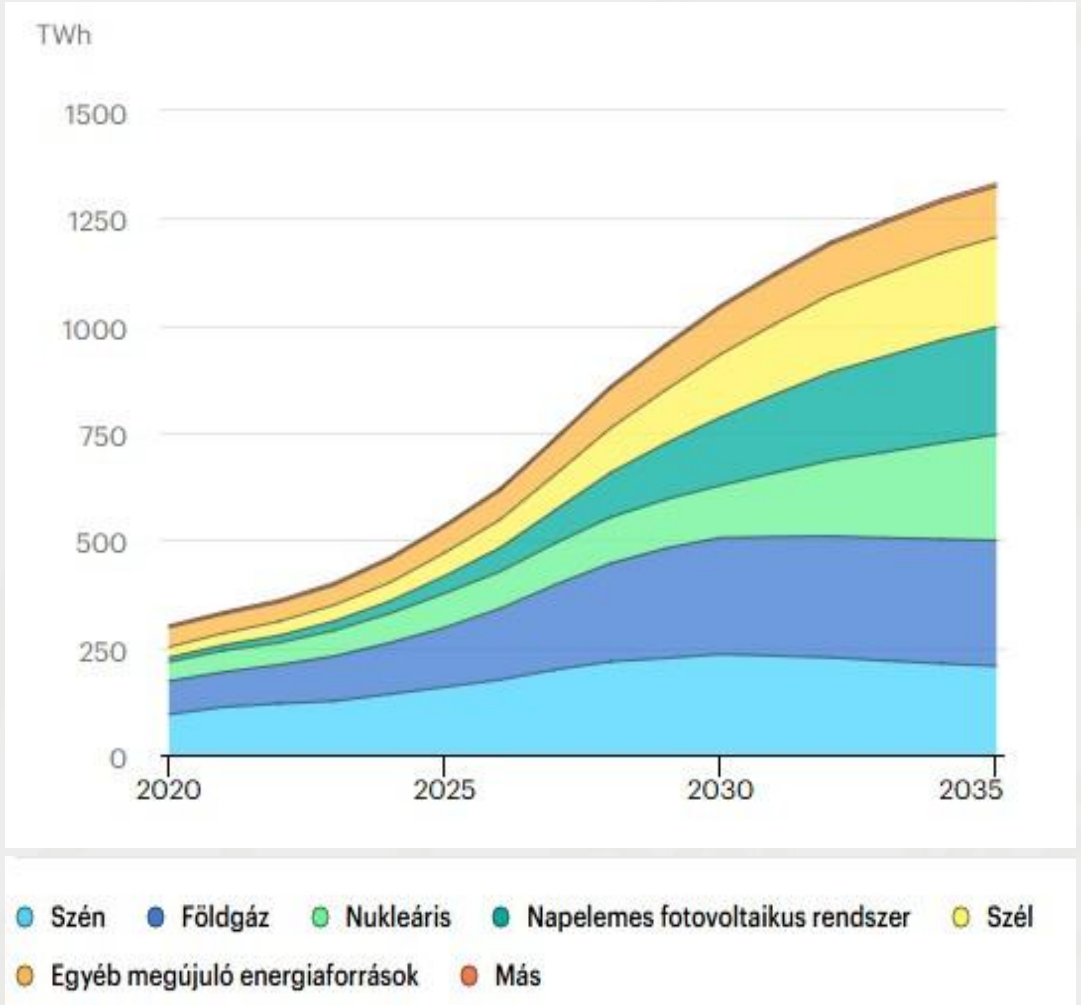
MINDENHOL !

mert egy, egy kontroll nem szolgáltató és/vagy technológia függő!

Környezetvédelem-AI

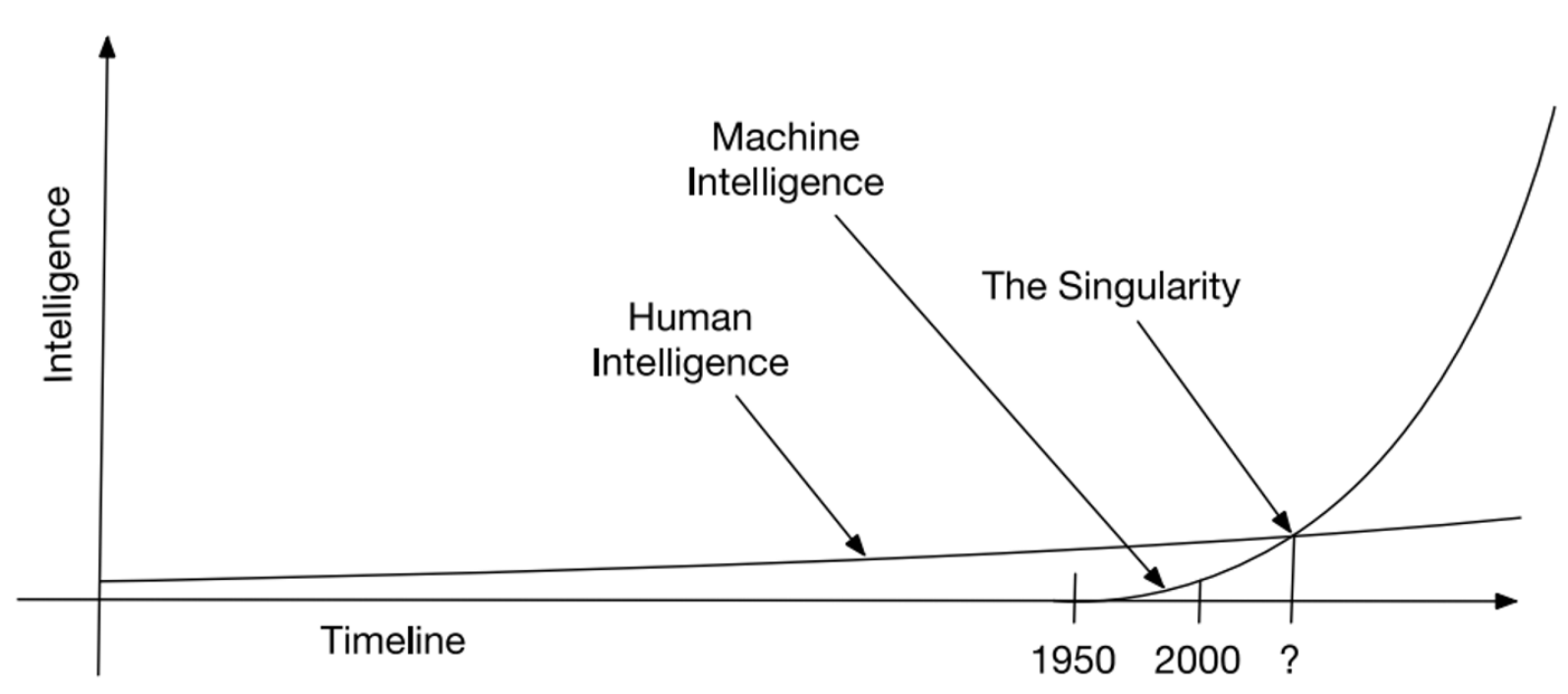
Adatközpontok energia igénye, és forrás mix:





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

Szingularitás?



Forrás: Microsoft Copilot

Fókusz a jövőben?

Kvantumbiztonság!

Kvantum+Felhő+AI?

1G--10G !!

Biológia + Matek processzorok

Mi és az MI

A jó MI magyarázható működésű, mert MI fejlesztettük, ezért tudjuk auditáltatni, és verifikálni is lehet.

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

Úgy működik ahogy a MI érdekünk!

(szuveneritás e nélkül nem lehet)

EIVOK ??

HTE 1949-ben alapítva nonprofit szervezet
https://www.hte.hu/fooldal

2018-ban jött létre az Információbiztonsági Szakosztály, az EIVOK. A közösség dinamikusan bővül taglétszáma 17 főről 400 főre emelkedett.

Minden hónapban legalább egy szakmai esemény/ konferencia
https://www.hte.hu/eivok
eivok@hte.hu

ISACA
Budapest Chapter

1 *AI / LLM Biztonsági Szolgáltatások - KPMG Magyarország*. https://kpmg.com/hu/hu/home/services/advisory/technology/cyber-lab/ai-llm-security.html
2 *What Are Adversarial AI Attacks on Machine Learning?*. https://www.paloaltonetworks.com/cyberpedia/what-are-adversarial-attacks-on-AI-Machine-Learning
3 *NIST Identifies Types of Cyberattacks That Manipulate Behavior of AI ....* https://www.nist.gov/news-events/news/2024/01/nist-identifies-types-cyberattacks-manipulate-behavior-ai-systems
4 *NIST Trustworthy and Responsible AI NIST AI 100-2e2025*https://csrc.nist.gov/pubs/ai/100/2/e2025/final
5 *ENISA THREAT LANDSCAPE 2025* https://www.enisa.europa.eu/sites/default/files/2025-10/ENISA%20Threat%20Landscape%202025.pdf
6 *Az Európai Parlament és a Tanács (EU) 2022/2555 irányelve (2022. december 14.) az Unió egész területén egységesen magas szintű kiberbiztonságot biztosító intézkedésekről, valamint a 910/2014/EU rendelet és az (EU) 2018/1972 irányelv módosításáról és az (EU) 2016/1148 irányelv hatályon kívül helyezéséről (NIS 2 irányelv) (EGT-vonatkozású szöveg) DIRECTIVE (EU) 2022/2555 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2555*
7 *Megjelent a NIST Mesterséges Intelligencia kockázatkezelési keretrendszere* https://nki.gov.hu/it-biztonsag/hirek/megjelent-a-nist-mesterseges-intelligencia-kockazatkezelesi-keretrendszere/
8 *ISO/IEC 42001:2023Information technology — Artificial intelligence — Management system* https://www.iso.org/standard/42001
9 *Az Európai Parlament és a Tanács (EU) 2024/1689 rendelete (2024. június 13.) a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról (a mesterséges intelligenciáról szóló rendelet) (EGT-vonatkozású szöveg)*
Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689
10 *Bankoknál, biztosítóknál vizsgálta a mesterséges intelligenciát, gépi tanulást az MNB* https://www.mnb.hu/sajtoszoba/sajtokozlemenyek/2023-evi-sajtokozlemenyek/bankoknal-biztositoknal-vizsgalta-a-mesterseges-intelligenciat-gepi-tanulast-az-mnb
11 *Regulation (EU)* **2016***/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)*
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
12 *Survey Shows How Generative AI Puts Organizational AI Maturity to the Test* https://www.gartner.com/en/documents/5373863
13 *A (kiber)biztonság gyakorlatAI: Inspiráló előadások és beszélgetések a jövőhöz vezető úton - EIVOK-59.* https://www.hte.hu/informaciobiztonsagi-szakosztaly-eivok/-/esemeny/1/4975549/eivok-59-a-kiber-biztonsag-gyakorlatai
14 *AI alapú dokumentum hitelesítés a gyakorlatban* www.hte.hu/documents/10180/4963359/I-03_Golda+Bence.pdf
15 *Code security audit based on regulations at a financial institution* https://ieeexplore.ieee.org/document/11030093
16 *AI Governance Framework - AI Governance Framework*. https://ai-governance.eu/
17 *MULTILAYER FRAMEWORK FOR GOOD CYBERSECURITY PRACTICES FOR AI*
https://www.enisa.europa.eu/sites/default/files/publications/Multilayer%20Framework%20for%20Good%20Cybersecurity%20Practices%20for%20AI.pdf
18 NIST *AI Risk Management Framework* https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
19 *REGULATIONS REGULATION (EU) 2022/2554 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011*
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2554
20 13 *Az EU AI Act és a GDPR kapcsolódásai* –https://dmp.hu/adatvedelem/az-eu-ai-act-es-a-gdpr-kapcsolodasai/
21*Megfelelés az AI Act-nek - gyakorlati példák az adatvédelmi feladatokra*. https://www.jogiforum.hu/blog-adatvedelem-10/2024/10/11/megfeleles-az-ai-act-nek-gyakorlati-peldak-az-adatvedelmi-feladatokra/
22 *A generatív AI adatvédelmi kockázatai – Jogászvilág*. https://jogaszvilag.hu/vilagjogasz/a-generativ-ai-adatvedelmi-kockazatai/
23 *How to Scale Your Finance AI Pilots* https://www.gartner.com/en/documents/6285783
24 *OWASP Top 10 for Large Language Model Applications* https://owasp.org/www-project-top-10-for-large-language-model-applications
25 *Navigate threats to AI systems through real-world insights* *https://atlas.mitre.org/*
26 *ML10:2023 Model Poisoning - OWASP Foundation*. https://owasp.org/www-project-machine-learning-security-top-10/docs/ML10_2023-Model_Poisoning
27 *Build a Zero Trust Framework for Secure AI Implementation*. https://www.microsoft.com/en-us/microsoft-365/business-insights-ideas/resources/build-a-secure-zero-trust-secure-foundation-for-ai
28 *How is AI Strengthening Zero Trust? | CSA*. https://cloudsecurityalliance.org/blog/2025/02/27/how-is-ai-strengthening-zero-trust
29 *Take Back Control and Trust Your Enterprise AI with ZeroTrusted.ai's AI Firewall / AI Gateway / AI Health Check* *ZeroTrusted.ai*. https://www.zerotrusted.ai/
30 *Hogyan segít kezelni az AI az ellátási láncban jelentkező kockázatokat ....* https://logisztika.hu/2024/10/14/hogyan-segit-kezelni-az-ai-az-ellatasi-lancban-jelentkezo-kockazatokat/

KÖSZÖNÖM A FIGYELMET!

uni-nke.hu