precognox

# A mesterséges intelligencia kettős szerepe az információbiztonságban: kockázat vagy a védelem eszköze inkább?

Hírközlési és Informatikai Tudományos Egyesület
Információbiztonsági Szakosztály szakmai fóruma

**2024.10.29.**

# Az MI körüli technológiai áttörések új információbiztonsági kockázatokat is hoztak

A mesterséges intelligencia lendületes fejlődése alapjaiban alakította át a nyelvtechnológiát, **új lehetőségeket** és **új információbiztonsági kockázatokat** is teremtve a természetes nyelvi feldolgozásban.

*Erről a kettősségről osztanám meg ma a leggyakoribb aggályokat, amik a döntéshozókban felmerülnek, amikor MI alapú asszintencia bevonásán gondolkodnak a munkafolyamataikat végző munkatársaik mellé.*

**Kása Károly**
CTO, Solution Architect
precognox

Az egyik alapítója vagyok a Precognox Kft.-nek, ahol a jogelődöket is számítva, több, mint 20 éve dolgozunk szövegbányászattal megoldható problémákon, természetes nyelvfeldolgozás és gépi tanulási technológiák alkalmazásával.

# **Bevezetés** - A nyelvtechnológiai és szövegbányászat eszközök és az MI kapcsolata

A **nyelvtechnológia** olyan technológiák és eszközök összessége,
amelyek a természetes nyelveket (például a beszédet vagy írott szöveget) dolgozzák fel.

A **szövegbányászat** olyan módszerek halmaza, amelyek célja értékes információk kinyerése nagy mennyiségű szöveges adathalmazból.

A **nyelvtechnológia** és azon belül a **szövegbányászat** régóta erősen **támaszkodik a mesterséges intelligencia (MI) eszközeire**, különösen a természetes nyelvfeldolgozásra (NLP) és a gépi tanulásra.

Ezek az eszközök lehetővé teszik, hogy a számítógépek értsék, elemezzék és feldolgozzák az emberi nyelvet, ami automatikus folyamatokat, mint például a fordítás, kategorizálás, és érzelemfelismerés, tesz hatékonnyá és gyorssá.

# A Mesterséges Intelligencia fejlődésének hatása

**Természetes nyelvfeldolgozás (NLP):** Az NLP az MI egyik ága, amely kifejezetten a természetes nyelvek számítógépes feldolgozására fókuszál. Az NLP technikák lehetővé teszik a szövegek megértését, fordítását, összefoglalását, valamint az automatikus válaszadást. Tipikus alkalmazási területek: gépi fordítás, chatbotok, nyelvtani ellenőrzés, beszédfelismerés, névelemek felismerése.

**Gépi tanulás (Machine Learning):** A nyelvtechnológiai alkalmazások gyakran használnak gépi tanulási algoritmusokat, amelyek képesek tanulni nagy szöveges adathalmazokból, és ezáltal felismerni mintákat. Ez lehetővé teszi a szövegkategorizálást, az érzelemelemzést (sentiment analysis), illetve az automatikus válaszgenerálást.

**Deep Learning és neuronhálók:** **A modern nyelvtechnológia jelentős előrelépéseket tett a mély gépi tanulás (deep learning) révén**, különösen a mély neuronhálók, például a transformer-alapú modellek (pl. BERT, GPT) alkalmazásával.

Ezek a nagy nyelvi modellek (LLM) képesek kontextuális nyelvi feladatok megoldására, például kérdés-megértésre vagy szöveggenerálásra.

# MI által generált új információbiztonsági kockázatok
## Bizalmasság (Confidentiality)

A bizalmasság célja, hogy *az információhoz csak azok férjenek hozzá, akik jogosultak erre.*

### Aggályok

Engedély és kompenzáció nélküli tartalom felhasználás történik az LLM fejlesztők által a tartalom előállítók kárára.

Rá lehet venni az LLM-eket promptolással, hogy olyat áruljanak el, amihez csak másnak van hozzáférése?

A Cloud alapú LLM-ekhez feltöltheti-e egy cég a dokumentumait, hogy kérdéseire választ kapjon anélkül, hogy illetéktelenek férjenek hozzá féltett információhoz?

# MI által generált új információbiztonsági kockázatok

## Integritás (Integrity)

Az integritás azt jelenti, hogy az információ nem módosítható vagy törölhető jogosulatlanul.

A változásokat csak engedéllyel lehet végrehajtani, és *fontos biztosítani, hogy a módosítások nyomon követhetőek legyenek.*

### Aggály

A nagy LLM gyártók nem transzparensek a módosításaik nyomonkövethetőségére

# MI által generált új információbiztonsági kockázatok

## Hitelesség (Authenticity)

A hitelesség annak biztosítása, *hogy az információk és kommunikációk valódiak* és nem módosultak az elküldéstől a fogadásig.

### Aggályok

1. hallucináció jelensége (a magabiztosság mögött csak statisztika alapú szógenerálás történik)
2. nincs már elég, jó minőségű tanítóadat és egyre több “körbejáró” szöveg lesz
3. kiemelten igaz ez a forráskódokra (gyengülhet a kódminőség)

# MI által generált új információbiztonsági kockázatok

## Adatvédelem (Data Protection)

Az adatvédelem a *személyes adatok kezelésére és védelmére vonatkozik, különösen olyan szabályozások keretében, mint a GDPR.*

A cél az egyének személyes adatainak jogosulatlan felhasználásának és kiszivárgásának megakadályozása.

### Aggály

A Bizalmasság (Confidentiality) részben már említett aggályok, kiemelten a személyes adatok hozzáférhetővé válása tükrében

# MI a kiberbiztonság szolgálatában

A kiberbiztonság területén a nagy adatmennyiségből minél korábbi és gyorsabb anomália felismerés igénye kapcsán - tudomásom szerint - régóta alkalmazzák az MI-t.

**Csak felsorolásszerűen néhány terület**

Anomáliadetektálás és betörésészlelés

**Kiberfenyegetések előrejelzése**

Automatikus válasz egyes típusú kiberincidensekre

**Adatvédelmi és hozzáférés-ellenőrzési rendszerek**

*Phishing és Social Engineering Támadások Felismerése*

Automatikus Víruselemzés és Malware Felismerés

*Szövegbányászat és Threat Intelligence*

# MI az információbiztonság szolgálatában
# A Precognox felhasználói esete

Data Protection Agreement (DPA) és Technical and Organizational Measures (TOM) auditálás támogatása.

Azt vizsgáljuk, hogy az ügyfelünk által korábban elvégzett többszáz auditjukból rendelkezésre álló historikus adatbázisuk alapján készíthető-e egy olyan szoftver asszisztens, amely előminősíti a szakértőknek az ügyfeleik által beadott dokumentumokat a DPA és TOM kérdéssorra.

**Megoldásunk**

- Nem vetettük el a klasszikus ML megközelítést sem, hisz rendelkezünk annotált, szakértők által címkézett, strukturált adattal.
- Ugyanakkor első körben LLM-re támaszkodó megoldást vizsgálunk, amelyben benne van a potenciál egy következő fázisra is, ahol indoklást kapunk.
- Illetve egy harmadik fázisra is, ahol a hiányos vagy hibás szövegezésre javaslatot ad a rendszer.

# Köszönöm a figyelmet!

**Kása Károly**
CTO, Solution Architect
precognox