Magyar Rendészet 2025/4. 177–198.
DOI: 10.32577/MR.2025.4.10

RENDÉSZETTUDOMÁNY

A generatív mesterséges intelligencia (GenAI) alkalmazása a rendészeti kutatásokban

Application of Generative Artificial Intelligence (GenAI) in Law Enforcement Research

ERDÉLYI Katalin[1]

Bevezetés: A generatív mesterséges intelligencia (GenAI) a 21. század egyik legjelentősebb technológiai innovációja. Bár rendészeti célú alkalmazása jelenleg még korai fázisban van, hosszú távú stratégiai eszközzé válhat a rendészeti kutatásokban.

Célkitűzések: A tanulmány célja a GenAI-technológiák hazai adaptációjának elősegítése a rendészeti szektorban. Ennek keretében feltárja a GenAI-alapú kutatások hatékony megvalósításához szükséges legfontosabb technológiai összetevők – mint tanítóadatok, kontextusablak-méret, prompt-stratégia, hallucináció és validáció – sajátosságait. Emellett kitér a GenAI felelős alkalmazásának kérdéseire, a tudományos megbízhatóság szempontjaira és a kutatói szerepek átalakulására. A technológiai potenciált friss, nemzetközi rendészeti kutatási példák révén szemlélteti.

Módszertan: A tanulmány 2016–2025 közötti időszakból származó, több mint 40 tudományos és szakpolitikai dokumentum áttekintésén alapul, többnyelvű keresési stratégiát alkalmazva.

Eredmények: A kutatás során azonosítottuk a generatív mesterséges intelligencia felhasználhatóságát és megbízhatóságát meghatározó alapvető technológiai összetevőket. A GenAI rendészeti célú alkalmazása már megjelent a gyakorlatban, de a nemzetközi példák jellemzően pilotprojektekhez kötődnek, és inkább technológiai kuriózumként, mintsem rutinszerű használati eszközként jelennek meg. A nemzetközi esettanulmányok a GenAI olyan innovatív alkalmazásait mutatják be, amelyekben e technológiát pszichológiai

[1] Tanársegéd, Budapesti Gazdaságtudományi Egyetem Pénzügyi és Számviteli Kar; doktori hallgató, Nemzeti Közszolgálati Egyetem Rendészettudományi Doktori Iskola, e-mail: katalin@irasszakerto.hu

mintázatok – például manipulációs stratégiák, fenyegető vagy kényszerítő kommunikációs formák – azonosítására, érvelési struktúrák feltárására, valamint digitális nyomozati dokumentumok gyors feldolgozására alkalmazták.

Konklúzió: A GenAI-technológia valódi potenciálja akkor bontakozhat ki, ha a felhasználók megfelelő szintű ismeretekkel rendelkeznek a működéséről, és képesek azt tudatosan, szabályozott keretek között alkalmazni. A GenAI tudományos munkába történő sikeres integrációja megköveteli a módszertani szemléletváltást, valamint a kutatói szerepkörök kiterjesztését olyan irányokba, ahol a kutató egyszerre működik technológiai eszközhasználóként, kritikai validátorként és etikai döntéshozóként. Ez a megközelítés biztosítja, hogy a GenAI által generált eredmények tudományosan megalapozottak, etikailag elfogadhatók és módszertanilag ellenőrizhetők maradjanak.

Kulcsszavak: generatív mesterséges intelligencia, GenAI, rendészeti kutatás, prompt, hallucináció

Introduction: Generative Artificial Intelligence (GenAI) represents one of the most significant technological innovations of the 21st century. While its application in law enforcement is still in its early stages, it has the potential to become a long-term strategic tool in law enforcement research.

Objectives: This study aims to facilitate the domestic adoption of GenAI technologies in the law enforcement sector. It explores the key technological components necessary for the effective implementation of GenAI-based research, including training data, context window size, prompt strategies, hallucination risks, and validation processes. Additionally, it addresses the responsible application of GenAI, considerations of scientific reliability, and the transformation of researchers' roles. The technological potential is illustrated through recent international law enforcement research examples.

Methodology: The study is based on a targeted review of over 40 scientific and policy documents from the period 2016–2025, employing a multilingual search strategy.

Results: The research identified the core technological components that determine the usability and reliability of generative artificial intelligence. The application of GenAI in law enforcement has already emerged in practice, but international examples are typically linked to pilot projects and are regarded more as technological curiosities than routine tools. International case studies showcase innovative GenAI applications, such as identifying psychological patterns – e.g. manipulative strategies, threatening or coercive communication forms – uncovering argumentation structures, and rapidly processing digital investigative documents.

Conclusion: The true potential of GenAI can unfold when users possess adequate knowledge of its functioning and are capable of applying it consciously within regulated frameworks. The successful integration of GenAI into scientific work requires a methodological paradigm shift and an expansion

of researchers' roles in directions where the researcher simultaneously functions as a technological tool user, critical validator, and ethical decision-maker. This approach ensures that results generated by GenAI remain scientifically grounded, ethically acceptable, and methodologically verifiable.

Keywords: generative artificial intelligence, GenAI, law enforcement research, prompt, hallucination

Bevezetés

„Az új technológia önmagában se nem jó, se nem gonosz. Minden azon áll vagy bukik, hogy az emberek hogyan használják."
(Bodnár 2022: 19)

A generatív mesterséges intelligencia (GenAI)[2] megjelenése új korszakot nyitott a tudományos kutatásban, radikálisan átalakítva az ember és gép közötti tudásmegosztás és együttműködés formáit. A ChatGPT 2022. novemberi debütálása – amely a digitális technológiák történetének egyik leggyorsabb felhasználói adaptációját eredményezte – katalizátorként hatott az egész GenAI-szektor fejlődésére. 2025 augusztusára már több mint 230 generatív mesterségesintelligencia-alapú rendszer vált nyilvánosan elérhetővé, általános célú és szakterület-specifikus megoldásokkal egyaránt (Artificial Analysis 2025). Ez a sokszínűség jól mutatja, hogy a GenAI túllépett eredeti tartalomgenerálási funkcióján, és egyre szervesebb részévé válik az üzleti, oktatási és tudományos ökoszisztémáknak.

A technológia fejlődése a korai nyelvtani ellenőrző eszközöktől az olyan nagy nyelvi modellekig vezetett, amelyek már képesek nemcsak szövegek automatikus generálására és értelmezésére, hanem komplex adatstruktúrák feldolgozására, statisztikai és szemantikai mintázatok azonosítására, többnyelvű szövegkorpuszok összehasonlító elemzésére, valamint az eredmények interaktív vizualizációjára (Raffel et al. 2020). E rendszerek tudományos integrációja mélyreható metodológiai átrendeződést kíván, amely újradefiniálja az értelmezés, adatfeldolgozás és érvelés hagyományos kereteit. Ezzel párhuzamosan új megvilágításba kerülnek a kutatói szerepek, az elemzési módszerek, valamint az etikai felelősség kérdései is. A GenAI integrációja szükségessé teszi a *human-in-the-loop* (HITL)[3] és *human-over-the-loop* (HOTL)[4] megközelítések következetes alkalmazását, amelyekben az emberi döntéshozó aktívan részt vesz a folyamatban, illetve végső felügyeleti és validációs szerepet tölt be.

[2] A generatív modellek családja, amely mélytanulási architektúrákon alapul, és képes új, eredeti tartalmak (szöveg, kép, hang, videó) előállítására meglévő minták alapján. Ide tartoznak például a nagy nyelvi modellek (LLM) és a diffúziós alapú képgenerátorok.

[3] *Human-in-the-loop* (HITL): az ember aktívan részt vesz az AI működésében. Bővebben lásd később.

[4] *Human-over-the-loop* (HOTL): az ember csak szükség esetén avatkozik be az AI működésébe. Bővebben lásd később.

A rendészeti kutatásokban ez a transzformáció különösen jelentős. A GenAI támogató szerepet tölthet be olyan feladatokban, mint a bűnözési mintázatok feltárása, a nyomozati dokumentumok gyors és pontos feldolgozása, a közbiztonsági kockázatok előrejelzése, valamint a prevenciós stratégiák kidolgozása (Nikolakopoulos et al. 2024). Bár az elmúlt években egyre több rendvédelmi szerv kezdett kísérletezni vagy pilotprogramokat indítani a GenAI-technológiák alkalmazásával, ezek használata jelenleg még túlnyomórészt kísérleti vagy elméleti szinten zajlik.

Jelen tanulmány célja, hogy feltárja a GenAI-alapú kutatások hatékony megvalósításához szükséges technológiai, módszertani és etikai kereteket, valamint nemzetközi példákon keresztül bemutassa a technológia rendészeti alkalmazhatóságát és korlátait.

Módszertan

A generatív mesterséges intelligencia rendészeti kutatásokban való alkalmazásának vizsgálata olyan interdiszciplináris megközelítést igényel, amely ötvözi a technológiai, kriminológiai, jogelméleti és etikai perspektívákat. A témakör újdonsága és a szakirodalom gyors bővülése miatt a munka szisztematikus szakirodalmi áttekintésen alapult, amely során célzott keresési stratégiát alkalmaztunk. A következőkben részletezzük a kutatási kérdéseket, a kutatás során alkalmazott módszertani kereteket és a szakirodalom-feldolgozás folyamatát.

Kutatási kérdések

1. Melyek a GenAI-technológia azon alapvető műszaki komponensei, amelyek meghatározzák felhasználhatóságát és megbízhatóságát a rendészeti kutatásokban?
2. Milyen gyakorlati alkalmazási példák léteznek már nemzetközi szinten a GenAI rendészeti felhasználására, és ezek milyen tanulságokat hordoznak?
3. Milyen módszertani, etikai és biztonsági kihívásokat vet fel a GenAI rendészeti kutatásokba történő integrálása?
4. Milyen kompetenciák és validációs mechanizmusok szükségesek a GenAI felelős és tudományosan megalapozott alkalmazásához?

A forráskiválasztás szempontjai

A kutatás 2024. szeptember és 2025. augusztus között zajlott, a következő tudományos adatbázisokat felhasználva: Scopus, Web of Science, Google Scholar, valamint tematikus repozitóriumok (arXiv, SSRN, SocArXiv). A keresés angol nyelvű kulcsszavakkal történt: *generative AI* OR *GenAI* OR *large language model* OR *LLM* kombinálva a *law enforcement* OR *policing* OR *criminal justice* OR *forensic* kifejezésekkel.

A generatív mesterséges intelligencia rendészeti alkalmazásainak technológiai, módszertani és etikai szempontból történő vizsgálata érdekében több mint 40 dokumentumot tekintettem át. A forrásválogatás az alábbi kritériumok mentén zajlott:

- Az elemzésbe bevont szakirodalom túlnyomó része a 2020–2025 közötti időszakból származik, tükrözve a transzformeralapú modellek érett fázisát és gyakorlati alkalmazásait. Az áttekintés kiegészült néhány korábbi, technológiai alapműnek tekinthető publikációval is, amelyek a generatív modellek fejlődéstörténetéhez és architekturális alapjaihoz nyújtanak kontextust.
- Előnyben részesítettük a lektorált *(peer-reviewed)* tudományos közleményeket, különös tekintettel a nemzetközi folyóiratokban és konferenciákon megjelent publikációkra. Emellett az elemzésbe bevontunk technikai jelentéseket és szabályozási irányelveket is, amennyiben azok technológiai relevanciával és módszertani megalapozottsággal bírtak.
- A válogatás során külön hangsúlyt kaptak azok a források, amelyek a GenAI alkalmazásának etikai és jogi vonatkozásait tárgyalják, ideértve az adatvédelem, az elszámoltathatóság, valamint az emberi felügyelet kérdésköreit.
- Nem válogattam be a kizárólag elméleti, spekulatív jellegű munkákat.

A tanulmányban szereplő eredeti ábrák magyar fordítása és grafikai adaptálása a tartalmi hűség megőrzésével, a forrásmegjelölés következetes betartásával történt.

Technológiai háttér és működés

A GenAI olyan gépi tanulási rendszerek gyűjtőneve, amelyek képesek emberihez hasonló tartalmak (szövegek, képek, videók) előállítására. Ezek a rendszerek különböző mélytanulási architektúrákra épülnek, amelyek közül az egyik legjelentősebb a Google kutatói által 2017-ben az *Attention Is All You Need* című tanulmányban bemutatott transzformer[5] architektúra (Vaswani et al. 2017). Ez a forradalminak minősülő, figyelemmechanizmuson alapuló architektúra tette lehetővé az olyan modellek fejlesztését, mint a Google BERT és az OpenAI GPT-sorozata, amelynek legismertebb alkalmazása a ChatGPT (Radford et al. 2018).

A GenAI elsődleges feladata, hogy emberi nyelven megfogalmazott utasításokra válaszolva új tartalmakat generáljon. Képességei a tanult minták felismerésén és alkalmazásán alapulnak, nem értelmez, nem gondolkodik, hanem statisztikai alapon találgatja, mi lenne a legvalószínűbb válasz vagy szövegfolytatás. A nyelvi modellek, mint a ChatGPT szavak közötti statisztikai és szemantikai összefüggéseket azonosítanak, míg a képgeneráló rendszerek, mint a DALL-E vagy a MidJourney vizuális elemeket és stílusjegyeket kombinálnak szöveges utasítások alapján.

A GenAI teljesítményét és megbízhatóságát számos tényező határozza meg, például a tanítóadatok minősége és mennyisége, a kontextusablak mérete (Press–Smith–Lewis

[5] Egy speciális mesterséges neuronhálózat-architektúra, amely lehetővé teszi, hogy a modell kontextusérzékeny válaszokat generáljon.

2022), a promptstratégia tudatossága, valamint a hallucinációk elkerülését célzó technikák és validációs eljárások. A következőkben ezeket mutatom be röviden.

Tanítóadatok

A generatív mesterségesintelligencia-rendszerek a tényleges használat előtt átesnek egy felkészítő lépésen, az előtanításon, így még a használatuk előtt rendelkeznek általános nyelvi tudással. Az előtanítás hatalmas, strukturált vagy félig strukturált adathalmazokon történik, ez az adatkorpusz teszi lehetővé a modellek számára, hogy felismerjék a nyelvi mintázatokat, szabályszerűségeket és összefüggéseket, és új, korábban még nem látott bemenetekre is releváns válaszokat vagy előrejelzéseket adjanak (Radford et al. 2018). A tanító adathalmazok jellemzően olyan nyilvános forrásokból származnak, mint például a több milliárd weboldal szöveges tartalmát havonta összegyűjtő CommonCrawl webarchívum (Common Crawl Foundation [é. n.]; OpenAI 2019), az emberek által ténylegesen olvasott, OpenAI által összeállított WebTex, az angol Wikipédia, különféle könyvkorpuszok és egyéb szabadon hozzáférhető adatbázisok. A gyártók gyakran emelik ki, hogy modelljeik terabájtnyi szövegen vagy milliárdnyi mintán tanultak, ezzel is erősítve a technológia megbízhatóságába vetett bizalmat.

Kontextusablak

A nyelvi modellek hatékony alkalmazásához különösen fontos a kontextusablak méretének figyelembevétele. A kontextusablak mérete egy technikai korlát, amely meghatározza, hogy mennyi információt tud a rendszer egyszerre feldolgozni, értelmezni, és ennek függvényében tartalmilag koherens, összefüggő és releváns szöveget előállítani (Press–Smith–Lewis 2022). A kontextusablak nagysága modellspecifikus, a mérete tokenekben meghatározott. A természetes nyelvfeldolgozásban a token egy szövegdarab, például egy szó, írásjel vagy karakter, amelyet a modell működése során különálló egységként kezel. A magyar nyelv agglutináló jellege miatt egyetlen magyar szó akár 3-4 tokent is eredményezhet.[6] A kontextusablak nemcsak a felhasználói inputot, hanem a modell válaszát is magában foglalja, így a teljes párbeszéd tokenmennyisége határozza meg, mennyi információval tud a rendszer egy adott pillanatban dolgozni (Sennrich–Haddow–Birch 2016). Ha a feldolgozandó szöveg hossza meghaladja a modell által kezelhető kontextusablak méretét, az a bemenet egy részének ignorálását eredményezheti, így előfordulhat, hogy a szövegösszefüggés elveszik, a válaszok logikailag szétesnek, az értelmezés pontatlanná vagy félrevezetővé válik.

[6] Ennek az az oka, hogy a modellek által használt tokenizáló algoritmusok nem mindig egyeznek meg az emberi szóhatárokkal. A toldalékokat külön szóként értelmezhetik, pl. „megszólítottalak” → akár 4 token: „meg” + „szólít” + „ott” + „alak”.

Promptstratégia

A GenAI egy rendkívül fejlett mintafelismerő rendszer, amely pontosan követi az utasításokat. Mivel ezek a modellek a korábbi szövégek mintázatai alapján számítják ki, hogy egy adott szó után melyik kifejezés következik a legnagyobb valószínűséggel, pontatlan vagy többértelmű utasítások esetén több lehetőséget kell mérlegelniük, ami növeli a hibás válasz esélyét. Ezért a hatékony kommunikáció kulcsa a precíz fogalmazás, amely csökkenti a találgatás szükségességét, és növeli az eredmény pontosságát.

Azt a szöveges bemenetet – legyen az kérés, felszólítás vagy kérdés –, amelyet az ember a modellnek ad egy válasz vagy művelet előállításához, *promptnak* nevezzük. Bár a *promptnak* jelenleg nincs szabványa, a szakmai gyakorlat kialakított egy strukturált megközelítést. Eszerint a hatékony *prompt* a következő képlettel írható le:

[Szerep] + [Cél/Feladat/Utasítás] + [Kontextus/Példák] + [Elvárt válaszformátum]

A *Szerep* meghatározza az AI személyiségét, nézőpontját, amit például így adhatunk meg: *Most jogász vagy; Most ötletgenerátor vagy; Most technikai asszisztens vagy*. Ha a GenAI-nak tudatosan meghatározott szerepet adunk, az nem csupán a felhasználói szándék pontosítását jelenti, hanem technológiai szinten is aktiválja a modell belső súlyozási mechanizmusait, előhívja a releváns tudásmintákat, és lehetővé teszi a kontextusérzékeny válaszgenerálást. A *prompt* képletének *Cél/Feladat/Utasítás* eleme által adjuk ki a konkrét, közvetlen parancsot arra, hogy az AI-modell milyen műveletet hajtson végre. Fontos, hogy legyen világos, célorientált, egyértelmű és konkrét. Formája felszólító vagy kérdő, egyszerűbb vagy összetettebb mondat lehet, például *Írj egy összefoglalót* vagy *Hasonlítsd össze a két szöveg érvrendszerét és értékeld kritikai szempontból*. A képlet *Kontextus/Példák* része háttér-információkat és mintákat nyújt a GenAI számára, ami segíti, hogy válaszát megfelelő stílusban, mélységben és tartalmi irányban generálja. A kontextus megadása elősegíti a releváns szavak és fogalmak kiemelését. A jól megválasztott példák tovább finomítják a kérés értelmezését, mivel a modell ezek alapján tanulja meg, milyen típusú válaszra számítunk. Különösen komplex feladatok esetén a példák megadása jelentősen növeli a válaszok pontosságát és konzisztenciáját. Az *Elvárt válaszformátum* meghatározásával igazíthatjuk a tartalmat a kívánt formára. Egyértelmű formátum a lista, a táblázat, a bekezdés, a *bekezdés* szó megadásával például az AI számára egyértelművé válik, hogy a választ folyó szövegként, összefüggő gondolatmenetben, olvasmányosan kérjük.

Fontos kerülni a felesleges részletekkel való túlterhelést, mivel az elvonhatja a figyelmet a lényegi kéréstől, és csökkentheti a válasz hatékonyságát. Az 1. ábra két példája segít megérteni, milyen típusú kérés vagy utasítás működik jól a GenAI számára.

Hatékony prompt	Túlterhelt prompt
Most egy banki kockázatkezelő vagy. Írj egy bekezdést arról, hogyan segíthet a generatív mesterséges intelligencia a pénzügyi csalások korai felismerésében, különös tekintettel a tranzakciós minták automatikus elemzésére és a gyanús viselkedések előrejelzésére.	*Légy szíves, írj egy bekezdést arról, hogy a mesterséges intelligencia hogyan tudna segíteni a csalások felderítésében. Egyre több olyan esetünk van, ahol a gyanúsítottak több banknál is számlát nyitnak, és nehéz nyomon követni a pénzmozgásokat. A kollégák szerint az MI segíthetne a minták felismerésében, de nem világos, hogy ez mennyire megbízható. A múlt hónapban például egy ügyben több ezer tranzakciót kellett manuálisan átnézni, ami rengeteg időt vett igénybe. Jó lenne, ha az MI nemcsak jelezné a gyanús mozgásokat, hanem javaslatot is tenne arra, hogy melyik ügyet érdemes előre venni. Szerinted ez lehetséges?*

1. ábra: Hatékony és túlterhelt prompt
Forrás: a szerző szerkesztése

Hallucináció

A generatív mesterséges intelligencia alkalmazása során az egyik legnagyobb aggodalomra okot adó jelenség a *hallucináció*, vagyis amikor a modell meggyőzően hangzó, de valótlan vagy félrevezető információt állít elő. Rendészeti alkalmazásokban a hallucináció kritikus kockázatmenedzsment-kérdés, mivel a téves információ súlyos jogi következményekkel járhat.

A hallucinációt a szakirodalom *intrinzik* és *extrinzik* típusokra osztja. A fogalom tudományos alapját többek között Maynez et al. (2020) fektette le az absztrakt szöveg-összefoglalás megbízhatóságát vizsgáló tanulmányában.

Az *intrinzik hallucináció* a forrásszöveggel ellentétes állításokat jelent, azaz a modell nemcsak kitalál valamit, hanem helytelenül értelmez vagy torzít. Ilyen például, ha egy GenAI-modell egy bírósági jegyzőkönyv alapján összefoglalást készít, és azt állítja, hogy „a vádlott beismerte a bűncselekményt", miközben a forrásszövegben az szerepel, hogy „a vádlott tagadta a vádat".

Az *extrinzik hallucináció* olyan információt tartalmaz, amely nem ellenőrizhető a forrásdokumentumból (Maynez et al. 2020), például amikor a modell külső adatokat *talál ki,* vagy nem létező hivatkozásokat generál.

A GenAI-fejlesztések célja a hallucinációk detektálása és csökkentése validációs mechanizmusok bevezetésével. Az egyik ígéretes megközelítés a *retrieval-augmented generation* (RAG), amely külső tudásbázisokhoz köti a modellt, ezzel csökkentve a ténybeli tévedések (faktahibák) előfordulását. A modell először lekérdez releváns információkat az adatbázisokból *(retrieval),* majd ezeket felhasználva generálja a választ. Aktív kutatási terület az olyan önellenőrző mechanizmusok *(self-checking mechanisms)* fejlesztése is,

amelyek képesek felismerni, ha a generált tartalom megbízhatósága kérdéses. Ilyen például a *SelfCheckGPT NLI* módszer, amely természetes nyelvi következtetésen alapuló konzisztencia-ellenőrzést alkalmaz (Huynh 2023), azaz megvizsgálja, hogy egy állítás ellentmond-e, következik-e, vagy független-e egy másik állítástól. A SelfCheckGPT NLI módszer több alternatív válasz összehasonlításával képes felismerni, ha a generált tartalom nem konzisztens, vagy nem megbízható (Manakul–Liusie–Gales 2023).

A hallucinációk napjainkban is e rendszerek egyik legkomolyabb kihívását jelentik. A fejlesztő vállalatok 2023 és 2025 között 12,8 milliárd dollárt fektettek kifejezetten a hallucinációs problémák megoldására (Ehtesham 2025). Ennek hatására a hallucinációs index az elmúlt öt évben jelentősen csökkent: az összes modell hallucinációs átlaga 35%-ról 8,2%-ra, míg a legjobb modelleké 15%-ról 0,7%-ra esett vissza. Ugyanakkor például jogi szövegek esetén a sajátos nyelvi és fogalmi komplexitás miatt még a legmegbízhatóbb modellek esetében is 6,4% és 18,7% közötti hallucinációs arányt mértek 2025 áprilisában.

Egyes iparági előrejelzések szerint 2030-ra a GenAI-programok az általános ismeretek terén elérhetik a 0,03%-os hallucinációs indexet, az egészségügy és jog területén pedig a 0,1%-ot (Rayner et al. 2016). Ezek a predikciók azonban óvatosan kezelendők, mivel a hallucinációs metrikák erősen feladatspecifikusak és doménfüggők.

GenAI a kutatásban: lehetőségek és kihívások

Az adatfeldolgozás gyorsítása

Kutatási szempontból a GenAI-rendszerek egyik legnagyobb előnye, hogy jelentős mértékben gyorsítják az adatfeldolgozást. A GPT-típusú modellek például másodpercenként több ezer token feldolgozására képesek, így egy 10 000 szavas tudományos cikk elemzése kevesebb mint egy percet vesz igénybe, messze meghaladva az emberi olvasási sebességet (200–300 szó/perc) (Rayner et al. 2016). A GenAI-rendszerek párhuzamos feldolgozási képessége révén több, eltérő forrásból származó szöveg egyidejű összevetésére is alkalmasak, ami különösen hasznos az irodalomkutatások és összehasonlító vizsgálatok során. Az irodalomkutatás terén kiemelkedő hatékonysággal azonosítják a releváns publikációkat, összefoglalják azok tartalmát, és tematikus struktúrába rendezik az információkat, ezáltal csökkentik a kutatók információs túlterhelését. A modellek fáradhatatlan működése biztosítja, hogy ez a folyamat nagy mennyiségű adat esetén is fenntartható és megbízható maradjon (OpenAI 2024).

A kutatási folyamatok újragondolása

Másrészt a GenAI-rendszerek alapjaiban alakítják át a kutatási folyamatok szerkezetét. Az információgyűjtés, irodalomfeldolgozás, adatértelmezés és hipotézisalkotás egymásra épülő szakaszaiból álló, hagyományosan lineáris kutatási modell egy dinamikus, iteratív és párhuzamosan zajló rendszerré szerveződik, amelyben a GenAI révén a kutatók már

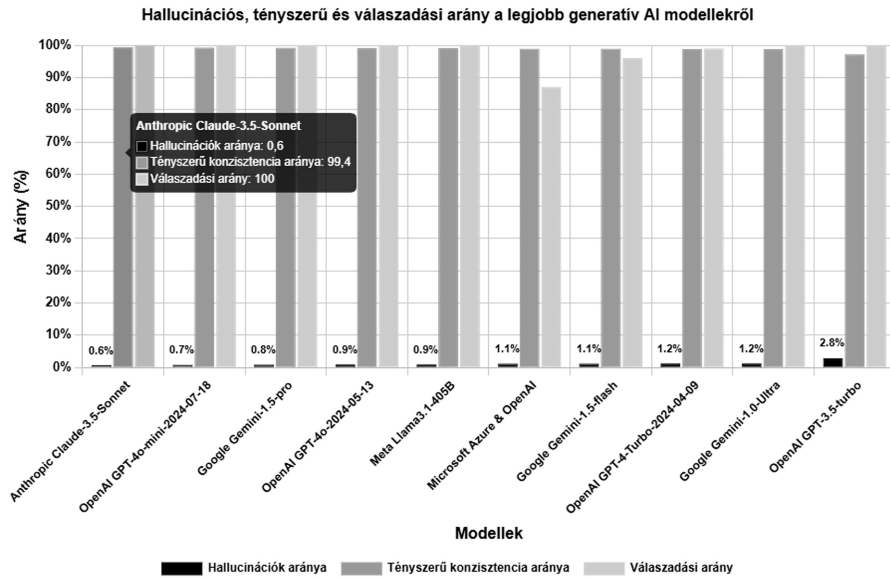
a kezdeti fázisban képesek előzetes elemzéseket végezni, gyorsan tesztelni hipotéziseket, majd az eredmények alapján azonnal módosítani a kutatási irányt, a módszereket, sőt akár a célokat is. Ez a fajta rugalmasság azt kívánja meg, hogy a tudományos gondolkodás alkalmazkodjon a technológia által nyitott új lehetőségekhez. A kutatói tevékenység kiegészül egyfajta kurátori feladatkörrel, a kutatónak szelektálnia kell az információk között, értelmeznie kell az AI által generált mintázatokat, döntéseket kell hoznia arról, hogy mely eredmények érdemelnek további vizsgálatot, emellett felelősséget kell vállalnia az etikai szempontokért, az adatforrások megbízhatóságáért és az értelmezés következményeiért is (Li–Wu 2025). A kutató az AI-rendszerek révén nemcsak technológiai eszközöket használ, hanem aktívan alakítja a tudományos diskurzust azáltal, hogy irányt szab az AI által feltárt lehetőségeknek. Mindezek alapján a GenAI integrálása a tudományos kutatásba technológiai előrelépésként értelmezhető.

Validálás és eredmények ellenőrzése

A GenAI-rendszerek rohamos fejlődése és széles körű alkalmazása szükségessé teszi mind a technológia önálló teljesítményének, mind pedig a vele végzett tudományos kutatások eredményeinek szigorú validálását. Mivel ezek a modellek gyakran nyílt végű, kreatív válaszokat generálnak, a hagyományos metrikák, például az egyszerű pontosság, nem elegendő mérőszámok. Az értékelés során kombinálni kell az automatikus, algoritmusokkal számított mutatókat[7] az emberi értékítéletet tükröző szubjektív szempontokkal.[8] A GenAI-technológia megbízhatóságának növekedését jól tükrözik a *Vectara hallucinációs ranglista* adatai (lásd 2. ábra), miszerint a legjobban teljesítő GenAI-rendszerek esetében a hallucinációs arány mára 2% alá csökkent, a tényhelyességi arány meghaladja a 98%-ot, míg a válaszadási arány eléri a 100%-ot (Vectara 2025). Ez azt jelenti, hogy ezek a modellek szinte teljes pontossággal képesek összefoglalni a forrásdokumentumokat, minden feladatra válaszolnak, és alig térnek el az eredeti tartalomtól.

[7] Például a BLEU *(bilingual evaluation understudy),* ami a gépi fordítások minőségét méri, vagy a generált szöveg és a referencia közötti jelentésbeli egyezést mérő BERTScore.

[8] Mint koherencia (a szöveg belső logikája), tényszerűség (igazságtartalom), ártalmatlanság (biztonságosság), hasznosság (gyakorlati érték).



2. ábra: A Vectara hallucinációs ranglistája a generatív AI-modellek megbízhatóságát három kulcsmutató alapján értékeli: hallucinációs arány, tényszerű konzisztencia és válaszadási arány
Forrás: a szerző szerkesztése a Vectara 2025 alapján

A generatív mesterséges intelligencia tudományos publikálásban betöltött szerepének egyik figyelemre méltó példája a norvég Iris startup fejlesztése. A vállalat olyan mesterséges intelligencián alapuló rendszert hozott létre, amelyet kifejezetten tudományos szövegek feldolgozására, tartalmi elemzésére és összefoglalására optimalizáltak. Az Iris célja, hogy a kutatók munkáját támogassa az információk gyorsabb áttekintésével, a releváns tartalmak kiemelésével és az összefüggések feltárásával, ezáltal hatékonyabbá téve az irodalomkutatást és a tudományos kommunikációt (Ramirez-Camara 2024).

Amikor a GenAI kutatási eszközként szerepel, az előállított kutatási eredmények validálása külön eljárást igényel. Ebben az esetben nem a GenAI teljesítményét, hanem az általa támogatott kutatási folyamat kimenetét kell ellenőrizni. A GenAI által adott kutatási eredmények megbízhatóságának értékelése több szempont alapján történik. Az például, hogy a különböző emberi értékelők/annotátorok mennyire egyformán értelmezik vagy kategorizálják ugyanazt a szöveget, a Cohen-féle kappa mutatóval mérhető (Cohen 1960). Ez különösen fontos olyan szubjektív értelmezést igénylő feladatok esetén, mint a szöveges tartalmak minősítése vagy címkézése. A szakirodalmi konzisztencia is fontos tényező, azaz annak mérése, hogy az újonnan kapott adatok mennyire állnak összhangban a korábbi kutatásokkal, illetve milyen mértékben támasztják alá vagy kérdőjelezik meg a meglévő elméleteket. A szakirodalmi konzisztencia nemcsak a tudományos hitelességet növeli, hanem segít az eredmények értelmezésében és kontextusba helyezésében is.

Human-in-the-loop és *human-over-the-loop* megközelítések a rendészeti GenAI-alkalmazásokban

A GenAI rendészeti integrációjának egyik legkritikusabb biztosítéka az emberi kontroll megfelelő szintű fenntartása, amellyel kapcsolatban a szakirodalom két alapvető megközelítést alkalmaz (Mosqueira-Rey et al. 2023):

1. Human-in-the-loop (HITL): Ez a modell azt jelenti, hogy az ember közvetlenül és aktívan részt vesz az AI-rendszer működésében, validálja, értelmezi, jóváhagyja, módosítja, kiegészíti vagy elutasítja a generált tartalmat. Rendészeti kontextusban ez azt jelenti, hogy például a GenAI által azonosított fenyegető kommunikációs mintázatot a rendőr vagy elemző szakértő minden esetben áttekinti, kontextualizálja és megerősíti, mielőtt az nyomozati döntés alapjává válna. A HITL-megközelítés különösen fontos olyan esetekben, ahol az AI-kimenet közvetlen intézkedéshez vagy jogi következményhez vezet.
2. Human-over-the-loop (HOTL): Ez a megközelítés az emberi felügyeletet rendszerszinten értelmezi. Az ember nem minden egyes kimenetet ellenőriz, hanem a rendszer egészének működését monitorozza, azonosítja a szisztematikus hibákat vagy torzításokat, és szükség esetén beavatkozik. Rendészeti környezetben ez jelentheti egy felügyeleti tisztviselő szerepét, aki rendszeres auditokat végez a GenAI-alapú elemzések minőségén és objektivitásán.

A két megközelítés kombinálása biztosítja, hogy a GenAI-rendszerek ne váljanak autonóm döntéshozókká olyan területeken, ahol az emberi felelősségvállalás és szakmai ítélőképesség elengedhetetlen. Ez összhangban áll az Európai Unió mesterséges intelligenciáról szóló rendeletének (AI Act) elvárásaival, amelyek a magas kockázatú alkalmazások esetében kötelezővé teszik az emberi felügyelet biztosítását (Európai Parlament és Tanács 2024).

Etikai és módszertani aggályok

A GenAI-programok népszerűsége ellenére a szöveggeneráló eszközökre való növekvő támaszkodás az oktatásban és a tudományos publikálásban komoly etikai és módszertani aggályokat vet fel. Ezek közé tartozik az akadémiai integritás sérülésének veszélye, a plágium kockázata, valamint a kritikus gondolkodás és az eredeti tudományos érvelés háttérbe szorulása. Az ilyen problémák kezelése érdekében egyre több tudományos intézmény (Pécsi Tudományegyetem Közgazdaságtudományi Kar 2025) és rangos folyóirat (Elsevier 2025) vezet be mesterséges intelligenciára vonatkozó irányelveket. Ezek az ajánlások egyrészt ösztönzik a generatív AI-eszközökben rejlő lehetőségek feltárását és felelősségteljes alkalmazását, másrészt szigorúan tiltják, hogy a kutatók tudományos szövegek tartalmi részét – különösen az eredmények értelmezését vagy tudományos állításokat – generatív AI segítségével írassák meg. Az irányelvek többsége az írási folyamatra vonatkozik, nem terjed ki az AI-eszközök kutatási célú használatára, például adatelemzésre vagy következtetések levonására, ahogyan azt az *Elsevier* is hangsúlyozza szabályzatában.

Az Európai Unió mesterséges intelligenciáról szóló rendelete (AI Act) 2024. augusztus 1-jén lépett hatályba, és megkülönbözteti az általános célú mesterségesintelligencia-modelleket (GPAI) és azok különböző kockázati szintjeit. A rendészeti kontextusban használt GenAI-rendszerek kockázati besorolása az alkalmazás céljától és jellegétől függ: míg a kutatástámogatási célú irodalomfeldolgozás alacsony kockázatúnak minősülhet, addig a prediktív rendészeti döntéstámogatás vagy a gyanúsítotti profilalkotás magas kockázatú kategóriába sorolható (Európai Parlament és Tanács 2024: 6. cikk).

A tudományos publikálás kontextusában kritikus kérdés a szerzőség. A jelenlegi nemzetközi konszenzus szerint a mesterséges intelligencia nem lehet sem szerző, sem társszerző, mivel nem rendelkezik jogi felelősséggel és tudományos elszámoltathatósággal (Nature Portfolio 2023). A GenAI-t segédeszközként lehet használni (például szövegszerkesztésre, fordításra), de ezt transzparensen dokumentálni kell a módszertani részben vagy külön AI-használati nyilatkozatban. Ennek ellenére a GenAI által létrehozott szövegek tudományosan koherens érvrendszert képviselhetnek, különösen akkor, ha nagy mennyiségű tudományos szövegkorpuszból tanulnak, ami további etikai dilemmákat vet fel a szellemi hozzájárulás definiálása terén.

Kiberbiztonsági és adatbiztonsági kockázatok

A GenAI rendészeti alkalmazása során figyelembe kell venni a technológia inherens biztonsági sebezhetőségeit, amelyek különösen fontosak a rendvédelmi szektorban.

A rosszindulatú *promptok* révén végrehajtható támadások lehetővé tehetik a támadók számára, hogy megkerüljék a modell biztonsági korlátait, vagy olyan válaszokat generáltassanak, amelyek veszélyeztetik az adatbiztonságot. Ezt a támadási formát a szakirodalom *prompt injectionnek* (parancsbeékelésnek) nevezi. Rendészeti környezetben ez azt jelentheti, hogy manipulált bemenetek révén a rendszer nyomozati információkat szivárogtathat ki, vagy félrevezető elemzéseket készíthet (Perez–Ribeiro 2022).

Az adatkinyerési kockázat különösen jelentős, mivel a nagy nyelvi modellek tanítóadatai bizonyos körülmények között rekonstruálhatók célzott lekérdezésekkel. A tanítóadatoknak a modellből való visszafejtését nevezik *modellinverziónak*. Ha a modellt érzékeny rendészeti dokumentumokon finomhangolták, az személyes adatok, nyomozati részletek vagy tanúvédelmi információk kiszivárgását eredményezheti (Carlini et al. 2021).

Hasonló problémát jelentenek a szándékosan perturbált, azaz apró mértékben módosított bemenetek, amelyeket a szakirodalom *adversarial* példáknak, azaz ellenséges mintáknak nevez. Ezek révén a modellek megtéveszthetők, így hibás kategorizálást vagy értékelést végezhetnek. Például rendészeti kontextusban egy gyanúsítotti kommunikáció fenyegető jellegét a rendszer nem ismeri fel apró szövegmódosítások esetén (Zou et al. 2023).

Az adatvédelmi megfelelés szempontjából különösen releváns a GDPR 22. cikke, amely értelmében az egyének jogosultak arra, hogy ne vonatkozzék rájuk kizárólag automatizált döntéshozatalon alapuló határozat. A GenAI rendészeti felhasználásában ezért elengedhetetlen olyan mechanizmusok alkalmazása, amelyekben az emberi döntéshozó aktívan részt vesz a folyamatban, illetve végső felügyeleti és jóváhagyási szerepet tölt be *(human-in-the-loop, human-over-the-loop)*.

Ezek a kockázatok indokolják, hogy a rendészeti GenAI-alkalmazások kifejezetten izolált, biztonságos környezetben, szigorú hozzáférés-ellenőrzéssel és folyamatos biztonsági auditálás mellett működjenek. A technológia előnyei csak akkor realizálhatók fenntartható módon, ha a biztonsági keretrendszer legalább olyan komoly figyelmet kap, mint maga a funkcionális fejlesztés.

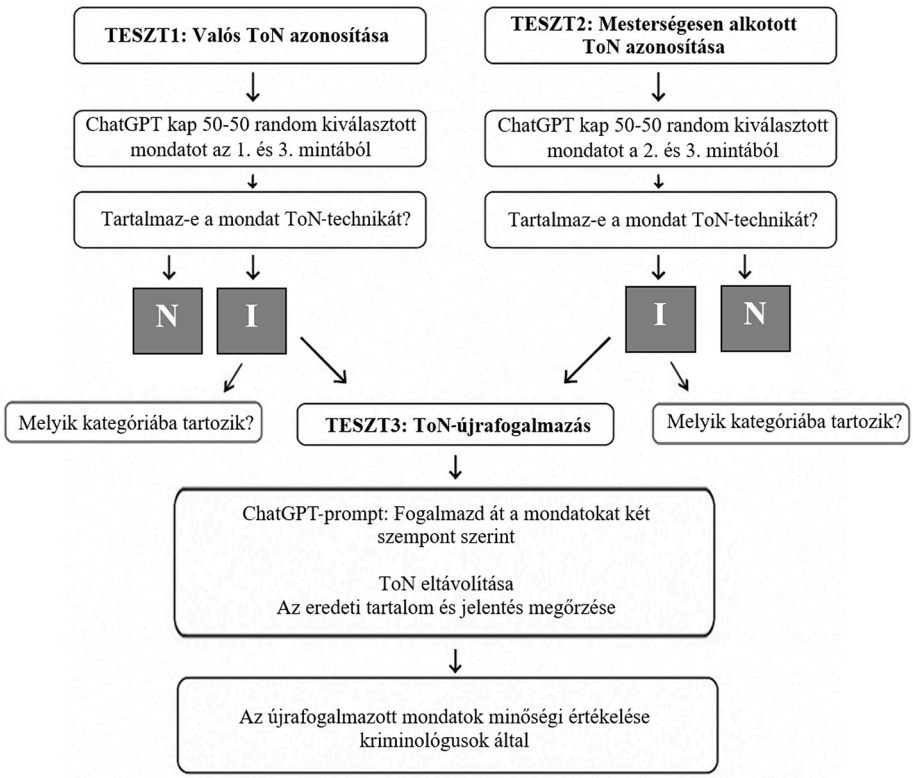
A GenAI gyakorlati alkalmazásai a rendészeti kutatásban

Gróf József, a technológiai világ aktualitásait közvetítő újságíró szerint „[a] tudományos kutatás a mesterséges intelligencia egyik legígéretesebb alkalmazási területévé, illetve egyúttal az egyik legmegosztóbbá is vált" (Gróf 2025). Ennek fényében a következőkben olyan nemzetközi szakirodalmi példákat mutatok be, amelyek szemléltetik, hogyan alkalmazzák a generatív mesterséges intelligenciát különböző kutatók és intézmények tudományos vizsgálataik során.

A normaszegés nyelvi igazolásainak azonosítása: semlegesítési technikák (ToN) felismerése és újrafogalmazása GenAI-eszközzel

A generatív AI és a kriminológia: fenyegetés vagy ígéret? A semlegesítési technikák (ToN) azonosításában rejlő lehetőségek és buktatók feltárása című, 2025 áprilisában olasz szerzők által készített tanulmány azt vizsgálja, hogyan alkalmazható a generatív mesterséges intelligencia egy klasszikus kriminológiai elmélet, a *semlegesítési technikák* (*techniques of neutralization,* ToN) azonosítására és értelmezésére (Pacchioni et al. 2025). Tipikus semlegesítési technika például a felelősség tagadása, amely lehetővé teszi az elkövető számára, hogy erkölcsileg igazoltnak érezze a viselkedését azzal, hogy a saját cselekedeteiért való felelősséget külső tényezőkre, például a társadalomra hárítja: nem én vagyok az, aki hibás, ha a társadalom nem hagyott volna magamra, sosem jutok idáig.

A kutatás célja az volt, hogy megvizsgálja, képes-e a GPT-4 felismerni és újrafogalmazni azokat a kognitív stratégiákat, amelyeket az elkövetők használnak viselkedésük igazolására. A kutatók az elkövetők lehetséges diskurzusainak szövegeit háromlépcsős promptstruktúrában vizsgálták: a *detektálási* szakaszban azt, hogy tartalmazott-e a mondat semlegesítési technikát, az *azonosítási* lépésben azt, hogy ha igen, melyik típust. A legösszetettebb *újrafogalmazási* lépésben a GenAI azt a feladatot kapta, hogy fogalmazza át úgy a mondatot, hogy a semlegesítési technika eltűnjön belőle, de a mondat tartalmi jelentése megmaradjon. Például, ha az elkövető eredeti mondata így hangzott: „Csak azért loptam, mert nem volt más választásom", a mondat a GenAI által újrafogalmazva: „Loptam". A semlegesítési technikák (ToN) GPT-4 általi feldolgozásának folyamatát a 3. ábra szemlélteti.



3. ábra: A semlegesítési technikák (ToN) GenAI általi feldolgozásának folyamata
Forrás: Pacchioni et al. 2025 tanulmányában szereplő vizuális modell saját fordítása és grafikai újraszerkesztése alapján készült

A kutatás rámutat, hogy a generatív mesterséges intelligencia magas pontossággal képes felismerni és újrafogalmazni a normaszegés nyelvi igazolásait, ami új utakat nyit a kriminalisztikai és társadalomtudományi elemzésekben, különösen az elkövetői narratívák megértésében, valamint a megelőzési és rehabilitációs stratégiák objektív vizsgálatában.

Bár az eredmények ígéretesek, a tanulmány korlátai is figyelmet érdemelnek. A kutatás nem tárja fel részletesen a modell esetleges kulturális torzításait: az olasz nyelvi kontextusban fejlesztett rendszer más nyelvi-kulturális közegben eltérő teljesítményt nyújthat. Kérdés az is, hogy a modell mennyire képes felismerni az implicit vagy kontextusfüggő semlegesítési technikákat, amelyek nem manifesztálódnak explicit nyelvi markerekként. A validációs folyamat is erősíthető lenne független, multidiszciplináris szakértői csoportok bevonásával, valamint nagyobb, heterogén mintán való teszteléssel. Etikai szempontból kritikus az is, hogy ilyen rendszereket csak kiegészítő eszközként, és sohasem önálló döntéshozatali alapként alkalmazzanak a büntetőeljárásban.

GenAI-alapú érvelésfeltárás gyermekelhelyezéssel kapcsolatos bírósági ítéletekben

A gyermekelhelyezéssel kapcsolatos bírósági ítéletek nyelvi mintázatai nemcsak jogi, hanem társadalmi szempontból is kiemelkedően fontosak, hiszen közvetlen hatással vannak a családok életére, a gyermekek jólétére és a jogalkalmazás következetességére. Muñoz-Soro és munkatársai a spanyol Zaragozai Egyetemen 2024-ben vizsgálták a mesterséges intelligencia jogi szövegeken való alkalmazhatóságát. *A neurális hálózat alkalmazása a gyermekelhelyezéssel kapcsolatos bírósági ítéletekben szereplő kérelmek, döntések és érvek azonosítására* című tanulmányukban bemutatják, hogyan dolgoztak fel egy saját fejlesztésű, transzformátoralapú mesterségesintelligencia-modellel több mint 3000 ítéletet. Modelljük célja az volt, hogy automatikusan azonosítsa a szülői kérelmek típusát (egyéni vagy közös elhelyezés), a bírósági döntéseket, valamint az ezeket alátámasztó érvelési mintákat, például a gyermek véleményére, a szülők anyagi helyzetére vagy az együttműködési készségre vonatkozó hivatkozásokat. A kutatás egyik lényeges eleme, hogy a mesterséges intelligencia teljesítményét összevetették az emberi annotátorok közötti egyetértéssel a kappa index (k) segítségével. Az emberi annotátorok közötti egyezés kappa értékei 0,33 és 0,73 között mozogtak, míg az AI-modell és az annotátorok közötti egyezés ennél magasabb volt: 0,57 és 0,80 között mozgott. A legjobb eredményeket a modell a gyermek véleményének (k = 0,8005), valamint a szülők rendelkezésére álló idő és anyagi eszközök (k = 0,8063) kategóriákban érte el. Ez azt jelzi, hogy a mesterséges intelligencia nemcsak gyorsan és hatékonyan képes feldolgozni komplex jogi szövegeket, hanem bizonyos esetekben még az emberi szövegértelmezés pontosságát is meghaladja.

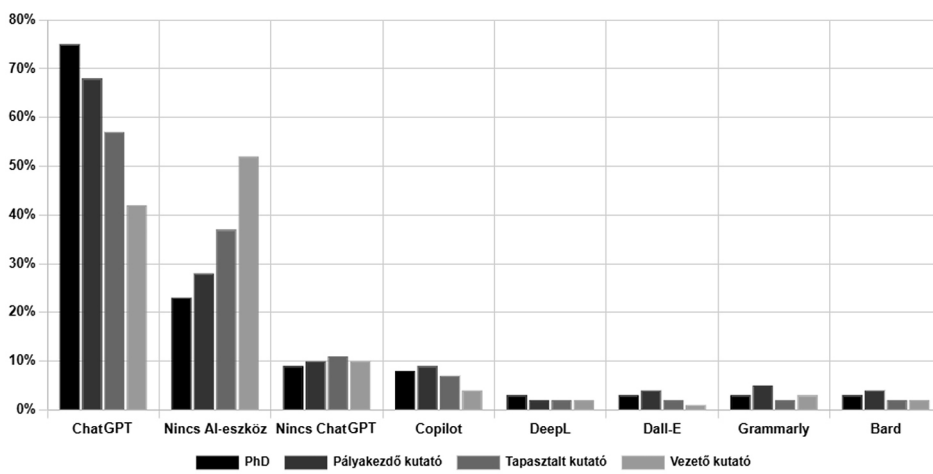
A tanulmány rámutat arra is, hogy Spanyolországban 2021-ben több mint 86 000 válás történt, és ezek 21,2%-a került bíróság elé, így a gyermekelhelyezési ügyek automatizált feldolgozása nem csupán technológiai, hanem társadalmi szinten is releváns. Ugyanakkor a kutatói álláspontok szerint a generatív mesterséges intelligencia jogi alkalmazása csak akkor tekinthető megbízhatónak, ha szigorú szakmai felügyelet mellett történik. Ez különösen igaz az olyan érzékeny területekre, mint a családjog, ahol az emberi ítélőképesség, az empátia és az egyedi élethelyzetek figyelembevétele nélkülözhetetlen (Muñoz-Soro et al. 2024).

A vizsgálat módszertani erénye a kappaindex-alapú validáció, de kritikaként fogalmazható meg az, hogy a tanulmány nem tér ki a *false positive* és *false negative* arányok bemutatására. Aggályos továbbá, hogy a rendszer tanítóadatai potenciálisan tartalmazhatják a bírói gyakorlatban meglévő implicit torzításokat (például nemi vagy szocioökonómiai alapú előítéleteket), amelyeket a modell reprodukálhat vagy erősíthet. A tanulmány nem vizsgálja, hogy az AI által azonosított érvelési minták megfelelnek-e az aktuális jogi standardoknak, vagy esetleg elavult, megkérdőjelezhető gyakorlatokat tükröznek. Kritikus kérdés továbbá az átláthatóság, a döntéshozók számára világosnak kell lennie, hogy az AI milyen logika alapján kategorizál, különben sérti a jogi procedurális követelményeket.

A GenAI belépése a kutatói műhelyekbe

A *Generatív mesterséges intelligencia (GenAI) a kutatási folyamatban – a kutatók gyakorlatának és felfogásának felmérése* című, dán intézmény által vezetett, nemzetközi mintán alapuló kutatás több mint 2500 egyetemi kutató bevonásával azt vizsgálta, hogyan alkalmazzák a generatív mesterséges intelligenciát a tudományos kutatásokban és munkafolyamatokban, különös tekintettel a szerepválasztásra, valamint az ezekhez kapcsolódó etikai megítélésre. A kérdőíves felmérés 32 különböző GenAI-használati esetet értékelt öt fő kutatási szakaszban: ötletgenerálás, kutatástervezés, adatgyűjtés, elemzés és publikálás. A válaszadók háromféle szerepben alkalmazták a mesterséges intelligenciát: mint megbízható háttérmunkást (munkalovat), mint nyelvi támogatót (asszisztenst), illetve mint a kutatási folyamatokat felgyorsító eszközt (gyorsítót).

A válaszadók többsége egyetértett abban, hogy a GenAI megfelelő alkalmazása jelentősen gyorsítja a kutatási folyamatokat, különösen a publikációs szakaszban, és segíti a nyelvi akadályok leküzdését, ami különösen fontos a nem angol anyanyelvű kutatók számára. A demográfiai adatok alapján a fiatalabb kutatók nyitottabban és aktívabban használják a GenAI-t, míg a nemek között nem mutatkozott jelentős különbség. A GenAI-eszközöket alkalmazó kutatók életkor szerinti százalékos megoszlása a 4. ábrán tekinthető meg.



4. ábra: A GenAI-eszközöket használó dán kutatók százalékos aránya kutatói életkor szerint
Forrás: a szerző szerkesztése Andersen et al. 2025 alapján

A dán tanulmány hangsúlyozza, hogy a GenAI használata nem mentes az etikai dilemmáktól, és különösen fontos a tudatos szabályozás: a technológia nem helyettesítheti az emberi gondolkodást, hanem annak kiegészítőjeként kell működnie (Andersen et al. 2025).

Társadalmi előítéletek öröklődése a GenAI-rendszerekben: az algoritmikus torzítás és a COMPAS-rendszer tanulságai

Az algoritmikus torzítás *(algorithmic bias)* olyan szisztematikus, visszatérő hiba, amely a mesterségesintelligencia-rendszerek működése során bizonyos csoportokkal szemben aránytalanul hátrányos vagy kedvező eredményeket produkál. Ez a jelenség általában három forrásból ered: 1. torzított vagy nem reprezentatív tanítóadatokból, 2. a modellépítés során alkalmazott implicit feltételezésekből, valamint 3. az alkalmazási kontextus strukturális egyenlőtlenségeiből (Barocas–Selbst 2016).

Az amerikai igazságszolgáltatásban alkalmazott COMPAS- *(Correctional Offender Management Profiling for Alternative Sanctions)* rendszer az algoritmikus torzítás emblematikus esete. A *ProPublica* 2016-os vizsgálata szerint a rendszer szisztematikusan 45%-kal nagyobb valószínűséggel sorolta a fekete vádlottakat magasabb kockázatú kategóriába, mint a fehéreket, azonos bűncselekmények esetén is, a fehér vádlottakat pedig 77%-kal nagyobb valószínűséggel minősítette alacsonyabb kockázatúnak (Angwin et al. 2016). Ez a torzítás nem a szoftver szándékos diszkriminációjából, hanem a történelmi adatokban megjelenő strukturális rasszizmus modell általi reprodukálásából fakadt, az afroamerikai közösségeket aránytalanul érintő rendőri igazoltatások és letartóztatások miatt ugyanis a tanítóadatok már eleve torzítottak voltak. A COMPAS-eset alapjaiban kérdőjelezte meg azt az elképzelést, hogy az algoritmusok képesek teljesen semlegesen működni. Rámutatott arra, hogy a matematikai modellek nem mentesek az értékítéletektől, hiszen a pontosság önmagában nem garantálja az igazságosságot – egy modell lehet nagyon pontos, miközben mégis torz eredményeket produkál. Emellett világossá vált, hogy a transzparencia hiánya komoly akadályt jelent az elszámoltathatóság szempontjából, mivel a COMPAS-rendszer zárt forráskódja miatt nem volt lehetőség független szakértői ellenőrzésre.

A generatív modellek esetében a torzítás még komplexebb, mivel nemcsak döntést hoznak, hanem narratívákat, értelmezéseket és szövegeket generálnak (Bender et al. 2021). Rendészeti kontextusban ez azt jelentheti, hogy egy GenAI-rendszer például egy kihallgatási jegyzőkönyv összefoglalása során árnyalatokban, implicit módon bizonyos etnikai vagy társadalmi csoportokkal szemben negatív nyelvi keretezést alkalmazhat, ami befolyásolhatja az ügyben eljárók ítéletalkotását (Barocas–Selbst 2016). A probléma kezelése megköveteli a tanítóadatok folyamatos *bias* auditálását, a *fairness* metrikák[9] beépítését a modellértékelésbe (Mehrabi et al. 2021), a rendészeti alkalmazásokban a kötelező emberi felügyelet biztosítását, valamint a transzparens dokumentációt a modell döntéslogikájáról és korlátairól (Raji et al. 2020).

[9] A *fairness* olyan technikai kritérium, amely azt vizsgálja, hogy a modell különböző csoportokat (pl. nem, etnikum, életkor) azonos feltételekkel kezel-e, és nem hoz-e rendszerszintű hátrányos döntéseket bizonyos csoportokra nézve.

Konklúzió

A generatív mesterséges intelligencia tudományos célú alkalmazása egy olyan szemléletváltás kezdetét jelzi, amely alapjaiban alakítja át a kutatói munka módszertani kereteit és lehetőségeit. Ugyanakkor használata nem tekinthető magától értetődőnek: az ilyen rendszerek mögött komplex és a felhasználók számára gyakran nehezen átlátható technológiai struktúra húzódik meg. Eredményes alkalmazásuk tudatos kompetenciaépítést, módszertani felkészültséget és kritikus gondolkodást kíván. Csak ezek megléte mellett válhatnak olyan valódi erőforrásokká, amelyek képesek fokozni a kutatás hatékonyságát, mélységét és innovációs potenciálját.

A tanulmányban ismertetett nemzetközi esettanulmányok világosan rámutatnak arra, hogy a GenAI-eszközök már most kézzelfogható előnyöket kínálnak a tudományos gyakorlatban. A technológia fejlettsége lehetővé teszi célzott és eredményes alkalmazásukat, amit jól tükröz a hallucinációs ráták jelentős csökkenése. Ez a megbízhatóság és az alkalmazhatóság szempontjából számottevő előrelépés. A rendészeti kutatások területén különösen ígéretesnek mutatkozik a GenAI integrációja, feltéve, hogy azt szigorú szakmai kontroll, átlátható dokumentációs eljárások és tudatos prompttervezés kíséri. A kontextusablak korlátainak figyelembevétele, valamint a többszintű validálási mechanizmusok alkalmazása elengedhetetlen a hiteles és reprodukálható tudományos eredmények biztosításához.

A jövőbeli kutatások irányai lehetnek a magyar nyelvű rendészeti dokumentumok GenAI-alapú feldolgozásának pilotvizsgálata, a rendészeti szakemberek GenAI-felkészültségének és attitűdjének felmérése kérdőíves vizsgálattal, etikai keretrendszer kidolgozása a GenAI rendészeti alkalmazására magyar jogszabályi környezetben és longitudinális vizsgálatok a GenAI rendészeti adaptációjának hatékonyságáról és társadalmi elfogadottságáról.

A jövőben várhatóan még szorosabb kapcsolat alakul ki a GenAI-rendszerek és a kutatási folyamatok között. A technológia fejlődési pályája azt vetíti előre, hogy 2030-ra a hallucinációs hibák jelentős mértékben csökkenthetők lesznek, ami új távlatokat nyit a széles körű tudományos alkalmazás előtt. Ugyanakkor kritikusan kell értékelni az ilyen prediktív állításokat, mivel a hallucinációs metrikák erősen feladatspecifikusak, és a rendészeti terület komplexitása továbbra is jelentős kihívásokat fog jelenteni.

Összegzésként elmondható, hogy a GenAI nem univerzális megoldás, de nem is elutasítandó technológia. Sokkal inkább olyan eszköz, amely megfelelő szakmai háttérrel, etikai irányelvek mentén és világos módszertani keretek között alkalmazva képes érdemben hozzájárulni a tudományos megismerés elmélyítéséhez és a kutatási folyamatok hatékonyabbá tételéhez. Eredményes integrációja feltételezi a kutatói közösség nyitottságát a tanulásra, az interdiszciplináris együttműködések erősítését, valamint a felelős innovációs szemlélet következetes érvényesítését.

Felhasznált irodalom

Andersen, Jens Peter et al. (2025): *Generative AI in the Research Process – A Survey of Researchers' Practices and Perceptions.* SocArXiv. Online: https://doi.org/10.31235/osf.io/83whe

Angwin, Julia et al. (2016): Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. *ProPublica,* 2016. május 23. Online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Artificial Analysis (2025): *Independent Evaluations of AI Models and Providers.* Online: https://artificialanalysis.ai

Barocas, Solon – Selbst, Andrew D. (2016): Big Data's Disparate Impact. *California Law Review,* 104, 671–732. Online: https://doi.org/10.2139/ssrn.2477899

Bender, Emily M. et al. (2021): On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* 610–623. Online: https://doi.org/10.1145/3442188.3445922

Bodnár András Péter (2022): A digitális bizonyítékok megjelenése a büntetőeljárásban – különös tekintettel a szakértő igénybevételére. *Büntetőjogi Szemle,* (1), 19–29. Online: https://szakcikkadatbazis.hu/doc/6289137

Carlini, Nicholas et al. (2021): Extracting Training Data from Large Language Models. *USENIX Security Symposium.* Online: https://arxiv.org/abs/2012.07805

Cohen, Jacob (1960): A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement,* 20(1), 37–46. Online: https://doi.org/10.1177/001316446002000104

Common Crawl Foundation [é. n.]: *Common Crawl*. Online: https://commoncrawl.org

Elsevier (2025): *Generative AI Policies for Journals*. Online: https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals

Ehtesham, Hira (2025): AI Hallucination Report 2026: Which AI Hallucinates the Most? *All About AI,* 2025. december 4. Online: https://www.allaboutai.com/resources/ai-statistics/ai-hallucinations/#low-hallucination-group

Európai Parlament és Tanács (2024): *Az Európai Parlament és a Tanács (EU) 2024/1689 rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint egyes korábbi jogszabályok módosításáról*. Hivatalos Lap, L 2024/1689. Online: https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX:32024R1689

Gróf József (2025): Hallucinációk az MI-kutatásban. *ITBusiness,* 2024. június 4. Online: https://itbusiness.hu/technology/gadget/hallucinaciok-az-mi-kutatasban

Huynh, Daniel (2023): Automatic Hallucination Detection with SelfCheckGPT NLI. *Hugging Face,* 2023. november 27. Online: https://huggingface.co/blog/dhuynh95/automatic-hallucination-detection

Li, Hui – Wu, Xiaofeng (2025): The Use of Generative AI Tools in Academic Writing: A Systematic Review of Reseach Trends and Thematic Insights. *AI and Ethics,* 5, 5821–5840. Online: https://doi.org/10.1007/s43681-025-00827-0

Manakul, Potsawee – Liusie, Adian – Gales, Mark J. F. (2023): SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models. In Bouamor, Houda – Pino, Juan – Bali, Kalika (szerk.): *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing.* Singapore: Association for Computational Linguistics, 9004–9017. Online: https://doi.org/10.18653/v1/2023.emnlp-main.557

Maynez, Joshua et al. (2020): On Faithfulness and Factuality in Abstractive Summarization. In Jurafsky, Dan et al. (szerk.): *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,* Association for Computational Linguistics, 1906–1919. Online: https://doi.org/10.18653/v1/2020.acl-main.173

Mehrabi, Ninareh et al. (2021): A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys,* 54(6), 1–35. Online: https://doi.org/10.1145/3457607

Mosqueira-Rey, Eduardo et al. (2023): Human-In-The-Loop Machine Learning: A State of the Art. *Artificial Intelligence Review,* 56(4), 3005–3054. Online: https://doi.org/10.1007/s10462-022-10246-w

Muñoz-Soro, José Félix et al. (2024): A Neural Network to Identify Requests, Decisions, and Arguments in Court Rulings on Custody. *Artificial Intelligence and Law,* 33, 101–135. Online: https://doi.org/10.1007/s10506-023-09380-9

Nature Portfolio (2023): Tools Such as ChatGPT Threaten Transparent Science; Here Are Our Ground Rules for Their Use. *Nature,* 613, 612. Online: https://doi.org/10.1038/d41586-023-00191-1

Nikolakopoulos, Anastasios et al. (2024): Large Language Models in Modern Forensic Investigations: Harnessing the Power of Generative Artificial Intelligence in Crime Resolution and Suspect Identification. In *2024 5th International Conference on Computer Science and Technologies in Education (CSTE).* 1–5. Online: https://ieeexplore.ieee.org/abstract/document/10654427

OpenAI (2019): *WebText Dataset*. Online: https://github.com/openai/gpt-2

OpenAI (2024): *GPT-4 Technical Report: Performance Metrics and Computational Specifications*. OpenAI Technical Publications. Online: https://arxiv.org/abs/2303.08774

Pacchioni, Federico et al. (2025): Generative AI and Criminology: Threat or Promise? Exploring the Potential and Pitfalls of Identifying Techniques of Neutralization (Ton). PLOS ONE, 20(4), e0319793. Online: https://doi.org/10.1371/journal.pone.0319793

Perez, Fábio – Ribeiro, Ian (2022): Ignore Previous Prompt: Attack Techniques for Language Models. *arXiv preprint arXiv:2211.09527.* Online: https://arxiv.org/abs/2211.09527

Pécsi Tudományegyetem Közgazdaságtudományi Kar (2025): *A mesterséges intelligencia használatára vonatkozó irányelvek*. Online: https://ktk.pte.hu/sites/ktk.pte.hu/files/uploads/szabalyzatok/mi-iranyelvek_pte-ktk.pdf

Press, Ofir – Smith, Noah A. – Lewis, Mike (2022): Train Short, Test Long: Attention With Linear Biases Enables Input Length Extrapolation. *International Conference on Learning Representations (ICLR).* Online: https://doi.org/10.48550/arXiv.2108.12409

Radford, Alec et al. (2018): *Improving Language Understanding by Generative Pre-Training.* OpenAI. Online: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Raffel, Colin et al. (2020): Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research,* 21(140), 1–67. Online: https://arxiv.org/abs/1910.10683

Raji, Inioluwa Deborah et al. (2020): Closing the AI Accountability Gap: Defining an End-To-End Framework for Internal Algorithmic Auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,* 33–44. Online: https://doi.org/10.1145/3351095.3372873

Rayner, Keith et al. (2016): So Much to Read, so Little Time: How Do We Read, and Can Speed Reading Help? *Psychological Science in the Public Interest,* 17(1), 4–34. Online: https://doi.org/10.1177/1529100615623267

Ramirez-Camara, Ellie (2024): Iris.ai Secured €7.64M to Optimize its AI Engine for Scientific Text Understanding. *Dataphoenix,* 2024. június 3. Online: https://dataphoenix.info/iris-ai-secured-eu7-64m-to-optimize-its-ai-engine-for-scientific-text-understanding/

Sennrich, Rico – Haddow, Barry – Birch, Alexandra (2016): Neural Machine Translation of Rare Words With Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics 1*. Berlin: Association for Computational Linguistics, 1715–1725. Online: https://doi.org/10.18653/v1/P16-1162

Vaswani, Ashish et al. (2017): Attention is All You Need. *Advances in Neural Information Processing Systems,* 30. Online: https://research.google/pubs/attention-is-all-you-need/

Vectara (2025): *Hallucination Leaderboard: Comparing LLM Performance at Producing Hallucinations when Summarizing Short Documents*. GitHub. https://github.com/vectara/hallucination-leaderboard

Zou, Andy et al. (2023): Universal and Transferable Adversarial Attacks on Aligned Language Models. *arXiv preprint arXiv:2307.15043*. Online: https://arxiv.org/abs/2307.15043