euro one

U.S. COMMERCIAL SERVICE
United States of America Department of Commerce

Cyber Summit Budapest 2025

Gondolatok a kiberbiztonságról (AI)

Oláh István (op)

HTE-EIVOK ., OTP-Bank Nyrt., NKE ., BME ., ÓE ., ELTE-OTP-Kiberlab,

Hewlett Packard Enterprise

NETWITNESS

paloalto NETWORKS

TREND MICRO

Témakörök

- miről lesz szó:
  - mióta létezik az MI
  - alapfogalmak
  - képességek
  - szabályozás
  - a „kulcs” kérdés
  - az AI kontrolljai
  - Az AI auditja
  - egyéb kérdések

Forrás: Microsoft Copilot

EIVOK

HTE 75 éves nonprofit szervezet
https://www.hte.hu/fooldal

2018-ban jött létre az Információbiztonsági Szakosztály, az EIVOK. A közösség dinamikusan bővül taglétszáma 17 főről 282 főre emelkedett.

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

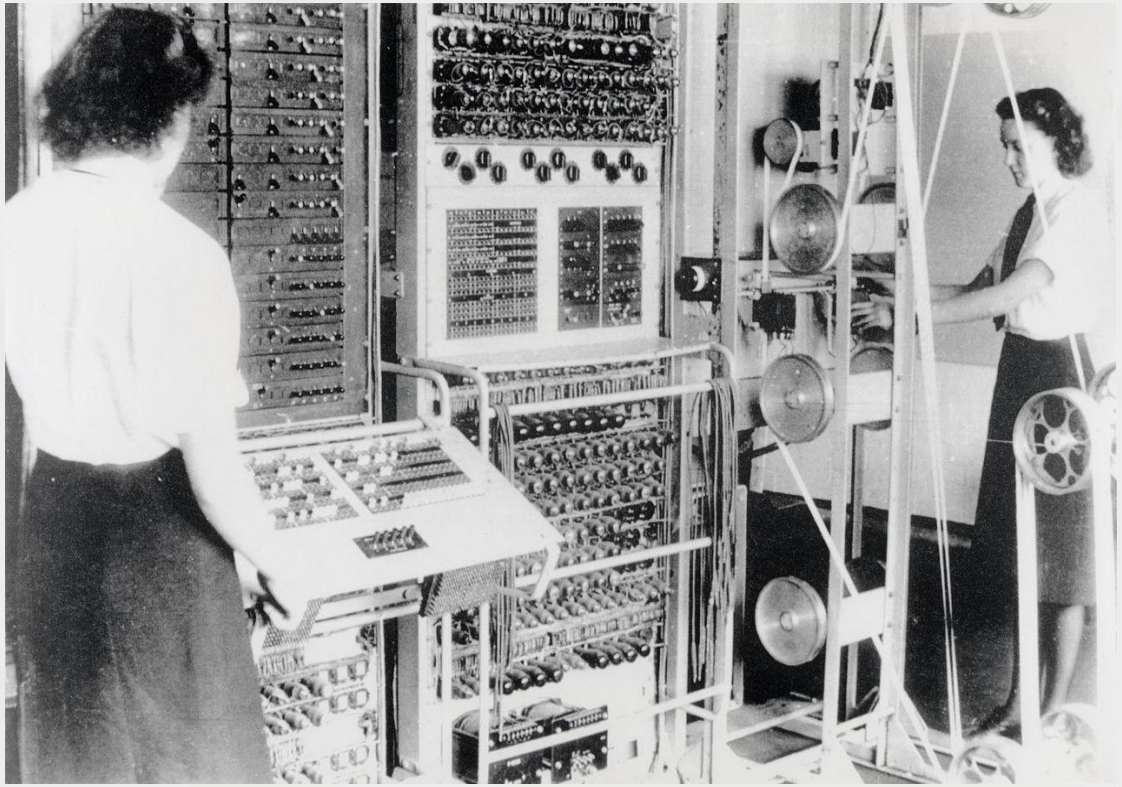
MI Mióta?

- 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”
- 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

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

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.

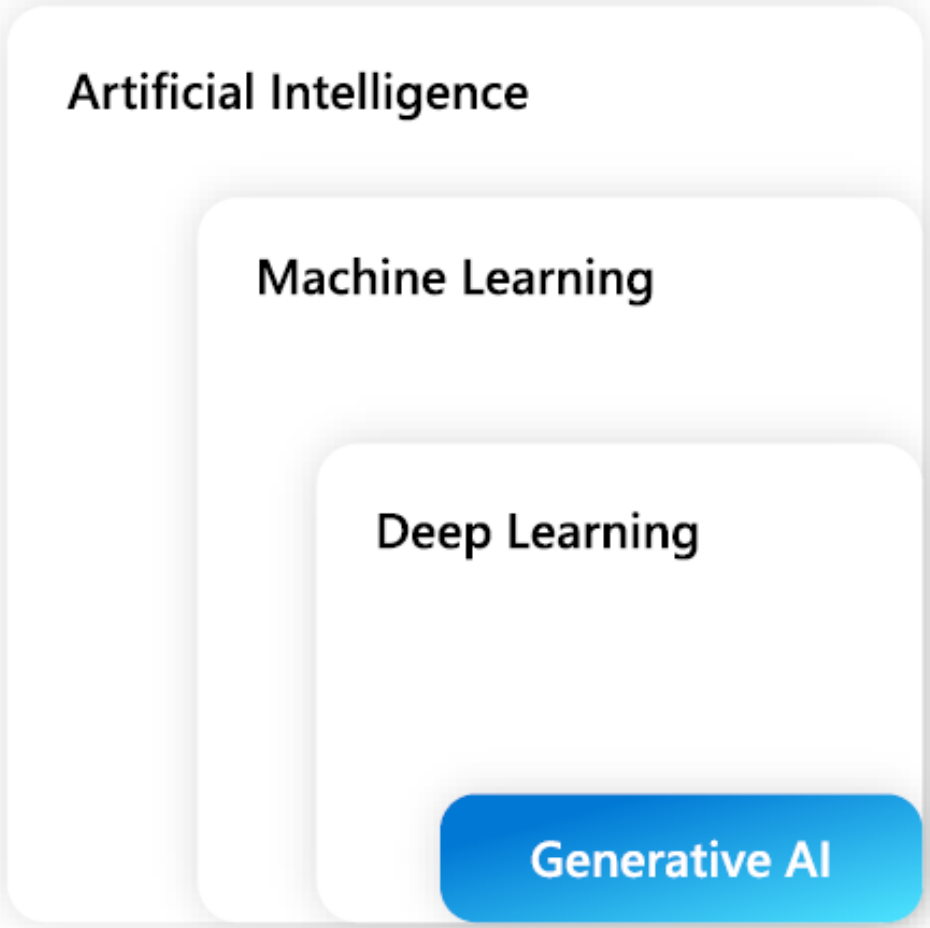
- **Deep Learning:** olyan gépi tanulási technika, amelyben neurális hálózatok rétegeit használják az adatok feldolgozására és a döntések meghozatalára.

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

Forrás: Tóth Márton EIVOK-40

Vezetőknek

A brief history of AI



1950s

Artificial Intelligence

the field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence.

1959

Machine Learning

subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions.

2017

Deep Learning

a machine learning technique in which layers of neural networks are used to process data and make decisions.

2021

Generative AI

create new written, visual, and auditory content given prompts or existing data.

Forrás: Microsoft

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?

- **Videó** – elemzés, generálás
- **Deepfake** – Arc csere
- **Hang csere**
- Mesterséges szaglás
- Emóciók felismerése
- KoDi

Forrás: Tóth Márton EIVOK-40

Forrás: Microsoft Copilot

Hack-el, és ... védekezik!

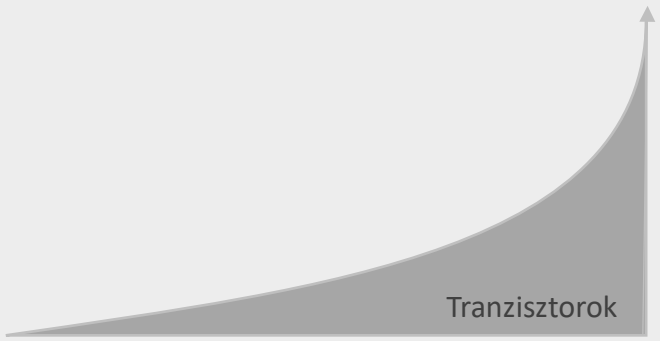
https://nki.gov.hu/it-biztonsag-kategoria/mesterseges-intelligencia/
https://nki.gov.hu/wp-content/uploads/2023/03/A-ChatGPT-kiberbiztonsagi-kockazatai_CTI_jelentes.pdf

Az adat- és technológiai időszakok megértése

Az adatmenedzsment fejlődésének üteme elmarad az adatmennyiség növekedésének ütemétől

Moore Törvény ideje

1971



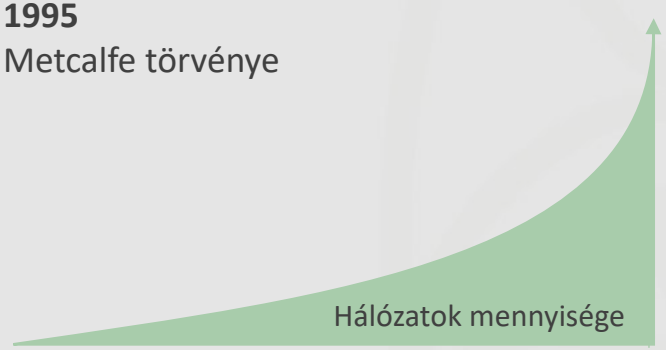
Metcalfe Törvény ideje

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

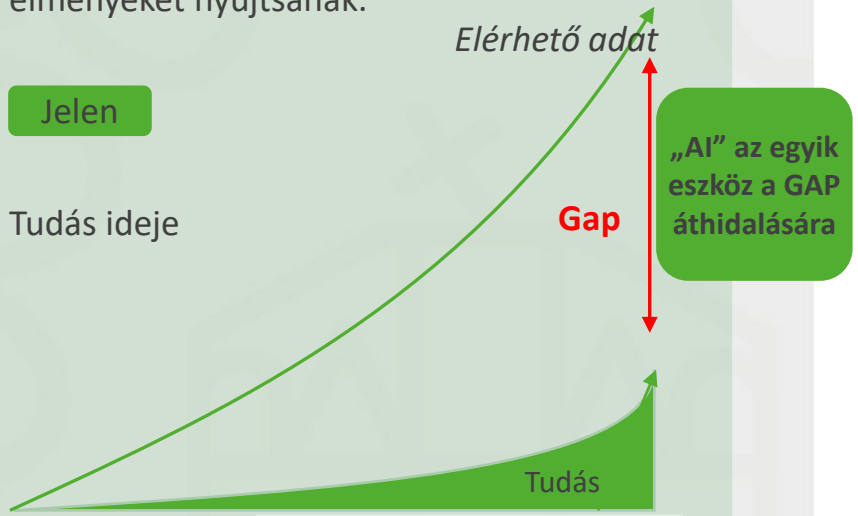
1995
Metcalfe törvénye



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.



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

2023. szeptember. 20. 09:49 · Utolsó frissítés: 2023. szeptember. 20. 10:06 · TECH

Megkezdődhet a chipek beültetése az emberek agyába, megkapta az engedélyt Elon Musk cége, a Neuralink

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

Forrás: Tóth Márton EIVOK-40

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

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

MATEK..

XXR, nem diszkrét!

Hogyan elemzünk?

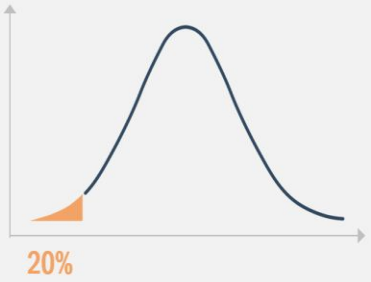
{#define SPACE_METRICS}

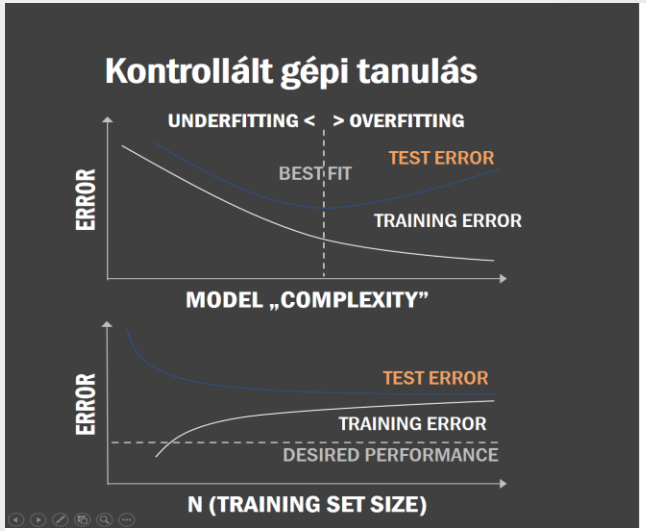
$$(1)\quad d_j(X_k, X_l) = \begin{cases} 0 & \text{if } X_k^j, X_l^j \text{ belong to the same class.} \\ 1 & \text{else} \end{cases}$$

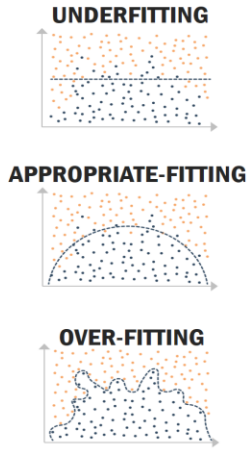
$$(2)\quad d_j(X_k, X_l) = \frac{|X_k^j - X_l^j|^{w_j}}{r_j^{w_j}}$$

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

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

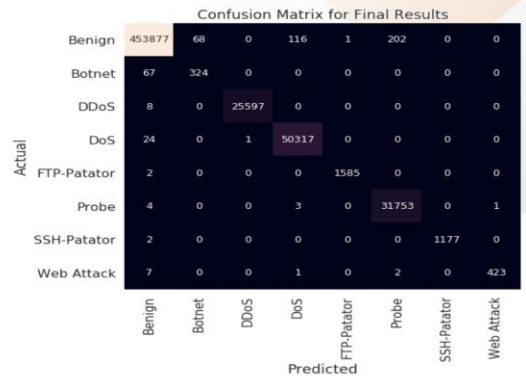






Eredmény scikit-learn használatával

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

EU AI ACT

Az AI Act a mesterséges intelligenciáról szóló európai törvényjavaslat, amely az első olyan átfogó törvény, amelyet egy jelentős szabályozó hatóság hozott a mesterséges intelligenciáról

- A törvény több kockázati kategóriába sorolja a mesterséges intelligencia alkalmazásait
- Először is, az olyan alkalmazások és rendszerek, amelyek **elfogadhatatlan kockázatot** jelentenek, mint például a társadalmi legitimáció nélküli szociális pontozás, tilalom alá esnek
- Másodszor, a **magas kockázatú** alkalmazásokra, mint például az álláspályázatokat rangsoroló önéletrajz-ellenőrző eszközre, különleges jogi követelmények vonatkoznak.
- Végül pedig a **nem kifejezetten tiltott vagy nem kifejezetten kockázatosnak** minősített alkalmazások nagyrészt szabályozatlanok maradnak.

Forrás: Tóth Márton EIVOK-40

AI Act – főbb pontok

- Műszaki robusztusság és **biztonság**
- Adatvédelem és adatkezelés

„szabályozási homokozó” (“regulatory sandbox”)

- **Emberi ügynökség és felügyelet**
- **Átláthatóság**
- Sokszínűség, megkülönböztetés mentesség és méltányosság
- Elszámoltathatóság, társadalmi és **környezeti jólét**

Forrás: Tóth Márton EIVOK-40

AI Act – a gyakorlatban

Elfogadhatatlan kockázatú, tiltott gyakorlatok

Tudatos észlelés, érzékelés nélküli, manipulatív rendszerek, jelentős társadalomformáló hatással bíró, vagy sebezhetőséget kihasználó megoldások, biometrikus jellemző alapján kategorizáló, illetve bizonyos azonosító rendszerek, amelyek sértik az egyének vagy bizonyos csoportok jogait, biztonságát

BETILTÁS, KIVEZETÉS

Nagy kockázatú rendszerek, erősen szabályozott feltételekkel

Alapjogokat érintő pl. foglalkoztatás, szolgáltatásokhoz történő hozzáférés, igazságszolgáltatás, határigazgatás, bűnüldözés

BIZTONSÁG, KOCKÁZATKEZELÉS, NYOMON KÖVETHETŐSÉG, ADATOK INTEGRITÁSA, EMBERI RÉSZVÉTEL

Korlátozott kockázatú rendszerek, átláthatóság biztosítása mellett

A transzparencia hiánya miatt hordoz kockázatot, ezért a tájékoztatáson nagy a hangsúly
pl. AI-jal folytatott chat, szintetikus médiatartalom, érzelemfelismerés

TÁJÉKOZTATÁS

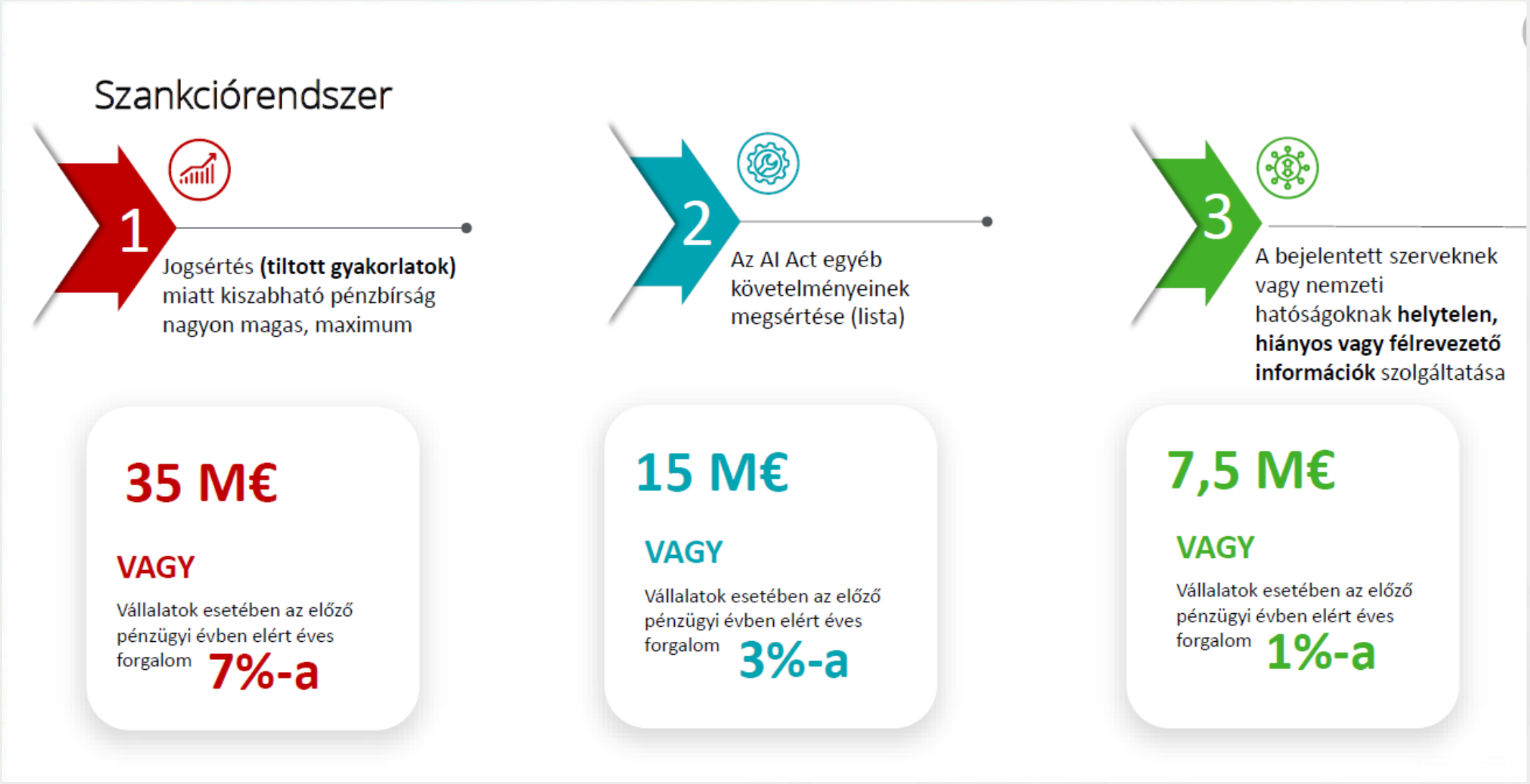
Alacsony, minimális kockázatú rendszerek, magatartási kódex használatával

A legtöbb megoldás ilyen, amelyek bizonyos folyamatokat támogatnak.

SZAKMAI MEGFELELŐSÉG BIZTOSÍTÁSA

Forrás: Nagy Zsombor EIVOK-59

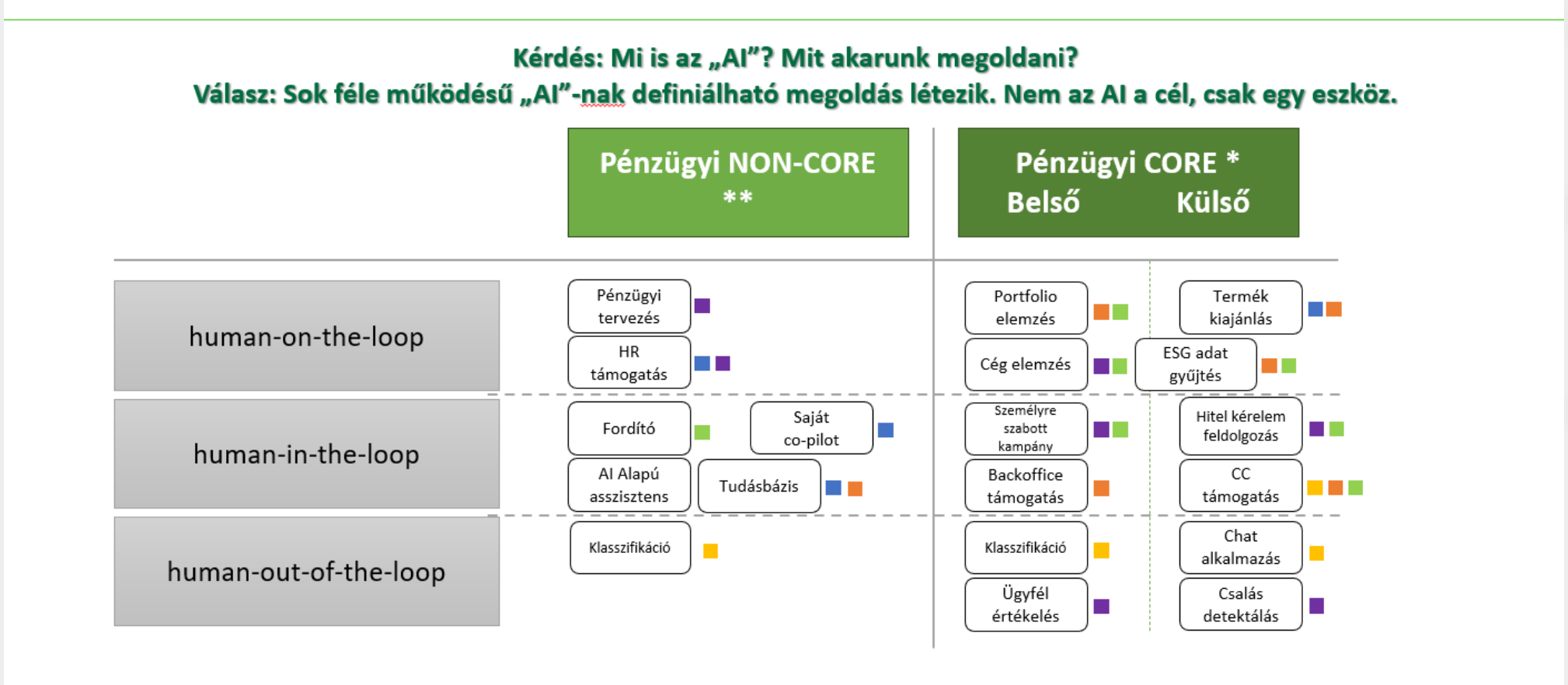
Potenciális szankció rendszer



Forrás: Nagy Zsombor
EIVOK-59

Mi a jó kérdés?

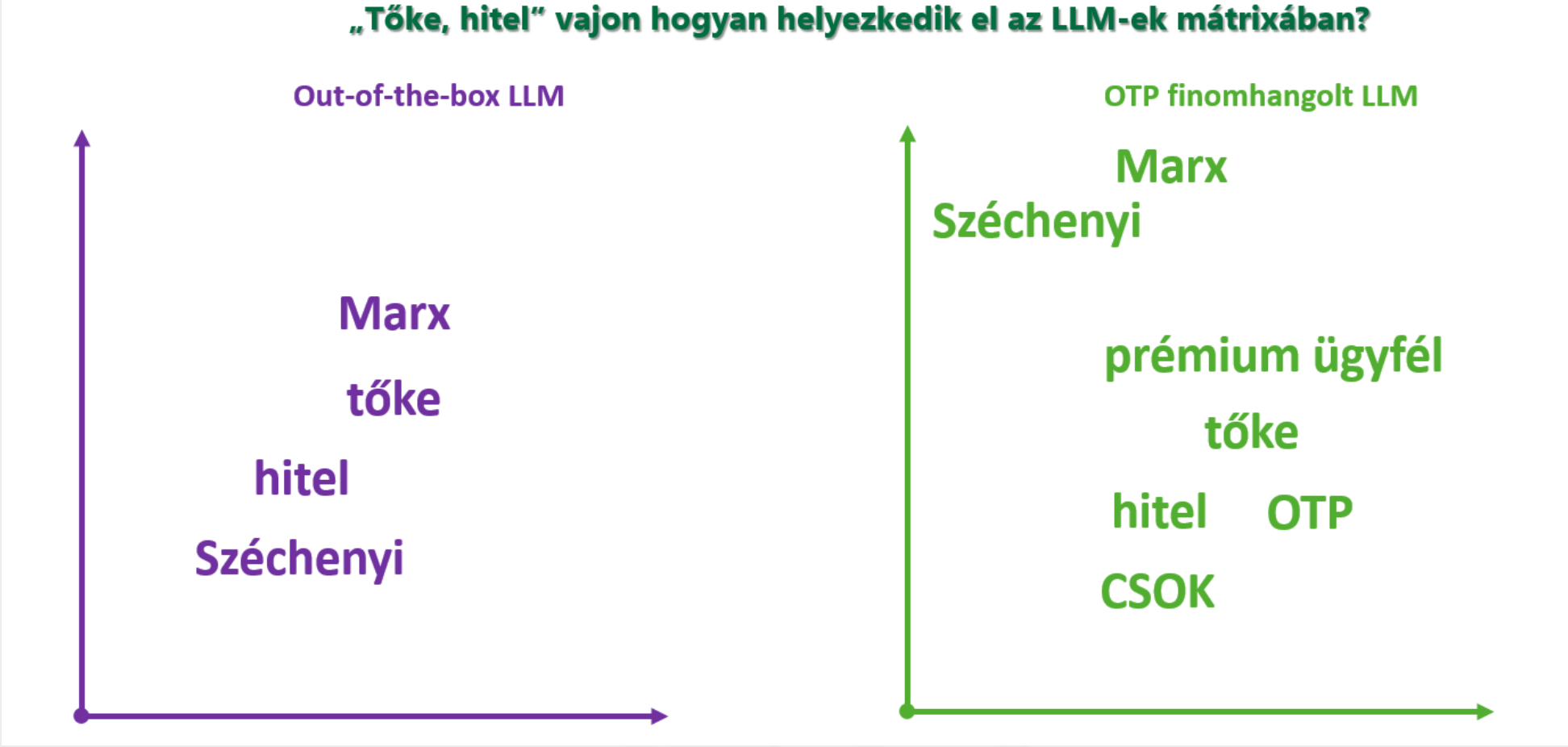
- **Kérdés:** Mi is az „AI”? Mit akarunk megoldani?
- **Válasz:** Sok féle működésű „AI”-nak definiálható megoldás létezik. Nem az AI a cél, csak egy eszköz.



Forrás: Nagy Zsombor
EIVOK-59

Miért érdemes finomhangolni, adaptálni LLM-eket?

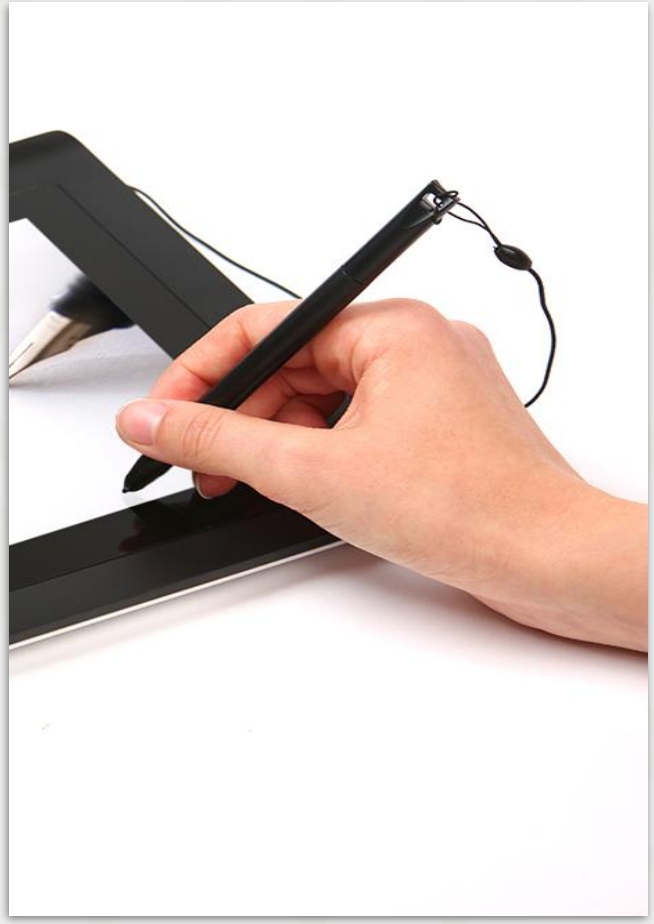
Out of the Box modellek, szolgáltatások is jók, de egyedi igényekre nem mindig optimálisak



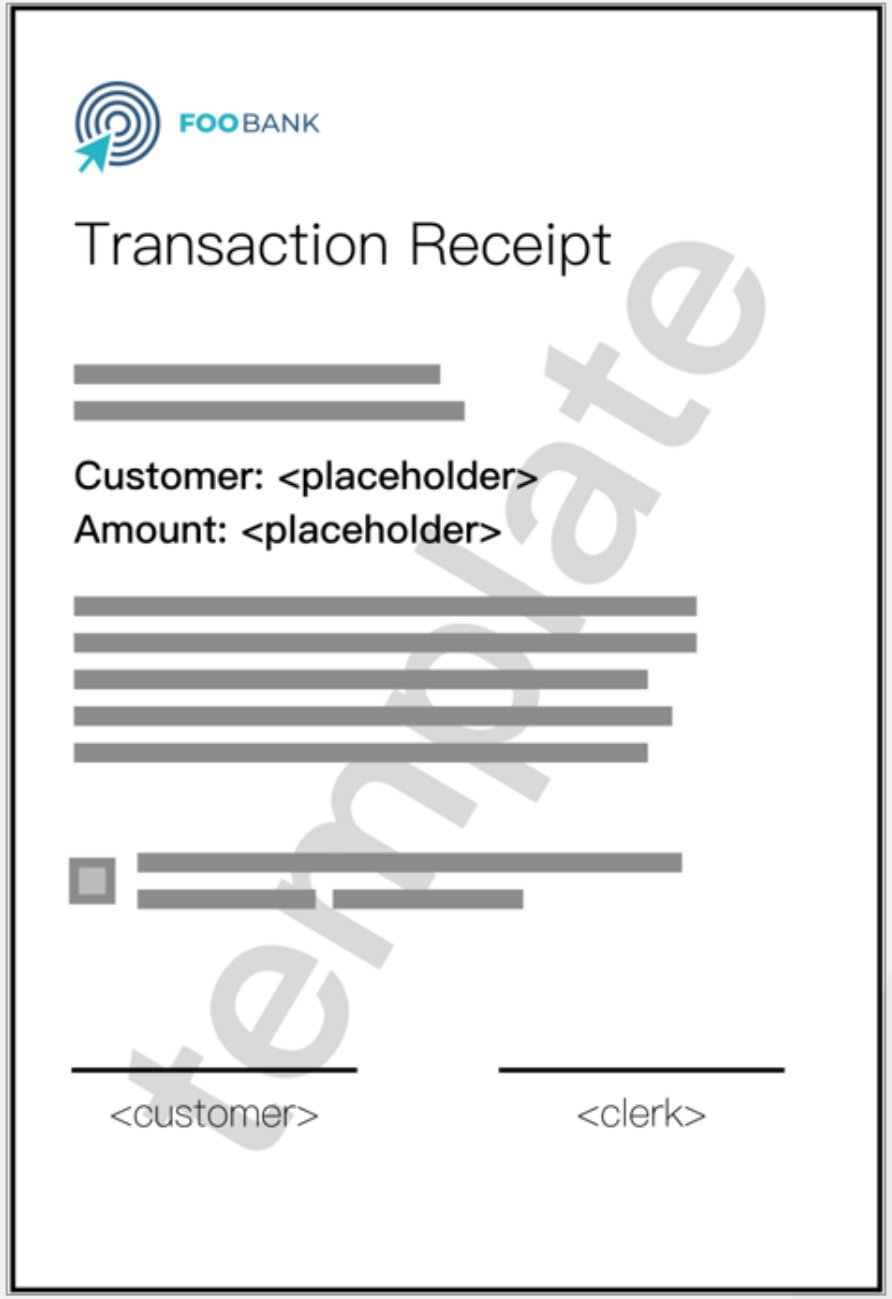
Forrás: Nagy Zsombor
EIVOK-59

OTP: Aláírópad projekt (SignoWise) 2015

- Dokumentumok digitális aláírása
- Minősített tanúsítványon alapuló fokozott biztonságú elektronikus aláírással ellátott dokumentumok (+ mobil)
- Személyes tanúsítvány nélkül
- Integráció
- Üzemeltetés támogatás



Forrás: Golda Bence EIVOK-59

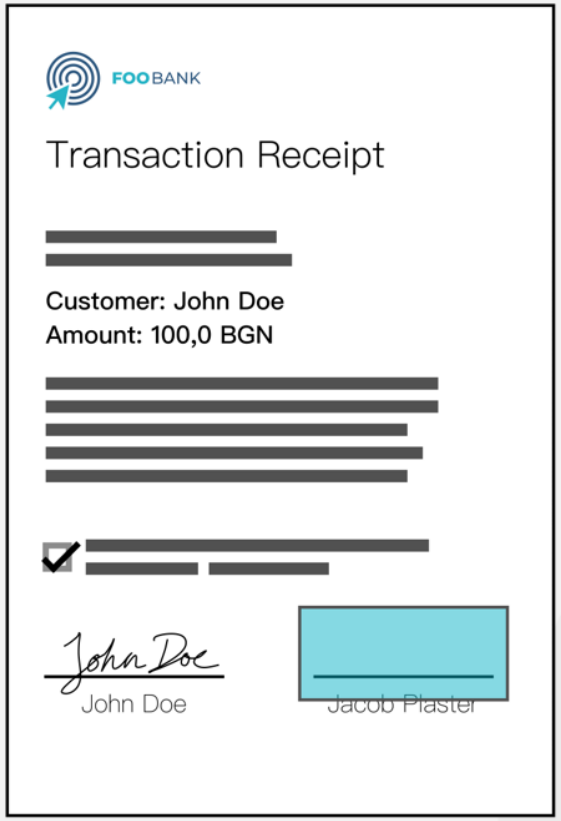


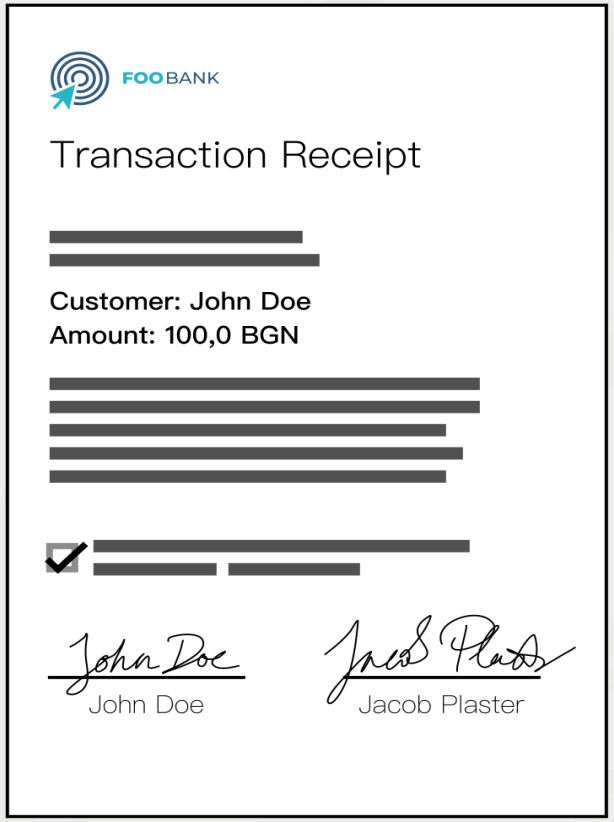
Hitelesítési alapok

- **+ metaadat**
- **+ interaktív mezők**

Forrás: Golda Bence EIVOK-59

Felek aláírása





Forrás: Golda Bence EIVOK-59

Banki környezet (2015)

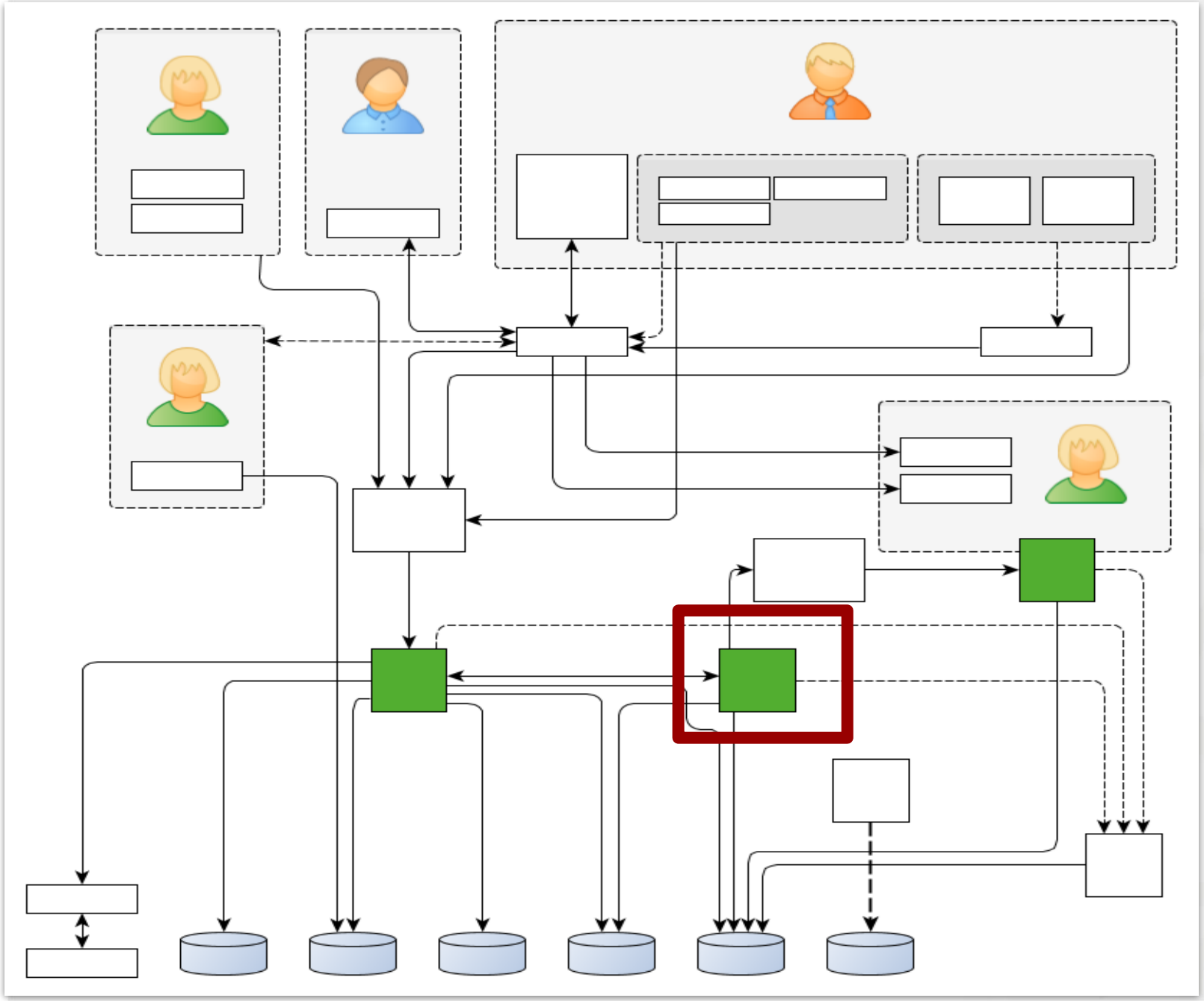
- IT biztonság
- Nagy, komplex rendszerek
- Fix méretezés
- Izoláció (konténerizáció)
- Audit
- HFWF (High Frequency Waterfall)
- f(bus) = 1

Nagy kockázatú rendszerek, erősen szabályozott feltételekkel

Alapjogokat érintő pl. foglalkoztatás, szolgáltatásokhoz történő hozzáférés, igazságszolgáltatás, határigazgatás, bűnüldözés

BIZTONSÁG, KOCKÁZATKEZELÉS, NYOMON KÖVETHETŐSÉG, ADATOK INTEGRITÁSA, EMBERI RÉSZVÉTEL

Forrás: Golda Bence, Nagy Zsombor
EIVOK-59



Integrált AI

Forrás: Golda Bence EIVOK-59

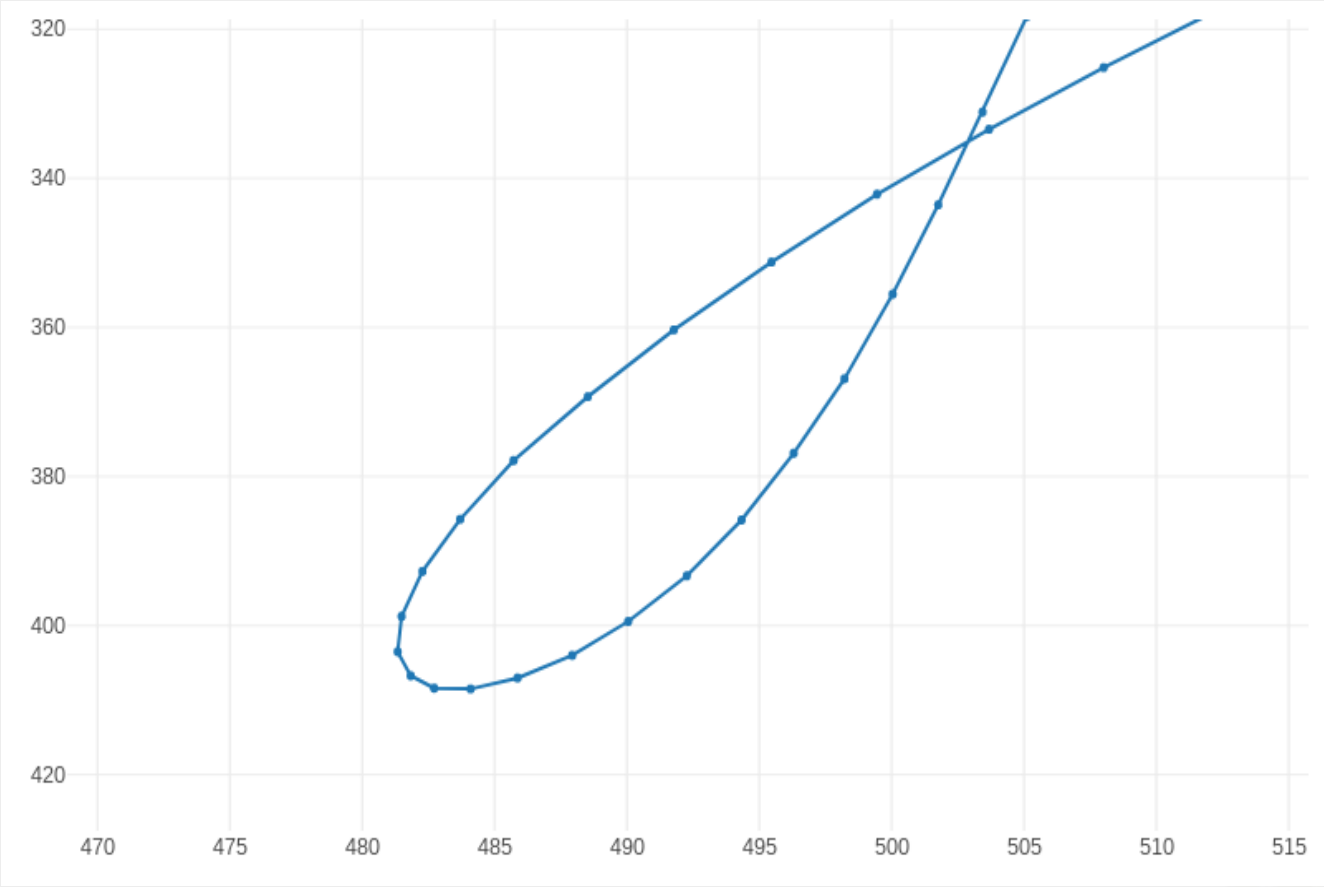
Fejlesztés

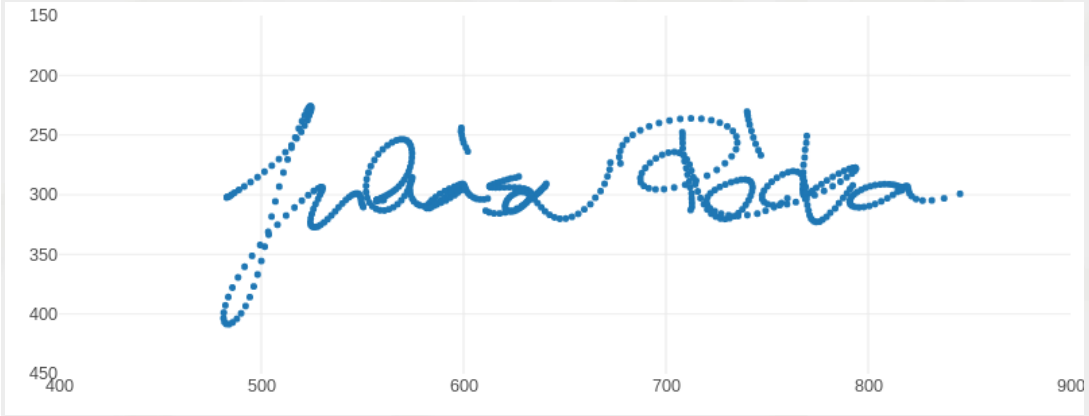
- Cél: kiértékelés hibájának minimalizálása
  - → Megfelelő méretű teszt adat / tanuló adat
  - → Legkisebb homogén csoport mérete meghatározó
  - → (Nagyon) sok adatra van szükség
  - → Megbízható adatgyűjtés

||: adatgyűjtés → ~tárolás → ~tisztítás → elemzés, tanulás → klasszifikáció, predikció → vizualizáció → refaktorálás :||

Forrás: Golda Bence EIVOK-59

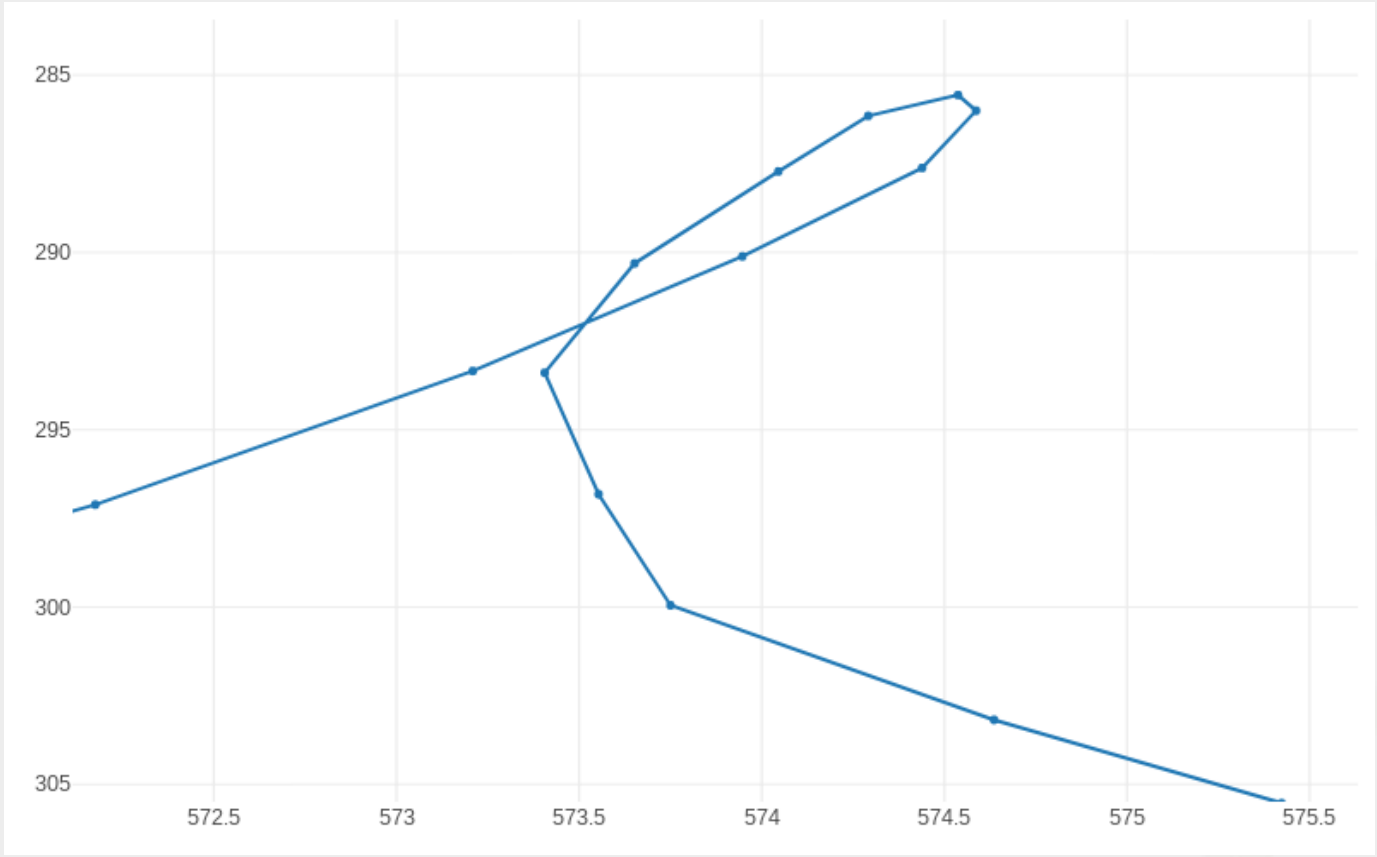
Példák – az ideális adat

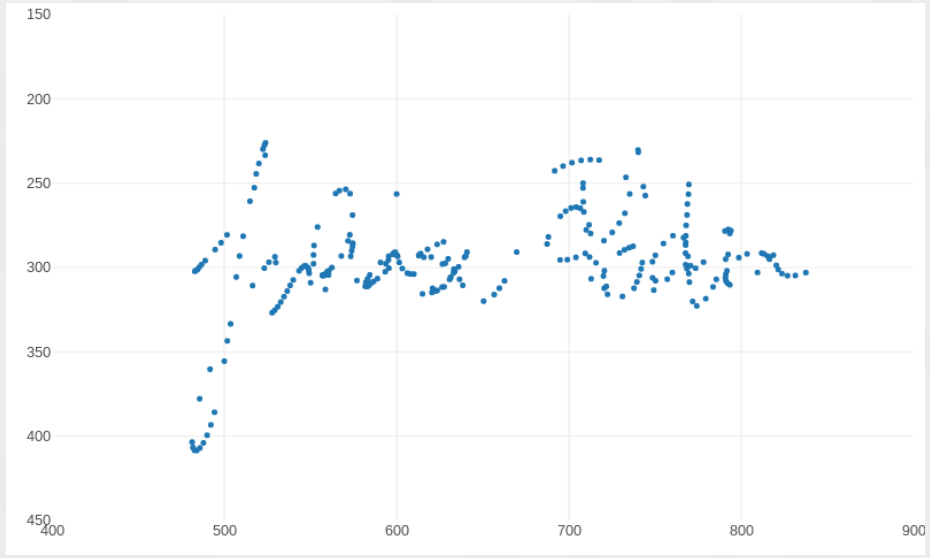




Forrás: Golda Bence EIVOK-59

Példák – valóság

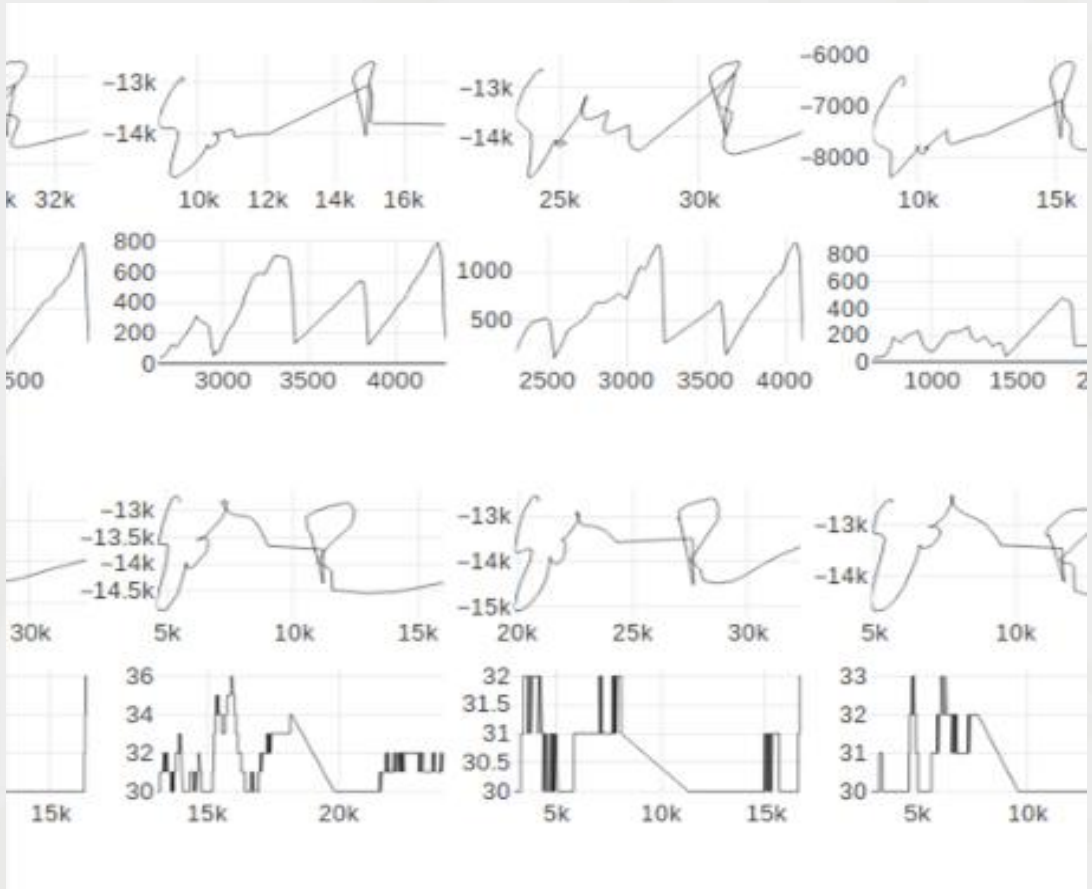




Forrás: Golda Bence EIVOK-59

Tapasztalatok

- Nincs univerzális megoldás
- 20% hiba – 80% adatelemzés
- Van 3-5% “outlier”

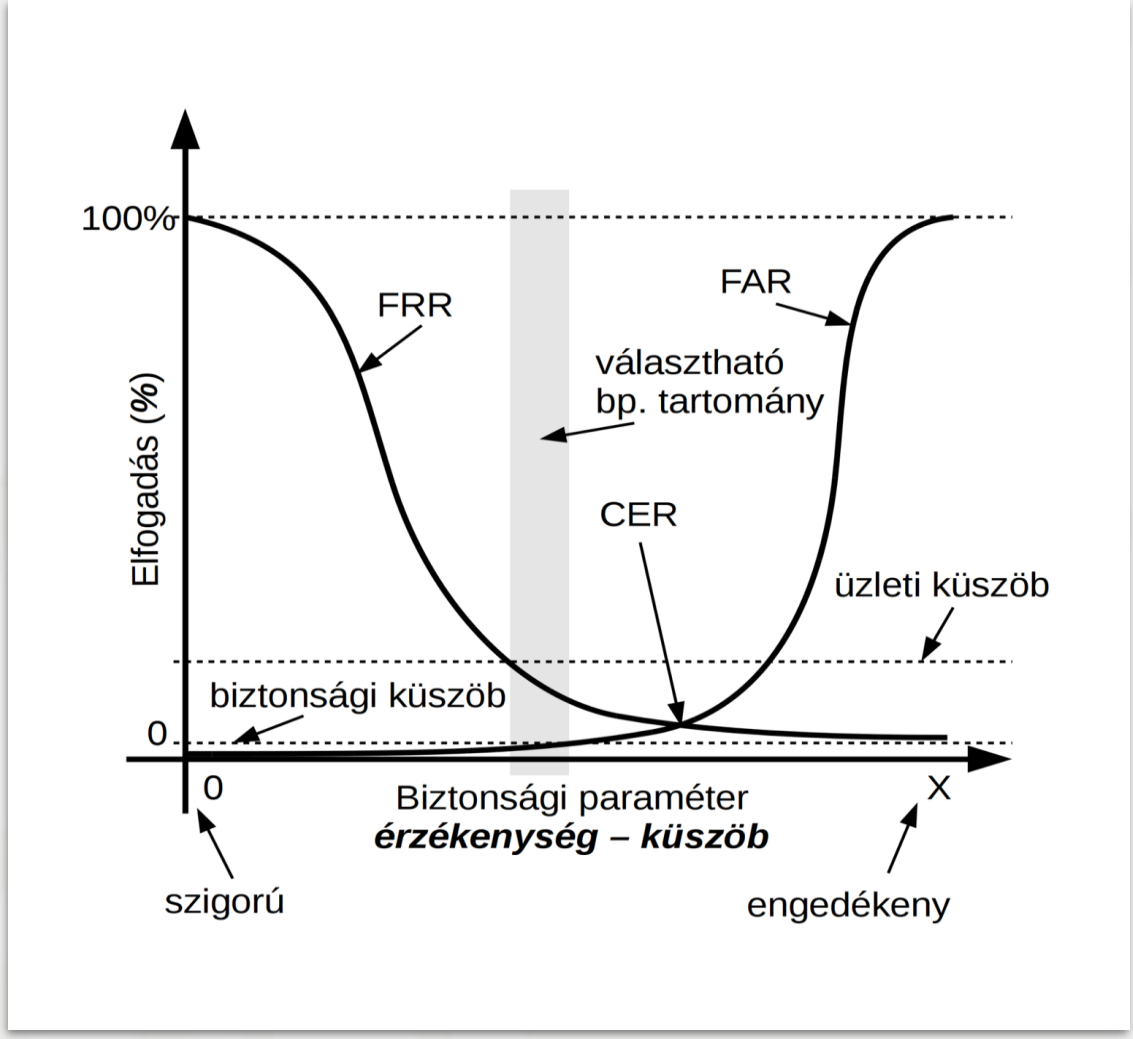


Forrás: Golda Bence EIVOK-59

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

Alkalmas még?

MFA
a
munka-
folyamatokba
építve !?

Forrás: Golda Bence, op EIVOK-59

Mi és az MI

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

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

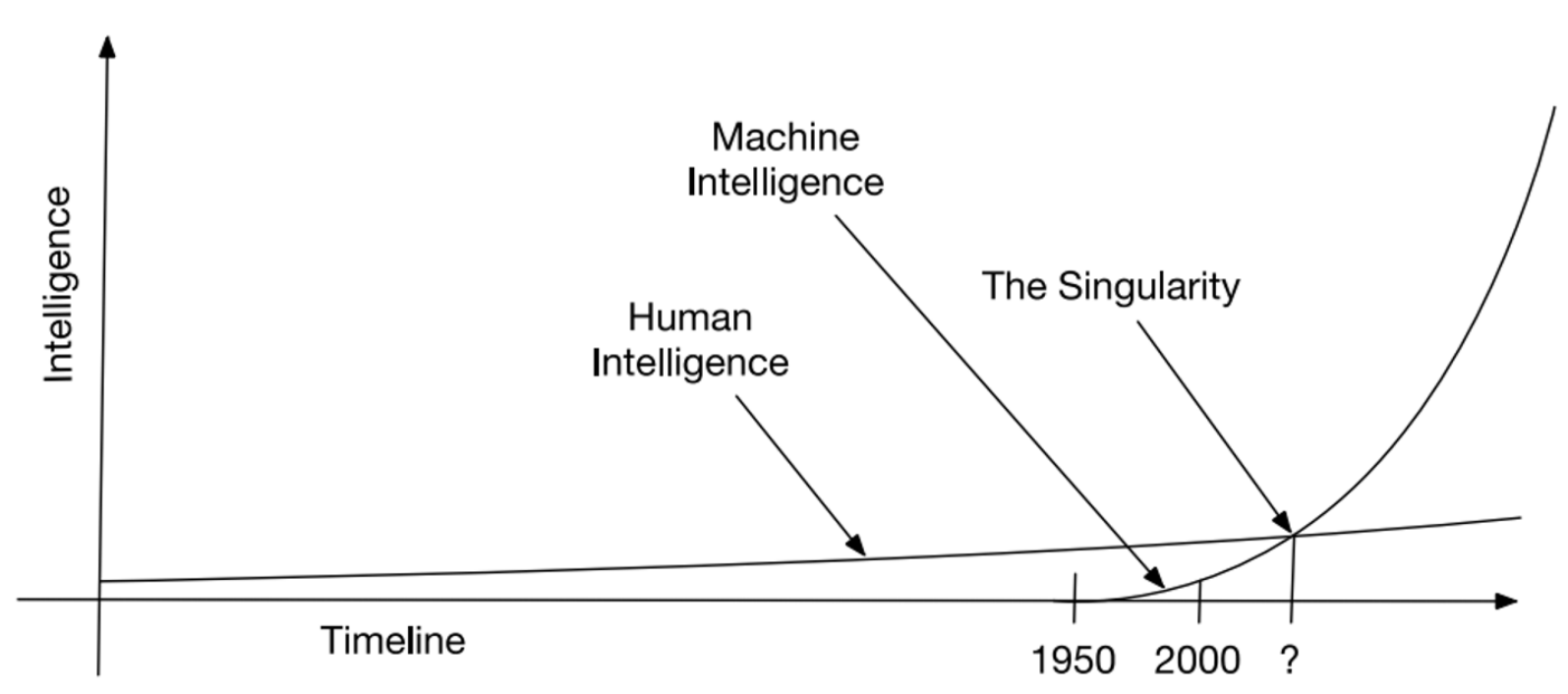
Forrás: Golda Bence / op EIVOK-59

Hol AI lehet a kontrollokat kialakítani?

MINDENHOL !

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

Szingularitás?

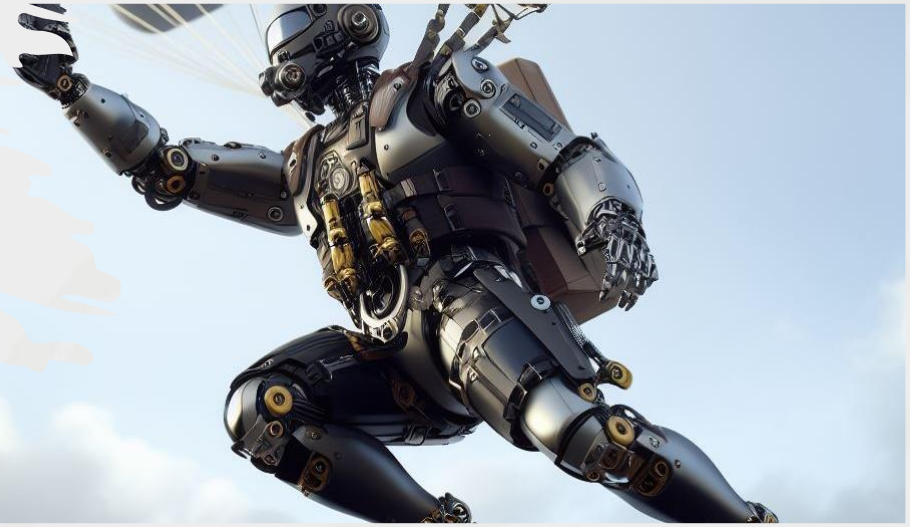


Forrás: Tóth Márton EIVOK-40

Forrás: Microsoft Copilot

AI és hadviselés

„A minisztériumnak cselekednie kell, hogy a mesterséges intelligenciát integrálja a kritikus funkciókba, a meglévő rendszerekbe, gyakorlatokba és hadgyakorlatokba, hogy 2025-re mesterséges intelligenciára kész haderővé váljunk.”



“The Department must act now to integrate AI into critical functions, existing systems, exercises and wargames to become AI-ready force by 2025.”

Forrás: Tóth Márton EIVOK-40

KIBER és hadviselés

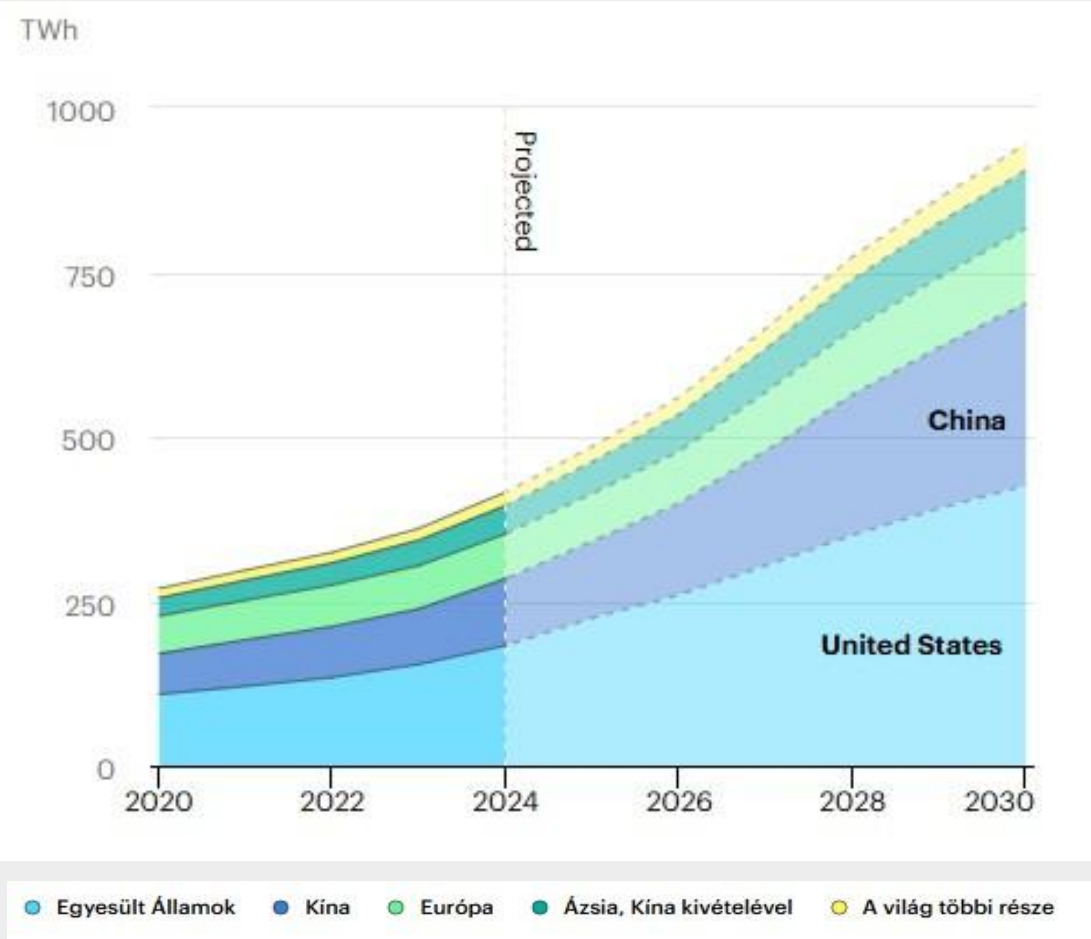
Hol lehet csatát vívni jogilag

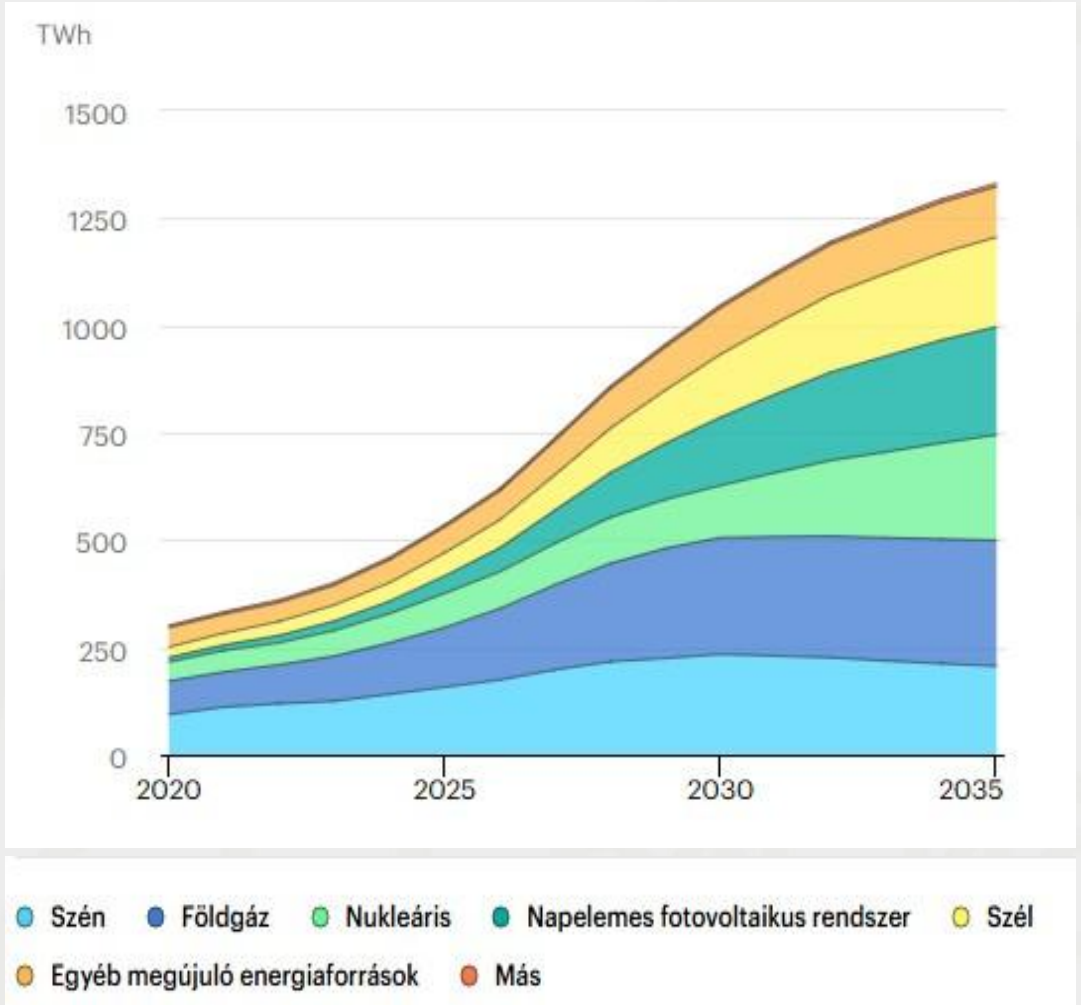
- Föld
- Víz
- Levegő
- Világűr

- Kiber? 2016. NATO V. cikkely kiegészítés

Környezetvédelem-AI

Adatközpontok energia igénye, és forrás mix:





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

Fókusz a jövőben?

Kvantumbiztonság!

Kvantum+Felhő+AI?

Biológia + Matek processzorok

KÖSZÖNÖM A FIGYELMET!

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

istvan.olah@hte.hu

olah.istvan.gyorgy@uni-nke.hu

olah.istvan@uni-obuda.hu

uni-nke.hu