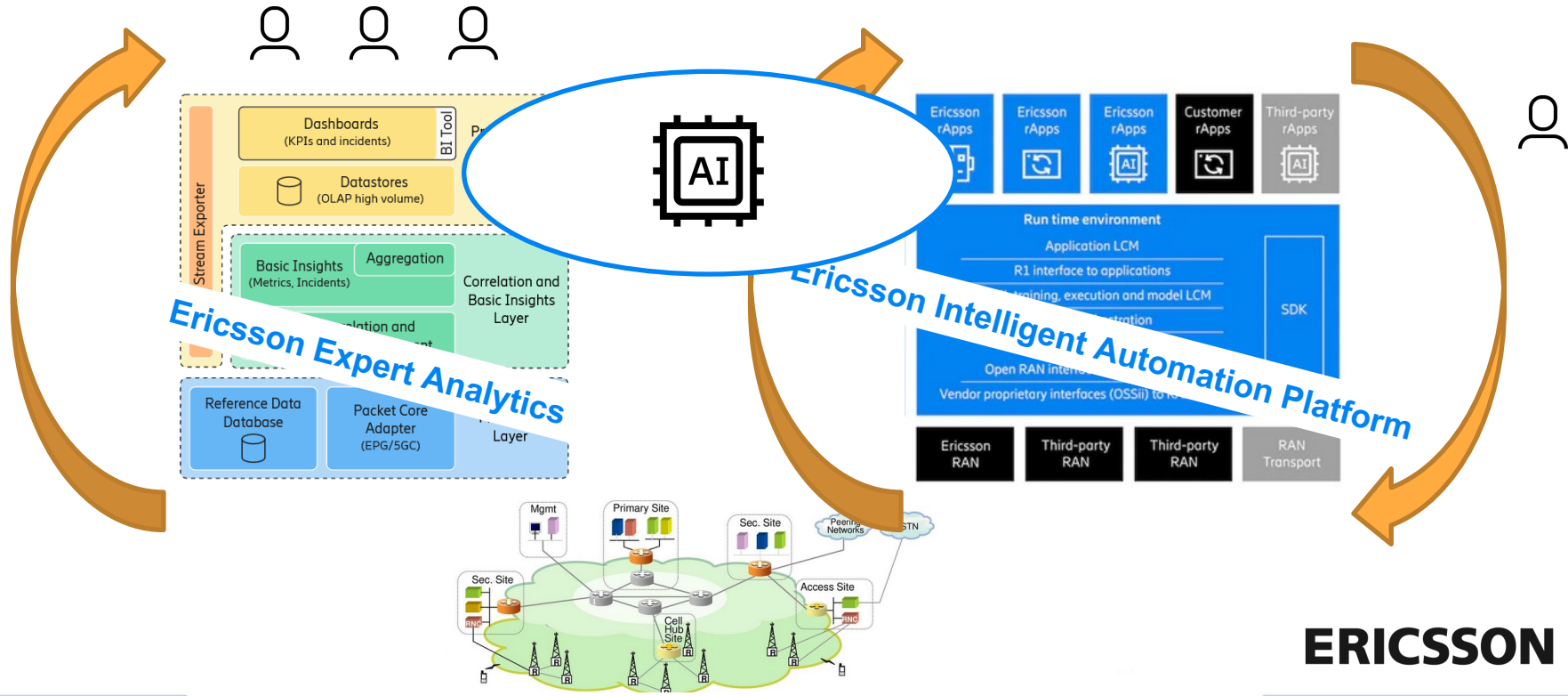


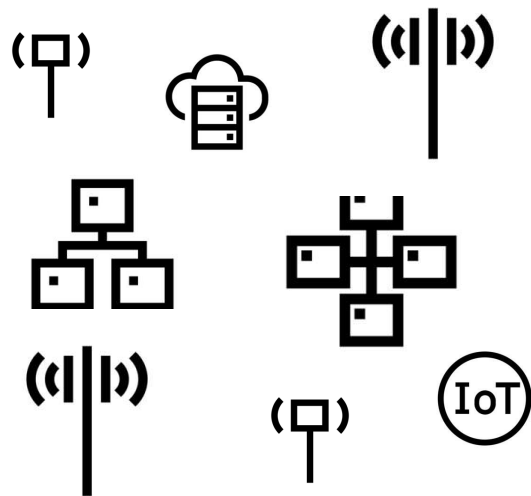
## Szintetikus adatgenerálás – az adathozzáférés Szent Grálja?

Dévai Gergely  
rendszermérnök  
Ericsson

## Analitika és zárláncú automatizálás



## Adathozzáférés



Távközlési hálózat

Személyes  
adatok

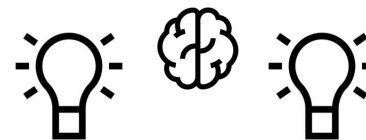
Üzleti  
titkok

GDPR

CCPA

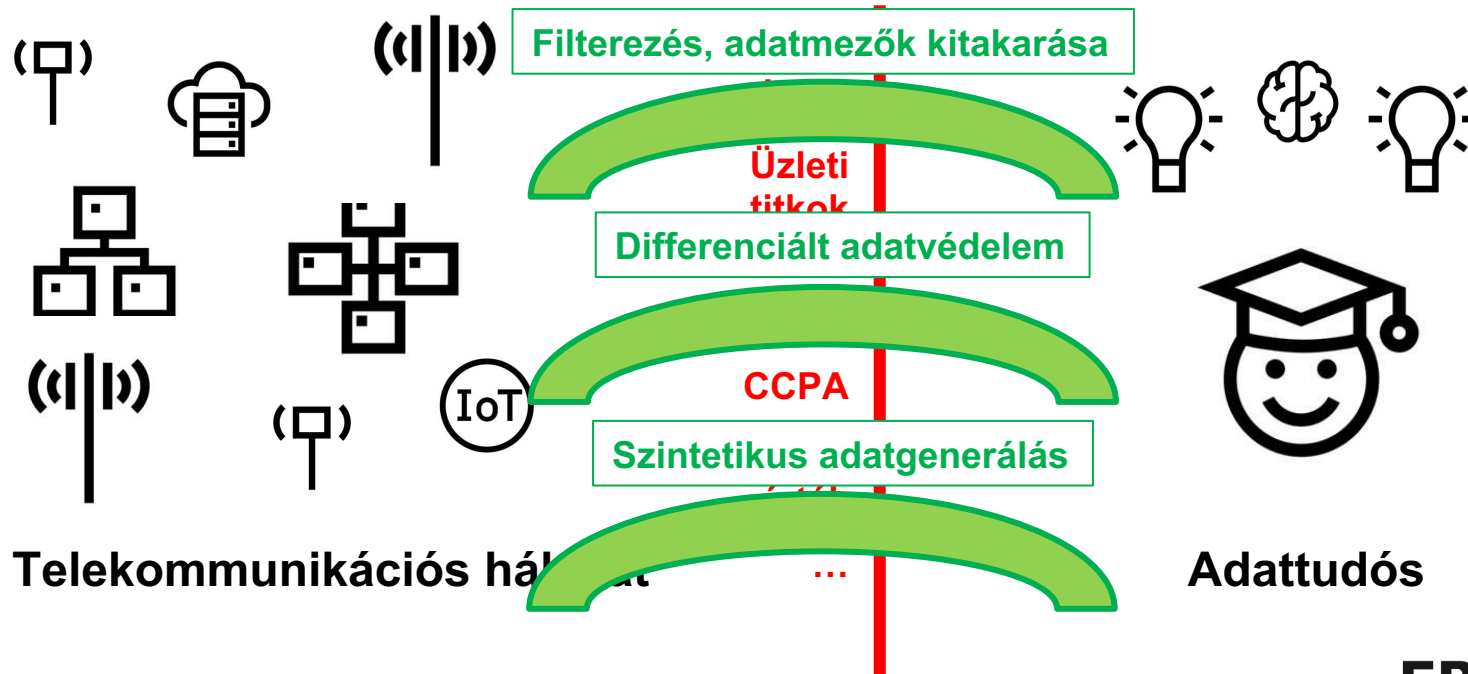
Az adat  
érték

...



Adattudós

## Adathozzáférés: Lehetőségek





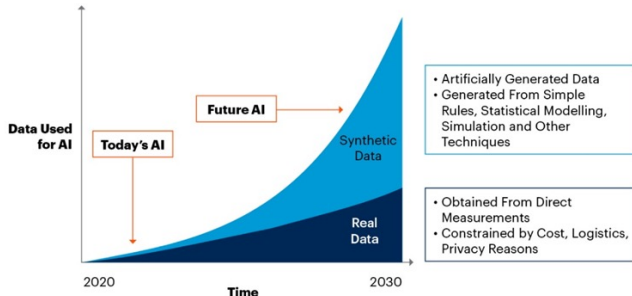
## Szintetikus adat

- Mesterségesen generált
- Forrás: Szimuláció vagy betanított generatív modell

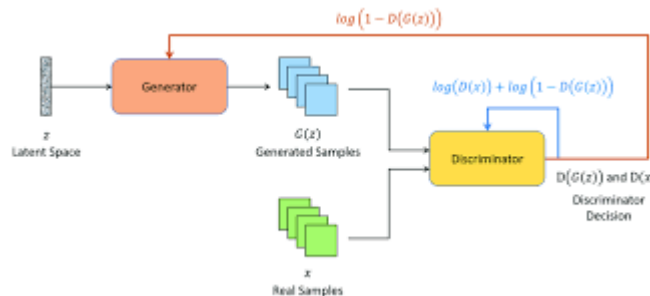


forrás: [thispersondoesnotexist.com/](http://thispersondoesnotexist.com/)

By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models



Source: Gartner  
750175\_C



forrás: [researchgate.net](https://www.researchgate.net)



## Módszerek, eszközök





## Kiértékelés: Generált jelentések

### SYNTHETIC DATA QUALITY REPORT

**489** # REAL DATA RECORDS

**489** # SYNTHETIC DATA RECORDS GENERATED

**11** # COLUMNS

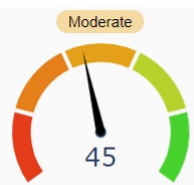
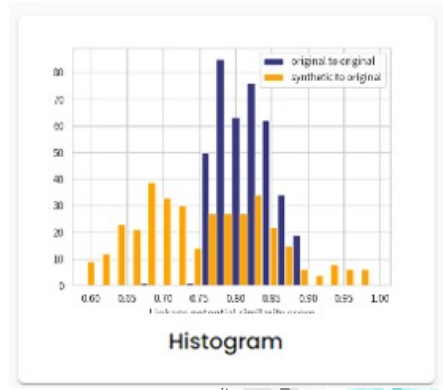
**0.81**

UTILITY SCORE

**0.92**

PRIVACY SCORE

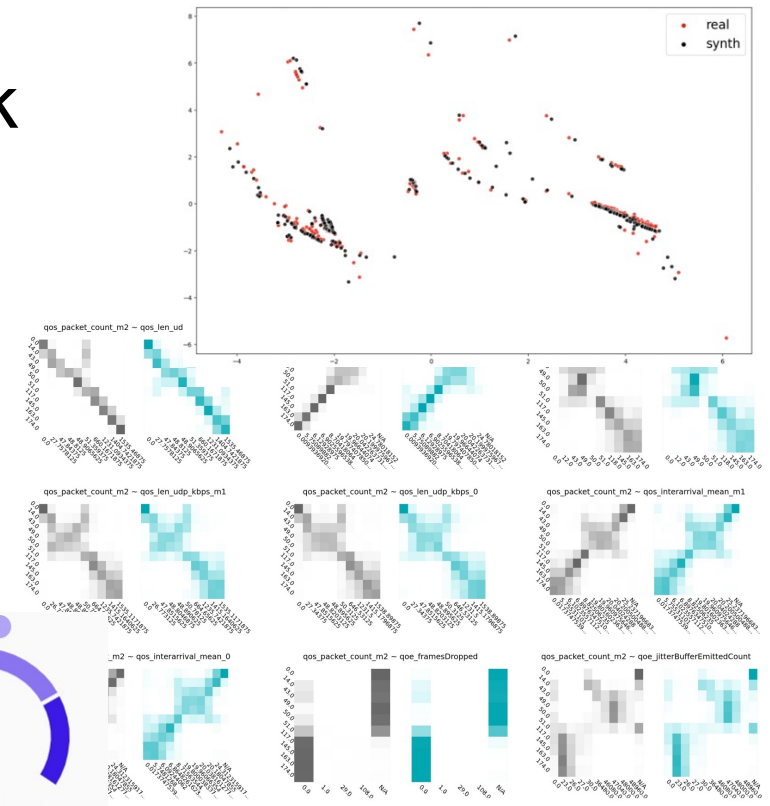
**0.96**



Synthetic Data Quality Score ⓘ

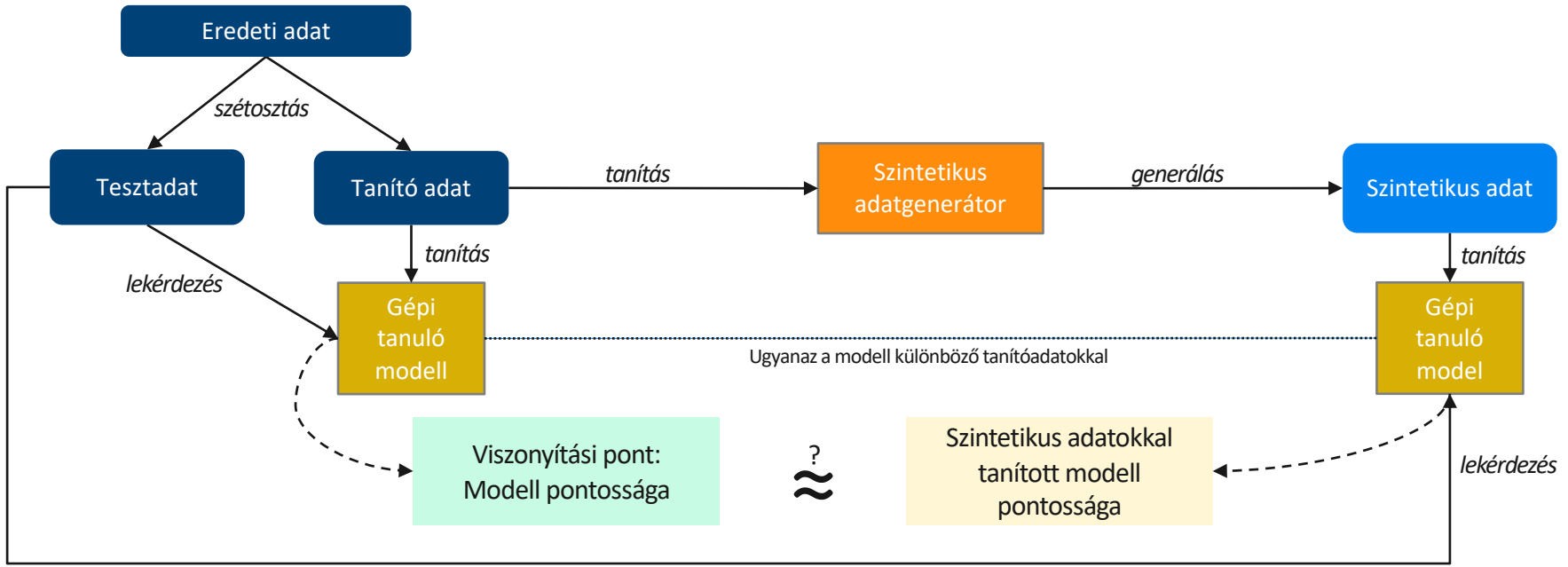


Privacy Protection Level ⓘ





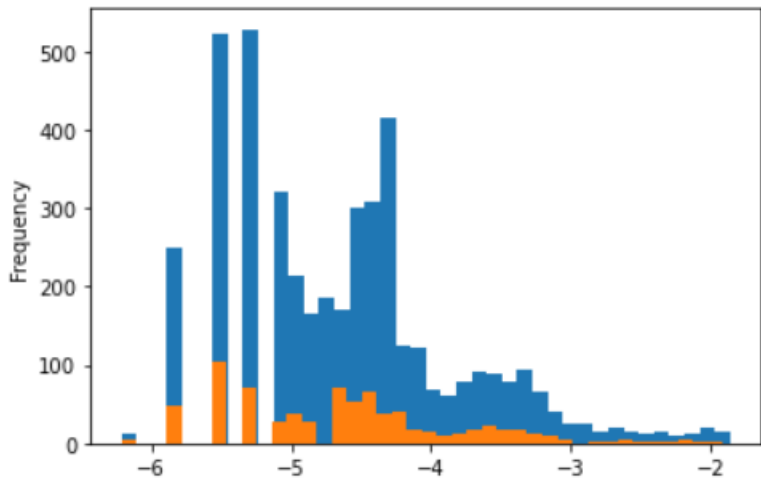
## Kiértékelés: Gépi tanulás szintetikus adatokon?







## Gépi tanulás szintetikus adatokon: eredmények



A modell által becsülendő érték hisztogramja  
(tanítóadat, tesztadat)

Eredeti adatokon tanított modell átlagos hibája:	$\pm 0.20$
---	------------

1. kereskedelmi eszköz szintetikus adatával tanítva:	$\pm 0.23$
---	------------

2. kereskedelmi eszköz szintetikus adatával tanítva:	$\pm 0.27$
---	------------

3. kereskedelmi eszköz szintetikus adatával tanítva:	$\pm 0.30$
---	------------

Nyílt forráskódú eszköz szintetikus adatával tanítva:	$\pm 0.39$
--	------------

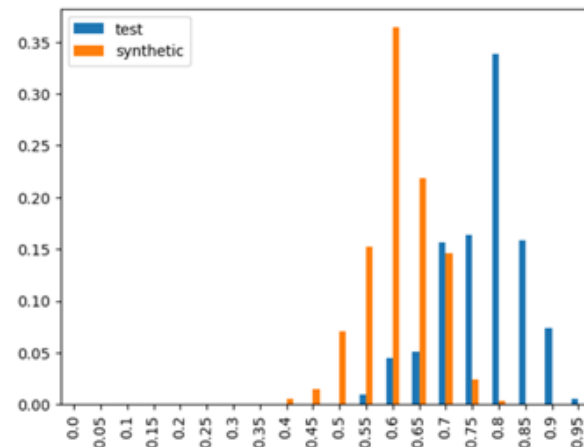
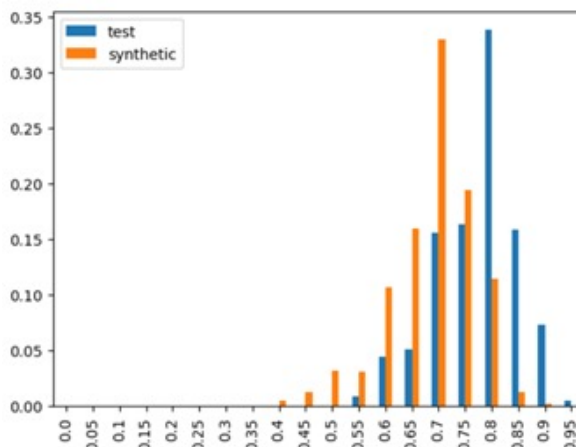
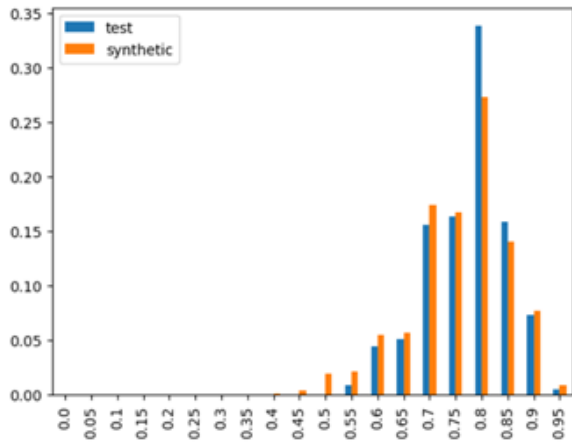
4. kereskedelmi eszköz szintetikus adatával tanítva:	$\pm 0.48$
---	------------



## Kiértékelés: Adatvédelem

„Összekapcsolhatóság”:

- Milyen közel van a generált rekordokhoz a leghasonlóbb eredeti?
- Mennyire hasonló az eredeti adat önmagához?

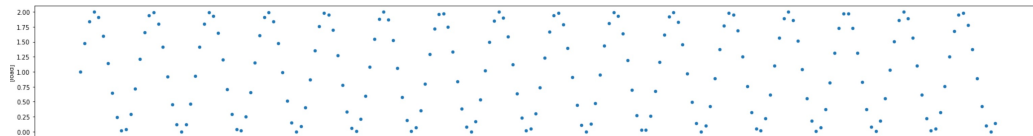




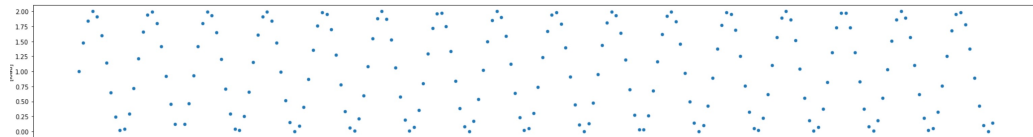
## Kihívás: idősoros adatok

„Időbélyeg” oszlop, „entitás” oszlop, „időfüggetlen” / ”időfüggő” oszlopok...

Eredeti adat:



Néhány eszköz jól „megtanulja”:



Más eszközök „kevésbé”:





## Szent Grál?

- A szintetikus adatgenerálás működik.
- Elfogadható pontosságú modellek taníthatók a szintetikus adatokon.
- Elérhetőek kényelmesen használható eszközök a piacon.
- Nagy különbségek vannak az eszközök között mind használhatóság, mind adatminőség szempontjából.
- Az idősoros adatok terén fejlődés szükséges.
- Vajon bizalmat kelt az adatgazdáknban?





Köszönöm a figyelmet!

Dévai Gergely  
gergely.devai@ericsson.com