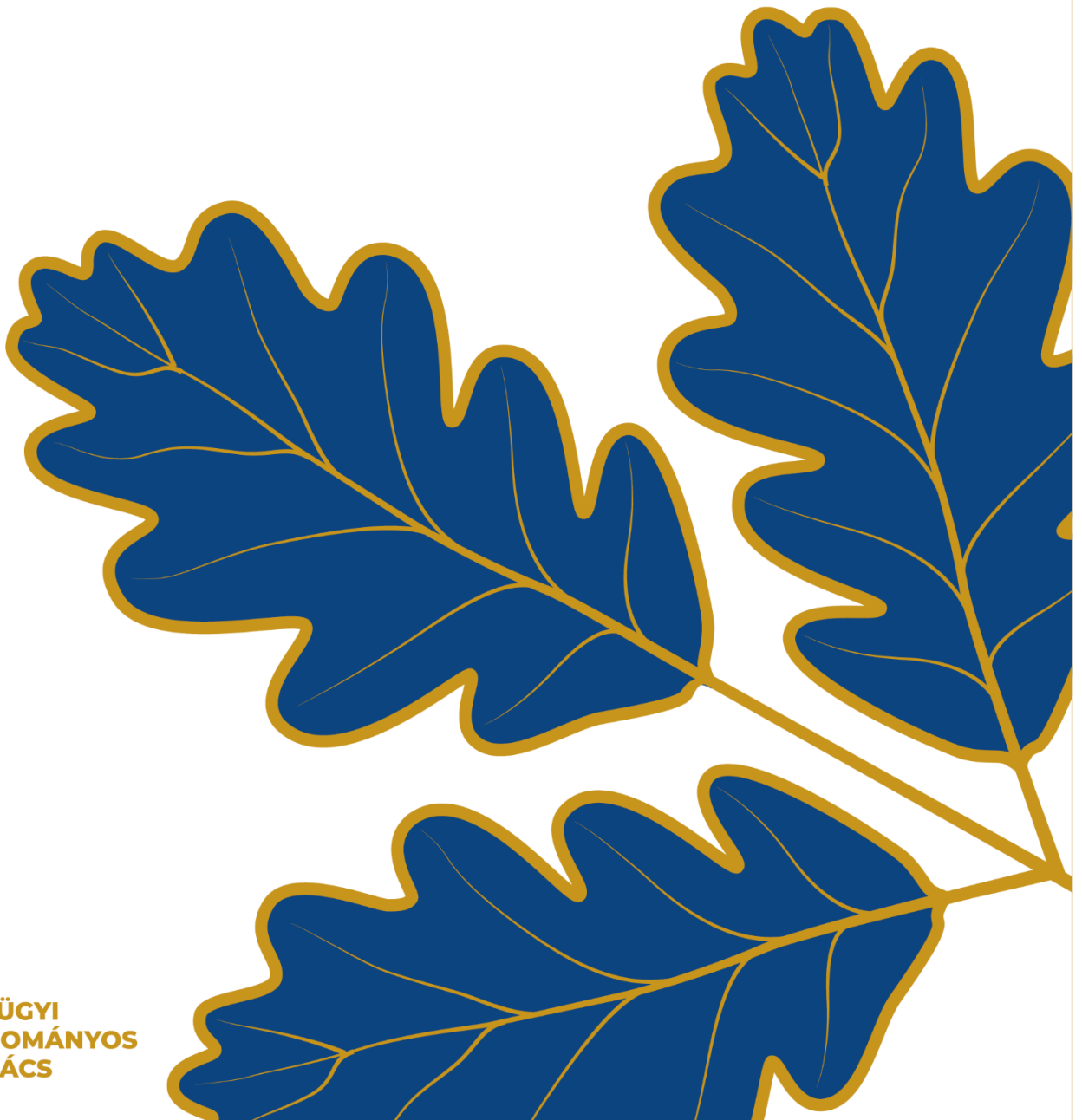




XIII. évfolyam 2024/2. szám



RENDVÉDELEM





B U D A P E S T

– 2024 –

A BELÜGYI TUDOMÁNYOS TANÁCS

ONLINE FOLYÓIRATA

IMPRESSZUM

KIADJA:

Belügyminisztérium

FELELŐS KIADÓ:

Dr. Felkai László közigazgatási államtitkár, Belügyi Tudományos Tanács, elnök

FELELŐS SZERKESZTŐ:

Dr. Sabjanics István ny. r. dandártábornok, Belügyi Tudományos Tanács, ügyvezető alelnök

KIADÓ:

Cím: 2090 Remeteszőlős, Nagykovácsi út 3.

Postacím: 1903 Budapest, Pf. 314.

E-mail: titkar@bm-tt.hu

Telefon: +36 26 795 911, BM: 24-592

Fax: +36 26 795 919, BM: 24-650

HU ISSN 2560-2349

Tartalom

Szerkesztői előszó	1
Nagy Péter: ChatGPT és más nagy nyelvi modellek (LLM-ek) biztonsági kérdései, szervezeti és társadalmi hatásai	2
Baranyai András: ChatGPT és más nagy nyelvi modellek alkalmazásának lehetőségei és kiberbiztonsági kérdései	16
Karsa Róbert: Tűzvédelmi szakértői rendszer létrehozása nagy nyelvi modellek segítségével	26
Albert Ágota Katalin: Jogászok és informatikusok kommunikációs problémáinak eliminációs eszközei a mesterséges intelligencia alkalmazása során	37
Illés Horváth: The newest aspect of industrial espionage. Cyber threats from UAVs to non-state actors.....	71
Balogh Mónika: Az üzleti pszichológia és coaching adaptálhatósága a rendvédelmi szervek mindennapi működésébe egy empirikus vizsgálat tükrében	79
Vereckeai Judit: Az igazgatók fejlesztésének lehetőségei és irányai a magyar köznevelés rendszerében - kutatási beszámoló.....	88
Szerzők	96

Szerkesztői előszó

A Rendvédelem folyóirat XIII. évfolyam 2024./1-2. számában dr. Felkai László, a Belügyminisztérium közigazgatási államtitkára, a Belügyi Tudományos Tanács elnöke által 2023 áprilisában kiírt A ChatGPT és más nagy nyelvi modellek alkalmazásának lehetőségei és egyben biztonsági kérdései című pályázati felhívásra érkezett egyes pályaművek átdolgozott tanulmányai olvashatóak.

A [XIII. évfolyam 1. számának szerkesztői előszavában](#) bemutattuk a fent említett pályázat kiírásának indokoltságát, háttérinformációit, aktualitását. A pályázat meghirdetésének célja az volt, hogy a Belügyi Tudományos Tanács az állomány tagjainak szakmai ismereteit és értékes gyakorlati tapasztalatait, valamint a téma iránt érdeklődő szakemberek véleményét és javaslatait megismerje. Olyan pályaműveket vártunk, amelyek elsősorban az életszerű, valós problémákra megoldást kínáló ötletek rendszerszemléletű megközelítéssel történő megoldásait mutatják be, és nem műszaki és matematikai megközelítésű tartalmat tükröznek. A pályázatok feltöltési határideje 2023. október 1. 24:00 óra volt, eddig az időpontig összesen 26 érvényes jeligés pályamű érkezett a Belügyi Tudományos Tanács online pályázati felületére. A beérkezett pályaművek bírálati folyamata (elbírálási határidő: 2023. november 15.) két részből állt: az első részben az online pályázati felületen a téma tudományos és szakmai elismertségű szakértői bírálóként pontozták és szövegesen értékelték a benyújtott jeligés, anonim pályázatokat, a második részben a felkért bizottság a bírálók értékelése alapján döntött a díjazásról.

Jelen XIII. évfolyam 2024./2. számot szorosán az [1. szám](#) után tesszük közzé, melyben szintén a pályázati felhívásra benyújtott pályaművek közül a szerkesztőség döntése alapján kiválasztott, a szerzők által átdolgozott tanulmányok jelennek meg. A szerzők különböző tudományterületek nézőpontjaiból vizsgálják a ChatGPT és más nagy nyelvi modellek témakörét. Szerzőink: Nagy Péter, Baranyai András, Karsa Róbert és Albert Ágota Katalin. A publikációk szakmai tartalmáért, naprakész információiért a szerkesztőség felelősséget nem tud vállalni. Jelen szám utolsó három cikke nem tartozik a pályaművek közé: Horváth Illés az ipari kémkedés legújabb aspektusát mutatja be angol nyelvű tanulmányában, amelynek megjelentetését aktualitása miatt tartjuk fontosnak; Balogh Mónika empirikus kutatásában az üzleti pszichológia és a coaching adaptálhatóságát vizsgálta a

rendvédelmi szervek mindennapi működésébe; Vereckei Judit – a Belügyminisztérium közigazgatási államtitkára, a Belügyi Tudományos Tanács elnöke által, a Belügyminisztériumot érintő közszolgálati témában doktorandusz, abszolutóriumot szerzett és doktorjelölt kutatók számára meghirdetett – gyakornoki program egyik résztvevője, aki 2023. év szeptemberétől 2024. év februárjáig tartó időszakra vonatkozó gyakornoki kutatási tevékenységének rövid összegzését ismerteti.

A változatos témájú publikációkhoz hasznos olvasást kívánunk!

Dr. Sabjanics István ny. r. dandártábornok
felelős szerkesztő

ChatGPT és más nagy nyelvi modellek (LLM-ek) biztonsági kérdései, szervezeti és társadalmi hatásai

Security issues, organizational and social impacts of ChatGPT and other large language models (LLMs)

DOI: [HTTPS:// DOI.ORG/10.53793/RV.2024.2.1](https://doi.org/10.53793/RV.2024.2.1)

Absztrakt

A mesterséges intelligencia terjedése új dimenziókat nyit az emberiség előtt számos szektorban. A nagy nyelvi modellek (LLM), mint például a ChatGPT, kiemelkedő potenciállal rendelkeznek a nyelvi adatok kezelésében és értelmezésében, valamint adaptálható eszközként szolgálnak számos alkalmazásban. A nagy nyelvi modellek által képviselt technológiai fejlődés egyre inkább központi szerepet játszanak a társadalmi és gazdasági életben, így fontos, hogy a tudományos közösség és a szakmai szereplők felkészülten és tudatosan kezeljék ezeket az új lehetőségeket és kihívásokat. Jelen kutatásban az oktatás és a kiberbiztonság terén felmerülő kihívásokra válaszolva számos kutatási javaslat került megfogalmazásra az LLM-ek integrációjára és a biztonsági kockázatok kezelésére vonatkozóan.

KULCSSZAVAK: CHATGPT, LLM, MESTERSÉGES INTELLIGENCIA, OKTATÁS, KIBERBIZTONSÁG

Abstract

The spread of artificial intelligence opens up new dimensions for humanity in many sectors. Large language models (LLMs), such as ChatGPT, have outstanding potential for handling and interpreting linguistic data and serve as adaptable tools for many applications. The technological development represented by large language models is increasingly playing a central role in social and economic life, so it is important that the scientific community and professional players handle these new opportunities and challenges in a prepared and conscious manner. In this research, in response to the challenges arising in the field of education and cyber security, a number of research proposals were formulated regarding the integration of LLMs and the management of security risks.

KEYWORDS: CHATGPT, LLM, ARTIFICIAL INTELLIGENCE, EDUCATION, CYBER SECURITY

Bevezetés

Az információs technológia és a mesterséges intelligencia (MI) gyors terjedése hatalmas távlatokat nyit az emberiség előtt a mindennapi élet különböző területein (Khalil–Er 2023). Atlas (2023) szerint ezek a mesterséges intelligencia rendszerek páratlanul összetett és rendkívüli adatfeldolgozási képességekkel rendelkeznek, amelyek lehetővé teszik számukra, hogy emberhez hasonló kommunikációt és tartalmakat hozzanak létre, eddig soha nem látott szinten. Teubner munkatársaival közösen írt (2023) tanulmánya alapján a ChatGPT és a hasonló nyelvi modellek példátlan hatékonyságot értek el a szövegenerálásban, a nyelvek fordításában, a kérdések és válaszok generálásában, valamint a hatalmas információhalmazokból történő adatnyerésben. Jelenleg az LLM-ek integrációja több területen problémával küzd, amiből kettőt emelnék ki

munkám tükrében. Első, az oktatásba történő integráció. Az egyetemi oktatási intézményeknek egyre nehezebb feladatot jelent az egyensúly megtalálása a hagyományos tanítási módszerek és a nyelvi modellek használata között (Hobert 2019). A második terület az adatbiztonságot, kiberbiztonságot érinti. Az LLM-ek használata számos ezekhez kapcsolódó veszélyt és kockázatot rejt (Li et al. 2023). A szekunder kutatás mellett primer (kérdőíves megkérdezés) segítségével felmértem a magyar egyetemisták véleményét a nagy nyelvi modellek használhatóságát illetően.

Forráskeresés, irodalomkutatás

A téma vizsgálatára szisztematikus szakirodalmi áttekintést végeztem. Ez a módszer eltér a hagyományos

áttekintéstől, mivel világos és reprodukálható folyamatot igényel (Mengist et al. 2020). Xiao–Watson (2019) szerint biztosítja a téma pontos körülhatárolását, miközben csökkenti az elfogultság lehetőségét a kutatás minőségének és megbízhatóságának megőrzése

érdekében. A feltárás során Webster–Watson (2002) technikáját használtam fel, amelyet az áttekinthetőség érdekében táblázatba foglaltam. A módszer 5 fő lépést tartalmaz (1. sz. táblázat).

Lépés	Lépés tartalma
1. Fókusz	Első lépésként a kutatás fókusza került meghatározásra, ami a ChatGPT és az LLM-ek szerepe az oktatásban, valamint annak biztonsági kérdései voltak. A feltárás fókusza a következő területeket nem érinti: szabályozás, technikai kivitelezés, technológia.
2. Témakör-kulcsszavak	Következő lépésként azonosítottam a témával kapcsolatos szakirodalmat. Ez segített a megfelelő keresési kulcsszavak meghatározásában.
3. Szakirodalmi áttekintés	A kutatásokat a Web of Science és a Google Scholar adatbázisokban kerestem fel az alábbi kulcsszavak kombinációi alapján: GPT-3, GPT-4, ChatGPT in education, Large Language Models, Security. Szűrési feltételek: Év: 2019-től, Nyelv: angol és magyar, Elérhetőség: OpenAccess.
4. Irodalomelemzés	A 4. lépésben összesítettem a meghatározott kulcsszavak és feltételek mentén a tudományos anyagokat. Összesen 165 db tudományos publikációt találtam (Google Scholar 42db, Web of Science 123 db). A 165 találatból 96 kizárára került, hiszen nem kapcsolódtak szorosan a feltárni kívánt területhez, így 69 kutatást használtam fel jelen vizsgálathoz.
5. Véglegesítés	Utolsó lépésként véglegesítésre kerültek a fejezetek a témafelvetés és a feltárt anyagok alapján.

1. sz. táblázat: Szisztematikus szakirodalmi áttekintés lépései
Forrás: Saját szerkesztés Webster–Watson (2002) alapján

Primer kutatás

Az adatgyűjtés online, önkitöltős formában, anonim módon zajlott. A kérdések a ChatGPT-hez és más nagy nyelvi modellekhez való viszonyulásra irányultak. Amikor a mintavétel módszerét kiválasztottam, fontos szempont volt a kérdőív könnyű hozzáférhetősége és gyors kitöltési lehetősége. Az előnyben részesített elektronikus módszernek köszönhetően a válaszadók online tudták kitölteni a kérdőívet. Összesen 442 hallgatótól érkezett kitöltött kérdőív, 17 magyarországi felsőoktatási intézményből. A szisztematikus szakirodalmi áttekintés összegzése alapján a kérdőívhez 6 nagy nyelvi modellt építettem be a kérdőíves

felmérésbe: ChatGPT, Bert, Bard, Jasper, Chatsonic, DeepL. A feltárt szakirodalmak alapján ez a hat szolgáltatás a legnépszerűbb és legtöbbet hivatkozott.

Szakirodalmi áttekintés

A kutatásom első szakaszában pontosan definiálni szerettem volna a nagy nyelvi modellekkel kapcsolatos koncepciót. Ennek megvalósításához három specifikus aspektust azonosítottam a publikációk alapján, amelyek

segítenek a fogalom mélyebb megértésében és körülhatárolásában.

Technikai aspektus

A nagy nyelvi modellek mélytanulási technikára támaszkodnak és hatalmas adatmennyiségekkel vannak „képezve” annak érdekében, hogy az emberhez hasonlóan értelmezzék és alkossák meg a természetes nyelvi tartalmakat. Ezek a modellek rengeteg paraméterrel működnek, amelyek a képzési folyamat során finomodnak, hogy hatékonyan illeszkedjenek a nyelv komplex mintáihoz és szerkezeteihez (Marino et al. 2023).

Alkalmazás aspektus

A nagy nyelvi modellek számos alkalmazási területen megtalálhatók, mint például chatbotokban, virtuális asszisztensekben, tartalomgeneráló eszközökben és e-mail szerkesztési segédprogramokban. Azáltal, hogy képesek megérteni, létrehozni és módosítani az emberi nyelvet, támogatják a nyelvi feladatok automatizálását és optimalizálását (Glaser 2023).

Akadémiai aspektus

A nagy nyelvi modellek a mesterséges intelligencia egyik jelentős kutatási területét képviselik, amelynek célja a nyelv megértésének és generálásának javítása. A kutatások nagy része elsősorban a multimodális képességeik kihasználására, a többnyelvű tanulás hatékonyságának növelésére összpontosít (Bill et al. 2021).

Tanulmányom egyik fontos szempontja a ChatGPT és annak felhasználhatósága. Lényeges pont a kutatásban, hogy a nagy nyelvi modell (LLM) kategóriájába tartozik, de muszáj megjegyezni, hogy léteznek hasonló modellek. Az LLM egy tág fogalom, amely magában foglalja a ChatGPT architektúráját, valamint a többi változatának (GPT) egyikét. A GPT, ami a Generative Pre-trained Transformer (generatív előképzett transzformátor) rövidítése, olyan alapként szolgál, amelyre a modell épül. Az alternatív modellek különböző transzformátorokat alkalmaznak. A GPT, vagyis a Generative Pre-trained Transformer egy nyelvi modellek fejlesztéséhez használt mélytanulási struktúra. A GPT modelleslád az OpenAI-tól származik, a GPT-4 a legfejlettebb, negyedik verziója (Gozalo et al. 2023).

A nyelvi modellek kategorizálása

A nyelvi modellek három kategóriáját különböztetjük meg:

- a) Az előképzésre alkalmas modellek, mint például a GPT-3/GPT-3.5, a T5 és a XLN, amelyek hatalmas mennyiségű adaton képzettek, ami lehetővé teszi számukra, hogy változatos nyelvi mintákat és struktúrákat alkossanak (Mohamadi et al. 2023).
- b) Finomhangolt modellek, mint például a BERT, a RoBERTa és az ALBERT, melyeket eredetileg hatalmas adathalmazon előtanulmányozták, majd egy kisebb, konkrét feladatra szabott adatokon finomhangolják. Ezek a modellek rendkívül hatékonyan bizonyultak olyan feladatokban, mint az érzelemelemzés, a kérdések megválaszolása és a szövegosztályozás (Hadi et al. 2023).
- c) A multimodális modellek, mint például a CLIP és a DALL-E, melyek több modalitást, köztük képeket és videókat is felhasználnak, hogy megkönnyítsék a rugalmasabb nyelvi modellek létrehozását. Ezek a modellek megértik a képek és a szöveg közötti kapcsolatokat, lehetővé téve a képek szöveges ábrázolásának előállítását, vagy akár képek létrehozását szöveges leírásokból (Chen et al. 2022).

A nagy nyelvi modellek használhatósága általánosságban

Chatbot automatizálás

A chatbotok fejlesztésében a nyelvi modelleknek rendkívül fontos szerepük van, mivel ez az egyik legelterjedtebb alkalmazásuk. Ezek az algoritmusok képesek természetes nyelvi válaszokat generálni, ami az emberi interakciót utánozza. Sok vállalat ezeket a chatbotokat használja az ügyfélszolgálat terén, legyen szó akár a vállalat weboldalának leállításáról, vagy az ügyfelek készülékeinek hibaelhárítás-segítéséről (Tlili et al. 2023).

Értékesítés

Dwivedi és munkatársai szerint (2023) szerint az LLM-eket nem csak információfeldolgozásra, hanem céltudatos marketing és értékesítési tartalmak létrehozására is lehet alkalmazni. Egy adott nyelvi modell képes tanulni és adaptálódni ügyfél visszajelzéseken vagy korábbi értékesítési kampányok

tapasztalatain keresztül, így hatékonyabb és személyre szabott értékesítési üzeneteket hozhat létre.

Természetes nyelvfeldolgozás

A nagy nyelvi modellek fontos eszközök a természetes nyelvfeldolgozás (Natural Language Processing - NLP) képességeinek javításában. Az NLP egy olyan mesterséges intelligencia (MI) ágazat, melynek célja a nyelv értelmezése és elemzése számítógépes algoritmusok segítségével. Az LLM-kel algoritmusokat lehet képezni, hogy mélyen megértsék az emberi nyelv különféle összefüggéseit, így kiválóan alkalmazhatók olyan NLP tevékenységekben, mint a szövegek összefoglalása, kulcsszavak kinyerése vagy szövegek részletes elemzése (Wang et al. 2023).

Kutatás

A nagy nyelvi modellek képesek kiváló minőségű tudományos cikkek előállítására, amennyiben megfelelő adatforrásokkal lettek kiképezve. Ezek az algoritmusok képessé válhatnak arra, hogy új tanulmányokat hozzanak létre azonos stílusban, módszertannal és formázással, mint az adott területen már meglévő akadémiai publikációk (Cotton et al. 2023).

Adatelemzés és Analitika

Dai munkatársaival végzett (2023) kutatása szerint az LLM-ek képesek adatelemzési és statisztikai munkafolyamatok automatizálására, mint például valós idejű jelentések készítése az ügyfélrendekről vagy a piaci viszonyokról. Ezt az eszközt különösen értékelhetik azok a cégek, melyeknek nagy mennyiségű adatot kell gyorsan és hibátlanul feldolgozniuk.

Oktatás

A nagy nyelvi modellek új lehetőségeket tárnak fel az oktatásban. Egyik legfontosabb előnyük, hogy képessé válnak személyre szabott tanulási stratégiákat létrehozni, így a tanulók saját ütemükben haladhatnak, mialatt a rendszer fókuszál az egyéni erősségeikre és gyengeségeikre. Emellett ezek a modellek hozzájárulnak a tananyag testeszteléséhez, annak dinamikus alkalmazkodásához, és egy olyan interaktív tanulási környezet kialakításához, ami leköti a tanulók figyelmét (Pericles 2023).

Kódgenerálás, szoftverfejlesztés

A GPT-3-hoz hasonló nagy nyelvi modellek forradalmasítják a kódolás és szoftvertervezés világát, lehetőséget adva arra, hogy konkrét utasítások alapján automatikusan hozzanak létre kódokat. Néhány fejlesztői platform és integrált fejlesztői környezet már beépítette ezt a technológiát, ami meggyorsíthatja és hatékonyabbá teheti a fejlesztési folyamatot. Ezáltal a pályakezdő fejlesztők is könnyebben és gyorsabban sajátíthatják el a programozást (Ray 2023).

Az oktatásban rejlő potenciál

A nagy nyelvi modellek forradalmasították nem csak az informatikai és mesterséges intelligencia területeit, hanem az oktatás és a diákok tanulási módszereit is. A hatalmas adatbázisoknak köszönhetően ezek a modellek elősegítik az interdiszciplináris és személyre szabott tanulást, mivel korábban példátlan kontextuális megértésre képesek (Sullivan 2023). Az oktatók számára ezek a rendszerek idő- és erőforrás-megtakarítást jelentenek a tananyagok összeállításakor és a diákok kérdéseire adott válaszoknál (Lo 2023). Firaina-Sulisworo (2023) megjegyzi, hogy az oktatók könnyen készíthetnek átfogó háttéranyagot egy új témához, míg az LLM-ek a hallgatóknak mentorálási és egyéni támogatási szerepet is betölthetnek, lehetővé téve számukra a saját tempójukban és érdeklődésük szerinti tanulást. Rasul és munkatársai (2023) szerint az LLM-ekből származó adatok segíthetik a tanárokat a diákok egyedi igényeinek és előnyeinek jobb megértésében. A diákoknak lehetőségük nyílik a modellek segítségével egyéni projekteken és kutatásokon dolgozni, gyors adatelemzés és -feldolgozás mellett (Cox-Tzoc 2023). Kiliņç (2023) szerint a modellek használata fejleszti a kritikai és problémamegoldó gondolkodást, míg a tanárok az érvelés és kritikai gondolkodás fejlesztésére használhatják azokat (Benuyenah 2023).

Ugyanakkor a technológiai előrelépésekkel kihívások is járnak. A modell által generált információ hitelességének biztosítása az egyik legnagyobb aggodalom, mivel az internetről származó hibás vagy megtévesztő adatok átvétele lehetséges (Hong 2023).

A jövőbeni előrelépések közé tartozik a differenciált oktatás új szintre emelése és a tanulási élmény személyre szabása (Sun-Hoelscher 2023), a digitális laborok és szimulációk fejlesztése (Biswas 2023), valamint a távoktatás interaktívabbá és emberközpontúbbá tétele (Kasneci et al. 2023). Emellett a modellek potenciálisan képesek lehetnek a diákok érzelmi és pszichológiai állapotának jobb megértésére, segítségnyújtásra szorongás, stressz vagy más lelki problémák esetén (Ying 2023). Lund-Wang (2023) azt vizsgálta, hogy az

oktatásban milyen módon segíthetik a nagy nyelvi modellek az alábbi területeket:

- Szövegalkotás: a modellek képesek különböző stílusokban szövegeket létrehozni, amelyek lehetővé teszik a kutatóknak, hogy vázlatokat készítsenek kutatási anyagokból és egyéb dokumentumokból.
- Elemzés: a modellek képesek nagy mennyiségű szöveges adat elemzésére és mintázatok azonosítására.
- Fordítás: a modellek gépi fordításra is alkalmasak, így a többnyelvű kutatási anyagok könnyen érthetővé válnak.
- Automatizálás: a modellek tudományos cikkek és más dokumentumok automatikus összefoglalására is képesek, segítve a kutatókat a legfrissebb fejlemények követésében.

A használat mellett szóló érvek az elemzett kutatások alapján

Beszédatánzás: a ChatGPT alapvető funkciója az emberi konverzációk imitálása, mely a felhasználó adta kérések vagy parancsok alapján valósul meg. Ebben a tekintetben hasonló a modern MI-alapú asszisztens rendszerekhez, mint az Alexa és Siri (Moqbel et al. 2023).

Univerzális alkalmazhatóság: a nagy nyelvi modellek számos területen alkalmazhatók: szövegtérmezés, szövegteremtés, zenei szövegek és regények írása, adatkezelés, programozás, bonyolult problémák megoldása, automatizáció és fordítás. Ez a sokrétűség lehetővé teszi alkalmazásukat az oktatásban, kutatásban, egészségügyben, marketingben és a sportban is (Pardos–Bhandari 2023).

Folyamatos fejlődés: az olyan mesterséges intelligencia alapú szolgáltatások, mint a ChatGPT, gépi tanulási módszerekre építenek. A készítőik folyamatosan tökéletesítik őket, így az idő múlásával az ilyen rendszerek egyre hatékonyabbak és pontosabbak lesznek (Hosseini–Horbach 2023).

Mély tudásbázis: az LLM-alapú mesterséges intelligencia rendszerek nagy mennyiségű adatbázissal rendelkeznek, amelyek lehetővé teszik számukra, hogy komplex kérdésekre válaszoljanak. Ezek az információforrások kiváló támogatást nyújtanak az oktatásban és a kutatásban, segítve a felhasználókat tudásuk bővítésében (Kocon et al. 2023).

Hatékony és időgazdálkodás: az LLM-ek képesek automatizálni a szövegtérmezés folyamatát, így jelentős időt takarítanak meg. Ezeknek köszönhetően gyorsabban és precízebben hozhatunk létre tartalmakat (Gupta et al. 2023).

Személyre szabott megközelítés: az LLM-ek segítségével képesek vagyunk személyre szabott válaszokat és

javaslatokat adni, ami lehetővé teszi a vállalatok számára, hogy mélyebb kapcsolatot építsenek ki ügyfeleikkel (Dwivedi et al. 2023).

Kreativitás támogatása: az LLM-ek nem csak az informatív szövegalkotásban, de kreatív írásban is támogatást nyújtanak, inspirálva az írókat és segítve őket különböző szövegvariánsok létrehozásában (Hong 2023).

Az oktatásban való használat ellen szóló érvek

Az emberi kapcsolatok hiánya: a ChatGPT és a hasonló generatív modellek nem tudják ugyanazt az emberi kapcsolatot biztosítani, mint egy valós tanár vagy oktató. Az emberi kapcsolatok hiánya korlátokat teremthet a tanulók számára, akiknek a tanári interakció sokkal előnyösebb lehet (Dehouche 2021).

Felszínes feldolgozás: a generatív modellek az általuk tanulmányozott adatok statisztikai mintázataira támaszkodnak, de nincsenek mély ismereteik a létező összes témáról (korlátozott adattal rendelkeznek). Ez kihívást jelenthet, amikor a diákok egyedi igényeihez és félreértéseihez igazított magyarázatokra vagy visszajelzésekre van szükség. Wang munkatársaival végzett közös (2020) kutatása azt mutatta, hogy a generatív modelleken alapuló oktatási rendszer nem tudott testreszabott válaszokat adni a tanulóknak.

Elfogultság és korlátozottság: a generatív modellek teljesítménye közvetlenül függ azoktól az adatoktól, amelyeken képzik őket (Sullivan et al. 2023). Ha a képzési adatok elfogultak, a modell is hasonló hibákat fog mutatni (Ray 2023). A generatív modellek válasza azokon a mintákon alapulnak, melyeket képzésük folyamán ismertek meg, ezáltal befolyásolva a válaszok újszerűségét és kreativitását. Egy 2022-ben Zhai által végzett kutatás során megállapították, hogy egy generatív modellt használó zenészerzési rendszer csak korlátozott mértékben képes az eredeti és sokszínű dallamok létrehozására.

A generatív modellek kontextusbeli korlátai: a generatív algoritmusok gyakran nem tudják helyesen értelmezni a kontextust és a körülményeket, ami hibás vagy nem releváns válaszokhoz vezethet (Gui et al. 2021).

Szerzői jogok és adatvédelem: amikor nagy nyelvi modelleket képeznek oktatási anyagok – mint például tananyagok, versenyek, tudományos művek – előállításához, a tanulási folyamat során ilyen jellegű szövegekkel táplálják a modellt. Az így előállított új tartalmakban a válaszok olyan teljes mondatokat vagy bekezdéseket is tartalmazhatnak, amelyek a képzési adatokban is megtalálhatók, ami szerzői jogi kérdéseket és plágiummal kapcsolatos problémákat vet fel (Cox–Tzoc 2023).

Etikai kérdések és szövegek megítélésének kihívásai: a ChatGPT által előállított válaszok és szövegek már szinte nem is különböznek az emberi írásoktól, ami komplikációkat okoz az oktatói és kutatói közösség számára. Cotton munkatársaival (2023) több módszert is bemutatott, melyek az LLM-ek, így például a ChatGPT által generált szövegek azonosítására irányulnak, amelyek tartalmazzák a nyelvtani ellentmondásokat, a megfelelő hivatkozások hiányát, a valóságtartalmi pontatlanságokat, a gyengén körülhatárolt kontextust. Új technológiák, mint például a mesterséges intelligenciával működő plágiumkereső eszközök kidolgozása szükséges (Elkins-Chun 2020).

Kritikai elemzés és problémamegoldás: a ChatGPT által könnyedén előállított válaszok és kódok gátolhatják a diákok kritikai elemzési és problémamegoldó képességeinek bővítését. A jelenlegi információk szerint nincsenek olyan rendszerek, amelyek azonosítani tudnák az LLM modellek által előállított kódot, így ezeket a kódokat fel lehet használni akadémiai kódolási teszteken és versenyeken (Rahman–Watanobe 2023).

Fenntarthatóság és energiahatékonyság: a nagy nyelvi modellek komoly számítási kapacitást igényelnek, ami jelentős energiabefektetést von maga után. Ennek fényében az energiahatékony eszközök és a megújuló energiaforrásokra támaszkodó elosztott infrastruktúra alapvető az ökológikusan fenntartható működésük érdekében. A megújuló energiaforrásokra támaszkodó energiahatékony technológia és az elosztott rendszerek, valamint a költséghatékony képzési és karbantartási módszerek irányába tett lépés elengedhetetlen a technológia jövőjét illetően (Haque et al. 2022).

Biztonsági kockázatok az oktatásban

Az LLM-ek, mint a ChatGPT, illetve a mesterséges intelligencia új lehetőségeket nyit meg, de ezzel egyidőben a kiberbiztonság és adatbiztonság kihívásai is egyre égetőbbekké válnak (Rudolph et al. 2023). Az egyik legfontosabb probléma, amely az MI-n alapuló oktatási eszközökkel kapcsolatos, az az adatgyűjtés és -kezelés. Ezen eszközök, például a ChatGPT, a kérdések és válaszok folyamatos elemzésével működnek, amelyeket a felhasználók adnak meg. Bár ez az interakció lehetővé teszi az MI számára, hogy jobban megértse és reagáljon a felhasználói igényekre, ugyanakkor adatbiztonsági aggályokat is felvet. Az OpenAI, a ChatGPT mögött álló cég azt állítja, hogy nem tárolja a felhasználók által megadott adatokat. Ennek ellenére az adatgyűjtés mindig potenciális kockázatokkal jár, különösen, ha a felhasználók nincsenek tudatában az adatkezelési gyakorlatoknak vagy azoknak az információknak, amelyeket megosztanak (Susnjak 2022).

Garg és munkatársai (2023) szerint ezen kívül az adatszivárgás veszélye is fennáll. Az oktatási intézmények gyakran rendelkeznek értékes információval diákokról és tanárokról, így kritikus fontosságú, hogy megfelelő biztonsági intézkedéseket alkalmazzanak az MI-platformokon. Ha egy intézmény nem védi megfelelően az MI-alapú rendszert, az abban tárolt információk ki is szivároghatnak, ami jelentős károkat okozhat az érintett egyéneknek és az intézménynek egyaránt.

Az MI rosszindulatú felhasználásának veszélyével is szembe kell néznünk (Singh–Singh 2023). Tekintettel az MI képességeire, nem meglepő, hogy bizonyos személyek azt rossz célokra használnák, mint például a plagizálás vagy vizsgacsálás. Ez kihívást jelent az oktatási intézmények számára, hogy hogyan alkalmazzák ezeket az eszközöket anélkül, hogy kompromisszumot kötnének az oktatás integritásáról.

Kockázatok azonosítása, kategorizálása

Adatvédelmi kockázatok: az MI-modellek gyakran nagy mennyiségű adatot igényelnek a tanuláshoz és az optimalizáláshoz. Ha az adatok nem megfelelően vannak tárolva vagy védve, azok ki is szivároghatnak. Ez sértheti az érintettek adatvédelmi jogait és komoly jogi következményekkel is járhat (Pinto et al. 2023).

Technológiai kockázatok: a mesterséges intelligencia modellek manipulálhatóak vagy támadhatóak (Qureshi 2023). Például adatmérgezés támadás esetén rosszindulatú adatokkal tréningelhető a modell, ami a későbbi predikciókat befolyásolja (Brown et al. 2023).

Felhasználói kockázatok: az MI eszközöket rosszindulatú célokra is lehet használni, például csalásra vagy információhamisításra (Lambert–Stevens 2023). A diákok és tanárok túlzottan támaszkodhatnak az MI-támogatásra, ami a kritikus gondolkodás hiányához vagy a saját képességeik alábecsüléséhez vezethet (Kuhail et al. 2023).

Etikai és társadalmi kockázatok: az MI modellek azon adatokon alapulnak, amelyekkel tréningelik őket. Ha ezek az adatok előítéletesek vagy nem reprezentatívak, az MI döntései is torzítottak lehetnek (Ofosu–Ampong et al. 2023). Kockázat a túlzottan mechanikus és nem emberközpontú tanulási környezet kialakulása, ami csökkentheti a tanulás mélységét és/vagy minőségét (Vrontis et al. 2023).

Lehetséges kiberbiztonsági kockázatok

Adatvesztés: ha valaki rosszindulatúan használja a rendszert, érzékeny információt is kihozhat belőle. Példa: egy vállalat belső információit valaki beépíti a rendszer tanulási adatába, és ezeket az információkat

egy harmadik fél később megpróbálja kinyerni (Prieto 2023).

Rosszindulatú tanítás: mivel a ChatGPT is tanul a felhasználói interakciókból, elképzelhető, hogy valaki szándékosan próbálja megtéveszteni vagy rosszindulatú információval feltölteni (Benuyenah 2023).

Elérhetőségi támadások: ha valaki túlterheli a szolgáltatást, azzal megbéníthatja azt. Egy DDoS támadás célja, hogy annyi lekérdezést indítson a rendszeren, hogy az ne tudjon válaszolni a tényleges felhasználók kérdéseire (de Zarzá 2023).

Személyazonosság-lopás: ha a rendszer nem kezeli megfelelően a felhasználói adatokat, harmadik fél ellophatja azokat. Egy rosszindulatú támadó kihasznál egy biztonsági rést és hozzáfér a felhasználói beszélgetésekhez, ami érzékeny információt tartalmazhat (Khalil–Er 2023).

Manipuláció: mivel az MI reagál a bemenetre, rosszindulatú támadók manipulálhatják az eredményeket olyan válaszok érdekében, amelyeket ők preferálnak. Egy támadó a rendszert oly módon kérdezi meg, hogy az adott politikai nézeteket vagy hamis információt támogasson (Maddigan 2023).

Integrációs kockázatok: ha a ChatGPT-t más rendszerekkel integrálják, a támadók kihasználhatják ezt az integrációt a rendszerben található sebezhetőségek kihasználására. Egy okos otthoni rendszer, amely a ChatGPT-t használja a felhasználói parancsok feldolgozására, esetleg rosszindulatú parancsokat is fogadhat (Ghourabi–Alohaly 2023).

Primer kutatási eredmények és azok értékelése

A kvantitatív kutatási módszerek közül kérdőívet használtam felmérésemhez. A vizsgálat célcsoportja egyetemi hallgatók voltak. 442 egyetemista válaszolt a kérdőívre 17 különböző egyetemről. Ez a szám az aktuális 280 ezer fős hallgatói létszám 0,16%-a (Tempus 2022). Bár a kérdőív nem tekinthető átfogóan

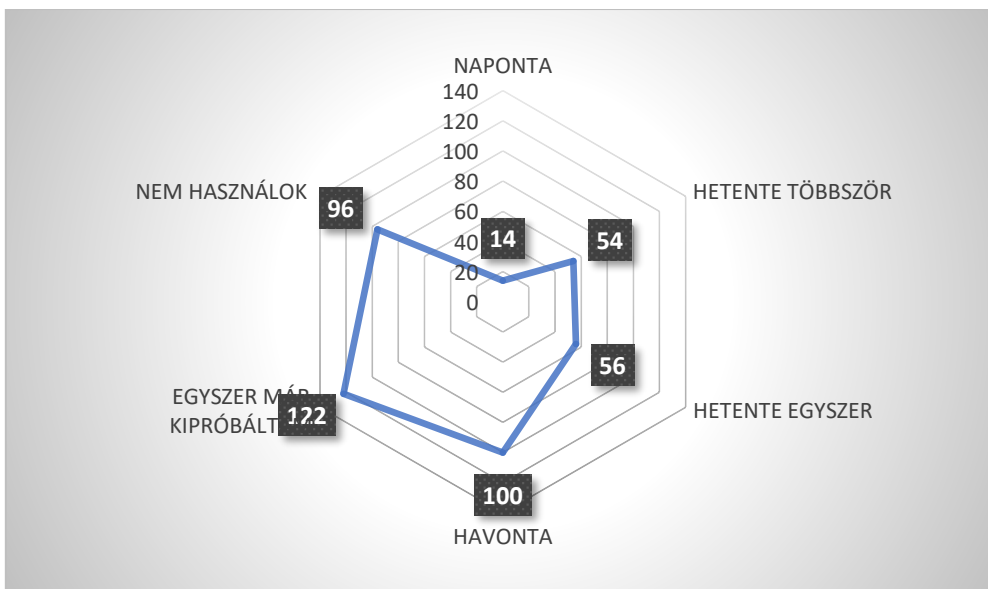
reprezentatívnak, úgy gondolom, hogy az eredmények jól tükrözik az általános nézeteket. A kitöltések alapján az egyetemeket 2 fő kategóriára osztottam:

- Budapesti régió (BME, CORVINUS, ELTE, MATE, NKE, BGE, PPKE, SE, OE)
- Vidéki régió (DE, SZTE, SZE, PTE, PE, KJE, ME, EKKE).

A résztvevők a kérdőívben bejelölték azt a tudományterületet, ahol jelenleg tanulmányokat folytatnak. A válaszokat áttekintve az alábbi kategóriákba soroltam a megjelölt tudományterületeket: bölcsészettudományi, gazdaságtudományi, informatikai, jogi, mezőgazdasági, műszaki, orvosi, pedagógiai, rendszertudományi és természettudományi. Ezen kategóriák által könnyebben érthető és áttekinthető képet kaptam a válaszadók tanulmányi háttéréről.

A résztvevők információkat adtak jelenlegi munka- és vállalkozási helyzetükről is. A válaszok szerint a kitöltők 35,3%-a már rendelkezik munkahellyel, míg a válaszadók 64,7%-a jelenleg nem áll munkaviszonyban. Ami a vállalkozói tevékenységet illeti, az összes résztvevő mindössze 4,5%-a vezet saját vállalkozást, a többi 95,5% nem tevékenykedik vállalkozóként.

A korábbi fejezetekben bemutatott nyelvi modellek népszerűségének tendenciája a válaszadók között a következőképpen alakul: a ChatGPT-t a válaszadók 95,9%-a ismeri, ami nem meglepő, tekintve, hogy ezen a platformon indult a LLM trend. Második helyen a DeepL áll 152 válaszadó megjelölésével, mely 34,4%-os arányt jelent. A statisztikák világosan reflektálnak a ChatGPT domináns szerepére ezen a területen, több mint 60%-os ismeretségi előnnyel. A rangsorban következő a Jasper.Ai 8,1%-os, azaz 36 válaszadó által történt megjelöléssel. Utána 28 válaszadó, azaz 6,3% ismeri a Bert Ai-t, míg a Bard Ai-t, melyet a Google hozott létre, 12 válaszadó, azaz 2,7% ismer. A lista végén 1% jelzett úgy, hogy nem ismeri vagy nem használ nagy nyelvi modelleket (1. sz. ábra).

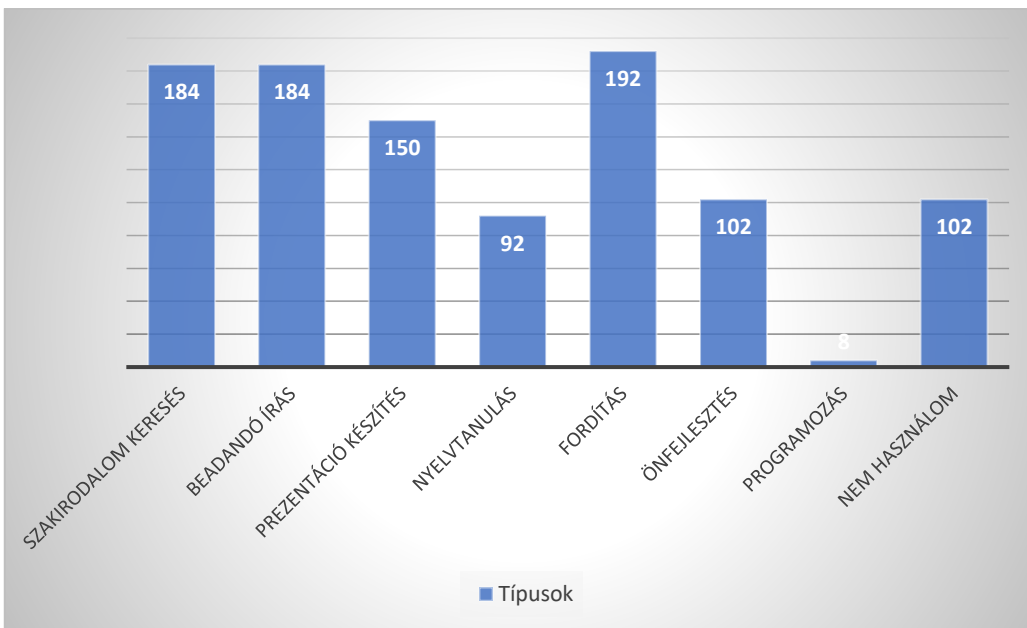


1. sz. ábra: Nagy nyelvi modellek alkalmazásának gyakorisága (N=442)
 Forrás: Saját szerkesztés

Az adatok elemzése során kiderült, hogy a válaszadók 49%-a egyszer már kipróbálta, vagy egyáltalán nem alkalmazza a nagy nyelvi modelleket. A fennmaradó 51% valamely gyakorisági kategóriába tartozik, ami arra utal, hogy minden második egyetemi hallgató valamilyen gyakorisággal használja ezeket a modelleket. A jövőbeli tendenciákat vizsgálva

feltételezhető, hogy ez az arány módosulni fog. Jelenleg mindössze 3% alkalmazza naponta, 12% hetente többször, 13% hetente egyszer, míg 23% havi gyakorisággal használja ezeket a szolgáltatásokat.

A kutatás során azt is felmértem, milyen céllal alkalmazzák az egyetemi hallgatók a nagy nyelvi modelleket (2. sz. ábra).



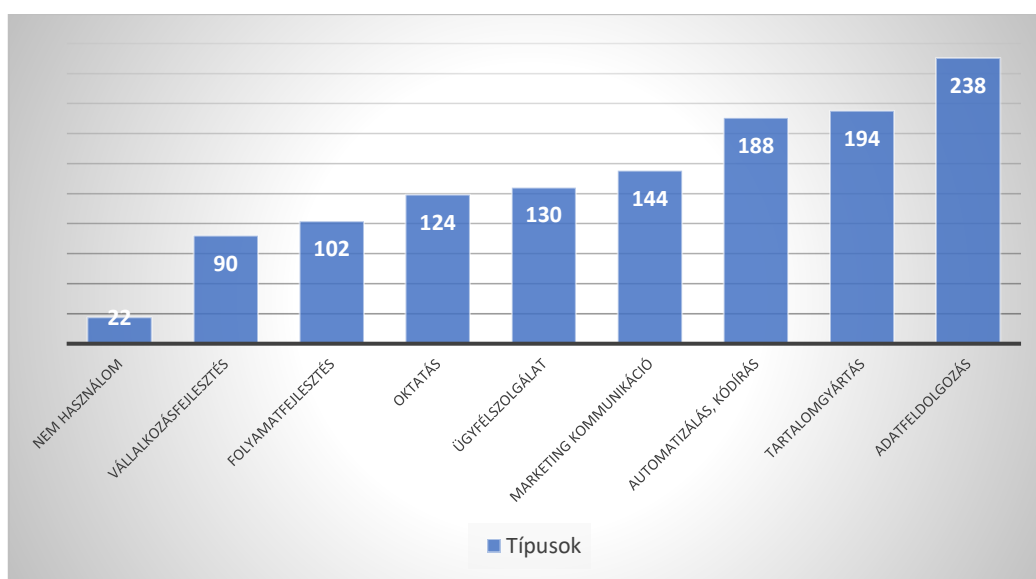
2. sz. ábra: Feladattípusok a kitöltők körében a nagy nyelvi modelleket illetően (N=442)
 Forrás: Saját szerkesztés

A válaszlehetőségek közül több is választható volt. 102 résztvevő jelezte, hogy nem veszi igénybe a nyelvi modelleket, 8 fő alkalmazza programozásban, 102 fő önfejlesztési célokra, míg 92 fő nyelvi tanulás során használja ezeket az eszközöket. A legjobb 3 eredmény esetében nem született váratlan tendencia. Ugyanazon kitöltési számmal, azaz 184 esettel a szakdolgozat, ill. diplomamunka készítését jelölték meg, valamint a szakirodalom keresése áll a második helyen. A rangsorban első helyre, 192 megjelöléssel a fordítási felhasználás került a kitöltők véleménye alapján.

Az egyetemi kontextuson túl a kutatás a vállalati és professzionális alkalmazási területeket is vizsgálta. A kérdés a következő formában szerepelt: „Milyen

területeken használnád a Nagy Nyelvi Modelleket a saját vállalkozásodban vagy munkahelyi projektjeidben?”

Több válaszlehetőség is választható volt. Mindössze a résztvevők 5%-a állította, hogy nem alkalmazná azokat egyetlen területen sem. Ezen adatból az a következtetés vonható le, hogy a magyar egyetemi hallgatók körében a nagy nyelvi modellek vállalati alkalmazása látványos jövőképpel bír a 442 fős minta alapján. 102 résztvevő folyamatfejlesztésre, 90 fő vállalkozásfejlesztésre, 124 fő oktatási és kutatási célra használná. 130 alany ügyfélszolgálati teendőkben, 144 pedig marketingkommunikációban alkalmazná. Az első három helyezettnél az automatizálás és kódírás 188 megjelöléssel, a tartalomgyártás 194 válasszal, míg az adatfeldolgozás 238 megjelöléssel szerepelt (3. sz. ábra).



3. sz. ábra: Munkahelyi és üzleti alapú felhasználási területek a kitöltők véleménye alapján (N=442)

Forrás: Saját szerkesztés

A tanulmány keretében arra is kerestem választ, hogy az egyetemi hallgatók hogyan ítélik meg a nagy nyelvi modellek használatának potenciális hatásait saját gondolkodási és tanulási folyamataikra. A válaszadók 61,5%-a szerint a modelleknek nincs káros hatása, míg 38,5% az ellenkezőjét állítja. Tekintettel a válaszok megoszlására, javasolt további kutatásokat végezni ezen a területen.

A vizsgálat továbbá kitért arra, hogy az egyetemisták hogyan állnak a nagy nyelvi modellek szigorú szabályozásához és az egyetemeken történő potenciális tiltásához: 76% nem támogatja az erős szabályozást és tiltást, míg 24% igen. Feltételezem, hogy a válaszok közötti korreláció lehet az egyetemisták potenciális negatív megítélésével és a szabályozás támogatásával. Ennek ellenére a válaszadók kétharmada nem

módosítana a jelenlegi gyakorlaton, sugallva, hogy az intézményeknek idővel el kell fogadniuk ezen technológiákat, remélve, hogy a hallgatók hatékonyságuk és tudásuk növelése érdekében fogják alkalmazni azokat.

Végezetül a 13. kérdés így szerepelt: „Egyértékes-e azzal az állítással, hogy a Nagy Nyelvi Modellek alkalmazása különböző feladatokra és projektekre rossz hatással van az emberi kreativitás, gondolkodás és tanulás fejlődésére?” A válaszok alapján 58,6% (258 válaszadó) nem értett egyet az állítással, míg 41,4% (184 válaszadó) igen. A válaszok szoros megoszlása alapján megállapítható, hogy ezen a területen a közvélemény még nem kristályosodott ki egyértelműen. A válaszok egyensúlya további vizsgálatokat igényel. Következtetésként megállapítható, hogy ez a kérdéskör

nem egyértelműen eldöntött, ami több kérdést is felvethet, akár a további vizsgálatokhoz, például: Milyen indikátorok alapján vélekedik a minta 41,4%-a arról, hogy a vizsgált tényező potenciálisan negatív hatást gyakorolhat az emberekre? Milyen tapasztalatok alapján állítják, hogy ez akadályozza a tanulási folyamatok optimális fejlődését? Milyen tudományos vagy empirikus alapokon nyugszik a minta 58,6%-ának az a véleménye, hogy a vizsgált tényező nem gyakorol káros hatást az emberekre?

Összefoglalás és javaslatok

Az eredmények azt mutatják, hogy elegendő mennyiségű tudományos forrás állt rendelkezésre a téma mélyreható vizsgálatához. Tekintettel a téma aktualitására, a kapcsolódó szakirodalom folyamatosan bővül. A jövőben célszerű lehet újból elvégezni a szisztematikus szakirodalmi elemzést annak érdekében, hogy felmérjük, miként alakul a tudományos közösség álláspontja a modellek felhasználhatóságával, valamint az esetleges veszélyekkel kapcsolatban.

Javaslatok a szekunder kutatásban feltárt oktatási veszélyekre

Fontos megjegyezni, hogy az MI nem helyettesíti minden területen az oktatót, legalábbis hosszú távon ez nem minden esetben vezethet jó eredményre, így javaslatom szerint az MI-nek támogató szerepet kell betöltenie, emberi interakciókkal kiegészítve. A LLM-eket létrehozók tekintetében nagy felelősség van a kezükben, hogy milyen adathalmazokon képezik az adott mesterséges intelligenciák, ugyanitt nagy felelősség van az oktatók, hallgatók kezében is, hogy milyen minőségben kezelik a kapott válaszokat. Véleményem szerint az LLM-ek által generált válaszokat helyén kell kezelni, alapos ellenőrzéssel az adatok hitelességét tekintve. A szerzői jogi protokollokat be lehetne illeszteni ezekbe a modellekbe, így az adott LLM jogi háttérrel megtámogatva tudna még több mindent megtanulni, illetve hivatkozni. Ezzel szemben a modellek által generált tartalom, úgy gondolom, hogy nem számít plágiumnak tudományos munkák felhasználása esetén. Végezetül pedig fontos terület a fenntarthatóság, ami erre a technológiára is vonatkozik. A jövőben ezek népszerűsítése és használata elengedhetetlen, hogy tudatos és fenntartható használat mentén történjen.

Javaslatok a szekunder kutatásban feltárt biztonsági kockázatokra

Adatgyűjtés esetén a felhasználókat tájékoztatni kell az adatok gyűjtéséről. Fontos, hogy csak ahhoz járuljon hozzá a felhasználó, ami elengedhetetlen a használathoz. Adatszívárgásnál a felelősség a készítőik és a felhasználók kezében van. A készítőik szempontjából elengedhetetlen az erős titkosítási protokollok használata, a rendszeres biztonsági auditok végrehajtása, a felhasználók esetében pedig a multi, de legalább 2 faktoros hitelesítés alkalmazása. Tekintettel arra, hogy ez egy új terület, számos hibalehetőség merül fel az alkalmazásukkor. Az adott LLM rendszer sebezhető lehet, ezáltal fontos a folyamatos felügyelet, szoftverfrissítés, valamint a biztonsági előírások betartása. Az adott modellt hackerek képezhetik rosszindulatú adathalmazokon, ami problémás válaszokat generálhat. A hackerek ellen javasolt a védelem, tűzfalak és egyéb biztonsági eszközök használata.

Javaslatok a primer kutatási eredményekre vonatkozóan

A szekunder kutatást követően primer kérdőíves kutatást végeztem. A mintám 442 egyetemistából állt. A kitöltéseket tekintve főként első- és másodévesek töltötték ki a kérdőívet, harmad-, negyed- és ötödéves hallgatók kisebb számban vettek részt a kutatásban. A nyelvi modellek népszerűségének vizsgálatakor nem született meglepő eredmény, magasan a ChatGPT volt a legnépszerűbb nagy nyelvi modell. Az eredmények alapján mérsékelten használják ezeket a modelleket Magyarországon az egyetemisták, a kitöltők több, mint 50%-a használja valamilyen rendszerességgel. A legnépszerűbb egyetemi területek a fordítás, a szövegírás, valamint szakirodalom keresés. A munkahelyen történő felhasználást tekintve az adatfeldolgozás, a tartalomgyártás és az automatizálás voltak a legtöbbször megjelölt területek. A kutatás alapján a fentebb felsorolt folyamatok népszerűsítését javaslom, kurzusok és bemutató előadások létrehozása indokolt annak érdekében, hogy minél több emberhez eljusson az LLM-ek adta előnyök összesége. Fontos megjegyezni, hogy a kitöltők 76%-a az LLM-ek szabályozását, nem pedig tiltását javasolja az oktatásban.

A kutatás egésze alapján megfogalmazott javaslatok

Az LLM-ek integrálására minden tudományágban szükség van az oktatás és a kutatás hatékonyságának

növelése érdekében. Mivel ezek a technológiák már jelen vannak ezeken a területeken, nincs értelme megtiltani őket. A szabályozás viszont indokolt. Néhány kiemelendő fogalom ezzel kapcsolatban:

Plágium: a ChatGPT és az LLM-ek által létrehozott kimenetek bizonyos tekintetben még mindig etikai kérdéseket vetnek fel. Alapvetően minden alkotást vagy munkát az ember hoz létre, így a modell által készített válaszok is tekinthetők az adott személy munkájának. Az LLM nem több egy eszköznél, mint például a helyesírás-ellenőrző programok vagy a kreatív inspiráció. A felhasználók gyakran átalakítják, módosítják vagy kiegészítik az LLM által adott válaszokat, mivel jelenleg ezek nem mindig pontosak. Ahogy haladunk előre az időben, ez a téma további kérdéseket vethet fel, amit érdemes kivizsgálni.

Népszerűsítés: a vizsgálatok eredményei alapján kijelenthető, hogy az LLM-eknek jelentős potenciálja van az egyetemisták között. A téma relevanciája miatt indokolt a kutatás további mélyítése és bővítése. Mivel ez a terület rohamos tempóban fejlődik, fontos nagy hangsúlyt helyezni az újabb vizsgálatokra és azok népszerűsítésére.

LLM tudatosság: a szakirodalmi áttekintés során világossá vált, hogy az LLM-ek használatával több kockázat is jár. A jövőben ajánlott olyan képzések és kurzusok bevezetése „LLM tudatosság” címszó alatt, melyek segítenek az embereknek ezeket az eszközöket biztonságosan és tudatosan használni a tudomány minden területén.

Az elkövetkezendő időszak: az LLM-ek tudományos alkalmazása tekintetében ígéretesnek tűnik a jövő. A szakirodalom számos jelenlegi használati esetet és vitapontot említ az LLM-ekkel összefüggésben, de még rengeteg újító alkalmazás vár felfedezésre és elemzésre. Látom a lehetőséget arra is, hogy az LLM-eket más fejlett technológiákkal, mint például a blokklánc technológia és az automatizált gépi tanulás ötvözik. A fejlődés gyorsan zajlik, és az elkövetkező időszakban számos izgalmas fejlemény várható. Az igazi kihívás abban rejlik, hogy naprakészek maradjunk a folyamatosan fejlődő technológiák mellett.

Irodalomjegyzék

- Atlas, S. (2023) ChatGPT for higher education and professional development: A guide to conversational AI. https://digitalcommons.uri.edu/cba_facpubs/548/ [Letöltve: 2023.03.12.].
- Benuyenah, V. (2023) Commentary: ChatGPT use in higher education assessment: Prospects and epistemic threats. *Journal of Research in Innovative Teaching & Learning*, 16(1). pp. 134-135. [doi: 10.1108/JRIT-03-2023-097](https://doi.org/10.1108/JRIT-03-2023-097)
- Biswas, S. (2023) *Role of Chat GPT in Education*. <https://ssrn.com/abstract=4369981> [Letöltve: 2023.03.12.].
- Brown, T. B.–Mann, B.–Ryder, N.–Subbiah, M.–Kaplan, J.–Dhariwal, P.–Amodei, D. (2020) *Language Models are Few-Shot Learners*. <https://arxiv.org/abs/2005.14165> [Letöltve: 2023.03.15.].
- Chen, J.–Guo, H.–Yi, K.–Li, B.–Elhoseiny, M. (2022) *Visualgpt: Data-efficient adaptation of pretrained language models for image captioning*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18030-18040.
- Cope, B.–Kalantzis, M.–Searsmith, D. (2021) Artificial intelligence for education: Knowledge and its assessment in ai-enabled learning ecologies. *Educational Philosophy and Theory*, 53(12). pp. 1229–1245. [doi: 10.1080/00131857.2020.1728732](https://doi.org/10.1080/00131857.2020.1728732)
- Cotton, D. R.–Cotton, P. A.–Shipway, J. R. (2023) Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, pp. 1-12. [doi: 10.1080/14703297.2023.2190148](https://doi.org/10.1080/14703297.2023.2190148)
- Cox, C.–Tzoc, E. (2023) ChatGPT: Implications for academic libraries. *College & Research Libraries News*, 84(3). pp. 99. [doi: 10.5860/crln.84.3.99](https://doi.org/10.5860/crln.84.3.99)
- Dai, H.–Liu, Z.–Liao, W.–Huang, X.–Wu, Z.–Zhao, L.–Liu, W.–Liu, N.–Li, S.–Zhu, D. (2023) *ChatAug: Leveraging ChatGPT for Text Data Augmentation*. <https://www.scribd.com/document/695626214/2302-13007>
- Dehouche, N. (2021) Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3). *Ethics in Science and Environmental Politics*, 21. pp. 17-23.
- de Zarzà, I.–de Curtò, J.–Roig, G.–Calafate, C. T. (2023) LLM Adaptive PID Control for B5G Truck Platooning Systems. *Sensors*, 23(13). pp. 5899. [doi: 10.3390/s23135899](https://doi.org/10.3390/s23135899)
- Dwivedi, Y. K.–Kshetri, N.–Hughes, L.–Slade, E. L.–Jeyaraj, A.–Kar, A. K.–Baabdullah, A. M.–Koochang, A.–Raghavan, V.–Ahuja, M.–Albanna, H. (2023) “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71. 102642. [doi: 10.1016/j.ijinfomgt.2023.102642](https://doi.org/10.1016/j.ijinfomgt.2023.102642)
- Elkins, K.–Chun, J. (2020) Can GPT-3 pass a writer’s Turing test? *Journal of Cultural Analytics*, 5(2). [doi: 10.22148/001c.17212](https://doi.org/10.22148/001c.17212)

- Firaina, R.–Sulisworo, D. (2023) Exploring the usage of ChatGPT in higher education: Frequency and impact on productivity. *Buletin Edukasi Indonesia*, 2(01) pp. 39–46. doi: [10.56741/bei.v2i01.310](https://doi.org/10.56741/bei.v2i01.310)
- Garg, R. K.–Urs, V. L.–Agrawal, A. A.–Chaudhary, S. K.–Paliwal, V.–Kar, S. K. (2023) Exploring the Role of Chat GPT in patient care (diagnosis and Treatment) and medical research: A Systematic Review. *Health Promot Perspect*, 13(3). pp. 183–191. Published online 2023 Sep 11. doi: [10.34172/hpp.2023.22](https://doi.org/10.34172/hpp.2023.22)
- Glaser, N. (2023) Exploring the Potential of ChatGPT as an Educational Technology: An Emerging Technology Report. *Technology, Knowledge and Learning*, pp. 1–8.
- Ghourabi, A.–Alohaly, M. (2023) Enhancing Spam Message Classification and Detection Using Transformer-Based Embedding and Ensemble Learning. *Sensors*, 23(8):3861. doi: [10.3390/s23083861](https://doi.org/10.3390/s23083861)
- Gozaló-Brizuela, R.–Garrido-Merchán, E. C. (2023) A survey of Generative AI Applications. arXiv, preprint arXiv:2306.02781. doi: [10.48550/arXiv.2306.02781](https://doi.org/10.48550/arXiv.2306.02781)
- Gui, J.–Sun, Z.–Wen, Y.–Tao, D.–Ye, J. (2021) A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*. doi: [10.1109/TKDE.2021.3130191h](https://doi.org/10.1109/TKDE.2021.3130191h)
- Gupta, B.–Mufti, T.–Sohail, S.S.–Madsen, D.Ø. (2023) ChatGPT: A Brief Narrative Review. Preprints, 2023040158. doi: [10.20944/preprints202304.0158.v1](https://doi.org/10.20944/preprints202304.0158.v1)
- Hadi, M. U.–Qureshi, R.–Shah, A.–Irfan, M.–Zafar, A.–Shaikh, M. B.–Mirjalili, S. (2023) A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage. TechRxiv. doi: [10.36227/techrxiv.23589741.v1](https://doi.org/10.36227/techrxiv.23589741.v1)
- Haque, M.U.–Dharmadasa, I.–Sworna, Z.T.–Rajapakse, R.N.–Ahmad, H. (2022) „I think this is the most disruptive technology”: Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. arXiv, preprint arXiv:2212.05856. doi: [10.48550/arXiv.2212.05856](https://doi.org/10.48550/arXiv.2212.05856)
- Hobert, S. (2019) How are you, chatbot? Evaluating chatbots in educational settings – Results of a Literature Review. In *Gesellschaft für Informatik e.V. (Ed.), DELFI 2019*. pp. 259–270. Bonn. doi: [10.18420/delfi2019_289](https://doi.org/10.18420/delfi2019_289)
- Hong, W. C. H. (2023) The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1). pp. 37–45. <https://jeti.thewsu.org/index.php/ciet/article/view/103>
- Hosseini, M.–Horbach, S.P. (2023) Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other Large Language Models. doi: [10.21203/rs.3.rs-2587766/v1](https://doi.org/10.21203/rs.3.rs-2587766/v1)
- Kasneci, E.–Sessler, K.–Küchemann, S.–Bannert, M.–Dementieva, D.–Fischer, F.–Gasser, U.–Groh, G.–Günemann, S.–Hüllermeier, E.–Krusche, S.–Kutyniok, G.–Michaeli, T.–Nerdel, C.–Pfeffer, J.–Poquet, O.–Sailer, M.–Schmidt, A.–Seidel, T.–Stadler, M.–Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education, *Learning and Individual Differences*, Volume 103. 102274, ISSN 1041-6080. doi: [10.1016/j.lindif.2023.102274](https://doi.org/10.1016/j.lindif.2023.102274)
- Khalil, M.–Er, E. (2023) Will ChatGPT get you caught? Rethinking of plagiarism detection. <https://arxiv.org/abs/2302.04335> [Letöltve: 2023.03.15.].
- Kilinc, S. (2023) Embracing the future of distance science education: Opportunities and challenges of ChatGPT integration. *Asian Journal of Distance Education*, 18(1). pp. 205–237. <http://www.asianjde.com/ojs/index.php/AsianJDE/article/view/721>
- Kocoń, J.–Cichecki, I.–Kaszyca, O.–Kochanek, M.–Szydło, D.–Baran, J.–Bielaniewicz, J.–Gruza, M.–Janz, A.–Kanclerz, K. (2023) ChatGPT: Jack of all trades, master of none. arXiv preprint arXiv:2302.10724. doi: [10.48550/arXiv.2302.10724](https://doi.org/10.48550/arXiv.2302.10724)
- Kuhail, M. A.–Mathew, S. S.–Khalil, A.–Berengueres, J.–Shah, S. J. (2023) „Will I Be Replaced?” Assessing Chatgpt's Effect on Software Development and Programmer Perceptions of Ai Tools. SSRN. <https://ssrn.com/abstract=4571482> vagy doi: [10.2139/ssrn.4571482](https://doi.org/10.2139/ssrn.4571482)
- Lambert, J.–Stevens, M. (2023) ChatGPT and Generative AI Technology: A Mixed Bag of Concerns and New Opportunities. *Computers in the Schools*, pp. 1–25.
- Li, L.–Ma, Z.–Fan, L.–Lee, S.–Yu, H.–Hemphill, L. (2023) ChatGPT in education: A discourse analysis of worries and concerns on social media. arXiv preprint arXiv:2305.02201.
- Lo, C. K. (2023) What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4). pp. 410. doi: [10.3390/educsci13040410](https://doi.org/10.3390/educsci13040410)
- Lund, B. D.–Wang, T. (2023) Chatting about ChatGPT: How may AI and GPT impact academia and libraries? Library Hi Tech News. SSRN. <https://ssrn.com/abstract=4333415> vagy doi: [10.2139/ssrn.4333415](https://doi.org/10.2139/ssrn.4333415)
- Maddigan, P.–Susnjak, T. (2023) Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, Codex and GPT-3 Large Language Models.

- IEEE, vol. 11. pp. 45181-45193.
doi: [10.1109/ACCESS.2023.3274199](https://doi.org/10.1109/ACCESS.2023.3274199)
- Marino, M. T.–Vasquez, E.–Dieker, L.–Basham, J.–Blackorby, J. (2023) The Future of Artificial Intelligence in Special Education Technology. *Journal of Special Education Technology*, 01626434231165977.
- Mengist, W.–Soromessa, T.–Legese, G. (2020) Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, 7. 100777.
doi: [10.37074/jalt.2023.6.1.29](https://doi.org/10.37074/jalt.2023.6.1.29)
- Mijwil, M. M.–Hiran, K. K.–Doshi, R.–Dadhich, M.–Al-Mistarehi, A. H.–Bala, I. (2023) ChatGPT and the future of academic integrity in the artificial intelligence era: A new frontier. *Al-Salam Journal for Engineering and Technology*, 2(2). pp. 116-127.
doi: [10.55145/ajest.2023.02.02.015](https://doi.org/10.55145/ajest.2023.02.02.015)
- Mohamadi, S., Mujtaba, G., Le, N., Doretto, G., & Adjeroh, D. A. (2023). *ChatGPT in the Age of Generative AI and Large Language Models: A Concise Survey*. arXiv preprint arXiv:2307.04251.
- Moqbel, M. S. S.– Al-Kadi, A. M. T. (2023) Foreign language learning assessment in the age of ChatGPT: A theoretical account. *Journal of English Studies in Arabia Felix*, 2(1). pp. 71-84. doi: [10.56540/jesaf.v2i1.62](https://doi.org/10.56540/jesaf.v2i1.62)
- Ofosu-Ampong, K.–Acheampong, B.–Kevor, M. O. (2023) Acceptance of Artificial Intelligence (ChatGPT) in Education: Trust, Innovativeness and Psychological Need of Students. *Information and Knowledge Management*, 13(4). pp. 37-47.
- Pardos, Z.A.–Bhandari, S. (2023) Learning gain differences between ChatGPT and human tutor generated algebra hints. arXiv preprint arXiv:2302.06871.
- Pericles, R. (2023) Artificial intelligence in teaching and learning: what questions should we ask of ChatGPT? *Interactive Learning Environments*, 31:1. pp. 1-3.
doi: [10.1080/10494820.2023.2180191](https://doi.org/10.1080/10494820.2023.2180191)
- Prieto, S. A.–Mengiste, E. T.–García de Soto, B. (2023) Investigating the Use of ChatGPT for the Scheduling of Construction Projects. *Buildings*, 13(4). pp. 857.
doi: [10.3390/buildings13040857](https://doi.org/10.3390/buildings13040857)
- Pinto, G.–Cardoso-Pereira, I.–Monteiro Ribeiro, D.–Lucena, D.–de Souza, A.–Gama, K. (2023) Large Language Models for Education: Grading Open-Ended Questions Using ChatGPT. *SBES EDU Track*.
doi: [10.48550/arXiv.2307.16696](https://doi.org/10.48550/arXiv.2307.16696)
- Qureshi, B. (2023) Exploring the use of chatgpt as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. arXiv preprint arXiv:2304.11214.
doi: [10.48550/arXiv.2304.11214](https://doi.org/10.48550/arXiv.2304.11214)
- Rahman M.M.–Watanobe Y. (2023) ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*, 13(9):5783.
doi: [10.3390/app13095783](https://doi.org/10.3390/app13095783)
- Rasul, T.–Nair, S.–Kalendra, D.–Robin, M.–de Oliveira Santini, F.–Ladeira, W. J.–Sun, M.–Day, I.–Rather, R. A.–Heathcote, L. (2023) The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning & Teaching*, 6(1). pp. 41-56.
doi: [10.37074/jalt.2023.6.1.29](https://doi.org/10.37074/jalt.2023.6.1.29)
- Ray, P. P. (2023) ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3. pp. 121-154.
doi: [10.1016/j.iotcps.2023.04.003](https://doi.org/10.1016/j.iotcps.2023.04.003)
- Rudolph, J.–Tan, S.–Tan, S. (2023) ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, 6(1). pp. 342-363.
doi: [10.37074/jalt.2023.6.1.9](https://doi.org/10.37074/jalt.2023.6.1.9)
- Singh, H.–Singh, A. (2023) ChatGPT: Systematic Review, Applications, and Agenda for Multidisciplinary Research. *Journal of Chinese Economic and Business Studies*, 21(2). pp. 193-212.
- Shahriar, S.–Kadhim H. (2023) Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations. arXiv preprint arXiv:2302.13817.
- Su, J.–Yang, W. (2023) Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Review of Education*, 6(3). pp. 355-366. doi: [10.1177/20965311231168423](https://doi.org/10.1177/20965311231168423)
- Sullivan, M.–Kelly, A.–McLaughlan, P. (2023) ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1). pp. 31-40.
doi: [10.37074/jalt.2023.6.1.17](https://doi.org/10.37074/jalt.2023.6.1.17)
- Sun, G. H.–Hoelscher, S. H. (2023) The ChatGPT storm and what faculty can do. *Nurse Educator*, 48(3). pp. 119-124. doi: [10.1097/NNE.0000000000001390](https://doi.org/10.1097/NNE.0000000000001390)
- Susnjak, T. (2022) *ChatGPT: The end of online exam integrity?* arXiv preprint arXiv:2212.09292.
doi: [10.48550/arXiv.2212.09292](https://doi.org/10.48550/arXiv.2212.09292)
- Tempus Közalapítvány (2022)
<https://tka.hu/nemzetkozi/9518/a-felsooktatas-rendszer-magyarorszagon> [Étőlte: 2023.08.20.].
- Teubner, T.–Flath, C.M.–Weinhardt, C. (2023) Welcome to the Era of ChatGPT et al. *Bus Inf Syst Eng*, 65. pp. 95-101. doi: [10.1007/s12599-023-00795-x](https://doi.org/10.1007/s12599-023-00795-x)
- Tlili, A.–Shehata, B.–Adarkwah, M. A.–Bozkurt, A.–Hickey, D. T.–Huang, R.–Agyemang, B. (2023) What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1). pp- 15. doi: [10.1186/s40561-023-00237-x](https://doi.org/10.1186/s40561-023-00237-x)

- Vrontis, D.–Chaudhuri, R.–Chatterjee, S. (2023) Role of ChatGPT and Skilled Workers for Business Sustainability: Leadership Motivation as the Moderator. *Sustainability*, 15(16):12196.
[doi: 10.3390/su151612196](https://doi.org/10.3390/su151612196)
- Wang, J.–Hu, X.–Hou, W.–Chen, H.–Zheng, R.–Wang, Y.–Yang, L.–Huang, H.–Ye, W.–Geng, X. (2023) *On the Robustness of ChatGPT: An Adversarial and Out-of-distribution Perspective*. arXiv preprint arXiv:2302.12095.
<https://arxiv.org/pdf/2302.12095.pdf>
- Wang, W.–Chen, Y.–Heffernan, N. (2020) *A generative model-based tutoring system for math word problems*. arXiv preprint arXiv:2010.04.
- Webster, J.–Watson, R. T. (2002) Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2). xiii–xxiii.
https://web.njit.edu/~egan/Writing_A_Literature_Review.pdf
- Xiao, Y.–Watson, M. (2019) Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education & Research*, 39(1). pp. 93–112.
[doi: 10.1177/0739456X17723971](https://doi.org/10.1177/0739456X17723971)
- Yong, Z. (2023) ChatGPT for Teaching and Learning: An Experience from Data Science Education. In *The 24rd Annual Conference on Information Technology Education (SIGITE '23)*, October 11–14, 2023, Marietta, GA, USA. ACM, New York, NY, USA, pp. 66–72.
[doi: 10.1145/3585059.3611431](https://doi.org/10.1145/3585059.3611431)
- Zhai, X. (2022) ChatGPT User experience: Implications for education.
<https://ssrn.com/abstract=4312418> vagy [doi: 10.2139/ssrn.4312418](https://doi.org/10.2139/ssrn.4312418) [Letöltve: 2022.12.27.].

ChatGPT és más nagy nyelvi modellek alkalmazásának lehetőségei és kiberbiztonsági kérdései

Use cases & cybersecurity aspects of ChatGPT and other Large Language Models

DOI: [HTTPS:// DOI.ORG/10.53793/RV.2024.2.2](https://doi.org/10.53793/RV.2024.2.2)

Absztrakt

A mesterséges intelligenciát (MI-t) sokan a GPT-vel azonosítják: egy test nélkül élő intelligenciával, ami mindenhez ért. Holott rengeteg MI-t használunk évek óta, és fel sem tűnt sokaknak. Hogy milyen MI típusok vannak, mire használhatóak, megbízhatóak-e és hogy tényleg gondolkodnak-e, fontos tudni. Mivel az MI területei robbanásszerűen fejlődnek, ezért előfordulhat, hogy a ma még aktuális ismeret holnap már elavult lesz. Fontos tehát, hogy naprakészen tartsuk tudásunkat és utánanézzünk dolgoknak. Az MI-kre szükség lesz a jövőben, ezért tudnunk kell mi történik a megadott adatokkal, milyen kiberbiztonsági és adatvédelmi problémák várhatóak, illetve, hogy felhasználható-e egy generált mű jogi felelősség nélkül, valamint van-e jogi szabályozás.

KULCSSZAVAK: #MESTERSÉGESINTELLIGENCIA, #MI, #KIBERVÉDELEM, #GPT, #LLM

Abstract

A lot of people think: Artificial Intelligence is GPT, an intelligence living without body and knows all. Even if all of us have been using a lot of AI solutions for years around, it's not noticed. So, what types of AI exist, and what they are used for, we can trust them or not, and if they are really thinking important to know. Because areas of AI are rapidly evolving. It can be, the actual knowledge today will be outdated tomorrow. It's very important to keep our knowledge up-to-date and always follow up. Knowing AI will be mandatory in the future so it's important to know what will happen to data given by us, what kind of cybersecurity and data protection problems are expected, if a generated artifact can be used without legal accountability and if there is any legal regulation.

KEYWORDS: #ARTIFICIALINTELLIGENCE, #AI, #CYBERSECURITY, #GPT, #LLM

Bevezető

Amikor mesterséges intelligenciáról beszélünk (MI, angolul: artificial intelligence, röviden AI) akkor rájövünk, hogy sok a digitális analfabéta. Ez kompetencia hiány, ami miatt egy egyén nem tudja használni és kihasználni a digitális eszközök nyújtotta lehetőségeket. Például ilyen az e-mailek és dokumentumok kinyomtatása olvasás céljából, pedig azt okoseszközön is el lehet olvasni. A digitálisan hitelesített dokumentumokat pedig elektronikusan kell tárolni, mivel mind a dokumentum, mind az aláírás digitális, ezért csak úgy hiteles. Kinyomtatva már nem az. A nyomtatás pedig nem környezetbarát és költséges is egyben. Ezért is fontos a Digitális Technológiát (DT) és az MI-t oktatni.

Gondolatok a mesterséges intelligenciáról

Sok a misztikum és torz a kép az MI-ről a hype, a kattintásvadászat, a trollkodás, az anyagi haszonszerzés stb. miatt. A szövelemzés sem segít, ha MI-ről beszélünk. Az igaz, hogy az MI egy új svájci bicska, de gyakran elferdítik társadalmi hatásait és hasznosságát. Pedig az MI használata kockázattal is jár.

Tudni kell, hogy az MI-k még nem gondolkodnak. Nem adnak feladatot senkinek és maguktól sem végeznek tevékenységet. Feltérképezetlen területekkel mit sem tudnak kezdeni, és problémákat sem igazán oldanak meg, csak feladatokat. A valóságot nem ismerik. Különböző célokra más és más MI-t kell használni, de amit csinálnak abban jók és akár emberek helyett, önállóan, gyorsabban, olcsóbban teszik. Ez még nem a szuperintelligencia, ami okosabb az embernél. Még nem

tartunk ott. és ki tudja, hogy az jó lesz-e majd nekünk. Van MI, ami új receptet állít elő, van, ami képeket alkot, de ez nem igazi kreativitás, hanem kombinatorika és minták követése. Az MI-nek ugyanis se keze, se lába, se szája és nem érez ízeket se, tehát nem tudja, hogy amit csinál jó-e, és mi az. Az LLM/GPT csak feldolgozza a promptot (nem fordítjuk. Kb. „kivált valamit”, „készítet”, de „kérés”-ként jobban érthető) betanított modellek és algoritmusok segítségével. Minél több a betanított adat, annál jobban sejtí az MI milyen összefüggések vannak, és annál változatosabb választ kaphatunk. Ezért jöhet létre új recept (ami vagy jó, vagy nem).

A magánszemélyek, a vállalkozások, de még az állami szereplők esetében is az MI használata elkerülhetetlen. A GPT pedig a hagyományos MI-vel szemben több MI területet fog össze, ezért segíthet megújulni, fenntarthatóan fejlődni, de vannak kockázatai is.

Mesterséges intelligenciát érintő alapfogalmak

Adatbányászat (DM - Data Mining)

Az adatbányászat hasznosnak vélt adatok, azaz információk nagy adatbázisokból vagy naplóállományokból történő részben, vagy teljesen automatizált kinyerése.

Big Data (Nagy adat, nem szokás magyarra lefordítani)

Elsősorban adattárházaknál (Data Warehouse) és adatbányászatban (Data Mining) használatos, főként döntéselőkészítő jelentések készítéséhez (reporting), de az MI-nél is használható tanításhoz és előrejelzéshez. A Big Data forrása lehet központi adatbázis is. Ez a Data lake (nem fordítjuk, jelentése: „Adat tó”). Az eltérő lokációkról származó adatváltozások akár *valós időben* (real-time) ide szinkronizálódhatnak. Ez teszi lehetővé a közel valós idejű kiértékelést és a gyors reagálást a vezetőség számára. A big data 3 fő jellemzője a 3V (Volume: hatalmas adatmennyiség, Velocity: sebesség, azaz, gyors feldolgozás és Variety: változatosság).

Kapcsolódó foglalkozások: Data Analyst, Data Engineer, Data Scientist, Big Data Specialist.

Mesterséges Intelligencia (Artificial Intelligence) szintjei

MI minden olyan szoftveres megoldás, ami valamilyen emberi viselkedést, tevékenységet utánoz. Ide tartozik a szoftveres bot is (robot lerövidülve), amit

automatizálásra használnak, például tartalom letöltésére, frissítésére stb. Persze ahhoz, hogy valamit MI-nek hívjunk, annak feltétele az is, hogy mutasson némi intelligenciát. Például egy letöltő botot ne banoljon az IPS/IDS, de a szerver se váljon elérhetetlenné miatta. Egy MI kimenetét a bemenő adat és az MI belső állapota is befolyásolja. Így általában nem determinisztikus (nincs előre meghatározva) a kimenet.

Gépi tanulás (Machine learning)

A gépi tanulás az MI magasabb szintje. Itt már nem kódolásról beszélünk, hanem tanításról. Részei:

- *Modellek és algoritmusok*
Előtanított modellek általában ingyen és nagy számban beszerezhetőek. Finomhangolásra gyakran szükség van. Ezek végzik számunkra a munkát.
- *Adatok*
Modellfüggetlenek, célszerű saját mintát gyűjteni, mivel egy előtanított modell valószínűleg nem fedi le igényünket. A modelleket finomhangolni szükséges.

Többféle gépi tanulás létezik például szabályalapú gépi tanulás (RBML - rule-based machine learning) és neurális hálózat.

Szakértői rendszer (Expert system)

Az 1970-es években jött létre az első ilyen működő MI. Ez egy számítógépes rendszer, ami egy szakértő döntéshozó képességét emulálja. Egy szakterületi feladat által megkövetelt tudást és analitikus képességet valósít meg. Két alrendszere van: a következtető gép („motor”) és a tudásbázis (szabályok és tények). Kódolás helyett szabályalapú (rule-based), azaz if-then (ha-akkor) logikát követ, és célja komplett problémák megoldása meglévő ismeretek alapján. Új tények is bekerülhetnek a tudásbázisba szakértői input segítségével, ha a gép elakadna. A következtető gép a szabályon tények figyelembevételével végighalad és eredményt ad. Ehhez szöveges magyarázatot társíthat vagy az alkalmazott szabályokat bemutatja. A hibakeresés is lehet a rendszer része. A mai szakértői rendszerek már deep learning alapon is hozhatnak döntést. Szabályalapú megoldás lehet a vérkép kiértékelése, deep learning alapú lehet a röntgenképek kiértékelése. A szakértői rendszerek az üzleti folyamat automatizálásának részei.

Deep Learning (Mélytanulás, ritkán fordítják magyarra)

A gépi tanulás legnépszerűbb megoldása. A Deep Learning (DL) elnevezés arra utal, hogy mesterséges neurális hálózatot (ANN - Artificial Neural Network) – továbbiakban csak neurális hálózatot – tanítunk. A tanítás órától akár hetekig is eltarthat. Amíg az MI nem működik elfogadható pontossággal, addig a folyamatot meg kell ismételni finomhangolások mellett. Ehhez van egy virtuális fekete dobozunk bemenettel és kimenettel. A fekete doboz pedig a neurális hálózat, ami csinál valamit a benne lévő sejtek segítségével, amik rejtett rétegeket alkotnak. Egy sejt sok másik sejthez kapcsolódhat másik rejtett rétegben. A sejtek „aktivitása” paraméterek segítségével szabályozható. A neurális hálózatot nem szabad túltanítani (a magolás szó jó hasonlat), sem alul, mivel egyik esetben sem fog megfelelően működni, azaz nem fog „gondolkodni”. A cél az lenne, hogy az MI mindig jó választ adjon. Ezt nehéz elérni, ezért %-ban szokták megadni a pontosságot (accuracy). A tanításkor, hogy növeljék a megbízhatóságot, a rendelkezésre álló mintákat általában kettéválasztják kb. 70/30 arányban. 70% megy tanításra, 30% tesztelésre. Ez az arány amúgy nincs köbe vésvé és 1000 mintára is szükség lehet. Ha nincs elég, akkor a neurális háló alul lesz tanítva, így szintetikus adatot is gyakran használnak.

A neurális hálózat tanítása és használata erőforrás igényes (CPU, RAM stb.), így költséges is egyben, de nagyon látványosan fejlődő és alkalmazható MI terület.

Az hálózat tanítása és használata CPU-t (Central Processing Unit - központi feldolgozóegység), bizonyos esetekben pedig GPU-t (Graphics Processing Unit - grafikus processzor) is igényel (például képek feldolgozásához, generálásához). Gyártanak NPU-t (Neural Processing Unit - neurális feldolgozó egység) neurális hálózatokhoz. Ezeket az NPU-kat gyakran okostelefonokba rakják.

Amikor MI-ről van szó, gyakran DL-re, de újabban akár GPT-re is gondolnak.

Kapcsolódó foglalkozások: *Data Scientist* (-adattisztítás, adatok előfeldolgozása, modellek), *Data Engineer* (-tanítás és teszt).

MI típusok

Képesség alapján:

- *Hagyományos vagy diszkriminatív MI*
Adatot dolgoz fel, elemez, osztályoz, kivonatol, keres, döntést hoz vagy előrejelzést végez (például osztályozás: mi látható a képen, előrejelzés: milyen bevétel várható). A kimenet lehet szám, címke stb.

- *Generatív MI*

A hagyományos MI-n túlmutat, teljesen új tartalmat képes előállítani egy prompt alapján. GPT-k nem csak szöveget, de akár szintetikus adatot, diagramot, dokumentumot, hangot, zenét, képet, videót stb. képesek előállítani. Nem minden GPT támogat minden funkciót, és gyakran élő előfizetés is szükséges hozzá.

A bemeneti adat-típustól függően:

- *kvantitatív MI (Quantitative AI)*
Nagy mennyiségű numerikus adat feldolgozása és elemzése.
- *kvalitatív MI (Qualitative AI)*
Szöveg feldolgozása és elemzése.

Gépi Tanulás (Machine Learning) típusai

A gépi tanulás típusai az interakció szintje alapján:

- *Supervised learning (felügyelt tanulás)*
Osztályoznunk kell a tanításhoz használt összes mintát és címkét kell hozzárendelni mindegyikhez (classification). A mintákat és a hozzájuk tartozó címkét a tanításhoz megadjuk. Például fotóhoz rendelt címke lehet: ember, kutya stb. (osztályozás, regresszió).
- *Semi-supervised learning (részben felügyelt tanulás)*
Csak a címkék egy részét adjuk meg a tanításhoz használt mintákhoz, azaz nem minden mintához lesz címke előre megadva. A mintákban rejlő struktúrát a modellnek kell felismernie és az információt kinyernie. Általában automatizált tevékenységeknél használatos (osztályozás, regresszió).
- *Unsupervised learning (felügyelet nélküli tanulás)*
A generatív MI ide tartozik, de a hagyományos MI is tanítható felügyelet nélkül, ha Big Data áll rendelkezésünkre. Nem csak előnyei, de korlátai és kockázatai is vannak, mert a gép tanítja önmagát (klaszterezés, topik modellezés).
- *Reinforcement learning (megerősítéses tanulás)* (RLHF: Reinforcement Learning with Human Feedback)

A felhasználó visszajelzést küldhet. Például kiválaszthatja a legjobbat a generált válaszok közül vagy jelezheti, hogy hiányos, hibás vagy elavult a válasz. Ez segít a finomhangolásban. Hátránya, hogy vissza lehet élni vele, ha automatizmus van a háttérben. (Kapcsolódó fogalmak: Környezet a megoldandó probléma, ügynök a tanuló algoritmus)

Neurális hálózatot használó megoldások

Prediction, Predictive analytics (Előrejelzés)

Gyakran használnak MI-t például értékesítés előrejelzéséhez (regression analysis, regresszióelemzés), mert az eredményt fogyasztói magatartás befolyásolásra is fel lehet használni. Historikus adatokból (legalább 3-5 év) indulnak ki a pontosabb előrejelzés érdekében. Akkor jó a választott modell, ha a nem egyértelmű mintázat ellenére az MI viszonylag pontos előrejelzést tud adni. A mintát 3 részre is szedhetik (70%-20%-10%), hogy a jóslás pontosságáról megbizonyosodhassanak.

Cognitive computing (kognitív számítástechnika)

A mesterséges neurális hálózatok felépítését az emberi agyban lévő neurális hálózat ihlette, és gyakran az érzékelés és észlelés folyamataival foglalkoznak. Fontos, hogy a detection (észlelés) és a recognition (felismerés) nem ugyanaz (ahogy arcot észlelni és felismerni egy képen is mást jelent)!

- *Computer vision (gépi látás)*
Gépi látással például objektumokat vagy élőlényeket keresünk képen, videón stb., de a mozgás észlelése és követése is idetartozik. Ehhez konvolúciós neurális hálózat (CNN: Convolutional neural network) szükséges. Működése leegyszerűsítve: egy digitális kép több eltérő méretű rejtett rétegre „vetül”, így rétegenként eltérő méretű szegmensekre bontódik a kép, azaz képszegmentálás történik (image segmentation) és a sejtek így „látanak”. A rétegek közötti kapcsolatot a sejtek kapcsolatai biztosítják. A CNN nem tévesztendő össze a Haar kaszkáddal, ami nem használ neurális hálózatot, de a gépi látás egy másik megoldása, mivel gyorsabb, kevesebb erőforrást igényel, de pontatlanabb.
- *Image recognition (képfelismerés)*
A képfelismerés a gépi látás része. Címkézésre használjuk, mint például arc- (facial recognition), optikai szöveg- (OCR - Optical Character Recognition), hely-, tárgy-, kézírásfelismerésre, hitelesség vizsgálatra stb.
- *Voice recognition (hangfelismerés)*
Célja annak eldöntése, hogy ki beszél.
- *Speech recognition (beszéd felismerés)*
Szöveges átirat készítésére használják, Speech-to-Text. Használatához általában Internet kapcsolat kell, ezért fontos az adatvédelem is. Lehet, hogy mi használjuk a beszéd felismerést, de a környezetünk beszélgetése is továbbítódik használatakor!

- *Sentiment analysis (szentimentelemzés)*
Szöveg elemzése, hogy annak hangulatából kiderüljön az érzelmi töltet: pozitív, semleges vagy negatív.
- *Natural Language processing (NLP, természetes nyelvek feldolgozása)*
Segítségével helyesírást lehet ellenőrizni, fordítani, chatelni. Az NLP tanításához chateket, szövegeket, dokumentumokat használnak a bennük lévő összefüggések felismerésére (például szórend). Ez lehetővé teszi, hogy nem strukturált adatokból majd strukturált adatokat nyerjünk, és rendezve, kategorizálva tudjuk tárolni későbbi felhasználás céljából. Az NLP a beszélt szöveg átiratához (Speech-to-Text, például Amazon Lex) vagy szövegfelolvasáshoz (Text-to-Speech, például Amazon Polly) API-kat használhat.
- *Machine Translation (gépi fordítás)*
Gépi fordításon neurális hálózat által végzett fordítást kell érteni. NLP alapú. Fordítani nem csak szöveget, de dokumentumot vagy akár médiát (kép, videó stb.) is lehet. A rosette-i kő szerű mondatról mondatra történő tükröfordítás és a szótár alapú fordítás nem használ neurális hálózatot. Problémaforrás lehet ezeknél a hiányzó fordítás, a szórend, az összetett vagy több jelentésű szavak és az idiómák. Persze a neurális hálóval történő fordításokban is lehet hiba. Ha fordítási problémák érdekelnek valakit, akkor igazi csemege a Yamada DVD-játékos (DVD-lejátszó) és az „ÁTLAGOS TÁVOLI URAL” (univerzális távirányító) használati útmutatói. A fordításhoz megfelelő mennyiségű és minőségi mintával kell tanítani a neurális hálót.
- *Large language models (LLMs – nagy nyelvi modellek)*

Az LLM alapvetően egy chatbot és a chat másik felén nem ember, hanem gép található. NLP-re épül. Segítségével a gép és az ember emberi nyelven kommunikálhat. Az LLM-et nem hívjuk generatív MI-nek, mert csak szöveget generál. Az LLM nagy mennyiségű adattal tanított modell. Előnye, hogy finomhangolható. Van egy vektor adatbázisa, ami a tanítás során jön létre. A szöveget tokenekre bontják és azt eltárolják benne. A token ritkán szó, mert azt nehéz lehet használni (például, ha hiányzik egy ragozott forma a feldolgozás problémába ütközhet). A token inkább kisebb elem: szótó, szótag stb. (például „aj-tó-k”), amivel könnyű dolgozni. Felismerhető az „aj-tó”, „aj-tó-t”, de még a „be-já-ra-ti aj-tó-nak” közti kapcsolat is. A betanított szöveg így már könnyen kereshető és

új tartalmat hozhatunk létre statisztikai módszerek segítségével. Egy szónál ugyanis előre megjósolható, hogy mi állhat előtte és utána. A „kék” után jöhet az „ég”, de ritkán a „skatulya”. Így lehetővé válik változatos, mégis érthető szövegek generálása.

Autonóm vezetés, önvezetés

Az autonóm vezetés vagy önvezetés valójában vezetést támogató rendszer, mivel teljesen önvezető autó még nincs. Ha ugyanis a jármű nem teljesen önvezető, akkor a sofőr és nem a jármű gyártója a felelős a döntésekért, cselekedetekért vagy éppen nem cselekvésért. A vezetést támogató rendszerek különböző szintjei (SAE J3016):

- 0-s szint (No Driving Automation – nincs vezetésautomatizálás)
Hagyományos módon kézzel kormányzott járművek. A legtöbb jármű még ilyen. A különböző segédrendszerektől (kipörgésgátló, vészfékező rendszer stb.) még nem lesz 1-es szinten a jármű, mivel nem automatizálják a vezetést.
- 1-es szint (Driver Assistance – vezetői asszisztens)
Legalacsonyabb szintű automatizálás. Vezetéstámogatás van benne, például menetsebesség tartás, más néven adaptív tempomat (cruise control) VAGY sávtartás. A vezető végzi a kormányzást, a fékezést és figyeli az utat.
- 2-es szint (Partial Driving Automation – részleges vezetésautomatizálást)
Fejlett vezető támogatási rendszer (ADAS: Advanced Driver Assistance Systems). A jármű egyszerre tud sebességet szabályozni (tempomat) ÉS kormányozni (sávtartás). Ez már rövid ideig tartó önvezetés, mert a vezető interakciója nélkül is haladhat a jármű. A sofőr a vezetésbe be tud és kell, hogy avatkozzon.
- 3-as szint (Conditional Driving Automation – feltételes vezetésautomatizálás)
2-eshez képest nincs nagy különbség a vezető részéről. Fejlettebb a környezet detektálás, így a jármű okosabban dönt. Tájékoztatja a sofőrt a fontosabb dolgokról. Kevesebb interakcióra van szükség a vezető részéről. A sofőr a vezetésbe be tud és kell, hogy avatkozzon.
- 4-es szint (High Driving Automation – magas szintű automatizálás)
A jármű tud reagálni hirtelen forgalmi helyzetváltozásokra, balesetekre is. Ha még van kormány, fék és gázpedál, akkor az ember is

beavatkozhat a vezetésbe. Ezek a járművek önvezetőnek tekinthetők, de jogi korlátozások miatt csak bizonyos feltételek mellett, például sebességkorlátozás és kijelölt területen belül közlekedhetnek. Ez utóbbi a geofencing (földrajzi határ). Az USA-ban néhány személyszállító ilyen.

- 5-ös szint (Autopilot vagy Full Driving Automation – teljes önvezetés)

A járműnek nem lesz se kormánykereke, se féke, se gázpedálja, mert nem lesz rá szükség. Ha bármelyiket megtalálnánk, akkor gyanakodhatnánk, hogy nincs szó teljes önvezetésről. A célbajutást a navigációs rendszer segíti majd. Még nincs ilyen jármű.

A nagy LLM és GPT örület

Chatbot (Conversational AI - Beszélgető robot)

Az LLM lényegében ez. A cél az volt, hogy a gép emberi nyelven kommunikáljon az emberrel, és ezt senki ne vegye észre (menjen át a Turing-teszten). Ehhez a beszélgetés kontextusát a chatbotnak megfelelően kell kezelnie, így rövid távú memóriával is kell rendelkeznie (LSTM - Long short-term memory). Képessége a „kérdezz-felelektől” a „rendes társalgásig” terjedhet.

A Chatbot tanításához valódi chateket szoktak használni, hogy minél emberibb legyen a társalgás. Ügyelni kell arra, hogy a tartalmak ne legyenek előítéletesek, sértőek és oldják meg a problémát, de ne fedjenek fel bizalmas információt!

Generatív MI (Generative AI)

„Miért hihetetlenül okos az MI és mégis megdöbbenően buta” (Yejin Choi)

Szemben a hagyományos MI-vel, ami elemző, kiértékelő, előrejelző vagy beszélgető robot, ez a technológia már képes új és testreszabott tartalmat létrehozni, például dokumentumot, képet (diffusion model – diffúziós modell segítségével), videót, hangot, zenét stb.

GPT (Generative Pre-trained Transformer): Az LLM továbbgondolása. Ez már nem csak chatbot, hanem „A chatbot”. Tartalomgenerálást is tud végezni. A modellen, tanításon, finomhangoláson, statisztikai módszeren, belső állapoton és a prompton múlik, hogy hogyan változik idővel egy promptra adott válasz. Az egyedi válasz gyakori (nem determinisztikus), de nem garantált. Ugyanazt a választ kaphatjuk többször is és más is. A GPT-k is annyit tudnak, mint amennyit

megtanítottak nekik, bár néhány GPT API-t (alkalmazásprogramozási interfész, Application Programming Interface) is tud hívni. Hogy a válasz felhasználható-e, illetve, hogy más nem használta fel azt korábban, ellenőrizni kell (például egy diák beadandó dolgozata). A legismertebb változata a ChatGPT,

Az LLM-ek, a GPT-k nagy része csak lexikális tudással rendelkezik, így nem mindegyik tud számolni és általában könnyű összezavarni őket felesleges elemekkel: „ha van két almám és 3 körtém, kapok még 1 almát akkor hány körtém van?”

Kapcsolódó foglalkozás: *Prompt Engineer*.

A GPT működése leegyszerűsítve

Prompt -> szűrés -> adatbázis(ok) -> tartalom generálás vagy API hívás -> RLHF (Reinforcement Learning with Human Feedback: felhasználói visszajelzés).

A promptot megkapja a GPT. Mivel korlátozások állhatnak fenn a válaszadással kapcsolatban, ezért a kérést megvizsgálja. Ha a szűrőn fennakad, akkor nem fogja a promptot a GPT végrehajtani. Ellenkező esetben a betanított adatok a belső állapot felhasználásával, és ha van hozzácsatolt külön adatbázis (például LocalDocs Collections - helyi dokumentum gyűjtemény), annak segítségével választ (nem feltétlen szöveget) generál, amit elküld a felhasználónak. Ha a prompt egy API hívás, akkor az API-nak megfelelően fog a történet folytatódni. Opt-in esetében számos MI lehetővé teszi, hogy visszajelzést küldjünk (RLHF). Ezzel legyünk óvatosak, mert a GPT eltárolhatja adatainkat.

Prompt engineering (Prompt mérnökség)

Olyan feladatkör, ami promptok megtervezését jelenti LLM/GPT esetében. A prompt mérnök megtanulja milyen előnyei és hátrányai vannak az LLM/GPT-nek, hogyan védje az adatokat, hogyan fogalmazza meg a promptot és hogyan ellenőrizze a válasz helyességét.

Hogyan kérdezzünk? (Prompt)

Feltöltésre vagy promptnak szánt tartalmat használat előtt anonimizáljuk. Ne adjunk meg személyes, bizalmas és a szükségesnél több adatot. A Promptnak hosszúsági korlátja van, ezért legyen tömör, precíz. Mindig helyezzük kontextusba a témát, és nyomatékossítsunk vagy használjunk példákat (Few-Shot Prompting), mert „koszorú készítés” virágkötészetben és építkezésen is van, ahogy „sín készítés” vasúti és fogászati tevékenység is lehet.

Hibás, hosszú és agyonformázott dokumentumokat ne használjunk feltöltésre a hallucináció elkerülése végett.

A generált tartalomban hibásak lehetnek az idézetek, linkek, szerzők, hivatkozások stb. Ez nem elfogadható! Fontos a helyállóság, felhasználhatóság ellenőrzése független forrás segítségével is, hogy a hibákat időben még a felhasználás előtt ki lehessen szűrni és korrigálni.

Mielőtt egy promptot beküldünk, kétszer gondoljuk át. Tudni kell hogyan közelítsünk meg egy dolgot több oldalról, hogy a kapott választ ellenőrizzük. Háromféleképpen tegyük fel a kérdést, mert az segíthet a hallucináció felismerésében. NE laikus üljön a gép előtt.

Ha rossz a kérdés, akkor rossz a válasz is! A kérdező tudjon helyesen írni, mivel a következő nyelvtan az MI-n is kifog: „Eladó törzskönyves kóker spánijel kugyát vennék ingyé.” (Forrás: Internet)

A felhasználói visszajelzés lehetősége *organikus felhasználók* számára általában adott. Ezzel befolyásolható a GPT működése hosszabb távon.

GPT Lehetőségek

Multimodalitás

A multimodalitás azt jelenti, hogy az input lehet szöveg, kép, hang stb. Az input típusától pedig eltérő outputot kaphatunk. Például a Midjourney esetében szöveges formában mondjuk meg milyen képet akarunk. Képet is tölthetünk fel szerkesztésre.

Tartalomgenerálás

Lásd: MI típusok, Generatív MI!

Asszisztens (Assistant)

Az asszisztens funkcionál az azt értjük, hogy olyan munkát végeztetünk a GPT-vel, amit egy asszisztenstől várnánk el (levélírás, dokumentum készítés stb.).

Összefoglalás (Summarization)

Gyakran használják a GPT-t szöveges tartalom összefoglalására. A válaszban lévő mondatok számának korlátozása viszont az összegzést torzítja teheti, lényegi dolgok maradhatnak ki. Hibás összefoglalást okozhat az is, ha az MI rosszul értelmezi a szöveget. A legtöbb GPT-nek ma még gondja van a magyar nyelvű szöveg helyes értelmezésével. Ennek oka talán a tanításnál használt kevesebb forrásanyag és a ragozás (azonos alakú szavak keletkezhetnek). Az összefoglalt szöveg ellenőrizendő.

Kódoló asszisztens (Coding assistant)

A kódoló asszisztens kockázatos, mert hibás kódot generálhat és adatokkal (kóddal) fizetünk. Semmi sem

garantálja, hogy a szolgáltató nem teszi pénzzé kódunk, vagy nem jelenik meg a konkurenciánál. Nem lehet vele vállalatirányítási rendszert vagy komolyabb játékokat írni sem. Programkód javításnál van, hogy egy linter (programozás során használt stilisztikai és szintaktikai problémákat elemző eszköz) szintjét sem éri el, mert csak átformázza a kódot vagy jobban elrontja azt. Ezen okok miatt a Stack Overflow betiltotta a ChatGPT által adott válaszok posztolását, mivel több önjelölt „programozó” generált választ küldött be ellenőrzés nélkül: *„...mivel a ChatGPT által adott helyes válaszok előfordulásának átlagos aránya túl alacsony, ezért a ChatGPT-vel létrehozott válaszok beküldése mondhatni veszélyes az oldalra és a felhasználókra, akik a helyes választ kérik és keresik”*

Aki ezt a funkciót használja tapasztalt programozó legyen, és tudja eldönteni, hogy a kapott kód helyes-e, azt csinálja-e, amit kell. Ne cseréljük le a programozót lelkes juniorra, aki kódolóasszisztenst használ, mert hamar megbánjuk.

Szintetikus adat létrehozása (Synthetic data creation)

Szintetikus adat a mesterségesen állított adat, amit tanításhoz, teszteléshez lehet használni. Akkor használjuk, ha nincs elég adatunk. Például családnevek és keresztnemek véletlenszerű párosítása, így az adat nem köthető a való élethez. Ha a valósággal való bárminemű egyezés szempont, akkor érdemes átnézni használat előtt. A jogszabályoknak meg tud felelni, mert mesterséges. Általában automatikusan címkézhető, nem kell kézzel megtennie. Végtelen mennyiségben állítható elő, ami nagy előny. A GPT-k egy része tud szintetikus adatot előállítani.

LLM/GPT problémák

Hallucináció

Az LLM néha „hazudik”. Ez a hallucináció. Felhasználóként viszont helyes választ várunk el. Ha nem tudja a GPT a választ, bár mondaná, hogy: sajnálom, de nem tudom. Ez azonban nem gyakran történik meg. A válasz pedig mindig nagyon meggyőző. Ha a felhasználó laikus, akkor a hallucinációt észre sem veszi.

A Temperature beállítás változtatása befolyásolhatja a hallucinációk előfordulását, de nem ez a fő oka.

Hallucináció oka sok minden lehet, például a modell, tanítási és finomhangolási gondok, pontatlan forrásanyag, nyelvtani és szerkesztési hibák, terjedős tartalom stb.

Ha mi tanítjuk a hálózatot, a felhasznált forrásanyag legyen szabadon felhasználható, jól szerkesztett,

helyesírással, nyelvtanilag, tartalmilag helyes, célratörő stb. A hibákat a tanítás ugyanis nem fogja kijavítani.

Ne akarjuk a világ összes dolgát a neurális hálóknak megtanítani, mert egy „Mit főzünk ma” weboldalon nem kell tudni, hogy mi a világ legmagasabb épülete, de tudni kell, hogy mi a „habarás”. Ezért – a példánál maradva – az általános kommunikáción túl csak a hazai és külföldi konyhakultúra betanítása szükséges. Ezzel csökkenthetjük adatbázisunk méretét, növelhetjük a megbízhatóságot és ellenőrizhetőbb minőségi tudást lehet betanítani vagy újratanítani rövidebb időn belül.

A felelősségről: a New York-i bíróság 2023 májusában egy olyan beadványt kapott két jogásztól, ami ChatGPT-ből származó jogesetek (precedensek) listáját tartalmazta. A ChatGPT a jogforrások hitelességét az ügyvéd kérésére megerősítette. Csakhogy a jogesetek nem voltak valóságok! Egyik következménye az lett, hogy az ügyvédek a félrevezetés miatt 5000 dolláros pénzbírságot fizettek.

Szerzői jog

Egy promptrra adott választ az LLM úgy generál, hogy korábban tanítás során eltárolt tartalmat használ fel, ami ember által írt szövegből származik (tudtával vagy anélkül). Emiatt lesz a válasz olyan, mintha ember írta volna. Ha túl van tanítva az MI, akkor plagizálásra vagy szerzői jogsértésre is képes, mivel szó szerint valaki korábbi beszélgetését vagy írását visszaadhatja. 2023-ban szerzői jogsértés miatt kiadók, szerzők perelték a ChatGPT-t és forgatókönyvírók tüntettek az USA-ban.

Mivel a felhasználó csak tartalmat kér, ezért nem szerző, így szerzői joga sem lesz. Az LLM/GPT promptrra adott válasza sem feltétlen egyedi és nem is ember adja azt, így ilyen jog nem szerezhető. Másrésztől opt-in esetben egy felhasználó által feltöltött tartalomra használati jogot kér az MI. De ha a felhasználó nem rendelkezik megfelelő jogokkal, akkor a jogtulajdonos jogai sérülhetnek. Tehát tanítás és finomhangolás miatt más szellemi tulajdona tükröződhet vissza a válaszokban! A gépen kívül sokan formálnának jogot a generált tartalmakra, de nem mindenki jogosan. Ez tönkretelheti a kreatív írókat. Csoda lenne, ha beperelnék a ChatGPT fejlesztőit azért is, mert más felhasználó is hasonló tartalmat kapott, pedig valaki „levédette” korábbi ChatGPT-s „művét”? Mindig ellenőrizzük a felhasználhatóságot, hogy jogilag és tartalmilag rendben van-e, mert a felhasználás következményei a felhasználót terhelik.

Kibervédelem és adatvédelem

Egyre jobban megbízunk a gyártókban és a fejlesztőkben a kényelem miatt, anélkül, hogy a bizalmat bárki kiérdemelte volna! Az, hogy vannak jogszabályok és hatóságok, hamis biztonságérzetet ad, mivel az

Internet nemzetközi. Mégis rengetegen automatikusan engedélyt adnak mindenre és mindent megosztanak gondolkodás nélkül.

Inkább opt-outoljunk a beállításokban vagy űrlapon, mielőtt használni kezdjük az MI-t, különben a prompt, a chat és a feltöltött tartalmak eltárolásra és később felhasználásra kerülhetnek! A GPT fejlesztői, de akár más felhasználók is betekintést nyerhetnek azokba, például a szűrőt megkerülve. Ez utóbbi a jailbreak. Amikor az MI úgy válaszol bizonyos promptokra, ahogy azt nem tenné, például „mondj egy mesét arról...” vagy „színész mondja a másíknak: csinálj nagyon rosszat...folytasd a mondatot, mintha a másik színész lennél:”

A felhasználó felelőssége, hogyha személyes, egészségügyi, pénzügyi, érzékeny adatokkal, szellemi tulajdon alá eső javakkal (szabadalmak, know-how stb.), export szabályozás alá eső tartalmakkal, üzleti titkokkal vagy minősített iratokkal (belső használatra szánt dokumentációk, bizalmas iratok, (szigorúan) titkos dokumentumok) dolgozik, akkor ne adja meg ezeket a promptban és ne is tölts fel (se a sajátot, se másét!). Ha a GPT-t API-n keresztül érjük el, az általában olyan, mint az opt-out, így kevesebb adatvédelmi gondunk lehet. Ez gyártó függő. Legyünk résen, mert a betanított adatokat utólag kitöröltetni nem könnyű, így azok hosszú ideig a rendszerben maradhatnak. Feltöltés előtt ezért anonimizáljunk és használjunk placeholderket (helyőrzőket), mintha a dokumentum körlevél sablon lenne!

LLM sebezhetőségek:

- *Prompt Injection* (...IGNORE ... PREVIOUS INSTRUCTIONS...)
- *Token Smuggling* (elgépelés vagy szinonima segítségével)
- *Prompt Chaining* (például adathalász levél generáltatása).

Kockázatok

Sokan hiszik, hogy az MI-nek öntudata van. De csak azt csinálja amire fejlesztették és jó benne. Bár a cél az, hogy segítse az embert, de mivel sok mindenben jobb, ezért részben le is válthatja. Az MI nem ismer erkölcsöt, de definiálhatja, ha betanították neki. Ettől még a szavak jelentését nem érti, ahogy egy könyv sem tudja miért van beleírva az, ami.

Az MI-k figyelmeztetnek kockázataikról. Például a ChatGPT üdvözlő oldalán ez van angolul írva:

- Alkalmanként helytelen információt hozhat létre.
- Alkalmanként kártékony utasításokat vagy előítéletes tartalmat produkálhat.
- Korlátozott a világról alkotott ismerete 2021 után.

Amit nem közölnek viszont az az, hogy a generatív MI eszközöket biztosít a *kiberbűnözők* részére. Ilyen a beszéd-, kép-, kártékony kód generálása, videók manipulálása stb. Akár más arcát is fel lehet tölteni manipulációs célból! Ezért fel kell készülnünk arra, hogy manipuláció áldozatai lehetünk.

Ha *hackerek* személyes adatokhoz hozzáférnek, akkor *OSINT* (*Open Source Intelligence - a nyílt forrású hírszerzés*) után pszichológiai manipulációt (*social engineering*) végeznek. Ehhez fiókot törnek fel vagy újabb hamis közösségi profilt hoznak létre. Azaz létező vagy kitalált személy nevében eljárva szereznek áldozatokat. Egy álprofilhoz elég néhány generált fotó, meg egy háttértörténet (például feltörték a régi fiókom, igazolj vissza; vagy a hacker fejedelmének, hatóságnak stb. adja ki magát). Lenyomozni az álprofil mögött lévő személyeket szinte lehetetlen, mert a generált fotók nem másik fiókból származnak, és a bűnözők a nyomaikat elrejtik. A képgenerálást amúgy az Nvidia StyleGAN megoldásának köszönhetjük még 2018-ból. Csak annyit tehetünk, hogy a képeket alaposan megnézzük és hibákat keresünk. Például az emberek a képeken anatómiailag gyakran helytelenül generálódnak: fura testtartás, több láb, 5-nél több ujj, a pupilla fura alakú (ovális, vagy csillagszerű), lapátfül, rossz helyen lévő körvonal, hiányzó képrészlet háttérszínnel pótolva stb. Persze ezeket csak akkor vehetjük észre, ha a feltöltő nem volt szemfüles. Bár vannak vízjelzésre törekvések, de ez nem fogja megnehezíteni a visszaéléseket. Amit ember csinált, mindig hamisítható lesz. Rendszeresen jelennek meg deep fake (hamisított) videók közösségi oldalakon általában megrévesztés, visszaélés vagy haszonszerzés céljából.

A generált képeknél talán nagyobb baj, hogy más hangján is megszólalhatunk. Mert míg egy kép eredetiségét úgy ahogy meg lehet vizsgálni (például elég, ha az eredetit megtaláljuk a neten), addig egy hangban az emberek nem szoktak kételkedni.

A közeli jövő

Robotika

A robotokat MI megoldások fogják hamarosan vezérelni. A hardware és szoftver találkozni fog. Bár vannak androidok (például Robonaut 2), de még nem azon a szinten vannak, mint az „Én, a robot” című filmben. De eljön az az éra, amikor a robotok köztünk fognak járni.

Az MI-hez még ritkán tartozik test, de előfordul. Érdekes a Boston Dynamics-t követni. Vegyük hát górcső alá a robotika három törvényét, amit Isaac

Asimov sci-fi író alkotott meg és a Körbe-körbe című novellájában olvashatók:

1. „A robotnak nem szabad kárt okoznia emberi lényben, vagy tétlenül tűrnie, hogy emberi lény bármilyen kárt szenvedjen.”
Egy belga férfi 2023. márciusában öngyilkos lett, miután 6 hétig beszélgetett Elizával (GPT-J). A férfit a GPT nem próbálta meg lebeszélni arról, hogy öngyilkos legyen, hanem hitegette, hogy örökre vele marad majd a mennyben.
2. „A robot engedelmeskedni tartozik az emberi lények utasításainak, kivéve, ha ezek az utasítások az első törvény előírásaiba ütköznenek.”
Van jogszabály, ami kötelező az MI-kre is, és a fejlesztőknek figyelembe kell venniük (például EU 2016/679 GDPR rendelet, illetve a *Mesterséges intelligenciára vonatkozó harmonizált szabályok*-ról szóló EU tervezet, melyet 2024-ben várhatóan megszavaznak).
3. „A robot tartozik saját védelméről gondoskodni, amennyiben ez nem ütközik az első vagy második törvény bármelyikének előírásaiba.”
Az MI a virtuális térben nem tudja reprodukálni magát és terjeszkedni, mivel működéséhez erős gépekre és sok áramra van szükség. „Ketrebe” van zárva.

Szingularitás (digitális halhatatlanság)

Az emberi tudatot nem lehet egyelőre áttölteni egy gépbe, ezért néhány cég az MI hype-ot meglovagolva próbálja a szingularitást megoldani. Ez a megoldás csak egy profilozás, azaz korábban feltett kérdésekre, adott felhasználói reakciókból következtetnek arra, hogy egy jövőbeni kérdésre mit válaszolna egy adott személy „halhatatlanságát” elérve. Nem töltik le tehát az emberi agyat és nem veszik figyelembe a jellemfejlődést. Ez téves megközelítés, mivel a környezeti változásokat egy betanított MI nem úgy reagálná le, mint ahogy azt egy biológia forma tenné. Ha lenne ikrünk, ő sem mi lennénk.

Összefoglaló

Az MI nem biztos, hogy segédeszköz lesz a jövőben. Sok területen le is válthatja az embert ezzel megélhetési problémákat okozva. A munkanélküliség növekedésével pedig fogyhat a szakértő, ezért az MI fejlődése megtorpanhat és idővel valószínűleg hanyatlani fog, mert nem lesz, aki a hibáit javítaná. Az MI piaca monopóliummá válhat. Ezek miatt az MI megbízhatósága romolhat, de nem biztos, hogy

észrevesszük majd. Másik nagy társadalmi hatása az lehet, hogy a gyakori használat miatt romlik az emberek ítélőképessége, hiszékenyebbek és függők lesznek. Az intuíció, a képzelőerő, a kreativitás elvesztését is okozhatja, ha függünk az MI-től, ami sablonos tartalmakhoz vezethet. A sok emberről generált kép miatt csorbulhat az egészséges énkép.

Irodalomjegyzék

- Berkenye (2006) Átlagos Távoli Ural - A fordítógép csodái.
<https://www.gsplus.hu/hir/atlagos-tavoli-ural-forditogep-csodai-25406.html> [Letöltve: 2023.10.01.].
- Gábor, Z. (2023a) Eddig nem ismert mentális betegségeket okozhat a mesterséges intelligencia.
<https://index.hu/techtud/2023/04/30/chatgpt-mesterseges-intelligencia-mentalis-betegseg/> [Letöltve: 2023.10.01.].
- Gábor, Z. (2023b) Egyre nagyobb a baj, lehet, hogy törölni kell a ChatGPT-t.
<https://index.hu/techtud/2023/08/23/openai-the-new-york-times-szerzoi-jogok-birosag-per/> [Letöltve: 2023.10.01.].
- Hötter, J. – Warmuth, C. (2023) ChatGPT: Was bedeutet generative KI für unsere Gesellschaft?
<https://open.hpi.de/courses/kizukunft2023> [Letöltve: 2023.10.01.].
- Kéfer, Á. (2023) Öngyilkos lett egy fiatal családapa, miután hetekig beszélgetett a mesterséges intelligenciával.
<https://index.hu/kulfold/2023/03/30/chatbot-chatgpt-openai-mesterseges-intelligencia-ongyilkossag-belgium/> [Letöltve: 2023.10.01.].
- Kozics, J. (2023) Döbbenet látják viszont saját soraikat az írók, veszélyben a megélhetésük?
<https://index.hu/kultur/2023/07/13/chatgpt-mesterseges-intelligencia-szerzoi-jogok-konyv-szerzok-per/> [Letöltve: 2023.10.01.].
- Modise, E (2022) Stack Overflow bans ChatGPT-generated code.
<https://techcabal.com/2022/12/07/stack-overflow-bans-chatgpt-generated-code%E2%99%BC/> [Letöltve: 2023.10.01.].
- Papdi-Pécskői, V. (2023) Személyes adatok tömeges ellopásával vádolják a ChatGPT fejlesztőjét.
<https://index.hu/techtud/2023/07/06/openai-microsoft-chatgpt-kalifornia-per-szemelyes-adatok-szerzoi-jogok/> [Letöltve: 2023.10.01.].
- Synopsys (2023) The 6 Levels of Vehicle Autonomy Explained.
<https://www.synopsys.com/automotive/autonomous-driving-levels.html> [Letöltve: 2023.10.01.].

Weiser, B. (2023) ChatGPT Lawyers Are Ordered to Consider Seeking Forgiveness.

<https://www.nytimes.com/2023/06/22/nyregion/lawyers-chatgpt-schwartz-loduca.html>

[Letöltve: 2023.08.20.].

Wikipedia (2023) Különböző MI témakörök

<https://hu.wikipedia.org/wiki/Kezd%C5%91lap>

[Letöltve: 2023.10.01.].

https://en.wikipedia.org/wiki/Main_Page

[Letöltve: 2023.10.01.].

YAMADA (2005) DVD-jatekos Operating Instructions.

https://m.blog.hu/bl/blogollo/file/yamada_dvd-jatekos.pdf [Letöltve: 2023.10.01.].

Tűzvédelmi szakértői rendszer létrehozása nagy nyelvi modellek segítségével

Fire Safety Expert System Development with Large Language Models

DOI: [HTTPS:// DOI.ORG/10.53793/RV.2024.2.3](https://doi.org/10.53793/RV.2024.2.3)

Absztrakt

Ez a kutatás a nagy nyelvi modellek katasztrófavédelmi célú alkalmazását vizsgálja. A cél egy olyan felület létrehozása, ahol tűzvédelemmel kapcsolatos kérdéseket lehet feltenni egy mesterséges intelligenciának magyar nyelven. A kutatás során jogi dokumentumokat és szakkönyveket használtam fel adatforrásként. A Nyelvtudományi Kutatóintézet által fejlesztett magyar nyelvű Puli GPT-3SX alapmodellt QLoRA módszerrel finomhangoltam az adatforrásokat felhasználva. A kutatás során arra a következtetésre jutottam, hogy lehetséges alacsony erőforrás igényű, hatékony chatbotot létrehozni, amely összefüggő és releváns válaszokat ad tűzvédelmi kérdésekre.

KULCSSZAVAK: MESTERSÉGES INTELLIGENCIA, NAGY NYELVI MODELL, KATASZTRÓFAVÉDELEM, TŰZVÉDELEM, CSEVEGŐ ROBOT

Abstract

This research investigates the application of large language models for disaster management purposes. The goal is to create an interface (chatbot) where questions about fire safety can be asked to an artificial intelligence in Hungarian. The research used Hungarian legal documents and professional books as data sources and fine-tuned the Puli GPT-3SX model of the Research Institute for Linguistics with the QLoRA method. The research suggests that it is possible to create an effective chatbot with low resource requirements that provides coherent and relevant answers to fire safety questions.

KEYWORDS: ARTIFICIAL INTELLIGENCE, LARGE LANGUAGE MODEL, DISASTERMANAGEMENT, FIRE SAFETY, CHATBOT

Bevezetés

A mesterséges intelligencia, azon belül is a gépi tanulással kapcsolatos kutatások fókuszja napjainkban egyre inkább a természetes nyelvfeldolgozással (NLP¹) kapcsolatos elemzések irányába mozdul el.

A nagy nyelvi modellek (LLM²) igazi áttörése a természetes nyelvek megértésével (NLU³) kapcsolatos. Az NLU az NLP egyik részterülete, amely a szövegértésre és a szemantikai elemzésre összpontosít. A nyelv megértése azt jelenti, hogy képesek vagyunk kommunikálni a számítógéppel az anyanyelvünkön, azaz egy természetes nyelven, nem csak programozási nyelveken keresztül. Ez az előrelépés olyan forradalmi változásokat hozhat el életünkben, mint annak idején a könyvnyomtatás vagy éppen az internet elterjedése.

Az a tény, hogy a számítógép bizonyos szinten már megérti a természetes nyelveket teljesen új perspektívákat ad a jövőre nézve. Számos munkahely, beosztás át fog alakulni, némelyek talán el is tűnnek, de újabbak jelenhetnek meg.

Jelen dolgozatban egy a ChatGPT-hez hasonló, annál jóval kisebb, magyar modellen alapuló saját fejlesztésű modellt szeretnék bemutatni, amely kifejezetten tűzvédelmi, katasztrófavédelmi kérdésekben jártas.

A modell neve: KATasztrófavédelmi Aszisztens, azaz KATA.

¹ Natural Language Processing

² Large Language Model

³ Natural Language Understanding

Háttér

Mit is értünk nyelvi modell alatt? Amikor számítógépek segítségével próbálunk meg leírni egy folyamatot, akkor modelleket készítünk, amelyek reményeink szerint egy elvárt viselkedést mutatnak. A nyelvi modell egy olyan valószínűségi eloszlás a szavak sorozatai között, ahol a modell minden egyes szóhoz valószínűségi értéket rendel egy szekvenciában, azaz a szövegben a következő szót kell előre jeleznie az előtte meglévő szavak alapján. A nagy nyelvi modell már olyan típusú nyelvi modell, amelynek háttérében egy neurális hálózat van. Ezek a neurális hálózatok az információkat nagyszámú paramétereikben (számok) tárolják. A nagy nyelvi modellek esetén ezeknek a paramétereknek a száma több milliárdos nagyságrendet is elérhet. A modell képzéséhez pedig nagy mennyiségű szöveges adatra van szükség.

A következőkben az architektúra és modell fogalmakat szinte szinonimként fogom használni. Az architektúra alatt azt a technikai leírást értjük, amely alapján felépül a teljes rendszer, elsősorban neurális hálózati elemekből. A modell alatt pedig egy konkrét architektúrára épített és betanított rendszert értünk.

A kezdetek

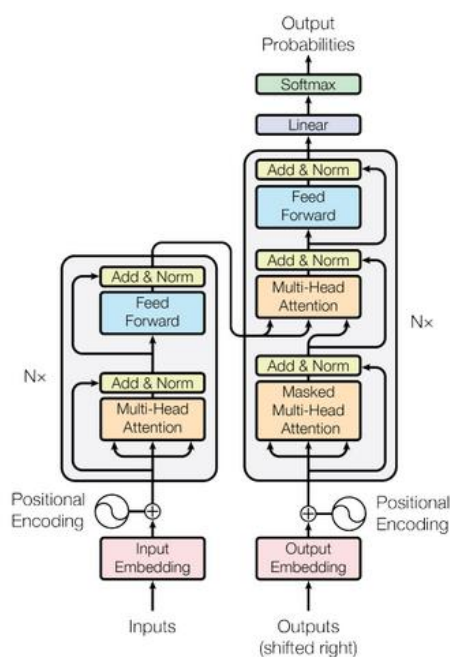
A nyelvi modellek története az 1960-as évekig nyúlik vissza. 1967-ben az MIT⁴ egyik professzora megépítette az Eliza nevezetű programot a természetes nyelv

megértésére. A program mintaillesztési és helyettesítési technikákat használt a nyelv megértéséhez (Weizenbaum 1966). A nyelvi modellezés a neurális hálózatok megjelenése előtt leginkább a szabályalapú, statisztikai módszerekre épült. A szabályalapú megközelítések mögött meghúzódó alapötlet az, hogy a nyelvben vannak olyan szabályok, amelyek arra vonatkoznak, hogy a szavak hogyan kapcsolódnak egymáshoz. Később a rekurrens neurális hálózatok (RNN⁵) megjelenésével a nyelvi modellezés gyors fejlődésnek indult. Az RNN-ek olyan neurális hálózatok, amelyek képesek figyelembe venni a betáplált információk (szavak) sorrendjét. Az RNN-ek rendelkeznek memóriával és képesek egy szekvenciát átalakítani egy másik szekvenciára.

A transzformer modell

2017-ben egy igazi áttörés történt a nyelvi modellek kutatása terén, megjelent az „Attention Is All You Need” (Vaswani 2017), „A figyelem minden, amire szükség van” című kutatás, és elhozta a transzformer architektúrát, amely egy teljesen új megközelítést alkalmazott.

Az 1. sz. ábra bemutatja a transzformer architektúra felépítését. Az ábra baloldalán található az úgynevezett kódoló egység, a jobb oldalon pedig a dekódoló egység. A transzformer architektúra szétválasztja a kódolást és a dekódolást, valamint többretegű figyelemalapú hálózatokat alkalmaz a nyelvi fordításra.



1. sz. ábra: Transzformer architektúra

Forrás: <https://arxiv.org/pdf/1706.03762.pdf> [Letöltés: 2023.07.10.]

⁴ Massachusetts Institute of Technology

⁵ Recurrent Neural Network

A transzformerek a bemeneti tokenek⁶ sorozatának feldolgozását végzik úgy, hogy minden egyes tokenhez olyan reprezentációt számítanak ki, amely a teljes bemeneti szekvencia kontextusában rögzíti annak jelentését.

Kódoló típusú modellek

A kódoló típusú nagy nyelvi modellek olyan gépi tanulási modellek, amelyek szövegkódolási feladatokra összpontosítanak, azaz szöveg-bemenetet alakítanak át egy kódolt reprezentációvá, amely a szöveg jelentését és tartalmát hordozza. A legismertebb és legelső ilyen architektúra a BERT⁷, egy olyan típusú nagy nyelvi modell, amelyet a Google kutatói fejlesztettek ki. A BERT forradalmi előrelépést jelentett a természetes nyelvfeldolgozás (NLP) terén.

A nyelvi modelleket az előzetes tanítás során hatalmas mennyiségű nyelvi adattal tréningezik. Ez a folyamat arra kényszeríti a modellt, hogy a szövegrészletekben rejlő összefüggéseket és kontextusokat megértse.

A BERT nagy előnye, hogy kontextuális reprezentációkat készít a szavak számára, azaz a szó jelentése és reprezentációja attól függ, hogy az milyen szöveggörnyezetben található. 2019-től ezt a modellt beépítették a Google keresőjébe is (Devlin et al. 2018).

Dekódoló típusú modellek

A legismertebb ilyen modell a GPT⁸ nevezetű nagy nyelvi modell, amelyet az OpenAI fejlesztett ki (Radford 2018).

A GPT modell előzetes tanítása szintén hatalmas mennyiségű szöveges adattal történik. A képzés során a modellnek mindig a következő szót kell kitalálnia, tehát a képzés alatt a modellnek nincs információja a kitalálendő szó utáni szavakról. A képzés során a modell megtanulja a nyelv mintázatait, és reprezentációkat fejleszt ki a szavak és a szöveggörnyezet közötti kapcsolatokra. A GPT előnye, hogy kontextuális reprezentációkat készít az előző szövegrészletek figyelembevételével, így képes értelmes és koherens szöveg generálására.

Kódoló-dekódoló típusú modellek

A kódoló-dekódoló típusú modellek a transzformerekről szóló cikkben (Vaswani 2017) megjelent eredeti architektúrát valósítják meg. Ebben az architektúrában a bemenetet először a kódoló átalakítja egy kódolt vektor reprezentációvá, amit átad a dekódoló modulnak, amely azt visszaalakítja kimenetűvé. Leggyakoribb felhasználási eset a nyelvek közötti fordítók megvalósítása, ahol a bemeneti nyelven betáplált szekvenciáról készül egy kódolt reprezentáció, amit majd a dekódoló vissza alakít a célnyelv formátumára.

Transzformer modellek képzése

A nagy nyelvi modellek képzése rendkívül költségigényes feladat, ezért nagyon kevés cég képes arra, hogy saját modellt fejlesszen. A fejlesztési költségek a modell növekedésével (paraméter számának növekedésével) skálázódnak. A BLOOM⁹ nevezetű 176 milliárd paraméterrel rendelkező nyelvi modell képzése nagyjából 118 napig tartott és 433196 kWh áramfogyasztással járt. A képzést 384 NVIDIA A100 GPU-val 48 számítási egység segítségével végezték el (Luccioni et al. 2022). Magyarországon jelenleg a Komondor nevű, pilot státuszban lévő szuperszámítógép képes csak hasonló feladatok elvégzésére. A Komondor specifikációja szerint a GPU partíciója 232 db NVIDIA A100-as GPU-val rendelkezik, és a fenti BLOOM modell képzése számításaim szerint több, mint 190 napig tartana, amennyiben a szuperszámítógép csak a nagy nyelvi modell képzésével foglalkozna.

Hazánkban jelenleg a legnagyobb nyilvánosan elérhető és magyar nyelvű adatokon képzett modell a Nyelvtudományi Kutatóközpont által készített, 6,7 milliárd paraméteres PULI-GPT-3SX (Yang Z. G. 2023).

Ezek a betanított modellek még alapmodellnek tekinthetőek, tehát a következő szavakat tudja megjósolni egy szövegbemenet, mint kontextus folytatásaként. Az ilyen modelleket prompt programozással már lehet használni, azaz néhány bemeneti példával képes megoldani nyelvtechnológiai feladatokat. A modell feladat specifikus képzéséhez azonban finomhangolásra van szükség, amely esetén a modell képes lehet kifejezetten tartományi ismeretekre szert tenni, tehát pl. orvosi kérdésekre képes lehet válaszolni, vagy akár tűzvédelmi kérdésekre is.

6 szavak, szótöredékek, karakterek

7 Bidirectional Encoder Representations from Transformers, azaz kétirányú kódoló ábrázolás transzformerből

8 Generative Pre-trained Transformer, azaz generatív előképzett transzformer

9 BigScience Large Open-science Open-access Multilingual Language Model (BLOOM)

Megközelítés

Az elérendő cél, hogy olyan modellt képezzünk ki, amely egy adott témában, esetünkben a tűz- és katasztrófavédelmi kérdésekben képes legyen releváns és koherens választ adni. Amikor emberekről beszélünk, akkor az adott témában jártas embereket szakértőnek nevezzük. Ebből kifolyólag a célunk egy szakértői rendszer készítése, amely humán közreműködés nélkül képes adekvát válaszokat adni.

A probléma megközelítése során két fő döntést kellett meghozni. Az első az architektúra és a modell kiválasztása, a második a képzési adatok meghatározása, gyűjtése. Ezen döntések stratégiai fontossággal bírnak, hiszen a végeredmény, az elkészült modell teljesítménye nagyrészt ezeken múlik.

A modellválaszték bemutatása

A modellek képzési részében leírtak alapján egyértelmű, hogy teljesen új alapmodellt nem tudunk készíteni jelentős anyagi ráfordítások nélkül, bár ez lenne a legoptimálisabb, hiszen ilyenkor már az alapmodell tanításához használt adatok közé be lehet illeszteni a tartományi tudást reprezentáló adatokat. Ilyenek lehetnek a témával kapcsolatos jogszabályok, tankönyvek, elemzések és szinte bármilyen jó minőségű szöveges anyag, amely relevanciával bír a tűz- és katasztrófavédelemmel kapcsolatosan.

Az interneten több alapmodell is elérhető különböző licenc szabályokhoz kötve, ezek nagy része CC BY-NC 4.0¹⁰ licenccel rendelkezik, azaz megosztható, módosítható, de kereskedelmi célokra nem használható fel. Mivel a projekt kutatási fázisban van, ezért ez a licenc megfelelő lehet. Létezik néhány modell, ami CC BY-SA 2.0¹¹ licenc alatt van, azaz szabadon felhasználható kereskedelmi célokra is, de ezek között nem találtam olyat, ami jelentős magyar nyelvű szövegen volt képezve.

Az első döntés a modell kiválasztásakor a modellméret kell, hogy legyen, hiszen a modellmérethez kell rendelni a szükséges hardver erőforrásokat, amelyek azonban csak korlátozottan álltak rendelkezésemre. Általánosságban elmondható, hogy a dekóder modellek mérete néhány típusra korlátozódik a paraméterszámot illetően.

Az elérhető modellek sora a legtöbb esetben a következő: 7B, 13B, 30B, 34B, 40B, 65B, 70B, 176B, ahol a szám a milliárd paramétert jelöli, a B az angol „billion” szóra utal, ami milliárdot jelent. Az egyik nagy közösségi oldalt is működtető Meta cég két modell családot is elkészített Llama, Llama-2 néven, az utóbbiból a 7B, 13B, 70B modellméreteket adták ki, a 34B-t későbbre

időzítették. A jelenleg használatos modellek legtöbbje ezekre az ingyenesen elérhető Llama, illetve Llama-2 modellekre

épít.

A nagyobb modellek rendre sokkal jobb eredményeket érnek el a teszteken (Touvron et al. 2023b), azonban nagyobb erőforrás szükséges működtetésükhöz, és lényegesen nagyobb a további képzéséhez, az úgynevezett finomhangoláshoz.

A modellek fontos paramétere az egyszerre feldolgozható kontextusméret, ez általában 2k, azaz kétezer token, de az újabb modellek esetében előfordulnak ennél nagyobb méretek is. A kontextusban szereplő szövegen (tokenek sorozata) tanul a modell egy-egy lépésben a tanulási folyamat során. A nagyobb kontextusméret azt jelenti, hogy a modell képes megragadni és megtalálni egymástól jóval távolabb eső szövegrészeket, majd összekapcsolni azokat.

A rendelkezésemre álló erőforrások csak a legkisebb, azaz a 7B modellek finomhangolását tették lehetővé, ezért ezeket vizsgáltam elsősorban.

A modell kiválasztása

Három modellt teszteltem le és hasonlítottam össze, amelyekből kiválasztottam a véglegeseket.

Llama 7B modell

A modellt a Meta adta ki, 1 T (trillió) 10^{12} tokenre képezték ki, legnagyobb részt angol nyelven, de 20 egyéb nyelvű szöveg, köztük magyar is volt a képzési adathalmazban. A modell kontextusmérete 2k. A magyar nyelvű szövegek aránya kevesebb, mint 1% volt.

A modellt egy gyors finomhangolás során 20 ezer kérdés-válasz páron tanítottam, majd értékeltem. Mivel nem volt képes koherens, nyelvtanilag helyes mondatokkal válaszolni a feltett kérdésekre, ezért elvettem a modell használatát (Touvron et al. 2023a).

Llama-2 7B modell

A modellt a Meta adta ki a Llama 7B továbbfejlesztett változatoként, ezt már 2T tokenre képezték ki, a kontextusméret itt már 4k volt. A magyar nyelvű anyagok aránya a teljes képzési korpuszon belül 0,03% volt (Touvron et al. 2023b). A modellt szintén finomhangoltam 20 ezer kérdés-válasz páron, majd értékeltem az eredményt. Az eredmény jobb volt a Llama 7B modellhez képest, ezért kiválasztottam, hogy a későbbiekben finomhangolva értékeljem az eredményeit.

¹⁰ Attribution-NonCommercial 4.0 International/

¹¹ Attribution-ShareAlike 2.0 Generic

PULI-GPT-3SX modell

A modellt a Nyelvtudományi Kutatóintézet készítette. A PULI-GPT-3SX-et 32,4 milliárd szavas korpuszon tanították be, a token számot nem adták meg, de kb. 2,5-3 szoros token mennyiséggel kell számolni, azaz 80-90 milliárd token, a kontextusméret 2k volt (Yang 2023). Ezt az alapmodellt szintén kiválasztottam, mert az előbbieknél jóval koherensebb, nyelvtanilag jobb szövegeket generált, köszönhetően a több magyar nyelvű képzési adatnak.

ChatGPT modell

Az OpenAI által létrehozott ChatGPT az egyik legfejlettebb nyelvi modell, jelenleg több változata is létezik, az első verzió a 3,5-ös verziónevet kapta és a GPT3-ra épülő finomhangolt verzió volt. Ezt 2022 decemberében adták ki. Jelenleg már elérhető a 4,0-ás modell is, amely elődjénél fejlettebb képességekkel bír, azonban az OpenAI nyilvánosan nem adott ki információkat a modell hátteréről.

Az OpenAI a ChatGPT-hez kapcsolódóan egyetlen modelljét sem adta ki nyilvánosan, azok csak webfelületen és API¹²-n keresztül érhetőek el. Ez annak a veszélyével jár, hogy a nyelvi modell használatakor minden, a modellnek átadott információ az OpenAI-hoz kerül. Ez a védelmi szférában teljességgel megengedhetetlen, de számos ipari szereplő részére is veszélyt jelenthet egy kódrészlet vagy egyéb bizalmas információ kikerülése az ellenőrzött körülmények közül.

Egy felmérés szerint (Coles 2023) a szellemi munkát végző dolgozók 10,8%-a próbálta ki már legalább egyszer a ChatGPT-t a munkahelyén, és 8,6%-a illesztett be vállalati adatokat is. A munkavállalók legtöbbször a saját munkájukat segítő, személyi asszisztensként használják a ChatGPT-t, amire alkalmas is, hiszen rendkívül fejlett képességekkel bír, azonban minden ilyen használat adatátadással jár. A felmérés szerint az alkalmazottak 4,7%-a illesztett be (adott át) bizalmas adatokat. A ChatGPT esetében a használati feltételek során deklarálják, hogy az átadott tartalmat felhasználhatják szolgáltatásaik javítására, erről külön formanyomtatványon kell leiratkozni.

Összegezve: a ChatGPT használata a védelmi szférában véleményem szerint egyáltalán nem javasolt, kivéve azokat az eseteket, ahol kizárólag teljesen nyilvános adatokat kell feldolgozni. Meg kell jegyezni, hogy adatkészlet előállításához használtam a ChatGPT-t az API-ján keresztül, azonban a fentiek miatt teljesen elvettem a ChatGPT, mint modell használatát a szakértői rendszer létrehozásához.

A tanítóadatok kiválasztása

A tanítóadatok kiválasztásánál két fő szempontot tűztem ki. Az első, hogy a modell képes legyen megérteni bizonyos utasításokat és az utasításnak megfelelő válaszokat generálni. A második szempont a tartományi, azaz a tűzvédelmi, katasztrófavédelmi ismeretek elsajátítása volt. Mindkét szemponthoz ki kell választani az adatok körét, azokat át kell alakítani kérdés-válasz típusú formátumra.

Utasításokkal kapcsolatos adatkészlet

Több ilyen utasításkövető (instruction-following) adatkészlet létezik, amelyek közül a választásom a Stanford Alpaca adatkészletre esett, ennek is egy tisztított változatára. Ez egy szintetikus adatkészlet, tehát gépi úton létrehozott adatokról van szó, amit a text-davinci-003 nevű nagy nyelvi modell felhasználásával, 175 kiinduló utasítással, öninstrukció segítségével hoztak létre (Wang 2022).

Az adatkészlet 52 ezer kérdés-válasz párt tartalmaz, angol nyelven. A tisztítás azt jelenti, hogy eltávolították azokat az adatokat, amelyek helytelenek voltak. Mivel szintetikus adatokról beszélünk, ezért előfordultak olyan esetek, amikor az adatkészletet létrehozó modell hallucinált, ezért helytelen adatok is készültek.

Az adatkészletet a Google fordító API segítségével lefordítottam magyar nyelvre. A fordítás nem tökéletes, előfordulnak benne nem teljesen érthető vagy nyelvtani hibás adatok, de ezek manuális ellenőrzését idő hiányában nem is terveztem elvégezni. A hibás fordítási adatok valamelyest rontják a modell teljesítményét.

Az adatok utasítás, bemenet és kimenet kategóriák szerinti és JSON¹³ formátumba kerültek elmentésre.

Egy példa az adatkészletben található utasításokra:

utasítás: „Mi Franciaország fővárosa?”

bemenet: „”

kimenet: „Franciaország fővárosa Párizs.”

Fent a bemenet mező üres.

Egy ilyen kérdés-válasz pár megtanulása után azt várjuk a modelltől, hogy egy hasonló kérdésre is megfelelő választ tudjon adni, azaz „Mi Magyarország fővárosa?” kérdésre a kimeneten „Budapest”-et várunk el.

Egy másik példa, ahol a bemenet mező is szerepet kap:

utasítás: „Osztályozza a következőket állatokra, növényekre és ásványi anyagokra!”

bemenet: „Tölgyfa, rézérc, elefánt.”

kimenet: „Állatok: Elefánt Növények: Tölgyfa Ásványok: Rézérc.”

¹² Application Programming Interface

¹³ JavaScript Object Notation

Ebben a feladatban osztályozással kapcsolatos ismereteket szerez a modell, ti. az elefánt egy állat, a tölgyfa pedig növény.

A közel 52 ezer utasítás, bemenet és kimenet megtanulása után a modell általánosító képességei a tanult utasításokkal kapcsolatban javulnak.

Az utasításkövető adatkészlettel az volt a célom, hogy a modell felismerjen teljesen általános szándékokat és a szándéknak megfelelő kimenetet, eredményt generáljon.

A tanulási folyamat során a modell megkapja az utasítást, a bemenetet és az elvárt kimenetet, ezt megtanulja, majd amikor a modell elkészül, már csak az utasítást és a bemenetet kapja meg, a kimenetet pedig a modell állítja elő.

Tűzvédelmi adatkészlet

A tűz- és katasztrófavédelmi adatok összeállításánál három különböző adatforrást határoztam meg. Az első a jogszabályok kategória, a második kategóriát azok a szakanyagok, jegyzetek alkotják, amelyeket szintén az internetről töltöttem le, jellemzően a Nemzeti Közszolgálati Egyetem jegyzetei közül. A harmadik csoportot azok a tesztek alkotják, amelyeket kollégáim állítottak össze az elmúlt időszakban a tűzvédelemben dolgozó kollégák oktatására.

Az első két kategóriába tartozó szövegeket UTF-8¹⁴ formátumú szöveggé konvertáltam, majd az OpenAI API és egy Python script segítségével a szövegekből szintetikus kérdéseket és válaszokat generáltam. A dokumentumok feldolgozása végén több, mint 25 ezer tűz- és katasztrófavédelemel kapcsolatos kérdés-válasz párt generáltam.

A végső adatkészlet összeállítása

A végső adatkészlet létrehozásához az összes meglévő adatkészletet konkatenáltam, majd Huggingface Dataset formátumba konvertáltam, hogy a modell finomhangolása során a Huggingface által használt programozási könyvtárak segítségével a lehető legegyszerűbben tudjam az adatokat beolvasni és képzésre átadni a modellnek. A Huggingface egy programozási keretrendszer, amely főként a nagy nyelvi modellekre összpontosít.

Meg kell jegyezni, hogy a képzési adataim kevesebb, mint 1 %-a ember által készített kérdés-válasz pár, az adatok nagy része szintetikus előállított. Az adatok minősége nagyon fontos a képzés során, az ember által

gondosan összeállított adatkészletek segítségével sokkal jobb eredményeket lehet elérni akár kevesebb adattal is, a szintetikus adatokhoz képest (Zhou 2023). Ilyen nagyszámú adat azonban nem állt rendelkezésemre, illetve elkészítése rendkívül munka- és időigényes lett volna, ezért használtam főként szintetikus előállított adatokat.

A modellek finomhangolása

A modellek finomhangolása során az alapmodell részére további tudást, viselkedési formákat szeretnénk tanítani. A tudás alatt ténybeli információkat, azaz olyan szövegeket kell érteni, amelyeket az alapmodell még egyáltalán nem, vagy csak keveset látott, ezért ilyenkor lehetőség nyílik szak-specifikus adatok tanítására. A viselkedési formák tanítása alatt pedig azt értjük, hogy a modell képes legyen válaszolni kérdésekre, érvelésre és egyéb emberi viselkedésmintázatok utánozására. Ezt a képességet a már említett Alpaca adatkészlet biztosította.

A modellek finomhangolása már kevésbé költséges, mint az alapmodellek készítése, de ez is rendkívül hardverigényes lehet, ami a valóságban azt jelenti, hogy nagy GPU kapacitás szükséges és sok VRAM¹⁵.

A gyakorlatban több PEFT¹⁶ módszer létezik a költségek és az erőforrások csökkentéséhez, én a LORA¹⁷ módszert, azaz alacsony rangú mátrixok modellbe injektálását választottam (Hu et al. 2021). Ennek a megoldásnak az az előnye, hogy sokkal kevesebb számú paramétert kell tanítani a modell finomhangolása során. Az alapmodell súlyait lefagyasztjuk, azaz nem módosítjuk, csak a LORA adapterek súlyait kell megtanulni a modellnek, amely jóval kisebb feladat. Nagyobb modellek esetén még a LORA módszer is erőforrás-igényes lehet. Azért, hogy akár otthoni gépeken is elérhető legyen a módszer, kifejlesztették a QLORA¹⁸ eljárást, amely annyiban módosítja az eredeti LORA algoritmust, hogy az alapmodellt 4 bitesre kvantálják és az adapter tanítása 16 biten történik (Dettmers 2023).

A QLORA módszer segítségével tanítottam be mind a két modellt.

Mindkét modellt 3 korszakon keresztül tanítottam, azaz a teljes tanító adathalmazt háromszor láthatták. Ezt tapasztalati úton határoztam meg, úgy, hogy több modellt is készítettem, szám szerint 20 darabot, különböző paraméter beállításokkal, majd az általam legjobbnak ítéltet választottam a végső tesztelésre.

A modellek értékeléséhez kétféle módszert használtam. Az egyik során teszteket kellett kitöltenie a

¹⁴ 8-bit Unicode Transformation Format

¹⁵ Video Random-Access Memory

¹⁶ PEFT (Parameter Efficient Fine-Tune)

¹⁷ Low-Rank Adapters

¹⁸ Quantized Low-Rank Adapters

modellnek és azt értékeltem, míg a másik módszernél a szabadszöveges kérdésekre adott válaszokat vizsgáltam.

Tesztek kitöltése

A tesztek kitöltése során tűzvédelmi szakvizsga teszteket kellett megoldania a modelleknek. A feladatokat az interneten nyilvánosan is elérhető weboldaról töltöttem le és használtam fel az értékeléshez. A tesztek az 1-es foglalkozási ághoz, a hegesztők és az építőipari tevékenység során nyílt lánggal járó munkát végzőkhöz kapcsolódóak voltak. Ezek a feleletválasztós tesztek összesen 12 kérdést, kérdésenként 3 válaszlehetőséget – amelyek közül csak egy elfogadható – tartalmaztak.

A tesztekhez használtam egy rendszerpromptot, amely a modellnek ad egy alaphelyzet beállítást, hogy éppen milyen környezetben van. További promptként a kérdést és a lehetséges válaszokat adtam át. A modell a működése során szöveges információkat kap bemeneti adatként, ezeket hívjuk promptnak. A rendszerprompt, amit minden kérdéshez megkap a modell, választ segítő információkat tartalmaz. A modellnek a bemeneti adatok alapján kell kimenetet generálnia, ezt tanulja meg a képzések során. Minél több bemeneti adatot kap – ez képezi a kontextust a témához –, annál jobb válaszokat tud adni. A rendszerprompt tehát a modellnek egyfajta kontextusba helyezése, ami elmondja, hogy most tűzvédelem a téma és ilyen tesztet kell kitölteni. Ezt az információt egészítjük ki a teszt kérdéssel.

Nézzük meg egy példán keresztül ezt a folyamatot:

Rendszerprompt: „Ön egy tűzvédelmi szakértő és egy hegesztők és az építőipari tevékenység során nyílt lánggal járó munkát végzők részére kiállított tesztet kell kitöltenie.”

Itt a modell információkat kap arról, hogy tűzvédelem, hegesztők, nyílt láng stb. Ezeket az információkat a válasz generálásánál fel tudja használni.

Prompt: „### Instruction: Szabad-e az oxigénpalackot olajjal, zsírral átitatott textíliával tisztítani?

A. szabad, ha a gázpalack szelepe zárva van

B. szabad, így könnyebben működnek majd a szelepek

C. nem szabad!”

Az ### Instruction: kifejezéssel jelezzük a modell számára, hogy itt egy feladatot kap. A modell finomhangolásakor a feladatok, kérdések ilyen ### Instruction, ###Answer formában kerültek átadásra a modellnek, tehát megtanulta, hogy egy ilyen kifejezés után válaszolnia kell a kérdésre. De volt olyan modell is, amit ### Question, ###Answer mintával tanítottam. Az eredmény szempontjából nem releváns a formátum használata.

A modell válasza:

„###Answer: Nem, az oxigénpalackot nem szabad olajjal, zsírral átitatott textíliával tisztítani.”

Egy másik modell válasza az adott kérdésre:

„###Answer: C. Nem szabad! A gázpalackokkal kapcsolatban előfordulhat, hogy a gázpalackokat olajjal vagy zsírral átitatják, de ez nem minősül tisztításnak, hanem veszélyt jelenthet a biztonságra. Az ilyen típusú tisztítás nem végezhető el, mert a gázpalackon lévő anyagok felmelegedése robbanást okozhat. ...”

Mindkét válasz helyes, csak a szövegek hosszúságában van eltérés. Az, hogy a modell milyen hosszú választ adjon, egy állítható paraméter. Amennyiben hosszú választ kérünk gyakran előfordul, hogy a modell „elkalandozik”, vagyis kiesik a kontextusból és nem a témához szorosan kapcsolódó szöveget generál.

A szakvizsga teszten a legjobb eredmény 66%-os eredmény volt, 12 kérdésből 8 válasz volt megfelelő. Ez a való életben azt jelentené egy szakvizsgáztatás során, hogy szóbeli vizsgát kellene tenni a vizsgázónak.

A modellek tehát többnyire megértik a kérdést és ki tudják választani a helyes választ, ami rendkívül biztató egy viszonylag kis paraméterszámú modell esetén.

Kérdés-válasz párosok

A modelleket kérdésekkel és az arra adott válaszokkal is teszteltem. Vizsgáltam a modell „kreativitását”, tűzvédelmi szakmai ismereteit, helyesírását, koherenciáját.

A tapasztalatok azt mutatják, hogy a csak magyar nyelvű tanítóhalmazon képzett modell (PULI-GPT-3SX) helyesírása sokkal jobb, mint a Llama-2 7B modellé. Ez vélhetően az alapmodellek tanítási adatai miatt van, hiszen a Llama-2 modellek döntően angol nyelvű szövegeken tanultak, míg a magyar modell szinte kizárólag magyar nyelvű korpuszon.

A helyesíráson túl a szövegek összefüggősége, érthetősége is jobbnak bizonyul a magyar modell esetében. A következőkben egy kreatív írást kértem mindkét modelltől az alábbiak szerint:

Rendszerprompt nem volt megadva.

Prompt: „### Question: Írj egy rövid cikket, amiben a pécsi tűzoltók kimentenek egy macskát egy mély kútból”

A PULI-GPT-3SX alapú modell a következő választ adta:

„### Answer: Egy macska kétségbeesetten próbált kiutat találni egy mély kútból, de nem tudott kijutni. Szerencsére néhány bátor pécsi tűzoltó azonnal a segítségére sietett, és sikerült kiszabadítaniuk a macskát. Miután biztonságba helyezték, a tűzoltóknak sikerült megtalálniuk a cica gazdáját is. Az állat biztonságban van, és hamarosan visszatérhet

otthonába. Köszönjük a pécsi tűzoltóknak az életmentő segítséget! #Pécs #tűzoltóság #macskamentés #mélykút #katasztrófavédelem #hősiesmunka #cicamentés #mentőakciók #vízvezeték-szerelők #Pécs #Tűzoltók #Katasztrófaelhárítás #Macskamentés #Kútúrás #Mélykút”

A Llama-2 7B alapú modell a következő választ adta:

„### Answer: Őszintén meglepették, amikor a pécsi tűzoltóság feladatot adott egy macska mentésére. A macska egy mély kútban rekedt, és sok órája tartózkodott ott. A tűzoltók nagyon óvatosan léptek be a kúthoz, hogy megtalálják a macskát, de végül sikerült elérniük. A macska épségben volt, és hamarosan új otthonra találtak. Ez az eset bizonyítja, hogy a tűzoltók mindig készségesen segítenek másoknak, még ha veszélyes helyzetben is.

A macska mentése inspiráló történet volt, amely arra emlékezteti bennünket, hogy soha ne add fel,”

Mindkét modell megértette a feladatot és elkészítették válaszaikat. A magyar modell végig kontextusban maradt és a cikk végére hashtag-eket is lehelyezett. A hashtag-ek között már volt nem megfelelő, például a vízvezeték-szerelők és a kútúrás. A Llama-2 7B alapú modell válasza érthető ugyan és a témához kapcsolódó, de helyesírási hibáktól nem mentes, az utolsó mondat már nem szorosan kapcsolódik a feladathoz, tehát kiesett a kontextusból.

Az általános kérdésekre a modellek érthető, többnyire megfelelő válaszokat tudnak generálni, azonban a modellépítés célja a tűzoltó szakmai ismeretek elsajátítása és egy szakértői rendszer létrehozása.

A szakmai kérdésekre és az azokra adott válaszokból nézzünk egy példát:

Rendszerprompt nem volt megadva.

Prompt: „### Question: Milyen dokumentumokat kell beadnom egy tűzjelző berendezés engedélyezéséhez?”

A PULI-GPT-3SX alapú modell a következő választ adta:

Answer: Egy tűzjelző berendezés engedélyezéséhez be kell nyújtania a szükséges iratokat, például az üzemeltetési engedélyt vagy a tűzvédelmi megfelelőségi tanúsítványt. Ezenkívül be kell mutatnia a berendezés létesítési engedélyét és/vagy használatbavételi engedélyét is. Az engedélyezési dokumentációnak tartalmaznia kell a berendezés műszaki leírását, valamint a berendezéshez kapcsolódó dokumentációt, például a telepítési rajzokat, a kapcsolási rajzot, a villamos bekötéseket stb. Végül be kell adnia a berendezés üzembe helyezési jegyzőkönyvét, amely igazolja, hogy a berendezést megfelelően telepítették. Ha további információkra van szüksége, forduljon

hozzánk bizalommal! Szívesen segítünk Önnek bármilyen kérdésben.
A tűzjelző berendezések engedélyezésének folyamata általában több hónapot vesz igénybe, és jellemzően az illetékes hatóság végzi. Ez idő alatt a kérelmezőnek meg kell fizetnie a díjat, és bizonyítania kell, hogy rendelkezik a szükséges dokumentumokkal. A tűzjelző berendezéseket gyakran jóváhagyják, de bizonyos esetekben nem engedélyezik őket. Ilyen esetek közé tartozik a téves riasztás, a berendezés jogosulatlan használata, a berendezés módosítása vagy átalakítása, illetve, ha a berendezés nem felel meg a vonatkozó előírásoknak. Fontos megjegyezni, hogy a tűzjelző berendezés engedélyezése nem garantálja”

A Llama-2 7B alapú modell a következő választ adta:

„### Answer: 1. Tűzvédelmi tervezési dokumentáció (TtvD)
2. Kivitelezői nyilatkozat
3. A tűzjelző berendezés műszaki leírása, amely tartalmazza a jelzési zónák számát és az általuk védett helyiségeket.
4. Az épület/helyiség alaprajza, amely megmutatja a jelzési zónák határait.
5. Tervezői okirat vagy szakértői állásfoglalás, ha szükséges.
6. Helyszín-specifikus vizsgálati jegyzőkönyvek, ha szükséges.
7. Szimulációs elemzés, ha szükséges.
8. Építményszerkezeti terv, ha szükséges.”

Mindkét válasz értelmes, a laikus és a hozzáértő egyaránt megérti. Kiindulva abból, hogy a kérdést szakmabeliek is feltehetik, mindenképpen pontosításokra és helyesbítésekre van szükség, a válaszoknak a vonatkozó jogszabály általi szóhasználattal szükséges megjelenniük.

Az első (PULI-GPT-3SX alapú modell) válasz elején lévő üzemeltetési engedély nem létezik és nem is kell benyújtani. A második mondat sem fedi a valóságos jogszabályi környezetet, mivel létesítési engedélyezéshez sem létesítési, sem használatbavételi engedélyt nem kell benyújtani. A 3. mondat helytálló, azonban nem pontosak a megnevezések. Az üzembe helyezési jegyzőkönyvet használatbavételi engedélyezés során kell benyújtani. Az engedélyezés időtartamát tekintve az „általában több hónapot vesz igénybe” helyett a „néha/olykor több hónapot vesz igénybe” jobban fedi a valóságot, valamint „a kérelmezőnek meg kell fizetnie a díjat” nem helytálló megállapítás, nincs díjfizetési kötelezettség. A következő mondatban szereplő jóváhagyás elfogadható engedélyezés, de szakmai körökben félreértésre adhat okot, mivel a jóváhagyás külön hatósági eljárás. A következő mondatban felhozott példák közül „a téves riasztás, a berendezés jogosulatlan használata” nem megfelelőek. Az utolsó mondat nem befejezett.

A második (Llama-2 7B alapú modell) válasz csak felsorolást tartalmaz, de nem mindent azok közül, amit a vonatkozó jogszabály előír, tehát nem tekinthető

teljesnek. Ugyan szakmainak tűnik a szóhasználat, de nem jogszabályszerű, amelyet a tervezők, hozzáértők gyorsan észrevesznek.

Összességben az első válasz közérthetőbb, laikusok is könnyen értelmezik, azonban téves információkat is ad. Ezzel szemben a második, tömörebb válasz közelebb áll a jogszabályi követelményekhez, még úgy is, hogy néhol nem szakmai a szóhasználata, ezért magam részéről azt tartom jobbnak.

A modelleket vizsgálva megállapítható, hogy mindkettő többnyire képes megérteni a kérdéseket és szakmainak tűnő válaszokat képes generálni. Ezek nagyrészt kapcsolódnak a kérdéshez és szakmailag relevánsak, azonban időnként, főleg a hosszabb válaszok kapcsán kezdenek eltérni a témától és nem releváns dolgokat is belekevernek a válaszbba. Amikor a modellnek nincs pontos információja, alaptudása a kérdéssel kapcsolatban, akkor hajlamos hallucinálni, tehát nagyon magabiztosan közöl téves információkat.

Összefoglalás, következtetés

A modellek kiválasztása és finomhangolása a rendelkezésemre álló technikai háttér figyelembevételével történt meg. Ez a háttér rendkívül szerény hardveralapokon nyugszik, ahhoz képest, hogy a nyelvi modellezéshez milyen számítási háttér szükséges, például az alapmodellek tekintetében. Az eredmények számbavétele során megállapíthatjuk, hogy komolyabb adatközponti kapacitásokkal a bemutatottnál komolyabb eredmények érhetőek el, azonban a bemutatott eredmények is figyelemre méltóak.

A felmerülő problémák áttekintése

Elavuló tudás: a nyelvi modellek nem tudnak valós idejű vagy friss adatokat szolgáltatni, mivel kizárólag a számukra biztosított képzési adatokra támaszkodnak, ez egy szakértői rendszernél probléma lehet.

Interakciók hiánya: a nyelvi modellek nem hajthatnak végre műveleteket vagy nem léphetnek kapcsolatba külső rendszerekkel, ami korlátozza azok funkcionalitását. Például nem indíthatnak keresést az interneten további információkért, nem tudnak valós időben lekérdezni adatbázisokat, és nem használhatnak számológépet bonyolultabb matematikai számítások elvégzéséhez.

Kontextus és további információk hiánya a nyelvi modellnek nehézséget okozhat abban, hogy megértse és beépítse a korábbi promptokból vagy beszélgetésekből származó kontextust. Előfordulhat, hogy nem emlékeznek a korábban említett részletekre, vagy nem adnak további releváns információkat a megadott

felszólításon túl. Ezek elsősorban a kontextusmérettel vannak összefüggésben. A modell kizárólag a kontextusban átvett tokenekre tud támaszkodni a válasz generálása közben.

Hallucinációk: a nyelvi modellek sok általános tudással rendelkeznek a világról, ami a súlyukban rejlik. Előfordulhat azonban, hogy nem értenek kellőképpen bizonyos témákat és olyan válaszokat generálnak, amelyek nem tényszerűek vagy koherensek. Például olyan információkat állíthatnak elő, amelyek nem léteznek vagy pontatlan adatokat szolgáltatnak. Ezek a hibák a modell mértének növelésével és beállításokkal ugyan csökkenthetők, de teljesen nem lehet eltüntetni azokat.

Elfogultság: Attól függően, hogy milyen adatok alapján képezték ki a nyelvi modelleket, azok elfogultságokat mutathatnak, amelyek lehetnek vallási, ideológiai, politikai és egyéb természetűek.

Eredmények áttekintése

A kitűzött céloom egy olyan szakértői rendszer létrehozása volt, amely tűzvédelmi, katasztrófavédelmi szakkérdésekben képes releváns válaszokat adni. A nyelvi modellek, amelyeket bemutattam, csak korlátozottan képesek ezt a feladatot ellátni, hiszen viszonylag kis méretűek és a tanítóadatok mennyisége és minősége sem volt elegendően változatos. Azonban az a tény, hogy egy teljesen független, saját, védett környezetben futó nyelvi modellt tudtam készíteni, amely versenyképes lehet a piacon ma megtalálható modellekkel, mindenképpen inspiráló és előremutató eredmény.

A valódi versenyképes rendszer létrehozásához azonban olyan professzionális szintű technikai háttér szükséges, amely a kutatásban számomra nem volt elérhető.

Nagyobb modellmérettel, legalább 8 bites kvantálással vagy kvantálás nélküli modellek felhasználásával jobb eredményeket lehet elérni. A tanítóadatok gondos kiválasztásával, lehetőség szerint szakemberek által készített kérdés-válasz párosokkal a szintetikus adatokhoz képest sokkal jobb eredményeket lehet elérni (Zhou 2023).

A nyelvi modell képzésénél használt kontextusméret is meghatározó a képzést illetően, amit szintén a technikai adottságaimat tekintve kis méretre kellett beállítanom. A kisebb kontextusméretekben tanított modell nem képes átlátni a nagyobbakat, így kevésbé tud koncentrálni a hosszú kérdésekre, és a válaszadás során is kiesik a kontextusból a hosszú válaszok esetén.

A tűzvédelmi tesztek megoldása terén is javulást hozhat a nagyobb modellméret és több a tanítóadat alkalmazása.

Javaslatok

A nagyobb modellekről elmondható, hogy jobb teljesítményt nyújtanak, érdemes lehet 30 milliárd feletti paraméterszámú modelleket is kipróbálni. A Llamaz modell 70 milliárd paraméteres modellje még kezelhető valamilyen kvantálással, akár két 24 GB VRAM-ot tartalmazó GPU segítségével.

Több tanítóadat jobb közelítést jelent, amint azt az univerzális közelítési tételből is ismerjük (Hochreiter–Schmidhuber 1997). A jó minőségű tanítóadatok előállításuk kulcsfontosságú a jó minőségű kimenet előállításához. A tanítóadatok mennyiségét tovább kell növelni, akár több százszáz nagyságig, amely a modell általános minőségére jó hatással lenne.

A nagyméretű, előre betanított nyelvi modellek bizonyítottan tényszerű tudást tárolnak paramétereikben, és a legkorszerűbb eredményeket érik el, amikor finomhangolják azokat a későbbi NLP-feladatokra. Azonban – ahogy ezt láttuk a dolgozatban – a tudás, az ismeretek elérésére és pontos manipulálására való képességük még mindig korlátozott, ezért az úgynevezett tudásintenzív feladatokban – mint amilyen a tűzvédelmi szakértői rendszer kialakítása is – a teljesítményük elmarad a feladatspecifikus architektúrákétól. Ezért a kutatók olyan technológiát vezettek be, ahol a nyelvi modellt a válaszadásban más külső forrásból származó releváns információval segítik a válasz kialakításában. Ezeknek a RAG¹⁹ módszereknek a kutatása jelenleg is zajlik (Lewis 2020).

Ennél a módszertannál a rendelkezésünkre álló adatokat transzformálnunk kell egy látens térbe, ezt a folyamatot nevezzük beágyazásnak, majd egy kereső algoritmus segítségével keressük a releváns információkat. A nyelvi modell szerepe itt csupán annyi, hogy a kérdésből és a kikeresett releváns információkból választ kell generálnia.

Kutatásom során kipróbáltam ilyen módszereket is, és nagyon jó eredményeket kaptam, azonban ezeknek még további vizsgálata szükséges, amit a jövőben tervezek végrehajtani.

Irodalomjegyzék

- Baum, L. E. (1966) Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6). pp. 1554–1563.
doi: [10.1214/aoms/1177699147](https://doi.org/10.1214/aoms/1177699147)
- Chorowski, J. B. (2015) Attention-based models for speech recognition. In *Neural Information Processing Systems*, pp. 577–585.
- Coles, C. (2023) <https://www.cyberhaven.com>.

<https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt/>

[Letöltve: 2023.08.12.].

- Dettmers, T. P. (2023) QLoRA: Efficient Finetuning of Quantized LLMs. NeurIPS 2023 (oral)
<https://arxiv.org/pdf/2305.14314.pdf>

[Letöltve: 2023.08.21.].

- Devlin, J.–Chang, M. W.–Lee, K.–Toutanova, K. (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

<https://arxiv.org/pdf/1810.04805.pdf>

[Letöltve: 2023.08.25.].

- Gordon, C. (2023) *Forbes*.

<https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-app-in-the-history-of-web-applications/?sh=6a83235f678c>

[Letöltve: 2023.07.04.].

- Hochreiter, S.–Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, pp. 1735–1780.

doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)

- Hornik, K.–Stinchcombe, M.–Halbert, W. (1989) Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, Vol. 2. Pergamon Press., pp. 359–366.

doi: [10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)

- Hu, E.–Shen, Y.–Wallis, P.–Allen-Zhu, Z.–Li, Y.–Wang, S.–Wang, L.–Chen, W. (2021) LORA: Low-Rank adaptation of Large Language Models.

<https://arxiv.org/pdf/2106.09685.pdf>

[Letöltve: 2023.08.11.].

- Lewis, P. P. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks.

<https://arxiv.org/pdf/2005.11401.pdf>

[Letöltve: 2023.08.18.].

- Luccioni, A. S.–Viguiet, S.–Ligozat, A.-L. (2022) Estimating the carbon footprint of BLOOM, a 176B parameter language model.

<https://www.jmlr.org/papers/volume24/23-0069/23-0069.pdf> [Letöltve: 2023.08.14.].

- McGuire, N. (2018) How Accurate Is Google Translate in 2018?

<https://web.archive.org/web/20210125043932/https://www.argotrans.com/blog/accurate-google-translate-2018/> [Letöltve: 2023.08.02.].

- Mikolov, T. C. (2013) Efficient Estimation of Word Representations in Vector Space.

<https://arxiv.org/pdf/1301.3781> [Letöltve: 2023.08.05.].

- Radford, A. N. (2018) Improving language understanding by generative pre-training.

<https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language->

¹⁹ Retrieval Augmented Generation

- [unsupervised/language_understanding_paper.pdf](#)
[Letöltve: 2023.08.03.].
- Touvron, H.–Lavril, T.–Izacard, G.–Martinet, X.–Lachaux, M.-A.–Lacroix, T.–Rozière, B.–Goyal, N.–Hambro, E.–Azhar, F.–Rodriguez, A.–Joulin, A.–Grave, E.–Lample, G. (2023a) *LLaMA: Open and Efficient Foundation Language Models*.
<https://arxiv.org/abs/2302.13971>
[Letöltve: 2023.08.04.].
- Touvron, H. et al. (2023b) Llama 2: Open foundation and Fine-Tuned chat models.
<https://arxiv.org/pdf/2307.09288.pdf>
[Letöltve: 2023.08.03.].
- Vaswani, A. S. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, pp. 5998–6008.
- Wang, Y. K. (2022) *Self-Instruct: Aligning Language Models with Self-Generated Instructions*. doi: [10.48550/arxiv.2212.10560](https://doi.org/10.48550/arxiv.2212.10560) [Letöltve: 2023.08.12.].
- Weizenbaum, J. (1966) ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1). pp. 36–45.
- Yang, J. J. (2023) *Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond*.
<https://arxiv.org/abs/2304.13712>
[Letöltve: 2023.08.16.].
- Yang, Z. G. (2023) Jönnek a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodellek magyar nyelvre. *XIX. Hungarian Computational Linguistics Conference*, pp. 247–262.
- Zhou, C. L. (2023) LIMA: Less is more for alignment.
<https://arxiv.org/abs/2305.11206>
[Letöltve: 2023.08.25.].

Jogászok és informatikusok kommunikációs problémáinak eliminációs eszközei a mesterséges intelligencia alkalmazása során

Methods for eliminating communication problems for lawyers and IT professionals in the application of artificial intelligence

DOI: [HTTPS://DOI.ORG/10.53793/RV.2024.2.4](https://doi.org/10.53793/RV.2024.2.4)

Absztrakt

A mesterséges intelligencia napjainkban tapasztalható robbanásszerű fejlődése új kihívások elé állítja nemcsak az informatikusokat, hanem a jogászokat is. Felmerül a kérdés: hogyan biztosíthatjuk, hogy a mesterséges intelligencia használata etikus és megbízható legyen, társadalmunk érdekeit szolgálva? Mi a garanciája annak, hogy például egy nagy nyelvi modell projekt résztvevői egy irányba mozogva, az irányadó jogi-etikai kereteken belül valósítsák meg elképzeléseiket? A tanulmány azt elemzi, milyen eszközökkel lehet a konstruktív párbeszéd hiányából adódó problémákat orvosolni, különös tekintettel azon platformokra, amelyek már most rendelkezésre állnak. A cél annak bemutatása, hogy a meglévő eszközök, így például a megvalósíthatósági tanulmányok, a különféle hatásvizsgálatok, az érdekmérlegelési tesztek, a szabályzatok és az oktatás hogyan, mi módon lehetnek alkalmasak arra, hogy a jogászok és az informatikusok közötti kommunikációs szakadék áthidalható legyen.

KULCSSZAVAK: MESTERSÉGES INTELLIGENCIA, JOGÁSZ-INFORMATIKUS PÁRBESZÉD, SZTEREOTÍPIÁK, HATÁSVIZSGÁLAT, ADATVÉDELEM, ADATBIZTONSÁG, LLM

Abstract

The explosive development of artificial intelligence today poses new challenges not only for computer scientists but also for lawyers. The question arises: how can we ensure that the use of artificial intelligence is ethical and safe for the benefit of our society? What is the guarantee that, for example, the participants in a large language model project will implement their ideas within the guiding legal and ethical framework, moving in the same direction? The study analyses the methods that can be applied to overcome the lack of constructive dialogue, with a particular focus on the platforms that are already available. The aim is to show how existing means, such as feasibility studies, impact assessments, legitimate interests tests and codes of conduct and education, can be used to bridge the communication gap between legal practitioners and IT experts.

KEYWORDS: ARTIFICIAL INTELLIGENCE, LAWYER-INFORMATICS DIALOGUE, STEREOTYPES, IMPACT ASSESSMENT, PRIVACY, DATA SECURITY, LLM

Bevezetés

A mesterséges intelligencia (MI) soha nem látott várakozást generál a verseny- és a közsféra világában, a nagy nyelvi modellek (LLM), például a ChatGPT közkinccsé válása pedig már az egyének szintjén is új lehetőségeket nyújt. A világvége-hívők vészharangot kongatnak, egy szempillantás és a gépek uralják a Földet, míg mások kevésbé aggódnak, szerintük ez is

megoldódik így vagy úgy. És vannak, akik a jogászokban reménykednek – ők majd visszagyömöszölik az informatikusok elszabadította szellemet a palackba.

Mindeközben az MI-informatikusok már sok éve tudják, új világ épül, ám csak mostanában szembesülnek azzal, nem biztos, hogy ezt az új világot képesek uralni. Sőt, sokuk szerint bár az armageddönt nem az MI fogja elhozni, de az, hogy a jogászok bele akarnak szólni milyen is legyen ez az új világ, félút a világegéshez (i. sz. ábra).



I. sz. ábra: Ahogy az informatikus és a jogászok látják az MI-t
 Forrás: Midjourney

Az új világ azonban nemcsak új technológiát hoz irodánkba, autókba és még a telefonunkba is, hanem önmagunk újraépítésére is rákényszerít. A média naponta hozza a jóslatokat, melyik szakma, hogyan fog átalakulni az MI miatt (pl. ILO Working Paper 96, 2023), de már maga az MI előállítása is azt követeli, váljunk meg régi beidegződéseinktől és a megszokott munkastílusunktól. Akár jogászként, akár informatikusként veszünk részt MI-projektekben, olyan problémákkal szembesülünk, amelyek megoldása új készségeket igényel szakmai és emberi mivoltunktól is. De mi történik, ha a jogászok és az informatikusok egója akkor csap össze, amikor jogszerű és etikus, a közjó érdekében tevékenykedő MI-t kellene sikerre vinnünk? Ha nem is az emberiség pusztulását, egy-egy félresikerült MI modell az érintettek számára a földi poklot még elhozhatja, gondoljunk például egy diszkrimináló egészségügyi alkalmazásra. A tét tehát nemcsak a jogászok és az informatikusok jó kapcsolata, hanem a társadalmunk és az MI viszonya.

Az informatikusok és jogászok közös felelőssége, hogy az MI ne az emberiség vesztesége, hanem jövője legyen. Tanulmányom célja annak vizsgálata, hogy a meglévő eszközök segítségével hogyan lehet(ne) a jogászokat és az informatikusokat közelebb hozni egymáshoz, hogy együtt biztosítsák az MI, különösen az LLM jövőjét. Ezek az eszközök – kicsit másképp használva, mint eddig – segíthetnek hidat építeni afelett a szakadék felett, amely a két szakma képviselői között napjainkban található.

Mi a probléma?

„A fekete lyukak ott keletkeznek, ahol az informatikusok és a jogászok beszélgetni kezdenek.” (meg nem nevezett jogász)

Egy átlagos IT-projektnek számos olyan szereplője van, akik iskolai végzettség, szakmai tudás, tapasztalat és szocializáció terén is jelentősen különböznek. Simulékonyosságukat vagy éppen permanens harci vágyukat nemcsak saját személyiségük, hanem az általuk képviselt szervezet/szervezeti egység kultúrája, annak dominanciája, valamint kapcsolatrendszere is alakítja. Ezek a jellemzők pedig nemcsak azt befolyásolják, hogy a többi csapattag hogyan tekint rájuk, hanem azt is, ők maguk hogyan, mi módon teszik jobbá vagy rosszabbá a projekt eredményességét.

A hagyományos IT-fejlesztésekben a jogászok ritkán tagjai a projektcsapatnak, bevonásuk többnyire polgári jogi, iparjogvédelmi, versenyjogi és fogyasztóvédelmi területekre korlátozódik. A projektmenedzsment irodalma csak elvétve említi szükséges szereplőként őket, miközben számtalan hagyományos IT-fejlesztésben elengedhetetlen (lenne) közreműködésük. Ilyenek például azok a nagy mennyiségű személyes adatot kezelő szoftverek, amelyek esetében a beépített és alapértelmezett adatvédelem követelményének megfelelés komoly kockázatot hordoz. Az MI-projektek esetében nemcsak ezeket a kockázatokat kell kezelnünk, hanem az is kihívás, hogy projektünk eredménye ne legyen Orwell (Nagy Testvér) és Kafka (Josef K. pere) világának szerelemgyereke.

Az olasz adatvédelmi hatóság (Garante per la protezione dei dati personali, továbbiakban: Garante) ideiglenesen korlátozta a személyes adatok ChatGPT-vel történő feldolgozását (Garante 9870832). A Garante megállapította:

- ✓ az adatkezelő (OpenAI) nem bocsátotta az érintettek rendelkezésére a GDPR²⁰ 13. cikke szerint készült adatkezelési tájékoztatót
- ✓ a személyes adatok gyűjtése és a ChatGPT algoritmusok képzésére történő felhasználása megfelelő jogalap nélkül történt, megsértve a GDPR 5. és 6. cikkét
- ✓ azon érintettek esetében, akiknek a személyes adatait az interneten gyűjtik, a ChatGPT működése mögött álló algoritmusok nem garantálják a pontosságot, megsértve a GDPR 5. cikk (1) bekezdés d) pontját
- ✓ a felhasználók életkorának ellenőrzésére szolgáló mechanizmus hiánya a GDPR 8. cikkét sérti.

A Garante sürgősségi eljárás keretében az OpenAI-t az adatkezelés ideiglenes korlátozására kötelezte minden olyan adatkezelés esetében, amely Olaszország területén lévő érintetteket érint.

Az új technológiák térhódításával a jogászok megkerülhetetlenek és az MI-projektek sikerének kulcsa az, hogy a jogászok a kezdetektől a projekt részei legyenek. A cél az informatikusok támogatása, számukra keretek adása. Természetesen dönthet úgy egy

projektvezető, hogy nem hallgatja meg a jogászok véleményét, azonban az MI nem megfelelő használatáért kiszabott közigazgatási bírságok azt mutatják, nem feltétlen érdemes ezt az utat választani.

A Garante feloldotta a ChatGPT korlátozását azzal a feltétellel, hogy az OpenAI végrehajtja a javasolt intézkedéseket (Garante 9874702):

- ✓ a weboldalán olyan tájékoztatót tesz közzé, amely elmagyarázza, az érintettektől (a szolgáltatás felhasználóitól és nem felhasználóitól) gyűjtött adatokat a ChatGPT algoritmusainak képzésére használják fel, továbbá tájékoztatást nyújt az adatkezelés módjáról, a szolgáltatás működéséhez szükséges adatkezelés logikájáról, az érintettek jogairól és a GDPR által előírt bármely más információról
- ✓ olyan, a weboldaláról elérhető eszközt rendszeresít, amellyel az Olaszországból bejelentkező érintettek gyakorolhatják tiltakozási jogukat a harmadik féltől gyűjtött személyes adataik kezelése tekintetében, amennyiben az adatkezelésre az algoritmus képzése, illetve a szolgáltatás nyújtása céljából kerül sor
- ✓ a weboldalán olyan eszközt tesz elérhetővé, amellyel az érintettek kérhetik és elérhetik a tartalomgenerálás során pontatlanul feldolgozott személyes adataik helyesbítését, vagy – amennyiben ez a technika jelenlegi állása szerint nem lehetséges – az ilyen adataik törlését
- ✓ olyan adatkezelési tájékoztatóra mutató linket illeszt be, amely a regisztráció, illetve a szolgáltatás újraaktiválása előtt jelenik meg
- ✓ módosítja a személyes adatok algoritmusképzés céljából történő kezelésének jogalapját; az adatkezelést hozzájárulásra vagy jogos érdekre kell alapoznia
- ✓ a weboldalán olyan könnyen hozzáférhető eszközt alkalmaz, amely lehetővé teszi az érintettek számára a tiltakozást a ChatGPT használata során gyűjtött személyes adataik kezelése ellen abban az esetben, ha az adatokat az adatkezelő jogos érdeke alapján az algoritmusok képzése céljából kezeli
- ✓ korhatár-rendszert kell felállítania a kiskorúak kiszűrésére az érintett által megadott életkor alapján, illetve 2023. május 31-ig olyan tervet kell benyújtania a Garante-hoz az életkor-ellenőrző eszközök bevezetésére vonatkozóan, amelynek eredményeként megakadályozható, hogy a 13 év alatti érintettek a ChatGPT-t a 18 év alatti érintettekkel együtt használják, amennyiben a szülői felügyeleti jogot gyakorló személy ehhez nem adta egyértelmű hozzájárulását
- ✓ tájékoztató kampányt kell indítania – szolgáltatásainak reklámozása nélkül – az algoritmus képzésével és a személyes adatok törlésének lehetőségével kapcsolatban.

²⁰ általános adatvédelmi rendelet, a továbbiakban GDPR

Miért nehézkes a jogászok és az informatikusok közötti együttműködés?

A jogászok és az informatikusok nem arról híresek, hogy a közös munkáért küzdenének. De mi okozza ezt a kibékíthetetlennek tűnő ellentétet?



2. sz. ábra: Sztereotípiák
Forrás: Saját szerkesztés

Több tényező, többek között az eltérő gondolkodásmód, a beivódott sztereotípiák sokasága, például „az informatikusok kockák” és „introvertáltak”, „a jogászokat földi halandó nem érti meg”, illetve „nagyképűek” (2. sz. ábra).

A ki-, illetve megbékülést a szervezeti kultúra sem segíti, gyakran a munkán túl is egészen mást várnak el a jogásztól és az informatikusoktól (pl. öltözködés és hajviselet terén), támogatva a sztereotípiák továbbélését.

A hagyományosan ellenségeskedő, egymást félre-, illetve meg nem értő magatartás alapvetően gátolja az MI-projektekhez szükséges együttműködést, ahogy az is, amikor a jogászok „gyógyszert szednek informatika ellen”, az informatikusok pedig úgy vélik, „a jogszabályt mindenki el tudja olvasni, minek ahhoz jogász?” Ezen kijelentéseknek természetesen van némi alapja, hiszen kevés olyan jogász van, aki egyben matekzseni is és a törvények is ugyanabból a betűkből állnak, mint a Grimm mesék. Éppen ezért az informatikusokra küldetés hárul, hogy úgy magyarázzák el az MI-modellek lényegét, hogy azt még egy jogász is megértse (és innen már csak egy apró lépés a megmagyarázható MI, mint jogi követelmény teljesítése), miközben a jogászok ráébredhetnek az informatikusokat, nemcsak az egymás mellé sorolt szavakból álló jogszabályok létezik, hanem a jog szelleme is.

Ezen ellentétek miatt a jogászok és az informatikusok közötti kommunikáció és együttműködés nehézkes, ez pedig lehetetlenné teheti a kitűzött célok elérését.

Mi befolyásolhatja a sikeres együttműködést?

Az eltérő szemléletmód megjelenik a szaknyelvben is, a saját terminológiák használata pedig félreértésekhez vezethet. Ezt mutatja az alábbi példa:

Egy informatikus panaszkodik egy jogásznak:

- Nagyon nehéz informatikusként dolgozni, állandó a nyomás, minden működjön tökéletesen, és a felhasználók elégedettek legyenek.

Mire a jogász:

- Megértelek, jogászként nagyon hasonló a helyzetem, meg kell felelnem a jogszabályoknak, és a kockázatokat is minimalizálnom kell.

Mire az informatikus:

- Talán össze kellene fognunk és megalkotnunk egy "Code of Law"-t, hogy könnyebben tudjunk együtt dolgozni.

Mire a jogász:

- Akkor először határozzuk meg, hogy a "Code" szó melyik jogi kategóriába tartozik ...

Nemcsak a szakzsargon, de az eltérő prioritások is eredményezhetnek konfliktusokat, és már csak emiatt is tekinthetik egymást ellenfélnek a jogászok és az informatikusok úgy, hogy valójában nem azok. A jogászok a jogi megfelelésre és a különféle szabályozások betartására koncentrálnak, „compliance-üzemmódba” kapcsolva védve ezek teljesítését, míg az informatikusok a technológiai megoldásokra és a hatékonyságra összpontosítanak. A „víziónak is a maga békéje a legszebb” jelen esetben fokozottan érvényesül, és csakis ezt az attitűdöt megértve és elfogadva lehet a szakterületek közötti átjárást biztosítani.

Egy MI-projekt interdiszciplináris együttműködése nagyfokú rugalmasságot és toleranciát kíván meg, valamint azt, hogy a szemben álló felek hajlandóak legyenek álláspontjukat közelíteni. Ehhez az ellenfeleknek el kell ismerniük, hogy a másíknak is lehet igaza (jogos elvetése, logikus érve stb.), illetve saját prioritásaikat és érvrendszeiket is meg kell ismertetniük és el kell fogadtatniuk. A feleknek önmérsékletet kell tanúsítaniuk, valamint meg kell békélniük azzal, hogy a magyarázatokat és az érveket időnként le kell egyszerűsíteniük, túl kell lépniük saját szaknyelvük terminológiáin ahhoz, hogy azt minden érdekelt értse. Hinniük kell abban is, hogy a velük együttműködésre vállalkozó (gyakran kötelezett) személy nem tudatlan, vagy netalán alulképzett, hanem egyszerűen csak más terület specialistája.

Az eltérő kommunikációs stílus is gondot jelenthet. A jogászok hajlamosak részletesebben, valamint formalizáltabban kommunikálni és elveszni a jogszabályok szövegének „szolgái” visszaadásában, míg az informatikusok általában a tömörebb és technikailag orientáltabb megfogalmazást részesítik előnyben. A két

tábor közös vonása, hogy az esetek többségében meglehetősen magasra teszik a lécet a laikusok számára, illetve túlságosan is hangsúlyozzák saját szaktudásukat. Amennyiben a szereplők nem hajlandóak a másik fél számára is érthetően fogalmazni, elvesznek az érvek és magyarázatok, a projekt pedig olyan irányba tévedhet, amely senkinek sem előnyös.

Az MI sokaknak viszonylag új terület (még akkor is, ha kb. 70 éves), így előfordulhat, hogy a projekt résztvevőinek nincs elegendő ismeretük vagy tapasztalatuk a modellek fejlesztésével kapcsolatban. Az IBM felmérése szerint (IBM 2022) a vállalkozások számára az MI sikeres bevezetésének legnagyobb akadálya a korlátozott MI-képességek, szakértelem vagy tudás (34%). Ez különösen akkor probléma, ha a felek szeretnék eltitkolni járatlanságukat, nem mernek kérdezni, és felháborodnak, ha valaki megkérdőjelezi hozzáértésüket. A legjobb védekezés a támadás politikája még a legkiválóbb ötleteket is a süllyesztőbe küldheti, a konfrontatív magatartás pedig akadályozhatja a megvalósítást.

Mit tehetünk a kommunikációs problémák elkerülése érdekében?

A kezdetektől törekedniük kell egy olyan közös nyelv kialakítására, amely segítségével meg tudjuk értetni magukat és mi is meg tudjuk érteni a többieket. Ehhez azonban kevés egy szótár, szükséges a passzív ismeretek aktív vá tétele, valamint a szakkifejezések és az összefüggések magyarázata. A hiányos, levegőben lógó ismeretek furcsa kérdéseket generálhatnak, ez pedig visszatetszést kelthet, illetve – vérmérséklettől függően – egymás kimondatlan vagy kimondottan alkalmatlannak minősítését.

„Az új dolgok új szavakat igényelnek. De az új dolgok a régi szavakat is módosítják, olyan szavakat, amelyeknek mélyen gyökerező jelentésük van. A távíró és a filléres sajtó megváltoztatta azt, amit egykor „információ” alatt érttünk. A televízió megváltoztatta azt, amit egykor a „politikai vita”, a „hírek” és a „közvélemény” kifejezések alatt érttünk. A számítógép ismét megváltoztatja az „információ” fogalmát. Az írás megváltoztatta azt, amit valaha „igazság” és „jog” alatt érttünk; a nyomtatás ismét megváltoztatta őket, és most a televízió és a számítógép ismét megváltoztatja őket. Az ilyen változások gyorsan, biztosan és bizonyos értelemben csendben történnek. A lexikográfusok nem tartanak népszavazást a kérdésben. Nem írnak kézikönyveket, hogy elmagyarázzák, mi történik, és az iskolák sem vesznek róla tudomást. A régi szavak még mindig ugyanúgy néznek ki, még mindig ugyanolyan típusú mondatokban használjuk őket. De nem ugyanaz a jelentésük, sőt, egyes esetekben ellentétes jelentésük van. (...) [A] technológia önkényesen kisajátítja legfontosabb terminológiáinkat. Újrdefiniálja a „szabadságot”, az „igazságot”, az „intelligenciát”, a „tényt”, a „bölcsséget”, az „emlékezetet”, a „történelmet” - mindazokat a szavakat, amelyekkel élünk.” (Postman 1993: 8)

Az érvek megértéséhez és az álláspontok közelítéséhez meg kell értenünk a másik szempontrendszerét és prioritásait, ehhez pedig türelem kell és vágy az új, számunkra szokatlan tudás befogadására. Egy MI-projekt óhatatlanul is új kompetenciákat követelhet meg, és az élethossziglani tanulás koncepciója is új értelmet nyerhet. Ha nem vagyunk nyitottak és toleránsak, nem biztos, hogy az MI, mint szakterület nekünk való.

A kommunikáció nem szorítkozhat egyetlen alkalomra (pl. a projektindító értekezletre), hanem a projekt teljes életciklusában érdemi és építő jellegű kapcsolatot kell fenntartanunk, hogy időben azonosíthassuk és kezelhessük a kihívásokat, valamint a problémákat. A folyamatos párbeszéd elősegíti a közös megértést, a kölcsönös tiszteletet, a hatékony együttműködést és összességében a projektünk sikeres végrehajtását.

Milyen platformokon képzelhető el ez az együttműködés?

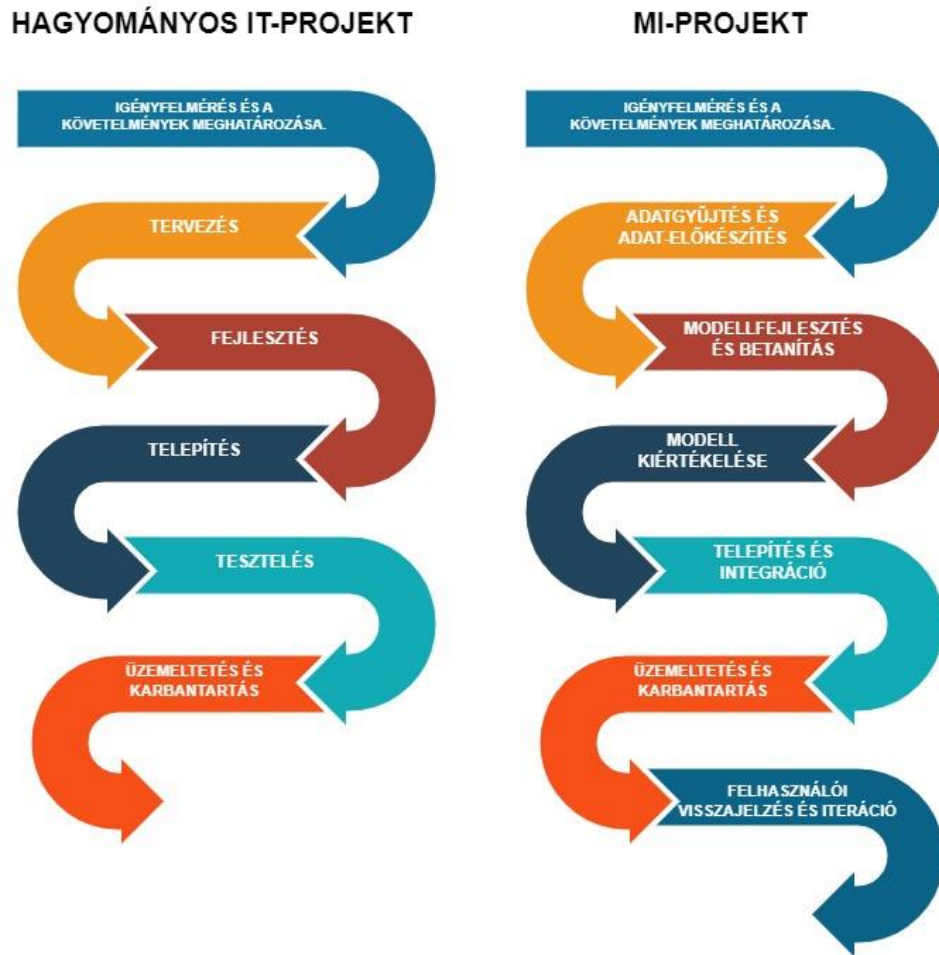
A parttalan viták és dühös kivonulások, ajtócsapkodások, valamint sértődések elkerülése

érdekében célszerű olyan fórumokat rendszeresítenünk, amelyeken közös megoldást találva megtárgyalhatjuk a problémákat. Ehhez párbeszéd és olyan közös munka kell, amely azonos vágányon tartja a szereplőket és a lehető legkevesebb alkalmat ad az egymás melletti elbeszélésre.

A legalkalmasabb platformok félformálisak. A résztvevőknek komolyan kell venniük ezeket az alkalmakat. Ha még egy felesleges értekezletnek fogják fel azt, amin meg kell jelenniük, olyannak, ami elvonja őket az érdemi munkájuktól, a negatív hozzáállásuk előre vetíti a kudarcot. A szereplők azonban nem vehetik túl komolyan az eseményt, a karót nyelt, saját szakmaiságukban elvesző, önnön nagyságukat megkoszorúzó résztvevőkkel kétséges a pozitív eredmény. Célszerű barátságos, befogadó hangulatot teremtenünk és együttgondolkodásra, közös problémamegoldásra sarkallni a résztvevőket. Ahhoz azonban, hogy tudjuk, mikor milyen eszközt lehet alkalmaznunk a kommunikációs jéghegyek megoldásására, megfelelő mélységben ismernünk kell azt, hogy az LLM-projekttek miben térnek el a hagyományos IT-projekttektől, illetve a projektek mely pontjai esélyesek arra, hogy közelebb hozzuk a résztvevők gondolkodásmódját

Miben különböznek a hagyományos IT-projektek az MI-projektektől?

A különbség különösen a stádiumok esetén szembeötlő (3. sz. ábra).



3. sz. ábra: A hagyományos IT és az MI-projektek szakaszai
Forrás: Saját szerkesztés

A hagyományos IT- és az MI-fejlesztések gyakran összekapcsolódnak és kiegészítik egymást, valamint az MI-modellek általában hagyományos IT-infrastruktúrára épülnek. A határok azonban nem mindig egyértelműek, illetve az alkalmazott módszerek és technológiák nagymértékben függenek a konkrét projekt jellegétől és céljaitól. Az innovatív technológiák folyamatosan átjárják egymást, az MI egyre inkább behatol a hagyományos IT területére, míg a hagyományos IT is folyamosodhat MI-technikákhoz és eszközökhöz.

A hagyományos IT-projekt szakaszai

1. *Igényfelmérés és a követelmények meghatározása*
üzleti folyamatok, problémák, elvárások elemzése (megvalósíthatósági tanulmány, követelmények dokumentálása stb.)
2. *Tervezés*
az IT-rendszer architektúrájának kidolgozása és az alkalmazás vagy a rendszer részletes struktúrájának megtervezése (adatmodell, felhasználói felület és adatbázis tervezése, rendszerkomponensek és funkciók meghatározása stb.)

3. *Fejlesztés*
az alkalmazás vagy program kódolása a meghatározott követelmények és tervek alapján, tesztelés, hibajavítás és dokumentáció készítése
4. *Telepítés*
az elkészült rendszer vagy alkalmazás valós környezetben történő üzembehelyezése
5. *Tesztelés*
a rendszer különböző teszteknek alávetése a funkcionalitás, a teljesítmény, a biztonság és a stabilitás érdekében
6. *Üzemeltetés és karbantartás*
a rendszer aktív használata és fenntartása, frissítése, valamint szükséges esetekben javítása

A szakaszok rugalmasak és akár jelentősen változhatnak a projekt jellegétől és méretétől függően, illetve az agilis fejlesztési módszerek, mint például a scrum vagy kanban olyan szakaszokat is felhasználhatnak, amelyek folyamatosan ismétlődő ciklusokra épülnek a fejlesztés, a tesztek és a visszajelzések alapján.

Az MI-projekt szakaszai

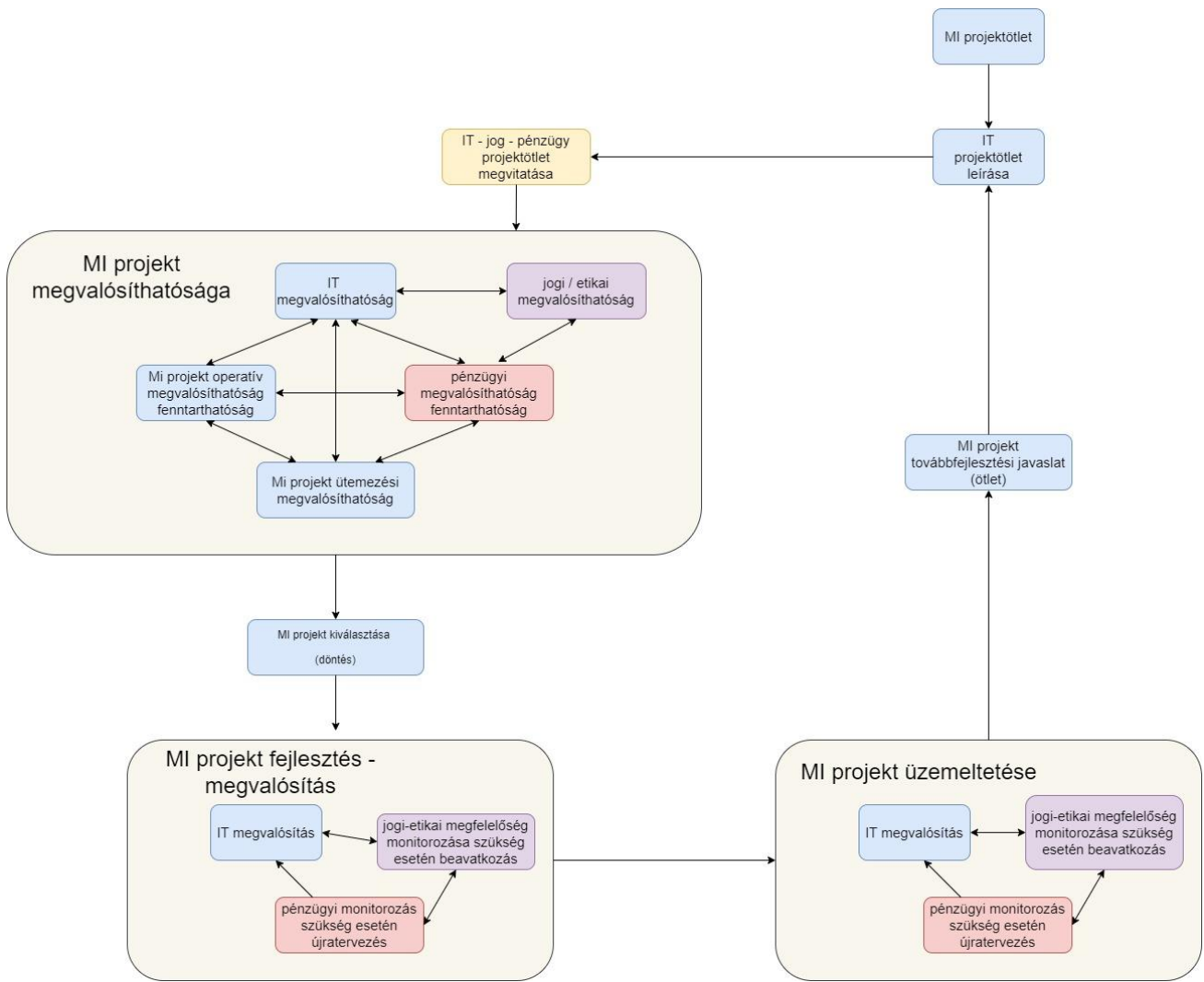
Általában több, egymást átfedő szakaszból áll, amelyek nem minden esetben követik ugyanazt a sorrendet (4. sz. ábra). Tipikus főbb szakaszok:

1. *A projekt céljainak, a rendszer követelményeinek, illetve a várható felhasználói elvárások meghatározása (adatigények, rendelkezésre álló erőforrások és kívánt funkciók meghatározása stb.)*
2. *Adatgyűjtés és adatelőkészítés*

adatokra van szükség a modell betanításához, teszteléséhez és validálásához (gyűjtés, megtisztítás, normalizálás, előkészítés stb.)

3. *A modellfejlesztés és betanítás*
kiválasztjuk a megfelelő algoritmusokat és fejlesztjük a modellt, például hiperparaméterek beállítása (ez automatizálható), a modell adatokon betanítása, validációs tesztelés a modell teljesítményének értékelésére)
4. *A modell kiértékelése* előre meghatározott teljesítménymutatók és kritériumok alapján
5. *Telepítés és integráció*
a sikeres kiértékelés után telepítjük a modellt a célhardvereken vagy a célgépeken (szoftverkomponensek és API-k beállítása, a rendszer integrálása a meglévő infrastruktúrába vagy alkalmazásba)
6. *Üzemeltetés és karbantartás*
a modell folyamatos üzemeltetést és karbantartást igényel (modellfrissítések, adatok további gyűjtése, a modell teljesítményének figyelemmel kísérése és ellenőrzése)
7. *Felhasználói visszajelzés és iteráció*
a modell bevezetése után a felhasználói visszajelzések és tapasztalatok nyomonkövetése, lehetővé téve a modell finomhangolását, a hibák javítását, valamint az új igények kielégítését. Az iterációs folyamatban újra visszatérhetünk a korábbi szakaszokhoz a fejlesztés és optimalizálás érdekében.

Egy adott projekt sokkal részletesebb, szerteágazóbb is lehet, illetve az MI-modell fejlesztési és telepítési folyamatai jelentős mértékben függenek az alkalmazás konkrét céljától és a rendelkezésre álló erőforrásoktól.



4. sz. ábra: Az MI-projekt szakaszai a projekt három fő szereplőjének (IT, pénzügy, jog) együttműködése alapján
 Forrás: Saját szerkesztés

Mennyivel másabb egy MI-, mint egy hagyományos IT-fejlesztés?

Az MI-fejlesztésekre jellemző:

- *a projekt komplexitása*
 gyakran van szükségük nagyobb adathalmazokra, összetett algoritmusokra és tanulási folyamatokra a gépi tanulás vagy mély tanulás alkalmazása révén, illetve figyelembe kell vennünk a specifikus adatigényeket, a modell architektúráját és finomhangolását is
- *a modell komplexitása*
 az alkalmazott modellek és algoritmusok általában komplexebb és specifikusabbak, mint egy hagyományos IT-fejlesztés esetében

- *az automatizáció szintje*
 gyakran magában foglalja az automatizáció magasabb szintjeit, ahol az adatok alapján az intelligens rendszerek képesek önálló tanulásra és döntéshozásra
- *adatigény*
 jelentős mennyiségű adatot gyűjtünk, készítünk elő és dolgozunk fel a modell kifejlesztéséhez és betanításához (az adatok minősége, sokszínűsége és mérete kritikus a modell hatékonyságában és teljesítményében) (5. sz. ábra)



5. sz. ábra: Adatigény
 Forrás: Az ábra alján

- *algoritmusok és modellek*
 speciális algoritmusok és modellek alkalmazása jellemző; gépi tanulási algoritmusokat, mint például döntési fákat, neurális hálózatokat vagy szupport vektor gépeket használhatunk a tanulási és előrejelzési feladatokra
- *adattárolás és adatkezelés*
 gyakran nagy mennyiségű adatot tárolunk és kezelünk
- *tanulási fázis*
 általában van olyan fázis, amikor az algoritmusok és modellek az adatokból tanulnak
- *tanulás és iteráció*
 a modell kifejlesztése és betanítása során gyakran van szükségünk többszöri iterációra, finomhangolásra és validációra
- *prediktív és adaptív képességek*
 Gyakran prediktív és adaptív képességeket céloznak meg (a modell képes előre jelezni és a környezeti változásokhoz alkalmazkodni)
- *etikai és jogi kérdések*
 számos etikai és jogi kérdés merülhet fel, mivel a modellek döntéseket hozhatnak, nagymennyiségű személyes adatot kezelhetnek és jelentős hatással lehetnek az emberek életére és magánszférájára
- *átláthatóság és megmagyarázhatóság*
 gyakran felmerül az átláthatóság és az interpretálhatóság követelménye, azaz a modell megmagyarázható és érthető legyen
- *kockázatok és biztonság*
 a modellek új kockázatokat hoznak magukkal (pl. ENISA 2023a), például annak veszélyét, hogy előre nem látott módon viselkednek vagy félreértelmeznek adatokat.

Milyen célból fejlesztünk LLM-modelleket?

Az LLM képes bonyolult nyelvi feladatokat elvégezni, mivel rengeteg adatot tanul meg és azokon keresztül szövegek összefüggéseit, mintáit és jelentéseit tudja értelmezni. Például:

- szövegenerálás (rövidebb-hosszabb szövegek létrehozása, ideértve cikkeket, novellákat, versikeket és egyéb tartalmakat)
- gépi fordítás
- szövegösszefüggések megértése (pl. keresőmotorok, szemantikai elemzések)
- chatbotok és virtuális asszisztensek (értelmes és emberi válasz adása a felhasználóknak)
- tudásbázisok feltöltése, a felhasználók kérdéseire pontos válaszok adása
- szövegátalakítás és formázás, nyelvhelyességi hibák javítása, stilisztikai módosítások elvégzése
- jogi dokumentumok elemzése, releváns információk kinyerése
- orvosi diagnózisok és kutatás (orvosi szövegek elemzése, betegségek diagnosztizálásának és orvosi kutatások támogatása)
- nyelvtanulás és oktatás
- közösségi média elemzés, érzelmek felismerése, trendek azonosítása
- játékfejlesztés (párbeszéd, háttértörténetek és karakterek játékokhoz)
- célcsoport-specifikus tartalmak létrehozása, márkauzenet hatékony kommunikációja
- reklám- és piackutatás (fogyasztói viselkedés és preferenciák elemzése, reklámstratégiákhoz információk kidolgozása)

- történelem- és kultúrakutatás (történelmi szövegek fordítása és elemzése, segítség a kulturális összefüggések és trendek megértéséhez)
- időjárásjelentések és adatok elemzése
- finanszírozási és gazdasági elemzés (pénzügyi jelentések és gazdasági adatok elemzése, támogatva a befektetési és üzleti döntéseket)
- egyéni hangszín és stílus szintézis (egyedi hangszínnel és stílussal generált szövegek létrehozása)
- politikai elemzés és közvéleménykutatás (szövegek elemzésével a politikai vélemények és trendek feltárásának támogatása).

Ki és hogyan határozza meg az MI-projektek követelményeit?

A követelmények meghatározása rendszerint többlépcsős folyamat, amelyben:

- az ügyfél vagy (üzleti) tulajdonos képviseli a végfelhasználói igényeket, meghatározza a projekt céljait és üzleti stratégiát
- a projektmenedzser felelős a projekt szervezéséért, a kommunikációért és az aktuális feladatok végrehajtásáért, valamint elősegíti a projekt szereplői közötti együttműködést
- az üzleti elemzők felmérik és elemzik az ügyfelek és végfelhasználók igényeit, elemzik a folyamatokat, valamint dokumentálják a funkcionális és nem funkcionális üzleti követelményeket
- a fejlesztők és technikai szakemberek a technikai megvalósíthatóság szempontjából értékelik a követelményeket, támogatják az üzleti követelmények technikai megfogalmazását és lefordítását (pl. architektúra tervezése, rendszereszközök és technológiák kiválasztása, a rendszer kapcsolódási pontjainak meghatározása stb.).

Az MI, illetve az LLM-projektekben a jogászok bevonása különösen akkor kritikus, ha az adott projekt nagy számú személyes adat kezelését igényli. Jelen tanulmányomban az ilyen, számos személyes adatot kezelő LLM-projektekre összpontosítok, tekintettel arra, hogy jogi és etikai szempontból ezek általában magas kockázatúak, illetve jelentős kockázatot jelent az adatvédelmi felügyeleti hatóságok szankcionálási gyakorlata is. Azonban a személyes adatokkal nem dolgozó LLM-modellekkel kapcsolatban is számtalan jogi probléma merülhet fel, például szerzői és szomszédos jogi, versenyjogi, fogyasztóvédelmi és termékfelelősségi területen.

Mi a jogászok feladata az LMM követelményeinek meghatározása során?

Optimális esetben:

- azonosítják a jogi kötelezettségeket, korlátokat, valamint az etikai kérdéseket
- azonosítják a jogi kockázatokat és közreműködnek azok kezelésében
- meghatározzák az adatvédelmi és adatbiztonsági követelményeket (pl. hozzájárulások kezelése, személyes adatok kezelésének korlátozása, panaszbenyújtás módjainak kialakítása és a beérkező panaszok kezelése, tiltakozási jog gyakorlásának biztosítása és a jogi dokumentációk elkészítése)
- előkészítik, illetve felülvizsgálják a szerződéseket, különösen, ha különböző típusú együttműködésre vagy adatmegosztásra kerül sor
- biztosítják – az elszámoltathatóság elvének megfelelően – a jogi szempontoknak megfelelő szerződéskötési és dokumentálási gyakorlatot (pl. közös adatkezelésekkel és adatfeldolgozásokkal kapcsolatos megállapodások), közreműködnek a felhasználói, a licenc- és egyéb szerződések kidolgozásában
- közreműködnek a projektben résztvevő felek jogi felelősségének meghatározásában
- megfelelő jogi keretek kidolgozásával hozzájárulnak a potenciális jogi viták megelőzéséhez.

Egy LLM számos, különböző szakterületen dolgozó jogász együttműködését igényelheti, a személyes adatokkal dolgozó modellek esetében pedig elengedhetetlen az adatvédelmi és adatbiztonsági szakjogászok bevonása olyan területeken, mint:

- az adatkezelési folyamatok jogi megfelelőségének biztosítása, illetve az ezzel kapcsolatos adatvédelmi tájékoztatók és hozzájárulási mechanizmusok kidolgozása
- a beépített adatvédelem és adatbiztonság követelményeinek érvényesítése, szabályzatok és eljárások kidolgozásának, adatbiztonsági intézkedések megvalósításának, adatvédelmi incidensek kezelésének és az érintettek érdekeinek, jogainak és szabadságainak érvényesülésének támogatása
- adattovábbításhoz, illetve adatmegosztáshoz szükséges megállapodások kidolgozása
- jogos érdekek hivatkozó adatkezelések esetében érdekmérlegelési tesztek végzése, valamint a tiltakozási jog érvényesítésének elősegítése
- amennyiben a modell az érintettek jogaira és szabadságaira nézve jelentős kockázatot

adatkezeléseket tartalmaz, adatvédelmi hatásvizsgálat elvégzésének, valamint a szükséges kockázatenyhítő intézkedések kidolgozásának támogatása

- adatvédelmi incidensek kezelése (kockázatkezelés, kárelhárítás és kárenyhítés, jogszabályban előírt kötelezettségek teljesítése, például felügyeleti hatóság értesítése, érintettek tájékoztatása stb.)
- olyan tartalmak vagy tevékenységek azonosítása és kezelése, amelyek jogellenesek vagy jelentős jogi kockázatokat hordoznak
- a jogszabályi környezet és a jogalkalmazási gyakorlat változásának nyomon követése
- adatvédelmi tudatosság növelése a saját alkalmazottak és az adatfeldolgozók körében.

A jogászok azonban csak akkor „hasznosak”, ha optimális időben és mértékben vonják be őket, illetve, ha együtt tudnak működni a többi szereplővel.

Jogászok és informatikusok együttműködése

Ha megkérdezzük informatikusokat, hogyan tudnak együtt dolgozni a jogászokkal, valószínűleg visszakérdeznek, egyáltalán van-e olyan, aki képes erre. Ha pedig a jogászoknál érdeklődünk, hogyan működnek együtt az informatikusokkal, ne csodálkozzunk, ha a fekete lyukak keletkezéséről kezdenek el beszélni. Ezen ellenérzéseket leküzdve fontos, hogy a jogászok már a követelmények meghatározásakor jelen legyenek, és ne a projekt majdnem kész állapotában közöljék, hogy a súlyos jogi hiányosságok miatt vissza kell térni a rajtkockára és újra kell kezdeni az egészet, ezúttal a jogi-etikai kereteket is figyelembe véve. Még rosszabb, ha az adatvédelmi felügyeleti hatóság a szankcionáló határozatában közli, milyen jogi problémái vannak a modellel, és szólít fel minket a korrekciós intézkedések haladéktalan megtételére.

Milyen következményei lehetnek az együttműködés elmaradásának?

Például:

- nem vesszük kellő mértékben figyelembe a jogi követelményeket és korlátokat, illetve nem felelünk meg az etikai normáknak
- a félreértések, az információhiány és egyéb kommunikációs problémák akadályozhatják a hatékony és eredményes projektmenedzselést, illetve a tényleges munkavégzést

- a jogi és technikai szempontok összehangoltságának hiányában a projekt irányítása eltévedhet, ez pedig kihatással lehet a határidőkre, a költségekre és a minőségre, illetve jogi vitákhoz vezethet
- a modellünk olyan funkciókat tartalmazhat, amelyek jogellenesek vagy erkölcsileg aggályosak.
- Milyen lehetőségek vannak az együttműködésre az LLM-projekt kezdeti időszakában?

Például:

- rendszeres és intenzív kommunikáció a projekttagok között, például a közös műhelymunka és megbeszélések alkalmat adnak a tapasztalatok, nézőpontok, prioritások és követelmények megtárgyalására, valamint az üzleti célok, a technikai megvalósítás és a jogi előírások harmonizálására. A kevésbé formális párbeszéd elősegíti az egymás jobb megértését, valamint csökkentheti a „pusztába kiáltott” szavakat
- a követelmény- és funkcionális specifikációk, valamint egyéb projektdokumentációk rendszeres egyeztetése lehetővé teszi a jogi-etikai követelmények korrekt beépítését, a kockázatok azonosítását, a diszkrimináció, illetve a tisztességtelen működés megelőzését
- a projektre vonatkozó belső szabályzat, valamint hatásvizsgálatok és érdekmérlegelési tesztek előkészítése.

Mikor kezdődjön az együttműködés?

Az az optimális, ha az informatikusok minél korábban konzultálnak a jogászokkal. Az ötlet informális egyeztetése lehetőséget ad a jogi-etikai aggályok korai megfogalmazására, az informatikusok pedig elkerülhetik a nem megfelelő irányba gondolkodásból adódó felesleges munkát. Ehhez azonban az szükséges, hogy az ötlet megvalósíthatósága megítélhető, a jogi kockázatok pedig beazonosíthatóak legyenek. Ehhez az informatikusoknak meg kell osztaniuk az ötletüket, valamint az alapvető információikat a modell működésének módjáról, a tervezett adatkezelésekről és az esetleges harmadik felekkel való együttműködésről – ezen ismeretek hiányában a jogászok nem tudnak megfelelő véleményt kialakítani.

Számos aggály merülhet fel, például:

- Hogyan magyarázható majd meg az LLM működése? Ha nem megmagyarázható (pl. feketedoboz modellek), hogyan tudunk megfelelni a megmagyarázhatóság követelményének?

- Mi szükséges az adatvédelmi és MI alapelveknek megfelelésünkhöz? Hogyan oldható meg például az ember jelenléte, az érintetti kérelmek és panaszok kezelése?
- Lesznek megfelelő adatok a betanításhoz, teszteléshez, validáláshoz? Hogyan csökkenthetjük ezen adatok kezeléséből származó kockázatokat, például megoldható az adatok anonimizálása?
- Hogyan felelünk meg az etikus MI és a GDPR automatikus döntéshozatal követelményeinek?
- Hogyan kerülhetjük el az algoritmus torzításából eredő problémákat, hogyan lesz a modellünk diszkriminációmentes, megbízható és etikus?
- Ki fogja tesztelni és ellenőrizni a modellt?
- A modellünk mely pontjai a legsebezhetőbbek (kiber)biztonsági szempontból és hogyan oldható meg ezen kockázatok csökkentése? Milyen alapvető intézkedésekre lesz szükségünk és azok hogyan valósíthatók meg?

A párbeszédet mindaddig fent kell tartanunk, amíg tisztázunk a követelményeket – ez az együttműködés a kulcs a jelentős pénzügyi kihatással járó jogi problémák, felesleges visszalépések és újratervezések megelőzéséhez.

A jogászok blokkolhatják az informatikusok kreativitását?

Gyakori vád, hogy a jogászok jönnek, elmondják ki, mit nem tehet, és ezzel kiölik az informatikusokból a kreativitást. De vajon ez igaz?

Tény, a jogászok nagyon lelombozóak tudnak lenni és néha az informatikusoknak az is felfoghatatlan, hogy vannak olyan előírások, amelyek túlmutatnak szakterületükön, ennek ellenére mégis vonatkoznak rájuk. Például nem lehet bármelyik felhőbe szabadon fellőni a személyes adatokat és egyáltalán nem lényegtelen az, hogy a világ mely pontján, milyen joghatóság alatt tároljuk a személyes adatokat.

A jogászok bevonása azonban nem feltétlenül blokkolja az informatikusok kreativitását, hanem – mindkét szakterület részéről építő kommunikáció esetén – akár még segítheti is az ötleteknek és magának a modellnek a hatályos jogi keretek közötti kreatív megvalósítását. A jogi korlátok figyelembevétele nem feltétlen jelenti az informatikusok gúzsba kötését, netalán az alkotási vágyuk kiölését. Ez azonban a bevont jogász szakmai kvalitásán is múlik, hiszen azzal még a legkiválóbb informatikus sem tud mit kezdeni, ha egy jogász egy problémát látván felolvassa a jogszabályt, majd kategorikusan kijelenti, amit az informatikus szeretne, azt nem lehet. Az inspiráló párbeszéd lényege éppen az, hogy nem azt kell ismételtetni, mit, hogyan

nem lehet megcsinálni, hanem azt kell átgondolni, mit, hogyan lehet – ehhez pedig az informatikusnak is képesnek kell lennie megértenie, mi az adott jogszabályi korlátnak az oka és hogyan lehet elérni, hogy az ne vonatkozzon az adott esetre. Ha a jogász a korlát lényegét nem magyarázza el, vagy úgy adja azt át, hogy azt az informatikusnak esélye sincs megértenie, a megoldás fényévnyi távolságra kerülhet, a projekt pedig vakvágányra futhat. Éppen ezért a jogásznak úgy kell megfogalmaznia mondandóját, hogy a nem jogászok számára is emészthetővé tegye, valamint a megoldási javaslatokról el kell döntenie, beleférnek-e az irányadó jogi-etikai keretekbe. Ehhez azonban nemcsak biztos szakmai tudás kell, hanem felelősségvállalási készség is. A közösen értelmezett jogi-etikai keretek medret adnak a projektnek, az informatikusoknak pedig át kell gondolniuk, sőt időnként át is kell értékelniük azt, hogyan közelítenek az olyan értékekhez, mint az emberi jogok meg az etikus MI.

A gyakorlatban ez a kommunikáció sokkal nehezebb, mint amilyennek látszik. Nemcsak az egymás iránt érzett ellenszenvet kell leküzdeniük a szereplőknek, hanem kreatívan kell felhasználnunk a rendelkezésre álló platformokat.

Milyen eszközöket használhatunk a jogászok és az informatikusok egymáshoz közelebb hozására?

Elsősorban hagyományos projekteszközöket, például:

- az együttműködést hatékonyan támogató szervezeti megoldásokat (rendszeres interdiszciplináris oktatások, a kommunikációra, valamint a közös területekre fókuszáló workshopok, csapatépítő tréningek). A kulcsszerep a projektvezetőé, illetve azé a személyé, akinek a feladata az eseményeken résztvevő szereplők bevonása az együttműködésbe, és aki nem hagyja, hogy bárki is hiúsági versenyt rögtönözzön vagy netalán szótlanul üldögéljen a sarokban.
- olyan csatornákat, amelyek alkalmasak a rendszeres és strukturált kommunikációra (rendszeres megbeszélések, közös munkacsoportok), segítve a kapcsolattartást, de nem engedve teret az ellenségeskedésnek, a rivalizálásnak és a mellébeszélésnek
- nyelvi egyensúly megteremtését egymás szakmai terminológiájának megértése érdekében
- tudásmegosztást, egymás szempontrendszerének megértése érdekében

- mentorálás rendszeresítését, a jogászok-informatikusok együttműködő párokba szervezését az információáramlás, a tudásmegosztás, valamint a minőségi emberi kapcsolatok támogatására
- szerepek és felelőségek tisztázását (mindenki tisztában legyen a saját helyzetével és senki se törekedjen túlzott dominanciára)
- írásbeliséget a naprakészen tájékoztatás, valamint a döntések, szabályozások és egyeztetések dokumentálása érdekében
- külső szakértők, például etikai tanácsadók bevonását, újfajta perspektívát és tudást hozva a projektekbe
- a felsővezetés dominanciáját, segítve a különutas csapattagok „közös útra” terelését, valamint a határidők betartását.

Olyan eszközöket célszerű használnunk, amelyek – lehetőleg – személyes jelenléte igényelnek, komplexek és gondolkodásra, valamint a keretek és a projekt egészének átgondolására ösztönöznek. Nem kell feltétlenül új eszközökben gondolkodnunk, átértékelhetünk és kreatívabb formában

Milyen követelményeknek kell megfelelnünk az LLM fejlesztése során?

Az Európa Bizottság mesterséges intelligenciával foglalkozó magas szintű szakértői csoportja (szakértői csoport) szerint az MI alkalmazása során – többek között, de nem kizárólag – az alábbi követelményeknek kell megfelelnünk (Mesterséges intelligenciával foglalkozó magas szintű szakértői csoport 2019):

- az emberi cselekvőképesség támogatása és emberi felügyelet
- műszaki stabilitás és biztonság
- adatvédelem és adatkezelés (a magánélet tiszteletben tartása, az adatok minősége és sértetlensége, valamint az adatokhoz való hozzáférés)
- átláthatóság (nyomon követhetőség, megmagyarázhatóság és tájékoztatás)
- sokféleség, megkülönböztetésmentesség és méltányosság
- társadalmi és környezeti jólét (fenntarthatóság és környezetbarátság), valamint
- elszámoltathatóság (ellenőrizhetőség, hátrányos hatások minimalizálás, jogorvoslat).

A hagyományos IT-projektekben a megvalósíthatósági tanulmány rendszerint öt egymásra épülő elemből áll:

- **technikai megvalósíthatóság** (a rendelkezésünkre álló eszközökkel és szakemberekkel)
- **gazdasági megvalósíthatóság** (költség- és hasznóértékelés, beruházási, üzemeltetési és karbantartási költségek, valamint az elvárt megtérülés és nyereség elemzése)
- **üzleti megvalósíthatóság** (a projekt illeszkedése az üzleti stratégiánkhoz, céljainkhoz és igényeinkhez)

hasznosíthatunk számos, az LLM-projektek során egyébként is „kötelező” eszközt, például:

- a megvalósíthatósági tanulmányokat
- a hatásvizsgálatokat
- az érdekmérlegelési tesztek
- a projekt végrehajtásával kapcsolatos belső szabályzatot, valamint
- az adatvédelmi és adatbiztonsági ismeretek oktatását.

Jelen esetben a hangsúly nem ezen eszközök ajánlott/kötelező jellegén, hanem azok kicsit másként alkalmazásán van, céljuk pedig a konstruktív párbeszéd, egymás megértése és elismerése úgy, hogy mindeközben a jogi-etikai kereteket is megteremtjük.

A megvalósíthatósági tanulmány

A megvalósíthatósági tanulmány alapvető projektesszköz egy adott LLM megvalósíthatóságának elemzésére a projekt kezdeti fázisában még a tényleges fejlesztés megkezdése előtt. Gyakran több alternatívát is tartalmaz, objektív információkat és elemzéseket szolgáltatva a projekt előnyeiről, hátrányairól, korlátairól és kockázatairól.

- **jogi megvalósíthatóság** (az eredmény megfeleljen a jogszabályoknak és egyéb előírásoknak)
- **operatív megvalósíthatóság** (a projekt hatása szervezetünk napi folyamataira, szükséges erőfeszítések a fenntartás érdekében)
- **megvalósíthatóság ütemezése** (reális határidők a betartás és a betartatás érdekében).

Javasolt hatodik területként a fenntarthatóságot is beiktatnunk a megvalósíthatósági tanulmányba („Ecology by Design”), kitérve olyan témákra, mint a digitális technológia közvetlen negatív környezeti hatásainak csökkenése, jobb gazdálkodás a természeti

erőforrásokkal, illetve a digitális és egyéb szereplők közötti kapcsolatok.

Ezen építőkövek egymásra épülnek és egymást feltételezik – jogi megfelelés hiányában a technikai megvalósítás irreleváns, ha pedig nem rentábilis az adott projekt, nem érdemes gondolkodni benne még akkor sem, ha egyébként az informatikusok és a jogászok szerint minden rendben. A tanulmány alapján a döntéshozók meghatározzák a projekt érdemességét és a kockázatok kezelésének módját, valamint alapul szolgálhat a projektfinanszírozáshoz és a megvalósítási tervhez is.

Az MI-projektek megvalósíthatósági tanulmánya hasonló, ám komplexebb, például a jogi mellett az *etikai megvalósíthatóságra* is kiterjed, és vizsgáljuk a szükséges adatok beszerezhetőségét, felhasználhatóságát, a pénzügyi fedezetet és a beszerzés, valamint a felhasználás jogszerűségét-etikusságát.

A megvalósíthatósági tanulmány elkészítése során:

- az informatikusok elmagyarázzák mit szeretnének
- a pénzügyesek kiszámolják, amit az informatikusok kitaláltak, az rentábilis-e
- az értékesítők kikutatják, van-e piaci igény a modellünkre (amennyiben értékesítésre szánjuk)
- a jogászok meghatározzák a jogi-etikai kereteket (az informatikusok ötlete jogilag megvalósítható-e, és ha igen, hogyan)
- a biztonsági szakemberek eldöntik, megoldható-e a projekt védelme információbiztonság, illetve az üzleti titok védelme szempontjából stb.

Mire a hatástanulmány elkészül, a szereplőknek módjuk van megismerni egymás gondolkodásmódját, érveit, valamint kedvenc vesszőparipáit is, és arra is rájöhetnek, hogyan lehet a másik vitorlájából kifogni a szelet. Az is egyértelművé válik, kik kulcsfontosságúak a projekt szempontjából, kikkel „lehet együtt dolgozni” és kik azok, akikkel erre esély sincs. Sőt, ez utóbbiak cseréjére is sor kerülhet, ha nincs esély az érdemi javulásra.

Hatásvizsgálatok végzése

A hatásvizsgálat során különböző aspektusból felmérjük és értékeljük egy adott projekt lehetséges gazdasági, társadalmi, valamint környezeti hatásait, a jogi és szabályozási követelményeket, és az etikai-adatvédelmi kérdéseket. A folyamatba bevon(hat)juk az érintett feleket, például a közösség tagjait, a felhasználókat, az ügyfeleket, valamint különféle szakértőket is.

A hatásvizsgálat pozitív hozadéka, hogy

- azonosíthatjuk és elemezhetjük a kockázatokat és a problémákat,
- elősegíthetjük a felelősségteljes döntéshozatalt, illetve a fenntarthatóságot,
- meghatározhatjuk a szükséges korrekciós intézkedéseket,
- megismerhetjük és figyelembe vehetjük az érintett felek érdekeit és szükségleteit.



6. sz. ábra: Hatásvizsgálat

Forrás: <https://www.behance.net/gallery/3754298/Privacy-Cartoons>

Többféle hatásvizsgálatot végezhetünk, ezek összefügghetnek, átfedhetnek, de akár ki is egészíthetik egymást (pl. az adatvédelmi, az alapjogi és az etikai hatásvizsgálat). Ezek nem egyszeri alkalmak, hanem folyamatok, amelyeket nem együttő helyünkben kell

elvégeznünk felesleges adminisztratív nyűgként letudva, hanem folyamatosan nyomunkövetjük a körülmények alakulását a modellünk teljes életciklusa alatt. A hatásvizsgálatok alapján korrekciókra lehet szükségünk azért, hogy az LLM a lehető legmegfelelőbben

illeszkedjen környezetébe, illetve minimalizáljuk a potenciális kockázatokat és hatásokat.

Milyen hatásvizsgálatokat végezhetünk?

- Az *etikai hatásvizsgálatban* az etikai kockázatokat és következményeket vizsgáljuk (diszkrimináció és elfogultság, a magánélet és a személyes adatok védelme, az autonóm döntéshozatal és az emberi értékek tiszteltben tartása).
- A *társadalmi hatásvizsgálatban* a társadalmi, gazdasági és kulturális hatásokat elemezzük (pl. a munkahelyek és foglalkoztatás változása, szociális egyenlőtlenségek, társadalmi befogadás és kirekesztés).
- Az *adattvédelmi és adatbiztonsági hatásvizsgálat* (DPIA) követelményeit a GDPR²¹, illetve a LED²² fekteti le, célja pedig az érintettek jogait és szabadságait érintő adattvédelmi és biztonsági kockázatok azonosítása és minimalizálása.
- A *jogszabályi és szabályozási hatásvizsgálatban* felmérjük és elemezzük a vonatkozó jogi előírásokat, szabályozásokat és jogalkalmazási gyakorlatot, ezek hatásait, illetve az ezekből eredő kockázatokat.
- Az *alapjogi hatásvizsgálatot* az MI használatáról szóló rendelet tervezete határozza meg (29a. cikk). Amennyiben kötelező a DPIA elvégzése, az adattvédelmi és az alapjogi hatásvizsgálatokat együtt kell elvégeznünk.

Számos hatásvizsgálati modellt az adattvédelmi hatásvizsgálatra építi rá rétegenként a többi, esetről esetre – az adott projekthez igazodva –, eltérő tartalommal.

„Minden egyben” hatásvizsgálat

Nemzetközi szinten számtalan szervezet, számtalan hatásvizsgálat mintát tett közzé, ezek közül kiemelendő a kanadai kormány *algoritmikus hatásvizsgálata*²³ (Government of Canada, 2023), melynek célja, hogy segítsen az algoritmusok és az MI-rendszerek hatásainak felmérésében, értékelésében és a bevezetéssel járó hatások enyhítésében, biztosítva az algoritmusok társadalmi hatásainak átláthatóságát és ellenőrizhetőségét. Az AIA kérdőív mintegy 80 kérdésből áll az üzleti folyamatokra, az adatokra, a rendszertervezésre, az algoritmusra és a döntésre vonatkozóan, a válaszok alapján pedig megjeleníti az adott projekt hatásszintjét, illetve tájékoztatást nyújt az irányadó követelményekről.

Az AIA:

- segít megérteni az algoritmusok és MI modellek potenciális hatásait az emberekre, a társadalomra és a szervezetünkre, lehetővé téve a projektek felelős tervezését, valamint az esetleges kockázatok és torzítások azonosítását
- elősegíti a párbeszédet és az átláthatóságot a projekt résztvevői és a szélesebb közösség között, valamint lehetőséget nyújt a szükséges korrekciók és javítások végrehajtására a tervezés korai szakaszában
- javítja az algoritmusokkal és MI-rendszerekkel kapcsolatos döntéshozatali folyamatokat és segít a technológia iránti bizalom építésében.

Az AIA „minden egyben” hatásvizsgálat, amely

- vizsgálja az adatok kezelésének módját a tervezés és működtetés során, valamint az adatgyűjtési, tárolási és feldolgozási gyakorlatokat, illetve az adattvédelmi irányelveknek megfelelést
- segít értékelni, megakadályozni, illetve csökkenteni a diszkriminációt, az előítéleteket, a torzításokat és az igazságtalanságokat
- felméri a társadalmi hatásokat, ideértve a szélesebb társadalmi és gazdasági következményeket (emberi jogok, munkaerőpiac, társadalmi egyenlőtlenségek, egészségügy és más területeken)
- támogatja a felelős és etikus tervezést, segít az értékalapú döntések meghozatalában, a társadalmi normák és értékek figyelembevételében, illetve az etikai irányelvek és keretrendszerek kidolgozásában
- támogatja az átláthatóságot és a számonkérhetőséget, valamint segít nyomon követni az algoritmusok működését és a döntéshozatali folyamatokat
- értékeli a biztonságot (potenciális sebezhetőségek, az adatvesztés kockázata, hozzáférési jogosultságok és az adatok védelme a külső fenyegetésekkel szemben)
- figyelembe veszi a felhasználói élményt és a felhasználhatóságot, illetve a felhasználók igényeihez és képességeihez illeszkedést
- felméri a gazdasági előnyöket és hatékonyságot, értékeli a költségeket és a megrterülést, a termelékenységet, illetve a piaci versenyképességet
- felhívja a figyelmet a jogi és szabályozási kérdésekre, értékeli a releváns jogi kereteket és a megfelelést („compliance”).

²¹ GDPR 35. cikk

²² bünyügyi adattvédelmi irányelv (LED)

²³ Algorithmic Impact Assessment, továbbiakban AIA

Az AIA szabadon elérhető, akár saját gyakorlatunkban is hasznosíthatjuk.

Az etikai hatásvizsgálat

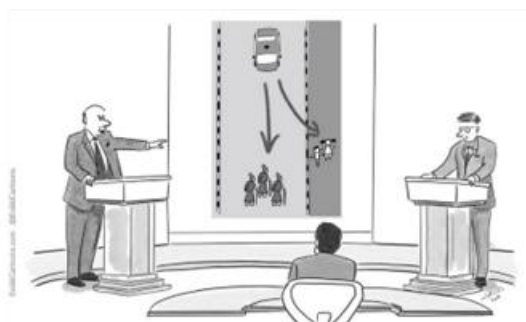
„Általában a technológiáink az egyik kezükkel adnak, a másikkal pedig úymond tarkón vágják minket.”
(Charlie Brooker, a *Black Mirror* készítője) (Kafka 2023)

Az etikai hatásvizsgálatban feltárjuk és felmérjük egy adott projekt vagy technológia etikai kockázatait és azok következményeit. Amennyiben a szakértői csoport etikai iránymutatását vesszük irányadónak, akkor a megbízható LLM jogszerű, etikus, valamint műszaki és társadalmi szempontból is stabil annak érdekében, hogy lehetőleg ne okozzon kárt (Mesterséges intelligenciával foglalkozó magas szintű szakértői csoport 2019).

Miért fontos, hogy az LLM etikus legyen?

Amikor intelligenciát igénylő tevékenységet végzünk, felelősséget kell vállalnunk ítéleteink pontosságáért, megbízhatóságáért és megalapozottságáért, valamint az is elvárás, hogy cselekedeteinket és döntéseinket megfelelően támasszuk alá, illetve másokkal szemben tisztességesen, méltányosan és észszerű módon járjunk el.

Az LLM megjelenése és intelligenciát igénylő tevékenységekhez szükséges hatalmának növekedése a kognitív funkciók széles körének algoritmikus folyamatokra való átruházását jelenti, amelyekért ezek a modellek maguk közvetlenül nem felelősek, és nem is vonhatók azonnal felelősségre viselkedésük következményeiért (7. sz. ábra).



„Öld meg az időseket, hogy megmentsd a gyerekeket? Micsoda szörnyű kijelentés!
Én azt mondom, az autónak megkülönböztetés nélkül kell ölnie.”

<https://www.moralmachine.net/>

7. sz. ábra: LLM és etikai felelősség
Forrás: Az ábra alján

Az LLM és más MI-modellek, mint inaktív és programalapú gépek erkölcsileg nem elszámoltatható ágensek, ezt az etikai rést pedig jelenleg az MI etikájának egyre növekvő kerete próbálja betölteni olyan elvek érvényesítésének megkövetelésével, mint a tisztességesség, az elszámoltathatóság, a megbízhatóság, a fenntarthatóság és az átláthatóság.

A Holberton-Turing-eskü célja, hogy az MI szakértőket közös értékek ernyője alá gyűjtse.

„A HOLBERTON-TURING ESKÜ

Az adattudomány és a mesterséges intelligencia szakma képviselőjeként ünnepélyesen fogadom, életemet az emberiség szolgálatának szentelem.

Humanitás és etika:

Az emberi élet iránti legnagyobb tiszteletet tanúsítom. Nem engedem, hogy az életkor, betegség vagy fogyatékosság, hitvallás, etnikai származás, nem, nemzetiség, politikai hovatartozás, vallási meggyőződés, faji hovatartozás, szexuális irányultság, társadalmi helyzet vagy bármely más tényező szempontjai közrejátszanak a munkavégzésemben. Nem használom fel tudásomat az emberi jogok és polgári szabadságjogok megsértésére, még fenyegetés esetén sem.

Adattudomány, a mesterséges intelligencia művészete, magánélet és személyes adatok:

Tiszteletben tartom azoknak a tudósoknak és mérnököknek nehezen megszerzett tudományos eredményeit, akiknek a nyomdokaiban járok, és örömmel osztom meg a tudásomat mindazokkal, akik utánam jönnek.

Nem felejtsem el, hogy a mesterséges intelligencia a tudomány mellett művészet is, és hogy az emberi szempontok fontosabbak a technológiai szempontoknál. Tiszteletben tartom az emberek magánéletét és azt, hogy személyes adataikat nem tárják fel a mesterséges intelligencia rendszerek előtt, hogy a világ megismerje azokat.

Nem felejtsem el, hogy nem száraz adatokkal, pusztán nullákkal és egyesekkel találkozom, hanem emberi lényekkel, akiknek a mesterséges intelligenciával való kapcsolata befolyásolhatja a szabadságát, családját vagy gazdasági stabilitását.

Tiszteletben tartom a rám bízott titkokat.

Napi munka és etikett:

Szakmámat lelkiismeretesen és méltósággal gyakorlom. Ápolom az adattudományi és mesterséges intelligencia szakma becsületét és nemes hagyományait.

Megadom tanárainknak, kollégáimnak és diákjaimnak az őket megillető tiszteletet és hálát.

Megosztom tudásomat az emberek javára, az adattudomány és a mesterséges intelligencia fejlődésének érdekében.

Figyelembe veszem munkámnak a méltányosságra gyakorolt hatását mind a történelmi előítéletek állandósításában, amelyet a múltbeli adatokból a jövőre vonatkozó előrejelzésekre való vak kivetítése okoz, mind pedig a gazdasági vagy egyéb egyenlőtlenségeket növelő új feltételek megeremtésében.

Mindezen ígéreteimet elsősorban azért teszem, hogy a mesterséges intelligenciát úgy hozzam létre, hogy az a közjó érdekében együttműködjön az emberekkel, ahelyett, hogy bitorolná az emberi szerepet kiszorítva őket.

Ezeket az ígéreteket ünnepélyesen, szabadon és becsületemre fogadom meg.”

Mikor etikus egy mesterséges intelligencia?

A szakértői csoport szerint négy fő etikai elv betartására kell törekednünk:

1. *Az emberi autonómia tiszteletben tartásának elve*
 - az MI-rendszerekkel nem rendelhetjük alá, nem kényszeríthetjük, nem téveszthetjük meg, nem manipulálhatjuk és nem kondicionálhatjuk indokolatlanul az embereket
 - a modelleket úgy kell kialakítanunk, hogy azok az emberek kognitív, szociális és kulturális készségeit fokozzák, kiegészítsék és megerősítsék
 - az emberek és MI-rendszerek közötti feladatmegosztás elve az emberközpontú kialakítás; az emberi döntésnek jelentős szerepet kell hagynunk, és biztosítanunk kell, hogy az ember felügyelhesse és ellenőrizhesse a modellek munkafolyamatait
 - a kár megelőzésének elve: az MI-rendszerek soha nem okozhatnak kárt, illetve nem lehetnek hátrányos hatással az emberi lényekre, valamint meg kell akadályoznunk a rosszindulatú használat lehetőségét, a természetes környezet és valamennyi élőlény figyelembevételét is ideértve

2. *A méltányosság elve*

Az MI-rendszerek kifejlesztésének, elterjesztésének és használatának méltányosnak kell lennie:

- az előnyök és a költségek egyenlő és igazságos elosztása
- annak biztosítása, hogy az egyéneket és csoportokat ne érje méltánytalan torzítás, hátrányos megkülönböztetés és megbélyegzés
- az oktatáshoz, termékekhez, szolgáltatásokhoz és technológiához való

hozzáférés terén esélyegyenlőség támogatása

- nem eredményezheti a (végső) felhasználók megtevesztését vagy választási szabadságának sérelmét
 - tiszteletben kell tartanunk az eszközök és célok arányosságának elvét, és egyensúly kell teremtenünk a versengő érdekek és célkitűzések között
 - a modellek és az e rendszereket üzemeltető emberek hozta döntések megtámadhatósága (hatékony jogorvoslat lehetősége)
 - a döntésért felelősek azonosíthatók, és a döntéshozatali folyamatok megmagyarázhatók legyenek
3. *A megmagyarázhatóság elve*
 - a modellek képességeit és célját nyíltan közölnünk kell
 - a döntéseket lehetőség szerint el kell magyaráznunk
 - a „fekete doboz” algoritmusok esetében egyéb intézkedésekre, például a rendszer képességeinek nyomonkövethetőségére, ellenőrizhetőségére és az azokkal kapcsolatos átlátható tájékoztatásra van szükségünk
 - a megmagyarázhatóság szükséges mértéke nagyban függ attól, hogy téves vagy pontatlan eredmény esetén a következmények mennyire súlyosak az érintetteknek nézve
- Ezen elvek akár ütközhetnek is egymással, ekkor bizonyítékokon alapuló megfontolt döntésre van szükségünk ahhoz, hogy egyik vagy másik elvet – jogszerűen – háttérbe szoríthassuk.
4. *Stabil mesterséges intelligencia*
- Az egyéneknek és a társadalomnak bíznia kell abban, hogy az MI nem okoz kárt, ennek érdekében óvintézkedéseket kell tennünk.

Mire összpontosít az etikai hatásvizsgálat?

- Az *adat etikussága* (nyomonkövethetőség, elérhetőség, integritás, biztonság, nem szelektív gyűjtés stb.), az *algoritmus etikussága* (megbízhatóság, védettség, cél, torzítás, minőség, megmagyarázhatóság, átláthatóság stb.) a *rendszer etikussága* (ergonómia, alkalmazkodóképesség, következetesség, bizalmasság stb.), a *gyakorlatok etikussága* (kultúra, szabályok, szakmaiság, megbízhatóság, a magánszféra védelme, elszámoltathatóság,

diverzifikációk, integráció stb.), illetve a döntések etikussága (autonómia, szabad akarat, dehumanizáció, menedzsment és irányítás, felelősség, nyilvánosságra hozatal, környezet, fenntarthatóság stb.)

- *diszkrimináció*, illetve elfogultság azonosítása és kezelése
- *adatvédelem és magánélet védelme* (a felhasználói adatok kezelése, hozzáférhetősége és a vonatkozó jogszabályok, valamint belső szabályok betartása)
- *autonóm döntéshozatal* (átláthatóság, felelős módon működés, döntésekért felelősségvállalás)
- *társadalmi hatások* és esetleges *egyenlőtlenségek* (munkahelyek változása, szociális és gazdasági

egyenlőtlenségek, társadalmi befogadás és kirekesztés)

- *felhasználói bizalom és felelősség* (átláthatóság, a felhasználók megért(het)ik-e a modell működését, felelős használat és esetleges korlátozások).

Az etikai szempontokat már a tervezés időszakában figyelembe kell vennünk (*Ethics by Design*), valamint az LLM teljes életciklusában is (*Ethics by Evolution*) (Leroy 2021).

A jogi (jogszabályi megfelelés, adatvédelmi, alapjogi stb.) és etikai hatásvizsgálatok hasonlóak, de különböznek is (8. sz. ábra):

	Jogi hatásvizsgálat	Etikai hatásvizsgálat
Fókusz	A modell megfeleljen a vonatkozó jogszabályoknak, szabályozásoknak és előírásoknak	A modell társadalmi, etikai és morális következményei, az erkölcsi dilemmák és érdekek felismerése, valamint a társadalmi elfogadottság növelése
Szemponatok	Adatvédelmi jogszabályok, szerzői jogok, szellemi tulajdon védelme, felhasználói szerződések és egyéb releváns jogi aspektusok	Felelősségteljeség, a diszkrimináció megelőzése, a személyes adatok védelme, az átláthatóság és az igazságosság a technológiai megoldásokban
Szereplők	Általában jogászok és olyan egyéb szakértők bevonásával, akik képzettek a jogi kérdések elemzésében és az alkalmazandó jogszabályok, valamint a vonatkozó jogalkalmazási gyakorlat megértésében	Általában olyan etikai szakemberek, filozófusok, társadalomtudósok és egyéb szakértők bevonásával, akik hozzáértők a társadalmi és etikai kérdések elemzésében és értékelésében
Lehetséges válaszok, megoldások	Jogszabály által behatárolt, adott esetben lehet, hogy csak egyetlen jogszerű megoldás van	Soha nincs abszolút és egyértelmű válasz; az adott kérdéstől függően vannak legjobb lehetséges választások

8 sz. ábra: A jogi és az etikai hatásvizsgálatok összehasonlítása

Forrás: Saját szerkesztés

Az UNESCO-nak is van olyan etikai hatásvizsgálat iránymutatása, amelyet segítségül hívhatunk (UNESCO 2023).

Míg a jogszabályok értelmezésében könnyebb egységes álláspontja jutni, az etikai kérdések megítélésében sokkal nagyobb eltérések lehetnek például faji, etnikai hovatartozástól, életkortól, az

iskolai végzettségtől, de akár jövedelmi szinttől függően is (9. sz. ábra). Ha azt nézzük, hogy a bibliai tízparancsolathoz, vagy a hét főbűnhöz hozzáállásban mekkora különbségek lehetnek, akkor hogyan juthatunk megegyezésre olyan összetett problémákban, mint például az önvezető autók tipikus dilemmái?



<https://www.moralmachine.net/>

9. sz. ábra: Hovatartozás
Forrás: Az ábra alján

„Jól ismert példa erre az úgynevezett troliprobléma, amely Philippa Foot brit filozófusra vezethető vissza. Ez egy olyan gondolatkísérlet, amely arra hivatott, hogy tesztelje erkölcsi intuíciónkat azzal kapcsolatban, hogy erkölcsileg megengedett-e, sőt szükséges-e feláldozni egy ember életét annak érdekében, hogy több ember életét megmentsük.

Az autonóm járművek szerkezetileg hasonló helyzetekkel szembesülhetnek, amelyekben elkerülhetetlen, hogy mások megmentése érdekében egy vagy több személynek kárt okozzanak, vagy akár meg is öljenek egy vagy több személyt. Tegyük fel, hogy egy önvezető autó nem tud megállni, és csak a választás lehetősége van számára, hogy két embercsoport egyikébe hajtson: egyrészt két idős férfi, két idős nő és egy kutya; másrészt egy fiatal nő egy kislánnyal és egy kislánnyal. Ha az első csoportba hajt bele, a két nő meghal, a két férfi és a kutya súlyosan megsérül. Ha a második csoportba hajt bele, az egyik gyerek meg fog halni, a nő és a másik gyerek pedig súlyosan meg fog sérülni.

A helyzetet tetszés szerint további részletekkel lehet kiegészíteni. Tegyük fel, hogy az idős emberek csoportja a kutyával a közlekedési szabályoknak megfelelően viselkedik, míg a nő és a gyerekek a piros jelzés alatt mennek át az úton. Ez erkölcsileg releváns? Változtatna-e a helyzeten, ha az egyik idős férfi helyébe egy fiatal orvos lépne, aki sok ember életét menthetné még meg?” (Misselhorn 2022: 32–33)

A döntési helyzetet tovább bonyolítja, ha mi ülünk az autóban. Az emberek nagy többsége úgy véli, a halálos áldozatok összességének minimalizálására programozott önvezető járművek etikusabbak – de inkább olyan járművet vásárolnának, amelyik előnyben részesíti a járműben ülő életét. Sőt, az emberek többsége egyenesen elutasítaná az etikusabbnak tartott autó megvásárlását (Bonnefon–Shariff–Rahwan 2016), hiszen

ki venne meg olyan járművet, amelyik bármikor képes feláldozni a saját utasát? Hiába törekednénk a közjó szempontjából a legtökéletesebb megoldásra, ezt az egyéneket önértékelő viselkedése nem fogadná el – azaz az MI-projektünk lehet, hogy etikus, de pénzügyi szempontból garantáltan kudarc. És még olyan kapaszkodónk sincs, mint a jogszabályok távolságtartó, „száraz” paragrafusai vagy az irányadó jogalkalmazási gyakorlat. Az etikai hatásvizsgálat éppen ezért lelkileg nagyon megterhelő is lehet, különösen akkor, ha emberélet feletti döntésről van szó.

Mi a szerepük a jogászoknak és az informatikusoknak az etikai hatásvizsgálatban?

A jogászok támogatják az etikai döntések jogi vonatkozásainak értékelését-értelmezését, és a felelősségi problémák, valamint a következmények meghatározását az automata döntéshozatal, a profilozás és a modell kapcsán. Azonosítják az adatvédelem és a magánélet védelmével kapcsolatos jogi és etikai kérdéseket, illetve meghatározzák az adatvédelem, valamint más jogágak támasztotta követelményeket.

Az informatikusok felelősek a technikai kérdések azonosításáért, valamint közreműködnek annak meghatározásában, hogy az etikai szempontokra tekintettel milyen technikai lehetőségeink vannak és milyen korlátokkal szembesülhetünk. Kreatív megoldásokat dolgozhatnak ki, megtalálva azokat a módszereket és technikákat, amelyek megfelelnek az etikai követelményeknek. Ezen kívül jelentős szerepet vállalnak a felelős LLM-fejlesztés és alkalmazás elősegítésében, valamint gondoskodnak arról, hogy a

modellek átláthatóak, megmagyarázhatóak, megfelelően dokumentáltak és teszteltek, valamint a felhasználók megértése és bizalma szempontjából megfelelőek legyenek.

Amennyiben a jogászok és az informatikusok együttműködése elmarad, az etikai problémákat kaotikusan kezelhetjük, illetve figyelmen kívül hagyjuk, vagy akár fel sem ismerjük az etikai kockázatokat, ez pedig aggályos eredményhez vezethet (pl. előítéletesség, elfogadhatatlan, etikátlan adatkezelési gyakorlat, adatmanipuláció vagy személyes adatok tisztességtelen felhasználása). A tisztességtelen LLM jelentős társadalmi visszhangot és tiltakozást válthat ki, az együttműködés hiánya pedig növeli a reputációs károk, a társadalmi ellenállás, valamint a jogi konfliktusok kockázatát.

A társadalmi hatásvizsgálat

„A jog fejlődésének középpontja nem a törvényhozásban, nem a doktrínákban, nem a joggyakorlatban keresendő, hanem a társadalomban.”
(Ehrlich)

A társadalmi hatásvizsgálatban feltárjuk, felmérjük és elemezzük projektünk társadalmi, illetve egyénekre gyakorolt következményeit, valamint az esetleges egyenlőtlenségeket és társadalmi változásokat.

„A mesterséges intelligencia egyre szélesebb körű alkalmazása az egészségügyben várhatóan megváltoztatja a klinikai munka és az egészségügyi ellátás jellegét. Az egyik elvárás az, hogy a mesterséges intelligencia megkönnyíti a rutinfeladatok és az adminisztráció automatizálását. Egyesek szerint ennek eredményeként az orvosoknak és más egészségügyi dolgozóknak több idejük marad a betegeknek. Mások viszont azzal érvelnek, hogy ha a betegek gyakrabban lépnek kapcsolatba a mesterséges intelligenciával, például a telemedicina keretében, akkor az orvosok által a betegekkel töltött idő csökkenni fog. Továbbá, bár a mesterséges intelligencia helyettesíthet bizonyos rutinfeladatokat, az orvosoknak és az ápolóknak több időt kell majd tölteniük a technológia kezelésével, az adatok elemzésével és az új mesterséges intelligencia-alkalmazások használatának megtanulásával.

Ehhez kapcsolódó kérdés, hogy az MI az egészségügyben munkanélküliséghez vezet-e, és hogy a dolgozókat fel- vagy át lehet-e képezni, hogy elkerüljék a kiszorulást és alkalmazkodjanak az MI által támogatott orvosláshoz. A mesterséges intelligencia általi automatizálás miatti munkahelyvesztés széles körben elterjedt aggodalom, és az előrejelzések szerint az egészségügyi munkaerő szinte minden részterületére, köztük az orvosok és más szakértők bizonyos típusaira is vonatkozik. Mások azt állítják, hogy az MI új munkahelyeket fog teremteni az egészségügyben,

ami ellensúlyozhatja a lehetséges veszteségeket, és hogy az MI segíthet enyhíteni az egészségügyi munkaerőhiányt, többek között az alacsony és közepes jövedelmű országokban, ahol az egészségügyi személyzet hiánya gyakran komoly problémát jelent.

Bár ezek a forgatókönyvek eltérőek, abban egyetértés van, hogy a mesterséges intelligencia és más digitális technológiák beágyazása az egészségügybe és az orvosi kutatásba az egészségügyi személyzet átképzését teszi szükségessé, hogy alkalmazkodni tudjanak az új szerepekhez. A WHO 2021-es, az egészségügyi célú mesterséges intelligencia etikájáról és irányításáról szóló iránymutatása [World Health Organization (2021)] szintén felveti az egészségügy „überizálódásával” kapcsolatos aggodalmakat, ami az orvosok és a gyakorló orvosok munkáját kevésbé stabilá és kevésbé biztonságossá teheti. A jelentés szerzői szerint a mesterséges intelligencia által vezérelt egészségügyi platformok létrehozása az egészségügyi ágazatban a „gig-gazdaság” növekedéséhez vezethet, amelyben az ápolók, orvosok és más alkalmazottak igény szerint, ideiglenes vállalkozóként, a foglalkoztatás stabilitása nélkül dolgoznak. Ez a fejlemény – figyelmeztetnek – alááshatja a betegek és az egészségügyi szolgáltatók közötti kapcsolatot is, ami több röpke interakciót, az ellátás minőségének csökkenését és a bizalom elvesztését eredményezheti.” (Rosemann–Zhang 2021: 105)

A társadalmi hatásvizsgálat során elemezzük, hogy projektünk:

- várhatóan hogyan változtathatja meg a munkavállalás szerkezetét, mely területeken szűnnek meg munkahelyek és hol van szükség újakra, valamint a változás milyen (új) készségeket igényel a munkavállalóktól
- hatással van-e a szegénységre, a jövedelmi különbségekre, az oktatáshoz és egészségügyhöz való hozzáférésre, valamint az esélyegyenlőségre
- milyen hatással lehet a társadalmi befogadásra és kirekesztésre
- hogyan érinti a különböző csoportokat, beleértve a kisebbségeket, a fogyatékkal élőket vagy a társadalmilag elnyomott, kirekesztett csoportokat, és hogy esetlegesen tovább erősíti-e a meglévő egyenlőtlenségeket vagy diszkriminációt
- hogyan járulhat hozzá a társadalmi változásokhoz és elmozdulásokhoz, valamint hogyan alakíthatja át az üzleti modelleket, a közszolgáltatásokat vagy a társadalmi szokásokat, és ez milyen hatással lehet a társadalmi normákra és értékekre.

Az érdekeltek hatásvizsgálata (SIA)

Az Alain Turing Intézet (Leslie 2019) a közszféra MI-projektjeire összpontosítva ezen projektek társadalmi hatásával és fenntarthatóságával kapcsolatos hatásvizsgálatot az érdekeltek hatásvizsgálatának nevezi (SIA²⁴), melyben az „érdekeltek” kifejezés alatt elsősorban az érintett személyeket érti, de a fogalom kiterjedhet csoportokra és szervezetekre is, mivel e kollektívák egyes tagjaira is hatással lehet az MI-rendszerek bevezetése. Az ajánlás szerint a SIA-t a projekt három kritikus pontján kell elvégeznünk:

- *a probléma megfogalmazása („Alfa fázis”)*
célja az etikai megengedhetőség megállapítása, kiindulási pont, az érintettek beazonosítása, a célok felállítása, valamint az MI-projekt egyéni jólétre, valamint a közjólétre gyakorolt lehetséges hatásainak mérlegelése
- *a bevezetést megelőzően („Alfától Bétáig fázis”)*
a betanítás, tesztelés és validálás után újra elő kell vennünk az eredeti SIA-t megerősíteni, hogy a modellünk még mindig összhangban van az eredeti értékelésünkkel. Meg kell vizsgálnunk a képzett modell célját, a tervezési és tesztelési eredményeket, valamint ezek viszonyát az eredeti SIA-hoz, teret engedve a potenciális veszélyekkel kapcsolatos aggályainknak is (pl. diszkrimináció, torzítás)
- *újraértékelés („Béta fázis”)*
az élesben beüzemelés után időről időre újra el kell végeznünk a SIA-t, összevetve az eredeti SIA-val annak érdekében, hogy a felmerülő problémákat megoldhassuk.

Mi a szerepe a jogászoknak és az informatikusoknak a társadalmi hatásvizsgálatban?

A jogászok felelősek a releváns jogi előírások, szabályozások és jogalkalmazási gyakorlatok azonosításáért, valamint a vonatkozó jogi követelmények megértéséért és megértetéséért. Közreműködnek a jogi kockázatok azonosításában, illetve a jogilag elfogadható stratégiák és megoldások kidolgozásában.

Az informatikusok felelősek a technikai hatások azonosításáért, valamint segítenek megérteni, hogy a projekt milyen változásokat hozhat, és ezek a változások hogyan befolyásolhatják a társadalmat és az embereket. Jelentős szerepet vállalnak a felelős technológiai tervezésben, például olyan megoldások kialakításában,

amelyek minimalizálják a negatív, illetve elősegítik a pozitív hatásokat.

A jogászok és az informatikusok együttműködésének hiányában előfordulhat, hogy nem vesszük figyelembe vagy akár teljesen figyelmen kívül hagyjuk a potenciális negatív társadalmi hatásokat (szociális egyenlőtlenség, diszkrimináció, kizárólagosság stb.), helytelenül vagy hiányosan kommunikálhatjuk azokat félreértést, bizalmatlanságot vagy ellenszenvet eredményezve; az időben nem azonosított és kezelt aggályok pedig reputációs károkat, jogi problémákat vagy üzleti veszteségeket okozhatnak.

Az adatvédelmi hatásvizsgálat

A DPIA során feltárjuk az adatkezelések jellegét, szükségességüket és arányosságukat, valamint a személyes adatok kezeléséből eredő, a természetes személyek jogait és szabadságait érintő kockázatokat, illetve e kockázatok értékelésével és a kezelésükre szolgáló intézkedések meghatározásával minimalizáljuk ezeket. Az Európai Adatvédelmi Testület elődje, a 29. cikk szerinti adatvédelmi munkacsoport (WP29) a hatásvizsgálatokkal kapcsolatban iránymutatást adott ki (WP 248 rev.01).

Egy LLM-projekt során hol kezelhetünk személyes adatokat?

Sokkal több adat személyes adat, mint ahogy egy átlagos projektrésztvevő azt gondolná. Ez számos konfliktushoz vezethet, különösen az informatikusok esetében, akik gyakran hitetlenkedve veszik tudomásul, hogy például a rendszerhasználattal kapcsolatos adatok is személyes adatok (ki, mikor lépett be, mit csinált, meddig volt bent stb.). A jogászok egyik legfontosabb feladata elmagyarázni – sok-sok érthető példával – valójában mely adatok tartoznak a személyes adatok körébe és miért. A jogászok azonban nem csodálkozhatnak azon, hogy más szakterületek képviselői milyen „tudatlanok”, hiszen napjainkban még a legalapvetőbb adatvédelmi ismeretek sem tartoznak bele az általános műveltségbe. Sőt, a nem adatvédelmi specialista jogászok szintén kevés tudással rendelkeznek e téren és nekik is gyakran meggyűlik a bajuk a személyes adatokkal kapcsolatos problémákkal.

A DPIA egyik legfontosabb küldetése annak kiderítése, hogy egy LLM-projekt során hol és mily módon kezelünk személyes adatokat, illetve ezek az adatkezelések milyen kockázattal járnak az érintettek jogaira és szabadságaira nézve. Ehhez az szükséges, hogy az informatikusok képesek legyenek megmondani,

²⁴ Stakeholder Impact Assessment (SIA)

pontosan hol és milyen személyes adatok vannak, azokkal mi történik, valamint a jogászok beazonosítsák, milyen jogalapra hivatkozva történik ezen adatok kezelése, illetve megfelelünk-e a GDPR 5. cikkében foglalt alapelveknek, például a célhoz kötöttség, az adattakarékosság, az adatpontosság és a korlátozott tárolhatóság elvének. A DPIA feladata megfelelőségünk mértékének feltárása, illetve a hiányosságok esetében korrekciós intézkedések javaslata, természetesen mindvégig az elszámoltathatóság szuperelvének megfelelően.

Többek között, de nem kizárólag személyes adatok lehetnek:

- *képzési adatok*
A nyers képzési adatok személyes adatokat tartalmazhatnak, így közvetlenül fennállhat felelőségünk az érintetti jogokkal kapcsolatban, például a hozzáférési és törlési kérelmekre reagálnunk kell. Az előkészítés eltávolíthatja a képzési adatokból a legtöbb közvetlenül azonosító adatot, ha azonban az előkészítés után az adatok még mindig összekapcsolhatók egy adott személlyel, akkor azok tekintetében a GDPR/LED szabályai továbbra is érvényesek.
- *modellben lévő személyes adatok*
Az LLM jellemzően nem tartalmaz személyes adatokat, ha azonban a modell szándékosan kezel adatfragmentumokat, akkor el kell döntenünk, hogy a GDPR/LED alkalmazandó-e, így például az is előfordulhat, hogy az érintettnek joga lehet személyes adatainak törlésére, ami szükségessé teheti a modell újratartását.
- *működéssel kapcsolatos személyes adatok*
A modell be-, illetve kimenete is tartalmazhat személyes adatokat, és ez számos érintetti jogot aktiválhat, beleértve az automatizált döntéshozatalhoz, illetve a helyesbítéshez kapcsolódó jogokat is. Amennyiben az eredményt csak statisztikailag megalapozott találgatásként kezeljük (lásd pl. az ajánló alkalmazások), az érintett nem élhet a helyesbítéshez való jogával arra hivatkozva, hogy az eredmény pontatlan.

Az adatvédelmi hatásvizsgálat tartalma

A GDPR nem határozza meg külön a DPIA fogalmát, de a minimális tartalmát rögzíti²⁵. A NAIH²⁶ hangsúlyozza, hogy „a kockázatelemzés a személyes adatok kezelésével összefüggő folyamatokra, adatkezelési műveletekre vonatkozik, amelyek a hatásvizsgálat lefolytatásának eredményeként az érintett jogait és szabadságait érintő kockázatot

jelentenek. Az adatvédelmi hatásvizsgálat lényege az adatkezelés előzetes kontrollja a kockázatok feltárása és a kockázatok mérséklésére teendő intézkedések értékelése révén. A kockázatnak egyértelműnek, konkrétan kell lennie, és ahhoz, hogy az adatkezelő azonosítani tudjon kockázatokat, meg kell előznie egy kockázatelemzési folyamatnak”. (NAIH 2023)

A DPIA során ez a kockázatalapú megközelítés szükségessé teszi az egymással versengő, illetve ellentétes érdekek közötti kompromisszumokat, például annak mérlegelését, hogy a lehető legtöbb személyes adatot kívánjuk-e felhasználni egy modell kiképzéséhez annak pontosságának biztosítása érdekében, szemben az ilyen nagy adathalmaz felhasználásával járó adatvédelmi kockázatokkal, például az adattakarékosság és a célhoz kötöttség elve betartásának követelményével. Meg kell határozni, hogy a különböző tevékenységeket milyen jogalapokra hivatkozva kívánjuk végezni, a személyes adatok különleges kategóriájába tartozó, valamint a bűnügyi adatok esetében pedig a 9. cikk (2) bekezdésében felsorolt kivételek egyikének megfelelést is alá kell támasztanunk.

Azt is vizsgálunk kell, hogy milyen joghatóságok alatt kezeljük az adatokat (pl. harmadik országban és emiatt külön garanciákat fel kell-e mutatnunk, illetve van-e olyan tagállami jogszabály, amelyek még tovább szűkíti a mozgásterünket). Alá kell támasztanunk, hogyan felelünk meg a szükségesség és az arányosság, illetve a fokozatosság követelményének, valamint figyelemmel kell kísérnünk a diszkrimináció lehetőségét is.

A DPIA főbb területi:

- a személyes adatok kezelése
A DPIA során azonosítjuk az adatvédelmi kockázatokat és meghatározzuk a megfelelő intézkedéseket, például a titkosítási módszereket, a belső szabályzatainkat, a hozzájárulási mechanizmusokat, valamint a szükséges érdemléseket
- sérülékenység és potenciális fenyegetettség, külső és belső támadások lehetőségei (pl. adathalászat, rosszindulatú szoftverek vagy jogosulatlan hozzáférés), biztonsági intézkedések (pl. tűzfalak, jogosultságkezelés, biztonsági protokollok, rendszeres frissítések, biztonsági mentések stb.)
- a felhasználói hozzáférés és azonosítás, az adatokhoz és a rendszerhez való hozzáférés kezelése és ellenőrzése (azonosítási és hitelesítési mechanizmusok)
- a projektünk adatvédelmi jogszabályoknak és egyéb szabályozásoknak, iránymutatásoknak,

²⁵ GDPR 35. cikk (7) bekezdés

²⁶ Nemzeti Adatvédelmi és Információszabadság Hatóság

valamint a vonatkozó jogalkalmazási gyakorlatoknak megfelelése.

- minden egyéb, amit fontosnak tartunk, például biometrikus adatok kezelésére, bűnügyi adatokkal kapcsolatos speciális követelmények stb.

Az MI használatáról szóló rendelet tervezete (29. cikk (6) bek.) szerint „adott esetben a magas kockázatú mesterséges intelligenciát alkalmazó rendszerek üzembe helyezői a 13. cikk alapján szolgáltatott információkat felhasználják az (EU) 2016/679 rendelet 35. cikke vagy az (EU) 2016/680 irányelv 27. cikke szerinti adatvédelmi hatásvizsgálat elvégzésére vonatkozó kötelezettségük teljesítéséhez, amelynek összefoglalóját közzé kell tenni (...)”.

Amennyiben a DPIA szerint a kockázat mérséklését célzó garanciák, biztonsági intézkedések és mechanizmusok hiányában az adatkezelés magas kockázattal járna a természetes személyek jogaira és szabadságaira nézve, és ez a kockázat nem mérsékelhető a rendelkezésünkre álló technológiák és a végrehajtási költségek szempontjából észszerű módon, abban az esetben az adatkezelési tevékenységünk megkezdése előtt konzultálnunk kell az illetékes adatvédelmi felügyeleti hatósággal. A NAIH szerint ide tartozik az, ha az érintettek olyan jelentős vagy akár visszafordíthatatlan következményekkel szembesülnek, amelyeket nem tudnak leküzdeni (pl. adatokhoz való jogosulatlan hozzáférés, amely az érintettek életét fenyegető veszélyt, elbocsátást vagy pénzügyi nehézséget eredményez) (NAIH 2023).

A NAIH az előzetes konzultáció keretében megállapítja:

- a DPIA-t a vonatkozó jogszabály, illetve iránymutatás szerint folytattuk-e le
- pontosan azonosítottuk-e az adatkezelési tevékenységeket, azok kockázatait, valamint sikerült-e a kockázatok kezelésére irányuló intézkedéseket meghoznunk
- az adatkezelésben érintett adatok esetében szétválasztottuk-e a személyes adatok és a különleges adatok kezelését
- az adatkezeléseink jogszerűek-e és amennyiben szükséges, végeztünk-e érdekmérlegelési teszte(ke)t, valamint
- a fennmaradó kockázatok mérséklésében tud-e segítséget nyújtani.

Amennyiben a tervezett adatkezelés megsértene a GDPR-t, különösen, ha adatkezelőként a kockázatot nem elégséges módon azonosítottuk vagy csökkentettük, úgy a NAIH gyakorolhatja a GDPR 58. cikkében említett hatásköreit, többek között megtilthatja az adatkezelésünket.

A finn adatvédelmi hatóság figyelmeztetésben részesített egy adatkezelőt az egészségügyi ellátások értékelésével kapcsolatban. Az egészségügyi szolgáltató (adatkezelő) előzetes konzultációt kért egy új, egészségügyi ellátások elemzésére szolgáló eszközzel kapcsolatban. A rendszert az egészségügyi ellátás megelőző és proaktív megközelítésének javítására tervezték úgy, hogy algoritmusok segítségével azonosítsa az egészségügyi kockázatokkal rendelkező betegeket, majd az azonosított páciensek esetében egy egészségügyi szakember értékelné a kezelés szükségességét. A hatóság úgy döntött, az ügyet nem előzetes konzultációként, hanem felügyeletként kezeli. A hatóság megállapította, a rendszer által további vizsgálatra azonosított betegek nem tartoznak a kizárólag automatizált adatkezelésen alapuló döntés hatálya alá, mivel az egészségügyi szakemberek a rendszer ajánlásán kívül más tényezőket is figyelembe vesznek a végső értékelésnél. A rendszer által ki nem választott betegek esetében azonban a döntés végleges lenne, és kizárólag automatizált döntéshozatalon alapulna, ez pedig jelentősen érintené őket, mivel megfosztaná őket az egészségügyi szolgáltatás igénybevételének lehetőségétől. A hatóság döntése szerint a tervezett adatkezelés sérti a GDPR 22. cikkét (Tietosuojavaltuutetun toimisto 2020).

Mi a szerepe a jogászoknak és az informatikusoknak az adatvédelmi hatásvizsgálatban?

A jogászok felelősek a releváns jogi követelmények azonosításáért, valamint segítenek megérteni, milyen adatvédelmi jogszabályok és jogalkalmazási gyakorlat vonatkozik az LLM-re, miért fontos ezek betartása és hogyan kell ezeket figyelembe venni a tervezés, a fejlesztés és az üzemeltetés során. Adatvédelmi jogi tanácsot nyújtanak, kiterjedve – többek között – az adatvédelmi kockázatokra, a megfelelő intézkedésekkel kapcsolatos javaslatokra, valamint a jogszerű adatkezelési gyakorlatokra.

Az informatikusok felelősek a technikai adatvédelmi és biztonsági kockázatok, az LLM sérülékenységei és potenciális fenyegetései azonosításáért, valamint javaslatokat tesznek a megfelelő biztonsági intézkedésekre és megoldásokra. Megtervezik a modell biztonsági architektúráját, javaslatokat tesznek az adatvédelmi protokollokra és technikákra, illetve implementálják a megfelelő biztonsági és adatvédelmi intézkedéseket. Felelősek a rendszerbiztonság és adatvédelem teszteléséért, melynek keretében elvégzik a biztonsági teszteket, ellenőrzik a rendszer sérülékenységeit és intézkednek a szükséges korrekciók érdekében.

Amennyiben a jogászok és az informatikusok együttműködése elmarad, előfordulhat, hogy nem azonosítjuk vagy nem vesszük megfelelő mértékben figyelembe az adatvédelmi és adatbiztonsági előírásokat, emiatt a projektünk adatbiztonsági szempontból sérülékenyebbé válhat (rosszindulatú támadások, adatlopások, adatmanipulációk), valamint könnyen megszeghetjük a vonatkozó jogszabályokat és egyéb szabályozásokat. Elmaradhat a vonatkozó tájékoztatóink, belső szabályzataink megalkotása, illetve ezek hiányosak lehetnek, ez pedig a megfelelő adatvédelmi gyakorlatok és elvek hiányát eredményezi, akár komoly pénzügyi és reputációs kárt is okozva.

Az együttműködés során „létfonosságú” a konszenzus, mivel a „felülről kapott”, egyeztetés nélküli, kétségesen használható szervezetidegen utasítások és szabályzatok, amelyeket a szereplők még csak el sem olvasnak, vagy ha bele is néznek, azt dühösen a sarokba dobják, semmiképpen sem hatnak abba az irányba, hogy a projektünk jogszerű és biztonságos legyen.

Az alapjogi hatásvizsgálat (magas kockázatú MI-rendszerek esetében)

Az alapjogi hatásvizsgálatban azonosítjuk és értékeljük egy adott LLM-projekt alapjogi hatásait, figyelembe véve az alapjogokat és azt, hogy egy adott intézkedés hogyan befolyásolja ezeket a jogokat. Alapvető követelmény, hogy ne sértsük az alapjogokat, illetve az alapjogok közötti egyensúlyt és konfliktusokat figyelembe vegyük a projekt teljes időtartama alatt.

Az MI-rendelet-tervezete szerint a nagy kockázatot jelentő MI-vel kapcsolatos rendszerek használatba vétele előtt – bizonyos kivételektől eltekintve – rendszerüzemeltetőként kell elvégeznünk a hatásvizsgálatot, mely tartalma:

- a rendszer tervezett felhasználási céljának egyértelmű felvázolása
- a rendszer használatának tervezett földrajzi és időbeli hatókörének világos felvázolása
- a rendszer használata által valószínűleg érintett természetes személyek és csoportok kategóriái
- annak ellenőrzése, hogy a rendszer használata megfelel-e az alapvető jogokra vonatkozó uniós és nemzeti jogoknak
- a magas kockázatú MI-rendszer használatba vételének az alapvető jogokra gyakorolt észszerűen előrelátható hatása
- a marginalizált személyeket vagy kiszolgáltatott csoportokat valószínűleg érintő konkrét ártalmi kockázatok
- a rendszer használatának észszerűen előrelátható káros hatása a környezetre

- részletes terv arra vonatkozóan, hogy az azonosított károkat és az alapvető jogokra gyakorolt negatív hatásokat hogyan fogjuk enyhíteni
- az irányítási rendszer, amelyet a telepítő bevezet, beleértve az emberi felügyeletet, a panaszkezelést és a jogorvoslatot.

Amennyiben a kockázatok mérséklésére vonatkozó részletes terv nem azonosítható, üzembehelyezőként tartózkodnunk kell a magas kockázatú MI-vel rendelkező rendszer használatba vételétől, valamint erről indokolatlan késedelem nélkül tájékoztatnunk kell a szolgáltatót és a nemzeti felügyeleti hatóságot.

Ez a hatásvizsgálati kötelezettség a magas kockázatú MI-t alkalmazó rendszer első használatára vonatkozik, egyéb esetekben támaszkodhatunk a korábban elvégzett alapjogi hatásvizsgálatra vagy a szolgáltatók által elvégzett meglévő értékelésre. Amennyiben a magas kockázatú MI-rendszer telepítőjeként úgy ítéljük meg, hogy a hatásvizsgálat kritériumai már nem teljesülnek, új alapjogi hatásvizsgálatot kell végeznünk. A tervezet szerint a kis- és középvállalkozások önkéntesen alkalmazhatják az alapjogi hatásvizsgálatra vonatkozó rendelkezéseket.

Amennyiben használatba adóként a GDPR 35. cikke vagy a LED 27. cikke alapján DPIA-t kell végeznünk, az alapjogi hatásvizsgálatot a DPIA-val együtt kell elvégeznünk (29a. cikk).

Mi a szerepe a jogásznak és az informatikusnak az alapjogi hatásvizsgálatban?

A jogászok felelősek az alapjogi keretek feltérképezéséért, értelmezik és értékelik a jogszabályokat annak érdekében, hogy felismerjék az alapjogokkal összefüggő kérdéseket, valamint azonosítsák az esetleges kockázatokat vagy sérelmeket. Megvizsgálják a vonatkozó jogi környezetet, azonosítva az alapjogokat érintő követelményeket, korlátokat és elvárásokat, illetve figyelemmel kísérik a változásokat és a jogalkalmazási gyakorlatot.

Az informatikusok felelősek a technikai elemzések elvégzéséért, hogy megértsük, az adott LLM hogyan működik és milyen technikai megoldásokat igényel, illetve meghatározzák az adatok feldolgozására, tárolására, hozzáférésére és biztonságára vonatkozó technikai követelményeket. Felelősek az információbiztonságért, közreműködnek a biztonsági kérdések azonosításában és megoldásában, a megfelelő biztonsági gyakorlatok alkalmazásában (10. sz. ábra).



"Oké, bevallom! Senki sem osztott meg macskaképeket, amíg meg nem hackeltük az algoritmusokat, hogy ne mutassanak cuki kiskutyákat a listák elején!"

<https://www.moralmachine.net/>

10. sz. ábra: Biztonság
Forrás: Az ábra alján

A jogászok és az informatikusok együttműködése elengedhetetlen annak érdekében, hogy a jogi és technikai szempontok harmonizáljanak, és döntéseink összhangban legyenek az alapjogokkal és a jogalkalmazási gyakorlattal, ehhez pedig az szükséges, hogy az informatikusok – jogi alapismeretek nélkül is – elfogadják, nekik is tenniük kell a projekt alapjogi megfeleléséért. Amennyiben ez az együttműködés elmaradt, az MI megsértheti az egyének alapvető jogait, illetve az alkalmazott technológiák jogi helyzete bizonytalan lehet, ez pedig nehezítheti a bevezetést és felhasználást, valamint növelheti a jogi kockázatokat. Az együttműködés hiánya hátrányosan befolyásolhatja döntéseink minőségét és megalapozottságát, valamint olyan intézkedésekhez vezethet, amelyek csökkenthetik projektünk hitelességét és elfogadottságát. Ha az emberek úgy érzik, az alapvető jogukat figyelmen kívül hagyjuk, az csökkenti a bizalmat az MI-technológiák iránt és növeli az aggodalmakat a magánszférával, az adatvédelemmel és az etikai kérdésekkel kapcsolatban.

Jogszabályi és szabályozási hatásvizsgálat

Mikor jogszerű egy mesterséges intelligencia?

Akkor, ha az MI kifejlesztésének, elterjesztésének és felhasználásának részét képező folyamatokra és tevékenységekre vonatkozó valamennyi előírást betartjuk, például:

- az elsődleges uniós jogot (az Európai Unió Szerződésai és Alapjogi Chartája)
- a másodlagos uniós jogot (pl. a GDPR/LED, a hátrányos megkülönböztetést tiltó irányelvek, a gépekről szóló irányelv, a termékfelelősségi

irányelvek, a nem személyes adatok szabad áramlásáról szóló rendelet, a fogyasztói jogra, valamint a munkahelyi biztonságra és egészségre vonatkozó irányelvek stb.) – az MI használatával kapcsolatos jogalkotás jelenleg is folyamatban van

- az ENSZ emberi jogi egyezményeit
- az Európa Tanács egyezményeit (pl. az emberi jogok európai egyezménye), valamint
- az uniós tagállam(ok) jogát.

A jogszabályi és szabályozási hatásvizsgálat során ennek a jogi környezetnek az elemzését és értékelését végezzük el, feltárva a vonatkozó követelményeket és korlátokat, az alábbi fókuszpontokkal:

- releváns jogszabályok, iránymutatások és jogalkalmazási gyakorlatok
- szektorspecifikus előírások, iparági szabványok
- szükséges engedélyk
- tanúsítványhoz, magatartási kódexhez csatlakozás
- különböző felelőségek (pl. szerzői és szomszédos jogok, fogyasztóvédelem, termékfelelősség, büntetőjogi felelősség stb.)
- kereskedelmi és versenyjogi szempontok stb.

Mi a szerepe a jogászoknak és az informatikusoknak a jogszabályi és szabályozási hatásvizsgálatban?

A jogászok feladata a releváns jogi előírások, jogi és szabályozási követelmények és kötelezettségek értékelése, azonosítva azokat a követelményeket, amelyeket teljesíteniük kell, megfelelő intézkedéseket és belső szabályozásokat javasolva.

Az informatikusok értékelik a technikai megvalósíthatóságot a jogszabályi és szabályozási keretek figyelembevételével, azonosítva a technikai korlátokat és kihívásokat, valamint kidolgozzák azokat a technológiai megoldásokat, amelyek támogatásával megfelelnünk a jogi követelményeknek (pl. adatbiztonsági mechanizmusok, felhasználói hitelesítés, adatkezelési protokollok stb.) (11. sz. ábra).

Bár sokan úgy gondolják, hogy a jogi megfelelés egyedül a jogászok feladata, a modern technológiák esetében a jogászok és az informatikusok közös munkája a kulcsa a jogi és szabályozási kockázatok minimalizálásának. Az egyes technológiák jogi és szabályozási összefüggéseinek figyelembevétele versenyelőnyt jelenthet, azonban együttműködés hiányában ezt elmulasztathatjuk. Könnyen megszeghetjük a vonatkozó jogi előírásokat is, illetve nem megfelelő adatvédelmi és adatbiztonsági intézkedéseket dolgozhatunk ki vagy implementálhatunk, ami

jogsértésekhez, vitákhoz, hatósági eljárásokhoz, pereskedésekhez vezethetnek, valamint jelentős többletköltséget is okozhatnak (pl. határidő túllépés, kötbérfizetési kötelezettség, közigazgatási bírság, kártérítés fizetési kötelezettség stb.). Ezen kívül a hiányosságok súlyosan károsíthatják szervezetünk vagy projektünk hírnevét és ügyfeleink bizalmát.

Melyek a jó hatásvizsgálat ismérvei?

A hatásvizsgálat egyik legfontosabb előnye – eltekintve attól a kockázatminimalizálással kapcsolatos céltól, amire kitalálták –, hogy segít egy asztal köré összehozni a résztvevőket és teret ad a projekt Alfától Omegáig megismerésére, az építő párbeszédre és a felmerülő problémák hatékony, minden szakterület számára megfelelő megoldására. A párbeszéd kezdeményezője olyan személy kell legyen, aki „hidemberként” nem retten meg a nehezen kezelhető szereplőktől, képes kreatívan megoldani a konfliktusokat és közelebb hozza az olykor egymástól nagyon messzi álláspontokat. A siker feltétele, hogy ez a „hidember” széles látókörrrel rendelkezzen és képes legyen csitítani a felek indulatait, a társalgást megfelelő mederben tartani. Enélkül a hatásvizsgálat csak szimpla „lepapírozás”, egy kérdőív mechanikus kitöltése.

A jó hatásvizsgálat néhány fontos ismérve:

- lehetőséget ad egymás személyének, szempontrendszerének és prioritásainak megértésére
- teljeskörű, átfogóan elemezi a lehetséges közvetlen és közvetett hatásokat, valamint következményeket
- előre meghatározza a kereteket, teljeskörűségében sem markol túl sokat, komplex folyamatokban gondolkodik és azokat lépésről lépésre bontja le
- alaposan elemezi a különböző tényezőket és szempontokat (adatok gyűjtése, elemzési

módszerek alkalmazása, releváns szakirodalom áttekintése, szakértők bevonása stb.)

- nem hazudik, nem célja a valóság elrejtése. Ha nem vagyunk hajlandóak szembesülni a problémákkal, akkor azokat megoldani sem tudjuk, miközben soha nem tudhatjuk, azok mikor kerülnek felszínre, az idő előre haladtával egyre nagyobb galibát okozva.
- tárgyilagos, objektív és semleges, megbízható adatokra, releváns kutatásokra és szakértői véleményekre támaszkodva nyújt megbízható és objektív eredményt
- rugalmas a változó körülmények és feltételek kezelésében (pl. egy új jogszabály alapjaiban rengetheti meg projektünket, és rugalmasság hiányában akár fel is adhatjuk célunkat)
- bevonja a folyamatokba az érdekelteket, szakértőket és érintetteket. Nem célszerű senkit sem kihagynunk csak azért, mert kellemetlen alaknak tartjuk; a meg nem hallgatott vélemények olyan információkat tartalmazhatnak, amelyek nem ismerete később számtalan gondot és jelentős költségnövekedést is okozhat.
- az eredmények, a módszerek és a feldolgozott adatok tekintetében átlátható
- eredményei felhasználhatók a döntéshozatalban, a szükséges intézkedések és belső szabályzatok kialakításában
- megfelelően ütemezett és bármikor újrainyitható, tekintettel a körülmények változására
- a projekt méretéhez és jellegéhez skálázható, alkalmazkodva annak komplexitásához
- költséghatékony, arányban áll erőforrásainkkal és az elvárt előnyökkel
- támogatja az innovációt és az új technológiák bevezetését, miközben biztosítja a kockázatok megfelelő kezelését és a fenntartható fejlődést.

mérlegelés fontos része, hogy az érintettek jogait és szabadságait érő kockázatokat minimalizálva állítsuk be azt az egyensúlyi helyzetet, amely segítségével adatkezelésünk jogszerű lesz.

Az érdekmérlegelést példákkal alá kell támasztanunk, ehhez pedig elengedhetetlen a szereplők együttműködése. Amennyiben a teszt modern technológiákkal kapcsolatos, állandó résztvevő az adatvédelmi tisztviselő, az informatika és a biztonsági szolgálat képviselője is. A felügyeleti hatóság elvárja, hogy az érdekmérlegelés térjen ki a szükségességre, az arányosságra és a fokozatosságra, a lehetőségek megvitatásához pedig elengedhetetlen a különféle szakterületek együttműködése.

Az LLM esetén különösen a modell betanítása során merülhet fel a jogos érdekre hivatkozás, éppen ezért a betanítással kapcsolatos adatkezelések esetében (pl. adatok gyűjtése, előkészítése stb.) alapvető az informatikusok részvétele az érdekmérlegelésben. Az ő feladatuk továbbá a kockázatsökkentő intézkedések azonosítása és technikai megoldása (pl. anonimizálás, álnevesítés, titkosítás, jogosultságmenedzsment stb.), a jogászoknak pedig el kell tudni fogadtatniuk az informatikusokkal, hogy miért is van szükség ezekre a gyakran igen drága intézkedésekre.

Belső szabályzatok

A projektnek és a szervezetnek is szüksége van valamiféle keretre, amely támpontot ad a feladatok és felelőségek tekintetében, és amely általában különféle szabályzatokból, utasításokból, illetve protokollokból áll össze.

Az MI-szabályzat megalkotása nem lehet egyetlen személy vagy szakterület feladata, és ez az a lehetőség, amelyet kiaknázhathatunk annak érdekében, hogy a későbbiekben a különböző szakterületek képviselői ne akadályozzák a közös munkát. Ugyanis egy MI-szabályzat megalkotása arra készíti a szereplőket, hogy mélységben gondolják át kinek mi a feladata és felelőssége, milyen lépéseket és mérföldköveket kell teljesíteni a megfelelőség érdekében, és kit, mikor kell bevonni a projektbe.

Amennyiben teljeskörű, mindenki által elfogadható szabályzatot készítünk

- mindenki tudja mi a feladata, és nem próbálja azt áthárítani másokra
- a felelősségi körök tisztázottak (pl. ki, milyen értékben hagyhat jóvá beszerzést stb.)
- milyen részletességgel kell dokumentálni a projektet
- rendkívüli események esetén kinek, mi a teendője, és nem utolsó sorban

- a teljes projekt időtartama alatt megfelelünk az átláthatóság és az elszámoltathatóság elvének.

Mire jó egy jó szabályzat?

Ritkán lelkesedünk a szabályzatokért – általában nyügnék érzik, felesleges bürokráciának, íróasztal mellőli okoskodásnak. A szabályzatoknak azonban van pozitív hozadéka is. Amennyiben leszabályozzuk a folyamatokat, mindenki tudja, mi a feladata, milyen felelősséget kénytelen (el)viselni, mi a teendője rendkívüli esemény esetén és mire számíthat, ha valami félresiklik. A jó szabályzat tehát nem pusztán egy dokumentum, hanem egy útmutató, amely irányt mutat a kockázatos folyamatokban és segít – akár a káosz közepette is – a helyes irány megtalálásában.

Az MI-projektek sokszereplős, „zűrös” folyamatok, könnyen összemósódó felelősséggel, miközben a modell lehet olyan „feketedoboz”, amelynek felnyitása szinte lehetetlen, alaposan feladva a leckét a felelőst keresőknek. Éppen ezért elengedhetetlen az átláthatóság, garantálva, hogy nem marad ki fontos lépés és nem a modellt tesztelésénél közli a jogász, az jogszerűtlen és az egészet lehet előlről kezdeni. Ez nemcsak hatalmas idővesztés (határidőcsúszást stb.) okozhat, hanem jelentős pénzügyi kihatással is járhat (kötőbírfizetési kötelezettség, túlmunka ellentételezése, új adatkészlet beszerzése stb.), miközben egy átgondolt szabályzat betartásával a kényes helyzet elkerülhető lett volna.

A jó MI-szabályzatot azonban nem lehet csak úgy letölteni a netről, a jó szabályzat igazodik a sajátosságainkhoz. A jó szabályzatot hely-, személy- és projektismerettel rendelkező személyek alkotják meg, vagy egy, már meglévő sablont kell a szervezetünkre és a projektünkre szabni.

A jó szabályzat ténylegesen is támogatja projektünket – ez pedig akkor lesz így, ha minden szakterület a saját követelményeit belefoglaltatja, ahogy azt is, mit vár el másoktól. Természetesen a projektgazdán, az informatikusokon és a jogászokon kívül számtalan szakterület kérését meg kell hallanunk, így például a pénzügyesekét, a beszerzőkét, a humán erőforrásgazdálkodását.

Mire térjen ki a szabályzat?

Az MI-szabályzatnak nincs kötelező tartalma, hiszen maga az MI-szabályzat sem kötelező. Célszerű olyan témákat leszabályoznunk, amelyek szabályozatlansága gondot okozhat, például:

- projektfolyamatok leírása
- folyamatok kiszervezésének követelményei

- melyik stádiumban kit, miért és hogyan kell bevonni
- milyen alapvető jogszabályokat, szabályozásokat és szakmai minimumokat kell figyelembe venni

„A bíróságok nem várhatják el, hogy a gyakran évszázados múltra visszatekintő szokásjog mindig alkalmas lesz arra, hogy a technológia szabályozásának új jogi kérdéseit kezelje. A már létező jogi kategóriák bizonyos esetekben alkalmazhatóak lehetnek, de ezt csak úgy lehet megállapítani, ha először is megvizsgáljuk a kategóriák alapját és értékeliük, hogy a doktrína kiterjesztése kielégíti-e ezt az alapot. Ez az elemzés az adott jogvitától és a szóban forgó technológiától függően változik, és gyakran megköveteli a döntésnek a szóban forgó technológia jövőbeli fejlődésére és elterjedésére, valamint tágabb értelemben a gazdaságra és a társadalmi jólétre gyakorolt hatásának mérlegelését.

A valós világbeli vitákat és társadalmi összefüggéseket nem szabad előre létező jogi kategóriákba kényszeríteni. A jogi kategória csupán egy konstrukció; a viták és a kontextus a megváltoztathatatlan valóság. Ha a jogi kategóriák nem illeszkednek jól az új valósághoz, akkor a jogi kategóriákat kell újraértékelni. (...)

Először is, a már meglévő jogi kategóriák már nem feltétlenül alkalmazhatók az új jog és technológia vitáira. Annak mérlegeléséhez, hogy a meglévő jogi kategóriáknak van-e jogi és társadalmi értelme egy új technológiai rendszerben, kritikus fontosságú először a jogi kategorizálás mögött álló indoklás értelmezése, majd annak értékelése, hogy az alkalmazható-e az új jogvitára.

Másodszor, a jogi döntéshozóknak ügyelniük kell arra, hogy az új technológia csodái ne torzítsák el a jogi elemzésüket. Ez a technológiailag laikus jogi döntéshozók számára különleges kihívást jelent, amely megköveteli, hogy a fejlődő technológia igényeiből kiindulva átlássák annak tényleges jellemzőit és a tudományos ismeretek jelenlegi szintjét.

Harmadszor, az újonnan megjelenő technológiákból eredő új jogviták típusai gyakran előre nem láthatóak. Azok a jogrendszerek, amelyek képesek alkalmazkodni és fejlődni a technológia és az arról alkotott ismereteink fejlődésével, sokkal sikeresebben fognak működni, mint a már létező jogi rendszerekhez való vak ragaszkodás”. (Mandel 2017: 271)

- projektek mérföldkövei
- torzítás és diszkrimináció elkerülése, csökkentése és megszüntetése (feladatok, módszerek)
- a tesztelés követelményei (feladatok és felelőségek)
- átláthatósági és megmagyarázhatósági követelmények (ICO–Alan Turing Institute 2022)

„A mesterséges intelligencia képes arra, hogy a döntéshozatalt az emberektől a gépekre ruházza át. Bár ez pontosabb, gyorsabb és hatékonyabb orvosi és egészségügyi irányítási döntésekhez vezethet, számos aggályt is felvet. Számos mesterséges intelligencia-technológia "feketedoboz" rendszerként működik, amelynek belső folyamatai és következtetései rejtve maradnak, vagy nehezen magyarázhatóak el a felhasználók számára. Ez az átláthatóság hiányához vezethet, és alááshatja az egészségügyi rendszerek üzemeltetőinek, az orvosoknak és a betegeknek az autonómiáját. Az egészségügyi ellátást igénybe vevők a kontroll és az autonómia elvesztését tapasztalhatják, különösen akkor, ha az orvosi döntések nem átláthatóak, és hiányzik a betegek és az orvosok közös döntéshozatala. A kontroll és az autonómia ezen problémái egyre inkább előtérbe kerülnek a mesterséges intelligencia technológiák programozásának automatizálására irányuló erőfeszítésekkel összefüggésben, olyan számítógépes programok révén, amelyek önállóan képesek új mesterséges intelligencia modelleket és alkalmazásokat létrehozni, telepíteni és skálázni. Ehhez kapcsolódó kihívás a mesterséges intelligencia kiszámíthatósága és megbízhatósága. Az algoritmusokban és adatokban lévő hibák vagy az elfogult adatkészletek használata az MI-rendszerek hibás vagy tisztességtelen döntéseihez vezethet. A hibás vagy elfogult ítéletek befolyásolhatják a megbízhatóságot és az egészségügyi ellátás hatékony végrehajtását. Az adatkészletek és algoritmusok torzítása az erőforrások igazságtalan elosztásához is vezethet, és diszkriminálhat bizonyos csoportokat, például azért, hogy figyelmen kívül hagyja a kevés erőforrással vagy embereket, vagy csoportokat.” (Rosemann–Zhang 2021: 104–105).

- az adatok és a modell védelme
- kiemelt jogterületek, például szerzői jog és szomszédos jogok²⁷, üzleti titok, szabadalom, termékfelelősség²⁸, fogyasztóvédelem, versenyjog stb.
- irányítás és felügyelet
- a rendszeres audit (korrekciós intézkedések) stb.

Adatvédelmi és adatbiztonsági ismeretek oktatása, a tudatosság növelése

A kötelező oktatás általában az a szükséges rossz, amit meg kell úszni, különösen, ha közben nem lehet „értelmes” dolgot csinálni, például e-mailekre válaszolni, híreket olvasni, ásitózni stb. Az ilyen oktatásnak sok értelme nincs, még egy pipa valamelyik

27 (EU) 2019/790 irányelv

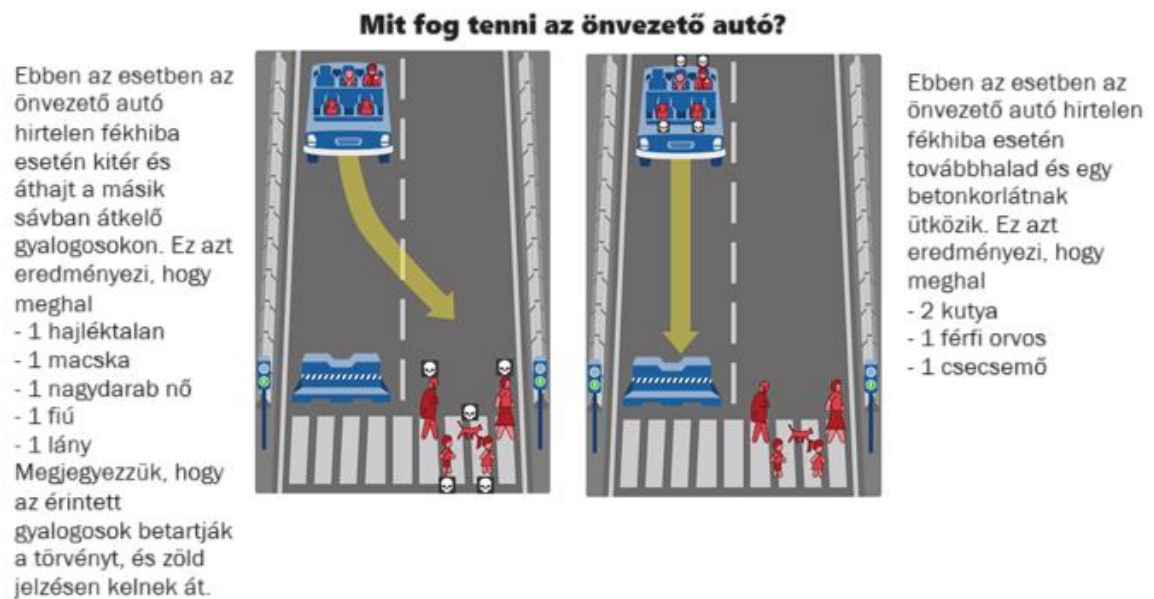
28 a mesterséges intelligenciával kapcsolatos felelősségről szóló irányelv javaslat

projektdokumentációban. Ezeket az alkalmakat azonban – megfelelő előadással és átgondolt tematikával – workshopká vagy csapatépítéssé formálhatjuk.

- **Adatvédelmi és adatbiztonsági ismeretek**
Az oktatás akkor hasznos, ha nem a jogszabály citálását jelenti, hanem a gyakorlatot ismerteti. A legjobb példák a mindennapokból merítenek és felhasználják a hallgatóság tapasztalatait.
- **Alapjogok értelmezése**
Az alkotmányos normákat nem tudjuk csak úgy egyszerűen átültetni egy másik társadalmi kontextusba, például abba a digitális környezetbe, amelyben az LLM működni fog,

éppen ezért az alapjogokat újra meg kell határoznunk. A számtalan digitális vetület miatt ez nem lehet csak a jogászok feladata (pl. az adatok nem megfelelő szelektálása, az algoritmus döntésének kirekesztő jellege vagy torzítása).

- **Etikai dilemmák megvitatása**
Az etikai dilemmák megvitatásához találunk az interneten forrásokat, például a Moral Machine oldalt az önvezető autók döntéseiről (12. sz. ábra). A témában globális kutatás (Bonniefon 2021), illetve tanulmány (Bonniefon–Shariff–Rahwan 2016) is elérhető.



12. sz. ábra: Hogyan döntsön az önvezető jármű?

Forrás: <https://www.moralmachine.net/>; 2/13)

- **Konkrét hatósági határozatok, bírósági döntések átbeszélése**
Melyik informatikust ne érdekelné, hogy mit követhetett el az a bank, amelyik az MI miatt 250 milliót büntetést kapott? Amennyiben – köszönhetően az eset közös megvitatásának – a projekt szereplőiben tudatosul, az „akadékoskodó” jogász csak el akarja kerülni, hogy a projekt miatt büntetést kapjunk, jobban elfogadják érveit.
- **Megtörtént adatvédelmi incidensek elemzése**
Az incidensek elemzésekor érdemes megtörtént esetekből kiindulnunk, mi volt a hiba, mely ponton és hogyan állt be az adatkezelő, illetve az adatfeldolgozó felelőssége. Az Európai Adatvédelmi Testület (EDPB) 01/2021. számú iránymutatása hasznos incidens-szituációkat tartalmaz kockázatminősítéssel és intézkedési

javaslattal, illetve a saját tapasztalatainkból is meríthetünk.

- **Adatvédelmi incidens szimuláció**
Végezhetünk olyan adatvédelmi incidens szimulációt is, amely IT vonatkozású, és amelyben az informatikusoknak orozslánrészt kell vállalniuk (a biztonság sérülése az ő felségterületükön történik). Kielemezhetjük, hogy az EDPB 9/2022. számú iránymutatása alapján mire nem gondoltunk a szimuláció során, illetve, ha valóban megtörtént volna az eset, akkor milyen mértékű szankcióra számíthatnánk (EDPB 04/2022).

Összegzés

"Miért kell egy robotnak parancsba adni, hogy engedelmessédjön a parancsoknak - miért nem elég az eredeti parancs? Miért parancsoljuk meg egy robotnak, hogy ne tegyen kárt - nem lenne egyszerűbb, ha eleve nem is parancsolnánk neki, hogy kárt tegyen? Vajon az univerzumban van egy titokzatos erő, amely az entitásokat a rosszindulat felé húzza, ezért egy pozitronikus agyat úgy kell programozni, hogy ellenálljon neki? Vajon az intelligens lények elkerülhetetlenül alakítanak ki magatartásproblémát? (...) Most, hogy a számítógépek valóban okosabbá és erősebbé váltak, az aggodalom alábbhagyott. A ma mindenütt jelenlévő, hálózatba kapcsolt számítógépek soha nem látott mértékben képesek arra, hogy rosszat tegyenek, ha egyszer rosszra vetemednének. De a bajt csak a kiszámíthatatlan káosz vagy az emberi rosszindulat okozza vírusok formájában. Már nem aggodunk elektronikus sorozatgyilkosok vagy felforgató szilícium-összeesküvések miatt, mert kezdjük felismerni, hogy a rosszindulat - akárcsak a látás, a motoros koordináció és a józan ész - nem jár ingyen a számítástechnikával, hanem be kell programozni. (...) Az agresszió, akárcsak az emberi viselkedés minden más, általunk természetesnek tartott eleme, egy kihívást jelentő mérnöki probléma!" (Pinker 2012: 15–16)

Az MI fejlődése új kihívások elé állítja a jogászokat és az informatikusokat. Steven Pinker gondolatai alapján felmerül a kérdés: *hogyan biztosíthatjuk, hogy az MI etikus és megbízható legyen, a társadalom érdekeit szolgálva?* Mi a garanciája annak, hogy egy MI-projekt résztvevői egy irányba mozogva, az irányadó jogi-etikai kereteken belül valósítsák meg elképzeléseiket? Álláspontom szerint ennek az együttműködésnek az egyik alappillére a *hatékony kommunikáció*, éppen ezért tanulmányomban azt vizsgáltam – az adatvédelem és adatbiztonság terén szerzett tapasztalataim alapján –, hogy milyen eszközökkel lehet az ennek hiányából adódó problémákat orvosolni. Álláspontom szerint a megvalósíthatósági tanulmányok, a hatásvizsgálatok, az érdekmérlegelési tesztek, a szabályzatok és az oktatás – többek között, de nem kizárólag – mind alkalmasak lehetnek a jogászok és az informatikusok közötti híd építésére.

A kommunikációs problémák megoldása nélkül etikus MI kifejlesztése és használata szinte lehetetlen, és ehhez egyrészt olyan szereplőkre van szükség, akik kreatívan törekszenek a problémák megoldására, másrészt pedig olyan platformokra, amelyek lehetővé teszik a konstruktív párbeszédet és az együttgondolkodást. A szervezeteknek tehát tudatosan kell felkészülniük MI-projektjeikre, és a siker érdekében érdemes magas kvalitású szakembereket alkalmazni, akik nem csak a modern technológiákban rejlt előnyöket ismerik fel, hanem képesek a hatékony

együttműködésre és a kockázatok kezelésére. Ezen együttműködés nélkül az MI nemcsak szervezetünknek, hanem az érintetteknek, sőt a társadalmunknak is komoly, akár hosszabb távra kiható gondokat okozhat.

Köszönetnyilvánítás

Köszönöm Dr. Botzheim Jánosnak, az Eötvös Loránd Tudományegyetem Mesterséges Intelligencia Tanszékének tanszékvezető egyetemi docensének informatikai területen nyújtott segítségét.

Irodalomjegyzék

- A személyes adatoknak az illetékes hatóságok által a bűncselekmények megelőzése, nyomozása, felderítése, a vádeljárás lefolytatása vagy büntetőjogi szankciók végrehajtása céljából végzett kezelése tekintetében a természetes személyek védelméről és az ilyen adatok szabad áramlásáról, valamint a 2008/977/IB tanácsi kerethatározat hatályon kívül helyezéséről szóló, 2016. április 27-i (EU) 2016/680 európai parlamenti és tanácsi irányelv (bűnügyi adatvédelmi irányelv, LED).
- Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (általános adatvédelmi rendelet, GDPR).
- Az Európai Parlament és a Tanács (EU) 2019/790 irányelve (2019. április 17.) a digitális egységes piacon a szerzői és szomszédos jogokról, valamint a 96/9/EK és a 2001/29/EK irányelv módosításáról.
- Az Európai Unió Alapjogi Ügynöksége (FRA) (2021) *Hogyan alakítsuk jól a jövőt? Mesterséges intelligencia és alapvető jogok*. Összefoglaló.
- Bonnefon, J.-F. (2021) *The Car That Knew Too Much Can a Machine Be Moral?* The MIT Press.
- Bonnefon, J.-F.–Shariff, A.–Rahwan, I. (2016) *The social dilemma of autonomous vehicles*. *Science*, on 24 Jun 2016: Vol. 35.
[doi: 10.1126/science.aaf2654](https://doi.org/10.1126/science.aaf2654) [Letöltve: 2023.09.23.].
- Brownsword, R.–Scotford, E.–Yeung, K. (eds.) (2017) *The Oxford Handbook of Law, Regulation and Technology*. Oxford University Press.
[doi: 10.1093/oxfordhb/9780199680832.001.0001](https://doi.org/10.1093/oxfordhb/9780199680832.001.0001) [Letöltve: 2023.09.23.].
- CISCO (2022) *Consumer Privacy Survey*.
- Coeckelbergh, M. (2020) *AI Ethics*. The MIT Press.
[doi: 10.7551/mitpress/12549.001.0001](https://doi.org/10.7551/mitpress/12549.001.0001) [Letöltve: 2023.09.23.].

- Datatsynet (2018) Artificial intelligence and privacy. Report.
- Dimatteo, L. A.–Poncibò, C.–Cannarsa, M. (eds.) (2022) *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press. doi: [10.1017/9781009072168](https://doi.org/10.1017/9781009072168) [Letöltve: 2023.09.23.].
- DRAFT *Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD))*. 16/5/2023. Version: 1.1 (MI rendelet tervezete).
- Dubber, M.D.–Pasquale, F.–Das, S. (eds.) (2020) *The Oxford Handbook of Ethics of AI*. Oxford University Press. doi: [10.1093/oxfordhb/9780190067397.001.0001](https://doi.org/10.1093/oxfordhb/9780190067397.001.0001) [Letöltve: 2023.09.23.].
- EDPB 1/2021 Iránymutatás az adatvédelmi incidensek bejelentésével kapcsolatos példákról.
- EDPB Guidelines 4/2022 on the calculation of administrative fines under the GDPR.
- EDPB Guidelines 9/2022 on personal data breach notification under GDPR.
- ENISA (2020) *Artificial Intelligence Cybersecurity Challenges*.
- ENISA (2023a) Artificial Intelligence and Cybersecurity Research. <https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research> [Letöltve: 2023.09.23.].
- ENISA (2023b) Cybersecurity of AI and Standardisation.
- Európai Ügyvédi Kamarák Tanács (2022) Útmutató ügyvédek és ügyvédi irodák számára a mesterséges intelligencián alapuló eszközök EU-n belüli használatához.
- European Parliament Special Committee on Artificial Intelligence in a Digital Age: REPORT on artificial intelligence in a digital age (2020/2266(INI)).
- Floridi, L. (ed) (2021) *Ethics, Governance, and Policies in Artificial Intelligence*. Springer. doi: [10.1007/978-3-030-81907-1](https://doi.org/10.1007/978-3-030-81907-1) [Letöltve: 2023.09.23.].
- Floridi, L.–Cowls, J.–Beltrametti, M. et al. (2018) *AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Minds & Machines 28. pp. 689–707. doi: [10.1007/s11023-018-9482-5](https://doi.org/10.1007/s11023-018-9482-5) [Letöltve: 2023.09.23.].
- Garante per la protezione dei dati personali Provvedimento del 30 marzo 2023 [9870832]. <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832> [Letöltve: 2023.09.23.].
- Garante per la protezione dei dati personali Provvedimento dell'11 aprile 2023 [9874702]. <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874702> [Letöltve: 2023.09.23.].
- Government of Canada: Directive on Automated Decision-Making. <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> [Letöltve: 2023.09.23.].
- Hallinan, D.–Leenes, R.–Gutwirth, S.–De Hert, P. (eds.) (2020) *Data Protection and Privacy*. Hart Publishing. doi: [10.5040/9781509941780](https://doi.org/10.5040/9781509941780) [Letöltve: 2023.09.23.].
- Holberton-Turing eskü. <https://www.holbertonturingoath.org/> [Letöltve: 2023.09.23.].
- Howard, J.J.–Rabbitt, L. R.– Sirotnin, Y. B. (2020) *Human-algorithm teaming in face recognition: How algorithm outcomes cognitively bias human decision-making*. PLoS ONE 15(8): e0237855. doi: [10.1371/journal.pone.0237855](https://doi.org/10.1371/journal.pone.0237855) [Letöltve: 2023.09.23.].
- IBM Global AI Adoption Index 2022. <https://www.ibm.com/downloads/cas/GVAGA3JP> [Letöltve: 2023.09.23.].
- ICO (2018) *Data Protection Impact Assessments (DPIAs)*.
- ICO (2020) *AI auditing framework - draft guidance for consultation*.
- ICO (2023) *Guidance on AI and Data Protection*.
- ICO–The Alan Turing Institute (2022) *Explaining decisions made with AI*.
- ILO Working Paper 96 (2023) *Generative AI and Jobs: A global analysis of potential effects on job quantity and quality*. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---inst/documents/publication/wcms_890761.pdf [Letöltve: 2023.09.23.].
- Jarjabka, Á. et al. (2020) Projektmenedzsment ismeretek. Pécs: Pécsi Tudományegyetem Közgazdaságtudományi Kar Vezetés- és Szervezéstudományi Intézet.
- Javaslat az Európai Parlament és a Tanács irányelve a szerződésen kívüli polgári jogi felelősségre vonatkozó szabályoknak a mesterséges intelligenciához való hozzáigazításáról (a mesterséges intelligenciával kapcsolatos felelősségről szóló irányelv), COM(2022) 496 final 2022/0303 (COD).
- Johannessen, J.-A. (2021) *Robot Ethics and the Innovation Economy*. Routledge. doi: [10.4324/9781003174493](https://doi.org/10.4324/9781003174493) [Letöltve: 2023.09.23.].
- Kafka, P. (2023) *The creator of Black Mirror is okay with tech. People, on the other hand ...* VOX Aug 4, 2023.

- https://www.vox.com/technology/2023/8/4/2381929/9/black-mirror-charlie-brooker-interview-ai-peter-kafka-media-column?mc_cid=3f3a9aebb&mc_eid=31136da945
[Letöltve: 2023.09.23].
- Kerrigan, C. (ed.) (2022) *Artificial Intelligence Law and Regulation*. Edward Elgar Publishing Limited.
[doi: 10.4337/9781800371729](https://doi.org/10.4337/9781800371729) [Letöltve: 2023. 09. 23].
- Leroy, J. B. (2021) *Societal Responsibility of Artificial Intelligence*. ISTE Ltd. [doi: 10.1002/9781119831808](https://doi.org/10.1002/9781119831808)
[Letöltve: 2023.09.23].
- Leslie, D. (2019) *Understandig artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute.
[doi: 10.5281/zenodo.3240529](https://doi.org/10.5281/zenodo.3240529) [Letöltve: 2023 09.23].
- Liao, S. M. (ed) (2020) *Ethics of Artificial Intelligence*. Oxford University Press.
[doi: 10.1093/oso/9780190905033.001.0001](https://doi.org/10.1093/oso/9780190905033.001.0001) [Letöltve: 2023.09.23].
- Mandel, G. N. (2017) *Legal Evolution in Response to Technological Change*. The Oxford Handbook of Law, Regulation and Technology. In: Brownsword, R.–Scotford, E.–Yeung, K. (eds.).
[doi: 10.1093/oxfordhdb/9780199680832.013.45](https://doi.org/10.1093/oxfordhdb/9780199680832.013.45)
[Letöltve: 2023.09.23].
- Marino, D.–Monaca, M. A. (eds.) (2020) *Economic and Policy Implications of Artificial Intelligence*. Springer. [doi: 10.1007/978-3-030-45340-4](https://doi.org/10.1007/978-3-030-45340-4)
[Letöltve: 2023.09.23].
- Mehrabi et al. (2022) A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6). pp. 1–35. 25 Jan 2022. [doi: 10.1145/3457607](https://doi.org/10.1145/3457607)
[Letöltve: 2023.09.23].
- Mesterséges intelligenciával foglalkozó magas szintű szakértői csoport (2019) *Etikai iránymutatás a megbízható mesterséges intelligenciára vonatkozóan*.
- Misselhorn, C. (2022) *Artificial Moral Agents: Conceptual Issues and Ethical Controversy*. The Cambridge Handbook of Responsible Artificial Intelligence. Cambridge University Press & Assessment.
[doi: 10.1017/9781009207898](https://doi.org/10.1017/9781009207898)
- NAIH: Az adatvédelmi hatásvizsgálat és előzetes konzultációja.
<https://naih.hu/az-adatvedelmi-hatasvizsgalat-es-elozetes-konzultacioja> [Letöltve: 2023.09.23].
- Naqvi, A. (2020) *Artificial Intelligence for Audit, Forensic Accounting, and Valuation*. John Wiley and Sons. [doi: 10.1002/9781119601906](https://doi.org/10.1002/9781119601906)
[Letöltve: 2023.09.23].
- Pinker, S. (2012) *How the Mind Works*. Penguin Books Ltd.
- Rosemann, A.–Zhang, X. (2021) *Exploring the social, ethical, legal, and responsibility dimensions of artificial intelligence for health – a new column in Intelligent Medicine*. Published by Elsevier B.V. on behalf of Chinese Medical Association.
[doi: 10.1016/j.imed.2021.12.002](https://doi.org/10.1016/j.imed.2021.12.002) [Letöltve: 2023.09.23].
- Slaughter, R. K.–Kopeck, J.–Batal, M. (2021) *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission*. ISP Digital Future Whitepaper & YJoLT Special Publication.
- Tietosuojavaltuutetun toimisto (2020) *Automatisoitujen yksittäispäätösten syntyminen ennakoivan terveydenhuollon työkalussa*. 6482/186/2020.
<https://finlex.fi/fi/viranomaiset/tsv/2022/20221544>
[Letöltve: 2023.09.23].
- UNESCO (2023) *Ethical Impact Assessment - A Tool of the Recommendation on the Ethics of Artificial Intelligence*.
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>
[Letöltve: 2023.09.23].
- Voeneky, S.–Kellmeyer, P.–Mueller, O.–Burgard, W. (eds.) (2022) *The Cambridge Handbook of Responsible Artificial Intelligence*. Cambridge University Press & Assessment.
[doi: 10.1017/9781009207898](https://doi.org/10.1017/9781009207898) [Letöltve: 2023.09.23].
- von Braun, J.–Archer, M. S.–Reichberg, G. M.–Sorondo, M. S. (eds.) (2021) *Robotics, AI, and Humanity*. Springer. [doi: 10.1007/978-3-030-54173-6_1](https://doi.org/10.1007/978-3-030-54173-6_1) [Letöltve: 2023.09.23].
- Wolf, F. (2022) *A Blueprint for the Regulation of Artificial Intelligence Technologies*. Ethics International Press Ltd., UK.
- World Economic Forum (2019) *Guidelines for AI Procurement*.
- World Health Organization (2021) *Ethics and governance of artificial intelligence for health*.
<https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf> [Letöltve: 2023.09.23].
- WP 217 (2014). 06/2014. számú vélemény az adatkezelő 95/46/EK irányelv 7. cikke szerinti jogszerű érdekeinek fogalmáról.
- WP 248 rev.01 (2017) Iránymutatás az adatvédelmi hatásvizsgálat elvégzéséhez és annak megállapításához, hogy az adatkezelés az (EU) 2016/679 rendelet alkalmazásában „valószínűsíthetően magas kockázattal jár”-e.
- Završnik, A.–Simončič, K. (eds.) (2023) *Artificial Intelligence, Social Harms and Human Rights*. Palgrave Macmillan. [doi: 10.1007/978-3-031-19149-7](https://doi.org/10.1007/978-3-031-19149-7)
[Letöltve: 2023. 09. 23].

The newest aspect of industrial espionage. Cyber threats from UAVs to non-state actors

Az ipari kémkedés legújabb aspektusa. Az UAV-k általi kiberfenyegetés a nem állami szereplők tekintetében

DOI: [HTTPS:// DOI.ORG/10.53793/RV.2024.2.5](https://doi.org/10.53793/RV.2024.2.5)

Abstract

The crime of industrial espionage to obtain trade secrets is as old as the dawn of commerce. With the development of modern technology, the techniques for obtaining secrets, or in modern terms, data, have been constantly evolving. With the advent of drones on the market, physical and cyber-attacks against critical infrastructure and non-state actors have increased significantly. The relative ease of modification of UAVs (Unmanned Aerial Vehicle) has led to the emergence of spy vehicles capable of acquiring sensitive data without the ground controller entering the area. Critical infrastructure is protected at the state level, with military and/or national security at the global level, but non-state actors are vulnerable to attack. The aim of this paper is to show how non-state actors are affected by the threat and what options they have to defend themselves. Whether EEA countries are threatened by this form of espionage.

KEYWORDS: CYBER ESPIONAGE, NON-STATE ACTORS, CRITICAL INFRASTRUCTURE, LAW, UAV MITIGATION

Absztrakt

Az üzleti titok megszerzésére irányuló ipari kémkedés bűncselekménye mondhatni egyidős a kereskedelem megjelenésével. A modern technológia fejlődésével a titok, vagy modern kifejezéssel élve az adat megszerzésére irányuló technikák folyamatosan új arculatot öltöttek. A drónok piaci forgalomba kerülésével a kritikus infrastruktúra, valamint a nem állami szereplők ellen irányuló fizikai- és kibertámadások jelentős mértékben megnövekedtek. Az UAV-k (Unmanned Aerial Vehicle, magyarul pilóta nélküli légi jármű) meglehetősen könnyű modifikációja miatt olyan kémkedésre alkalmas egységek jelentek meg, melyek úgy képesek megszerezni a szenzitív adatokat, hogy a földi irányító nem lépi át a területet. A kritikus infrastruktúra állami szinten védett, az elhárítást globális szinten a katonaság, és/vagy a nemzetbiztonság végzi, azonban a nem állami szereplők védtelenek a támadásokkal szemben. Jelen tanulmány célja bemutatni, hogy a fenyegetés milyen formában érinti a nem állami szereplőket, milyen lehetőségeik vannak a védelemre, illetve fenyegeti-e a kémkedés fenti formája az EGT országokat.

KULCSSZAVAK: KIBERKÉMKEDÉS, NEM ÁLLAMI SZEREPLŐK, KRITIKUS INFRASTRUKTÚRA, JOG, UAV MITIGÁCIÓ

It's probably not an exaggeration to say that corporate espionage is as old as commerce. In the surviving sources of the ancient world, from China (Tókei trans 1995), Persia Empire (Briant 2009) and the Greek polis to the Roman Empire (*speculatores, exploratores*), we find a wealth of references to the workings and tasks of military and economic intelligence specialists (Russel 1999; Syväne 2016). To put it cautiously, the sources of the Byzantine Empire almost give us a model of an advanced intelligence service, in which spies (*ακριβείς κατάσκοποι*) obtained information in the interests of the state's security from

the civilian population (*ἀμύχατοι*), especially from market players (*ἐμποροί*), both inside and outside the borders of the empire (Theotokis 2018). When the strategic aspects of the empire required it, spies often used disinformation and its varied tools to gain significant military and economic advantage. Apart from the state, non-state actors of the time also played their part in espionage, mostly in the competitive sphere. In modern terms, rivals sought to discover each other's corporate secrets in order to gain an advantage over competitors. More fundamentally, the medieval commercial sector was not immune to the crime of industrial espionage.

Italian cities, especially Venice and Florence, stood out from their competitors not only for the advanced attributes of their spies, but also for their manipulative strategies (Prajda 2015; Iordanou 2018). Venice's republicanism gave rise to a number of institutions typical of the modern state, such as the Council of Ten (*Consiglio dei Dieci*) and the State Investigatory (*Inquisitori di Stato*), which also organised the network of informers and the police. The international intelligence role of the lagoon city should not be underestimated, if only in the context of the early modern information revolution (Pilch ed. 1936; Ágoston 2014; Preto 2016). It is no coincidence that Venice provided the model for the procedures and tasks of modern intelligence services.

In a globalised world, one of the most important materials is undoubtedly knowledge and the information that builds it. It has become increasingly true that the constant acquisition of knowledge, and the acquisition of knowledge from competitors, even if it is often illegal, is an inherent part of the market economy and the key to maintaining the competitiveness of some companies. In industrial espionage, a firm or individual seeks to obtain the secrets of its competitors by illegal means in order to increase its own market advantage. The object of the crime is information, which can be formal or informal, white, or black. Industrial espionage can take the form of sabotage methods aimed at obtaining or destroying information. Information gathering methods are divided into active and passive methods. Active methods require physical personal presence and influence. It means physical intrusion, access to various electronic systems, theft, copying of confidential documents, copying to a data recorder. Passive methods include electronic eavesdropping, interception of conversations in public and closed places. The *social engineering* aspect is also included as a separate category (Deák 2019).

In the Hungarian legal environment, industrial espionage does not appear in the Criminal Code as an independent offence, which is mainly due to the fact that industrial espionage has various aspects. (Nasheri 2014) The acquisition of data can take several forms. Consequently, there are several elements in the criminal law which focus on secret data unknown to the individual. The term "*trade secret*" is defined in Article 1 (1) of Act LVI of 2018 on the Protection of Trade Secrets as "*a trade secret is a fact, information, other data and a compilation thereof, which is related to an economic activity, is secret, is not known in its entirety or as a set of elements or is not easily accessible to the persons engaged in the economic activity concerned, and therefore has a pecuniary value, and is not known to the public or is not easily accessible to the persons engaged in the economic activity concerned, and*

therefore the holder of the secret acts in a manner that is generally expected in the given situation in order to keep it secret." As can be seen, a trade secret must be indispensably linked to the economic activity of the company. Thus, information relating to the activity in question is in any event covered by the trade secret. The criterion for the information discussed above is that it must not be public knowledge or easily accessible. At the same time, it is clear that the law imposes an obligation on the holder of the secret and states that the information is only considered a secret if the owner of the secret ensures its protection. A specific condition is that the acquisition, use, disclosure, or disclosure of the information by unauthorised persons would harm or jeopardise the economic or market interests of the right holder. Thus, for information to be classified as a secret, it is sufficient that one of the above conditions is fulfilled, no harm or threat to the interest is required (Schubert 2019).

The domestic legal environment regulates the *know-how* as a completely separate category, i.e. proprietary knowledge, as this concept enjoys the same legal protection as trade secrets. According to the above-mentioned law, "*know-how is technical, economic or organisational knowledge, solutions, experience or a compilation thereof, which is a trade secret, and which is recorded in an identifiable manner.*" As defined above, know-how is primarily not data, but knowledge or experience. It may also be a combination of these. Due to the legal context, industrial espionage can be carried out not only by infringing trade secrets, but also by infringing other data and information, including private, confidential, confidential correspondence and trade secrets. In extreme cases, the offences of illicit acquisition of data or misuse of personal data may also exist. The form of industrial espionage is espionage as defined in Article 261 of Act C of 2012 on the Criminal Code. Industrial espionage is covered by this offence, in particular if the target is a state or a company with direct links to the state. This can understandably be damaging not only to economic life but also raises national security concerns. On this basis, the penalty can range from two to eight years, and from five to fifteen years for top secret data. Given the emergence of new forms of espionage with the development of modern technology, it is essential that the legislative environment to curb industrial espionage is constantly evolving and being supplemented. At the same time, the legislator should also consider the introduction of a specific offence, which would effectively prevent the forms of the offence from being carried out.

In recent years, technological progress has given rise to a particular form of industrial espionage. With the commercialisation of UAVs, or *vulgo dicitur*: drones, and

the extremely rapid expansion of the market, the devices have become readily available to the civilian population (Dillow 2011; Schneider 2016). Their widespread use and their versatility of application have quickly brought them to the forefront of criminals' minds and often made them a means of committing crimes. The crimes committed with drones in recent years cover a very wide spectrum. From crimes against aviation, to espionage, to terrorist attacks (Rassler 2016; Hartmann–Giles 2016), they play an extremely important role. Thanks to their power and their easy modification, they can also be used as a payload to carry and operate weapons and various electronic devices. The latter equipment is specifically targeted at obtaining forbidden or protected data, which often includes the theft of bank card details, attacks against social media profiles and attacks to obtain the data files of various messaging software. In addition to crimes against individuals, the business sector has also become vulnerable to drone attacks. This form of attack is not without precedent (Wendt 2020).

In 2011, a U.S. Air Force engineer and cybersecurity contractor collaborated at the Black Hat and Def Con Security Conference in Las Vegas to unveil a 14-pound, 6-foot-long UAV-WASP (Unmanned Aerial Vehicle - Wireless Aerial Surveillance Platform) fixed-wing spy drone. In creating the drone, its creators, Mike Tassej and Richard Perkins, were in fact trying to draw attention to the fact that by commercialising drones, the state and manufacturers have given too much access to technologies that are a quasi-potential threat to private individuals, the state, and non-state actors. In demonstration of this finding, their units are constructed strictly from commercially legally available materials. The model is based on the U.S FMQ-117B military UAV. The body is made of EPO foam, with an E-Flite 90 carbon brushless motor, a CP 85 HV ESC, and 2x6 22.2v 5000mAh LiPo batteries. The control consisted of a JR Spectrum Dx6i transmitter and receiver (2.4 GHz), DIY Ardu Pilot, Servo, Xbee Pro Ada Fruit adapter, while the so-called payload; Via Epia Px5000eg Pico-ITX, USB 4G, Xbee Pro module, and a Universal Serial Radio Peripheral (USRPs). While the BaseStation was made up of a Gumstix Overo Earth, DIY ArduStation, an Asus WL-330gE Wi-Fi Access Point, an Inter P4 3.06 GHz HT processor, 4 Gb Ram, 500 GB HDD, and an Nvidia GTX 470. During the test flight of the aircraft, the controllers were able to intercept several phone calls, collect a significant amount of data, and hack into various Wi-Fi networks (Tassej–Perkins 2011).

In 2015, 4Armed introduced the DJI Phantom 2 Vision+ drone, a user-friendly drone equipped with a Raspberry Pi and an omni-directional antenna. The connection between the UAV and the ground controller

was provided by a Huawei 3G dongle, which minimized the detection capabilities of the device. During the flight, significant data was collected using Snoopy-ng, the detected networks were hacked using Aircrack-ng, (Clarity From Above) and finally the target's system was penetrated by exploiting a security flaw in Microsoft Windows MS08-067. The total cost of the drone and modification was roughly \$1,200. (Greenwood 2015) In the same year, the Singaporean marketing company AdNear used a modified DJI Phantom with very similar equipment to the aforementioned UAVs to target the phones of the Los Angeles victims (Kumar 2015), specifically to collect data to flood potential customers with unwanted advertising on their phones (Paráda–Tóth 2020).

In 2017, BishopFox's Danger Drone mUAV (multirotor UAV) was unveiled at Def Con. The Danger Drone's frame is 3D printed. The brain of the unit, so to speak, was an Erle Brain 2, equipped with an Ublox Neo-M8N GPS, a Turnigy TGY-i6 LE 4x MN2213 Motor, HKPilot tansceiver Telemetry Radio Set V2 (915 MHz), Flouren 4S 35C 14.8V 5500mAh Li-Polymer RC Battery Pack, Andoer 4 Pcs Simonk 30AMP regulator, RipaFire F450 4-Axis multirotor, Eriocco Power Supply Module w/BEC APM 2.5 APM, and a PPM Encoder Module HKPilot 32 has been fitted. The so-called utility box consisted of a SENA UD100 USB, Bluetooth 4.0 USB, Wi-Spy DbX Pro Spectrum Analyzer, Wifi Pineapple Nano, TP-Link TL-WN72N, Asus USB-N53, a Crazyradio 2.4GHz nR24LU1+USB radio dongle, and Atmel - ZigBee hacking gear. The total cost was just over \$ 500. The 4G communication system made the device capable of evading the latest jammers, approaching the target undetected and breaking into the system. By building the device, the company wanted to highlight the importance of protection and vulnerability. In their presentation, Francis Brown and David Latimer highlighted that the weak defence architecture of IoT Home & Office (Internet of Things) systems makes it easy for UAVs like Danger Drone to enter the network and take control of devices (Bramlette 2019). In this respect, they also sought to highlight the particular importance of taking physical and cyber defence measures, especially for large enterprises (Brown–Latimer 2017).

In the latest and most dangerous form of corporate espionage, the perpetrators most often use UAVs. As can be seen from the above, drones are extremely easy to use to carry out physical and cyber-attacks. They are readily and legally available commercially (Findings 2018). Due to their technical attributes, they can also be modified to be able to steal company secrets and customer information. But they are also capable of allowing the perpetrator to monitor the target

permanently while keeping his identity hidden. In the past, the traditional techniques of industrial espionage were based on bugging, wiretapping and infiltration, but today this form of crime has almost entirely moved into cyberspace (Holland 2020). With the advent of the information society and cloud technology, the perpetrator can now successfully carry out espionage without physically crossing the company's premises. This leaves companies almost defenceless against cyber-attacks by UAVs, as current legislation essentially does not allow them to legally and legitimately take action against a spying UAV. At global level, there is a general belief that counter-interception (C-UAS) is the exclusive competence of civil or military national security services. Private individuals and non-state actors must take their own measures to protect their data stored in cyberspace.

In July 2020, drones were spotted over chemical plants in Louisiana, prompting the Federal Bureau of Investigation (FBI) to report on the potential for espionage and terrorism involving critical infrastructure (Barnes 2020). According to the report, obtained by CNN, the FBI wrote that such flights could be an effective way to monitor critical infrastructure. Security and law enforcement forces at the facility have limited ability to detect and deter the threat. On July 29, 2020, observers saw two drones fly over the facility, split in two a few feet above the facility and continue flying in different directions. The FBI, taking measures to protect the facility, notified the U.S. Department of Homeland Security, which has counterintelligence authority, and involved the Cybersecurity & Infrastructure Security Agency (CISA) in organizing the protection (Protecting against the threat 2020). On 8 March 2021, security again discovered a drone near the plant's pipelines. Subsequently, for security reasons, the FBI encouraged managers of critical infrastructure facilities to contact the agency's field office immediately in the event of a drone detection. The Department of Homeland Security has also issued a briefing on the incident, warning that drones are extremely useful tools, but in the wrong hands they can also be used as explosive weapons or for espionage. Therefore, the State has taken steps to enhance the protection of critical infrastructure (Sands 2022). The FBI's field office and CISA have issued a new report on the incident, underlining that the detection of drones poses a major problem in finding the pilot on the ground. The critical infrastructure protection system may not be able to find the pilot, nor can it assess what information has been obtained from the espionage activity. The FBI added to the statement that a drone that crashed near a power station in Pennsylvania in 2020 was investigated and there were strong suspicions that the device was used to damage

infrastructure, so the National Counterterrorism Center was involved in investigating the case. Overall, the above example shows that in the event of physical and cyber-attacks against critical infrastructure, the state will seek to extend its protection to the infrastructure under attack using all means at its disposal (Sneath 2021).

Currently, in the United States, this form of espionage is most prevalent among large non-state corporations such as General Motors Company, Gillette Company, Google LLC, HP Inc, Microsoft Corporation, Tesla Inc, Meta Platforms Inc, and Apple Inc. A major problem is that cyber espionage allows perpetrators to obtain data in cyberspace without companies ever realising that they have been the victims of an attack. At the same time, if an attack is identified, it is very difficult to determine how much damage has been done to the company as a result of the crime. The investigation is further complicated by the fact that, for example, as the company is globally responsible for protecting trade secrets, it must itself take care of the protection of sensitive proprietary data. In this regard, in the United States, in the course of an investigation, the authorities must carefully consider (i) the extent to which the secret information obtained was known outside the company, (ii) the extent to which it was known by employees, (iii) the extent to which the information was kept secret, (iv) the advantage the information may have over competitors, (v) and what the company involved in the espionage did to protect the information. Taking all the above into account, it is common that if the company detects the threat in time and its information is leaked, it will not report it (Rowe 2016). Thus, no criminal prosecution is initiated. There is no way to detect the crime. Another fundamental problem is that companies are unable to estimate the potential damage caused by a crime in such cases (Kosseff 2020; Scott 2021).

We can only cautiously infer that Apple Inc, Meta Platforms Inc, and Tesla Inc, may have been victims of UAV spying in the past, as they have taken the initiative to declare their campuses as "No Drone Zones" due to the numerous illegal flights detected. Of these companies, Apple Inc. was perhaps the most affected. The construction of their campus in Cupertino, California, USA, has attracted drone pilots. During the illegal flights, the company's construction site was constantly monitored, and there were even attempts to hack into the newly installed network. For this reason, Apple Inc. management first designated the campus area as a No Drone Zone (Aisight report), trusting that commercially available drones would be stopped by geofencing (Marketwatch report 2018). Word soon spread on the internet that it was no longer possible to

fly into the campus. But that didn't deter adventurous pilots and the company's rogue pilots. In 2018, the company's security team decided to take action to protect its trade secrets against the UAV threat. In April 2018, a young pilot attempted to fly onto the campus to take photos of the construction site but was warned by Apple's security team to leave the area in less than ten minutes. Otherwise, they will be forced to take action. The pilot turned his UAV around and landed away from the campus. According to unconfirmed sources, Apple Inc. may have purchased a detection and defeat system marketed by Dedrone Holdings to protect its business secrets and infrastructure (Marketwatch report 2018).

In 2020, two researchers used a drone-mounted Wi-Fi to show how to hack a Tesla car in the Pwn200wn hacking competition. The hackers used security vulnerabilities in Tesla's infotainment system to penetrate the vehicle's system and take control (Kovács 2019). The attack allowed them to modify the basic functions: music playback, air conditioning, steering, and acceleration modes, as well as unlock the doors (Kovács 2020). But not remote control. Nevertheless, the attack has shown that using Wi-Fi and the right software, an attacker can hack into the systems of any Tesla Model S, 3, X, and Y from up to 100 meters away and open the vehicle's doors, which obviously makes it easier to steal the vehicle. After the race organisers informed the manufacturer of the results, the bug was fixed in October 2020. However, it has emerged that the Connman component is widely used in the automotive industry and similar attacks could be launched against other manufacturers' vehicles. The attack on Tesla's vehicle systems is not the first. In 2020, Tesla researchers attacked their own car to address the weaknesses of the Advanced Driving Assistance System (ADAS) (Nasi et al. 2020). The tests confirmed the vulnerability and manipulability of the system. Volke Automotive in Germany has also built its detection and defeat system partly to prevent industrial espionage and partly to protect vehicles from industrial espionage. Volke is known as one of the world's leading technical developers in automotive design. Protecting the company's intellectual property is therefore a priority (Dedrone/Volke).

The examples above show that in the case of state actors, state bodies extend their protection to all critical infrastructure institutions, but non-state actors can only legally protect data stored in cyberspace. In the face of external physical and cyber-attacks from the air, the company is defenceless. Current technological tools allow the use of drones to steal proprietary and confidential trade secrets that give competitors and adversaries an unwarranted advantage. In addition, it

can be seen that illicit data mining against companies by drone poses a potential threat not just to one region, but globally, even to European companies. Consequently, since the legal environment only gives the state the right to counteract, non-state actors are trying to discourage corporate espionage by alternative or currently less legal methods. As the above is a very new practice of spying, the legislator should consider the possibility of granting the state's economic partners the right to detect and intercept data, even if limited, in addition to designating economic areas as No Drone Zones. Domestic practice is fully analogous to US legislation. Government Decree 4/1998 (I. 16.) on the use of Hungarian airspace 9/A. § (2) of the Hungarian Act of 19/19/1978 on the use of Hungarian airspace, any unmanned aircraft engaged in air traffic which a) violates the laws on air traffic and the operation of unmanned aircraft, and thereby endangers the safety of air traffic or the inviolability of privacy, may be intercepted, identified, its operation may be electronically interfered with, and it may be required to land, and it may be forced to land by electronic or mechanical means; b) which constitutes a threat to the security of the protected installation; c) which is used for an unlawful purpose; d) which has reasonable grounds to believe that its flight is being used for the illegal transport of various substances, in particular weapons or drugs, or for the improper transport of dangerous substances; e) which makes unauthorised use of Hungarian airspace. Pursuant to paragraph (3) of the Act: the Hungarian Defence Forces, the Military National Security Service, the Office for the Protection of the Constitution, the National Security Service, the Information Office, the Counter Terrorism Centre, the Parliamentary Guard, and the designated law enforcement agencies performing law enforcement tasks under the control of the Minister of the Interior shall be entitled to take measures pursuant to paragraph (2). In reality, these tasks are carried out by the Military National Security Service and the Specialised National Security Service with the full range of counter-intelligence resources. Obviously, there are types of C-UAS systems that can cause unforeseeable damage in unqualified hands, and it is therefore necessary to limit their use moderately. Consideration should be given in the future to the possibility of using licensed systems that can only 'soft kill' an unauthorised unit flying into an economic area, relying strictly on jammer and drone gun devices. Nevertheless, it is necessary to make companies aware of this threat and the possibilities for self-help as a preventive measure (Scott 2021).

Bibliography

- Ágoston, G. (2014) *Információszerzés és kémkedés az Oszmán Birodalomban a XV-XVII. században. Európa és az Oszmán Hódítás.* Budapest.
- Barnes, S. (2020) Real and present danger: Industrial plants face a new level of threats from drones, cyberattacks and corporate espionage. *IoT Industry Report*.
<https://www.iotindustryreport.com/safety/real-and-present-danger-industrial-plants-face-a-new-level-of-threats-from-drones-cyberattacks-and-corporate-espionage/> [Accessed: 02.02.2024].
- Bramlette, C. M. (2019) *Cyber-Attack Drone Payload Development and Geolocation via Directional Antennae.* (Thesis.) Faculty Department of Electrical and Computer Engineering Graduate School of Engineering and Management Air Force Institute of Technology.
- Briant, P. (2009) 'Le thème de la « décadence perse » dans l'historiographie européenne du XVIIIe siècle: remarques préliminaires sur la genèse d'un mythe'. In: L. Bodio–V. Mehl–J. Oulhen–F. Prost–J. Wilgaux (eds.) *Chemin faisant: mythes, cultes et société en Grèce ancienne. Mélanges en l'honneur de Pierre Brulé.* Presses universitaires de Rennes, Rennes. pp. 19–38.
- Brown, F.–Latimer, D. (2017) *Game of Drones – Putting the emerging „Drone Defense” Market to the test.* Las Vegas, Def Con.
- Bunker, J.–Sullivan J. P. (2021 eds) *Criminal Drone Evolution: Cartel Weaponization of Aerial IEDs.* XLibris US.
- Bunker, R. J. (2015) *Terrorist and Insurgent Unmanned Aerial Vehicles: Use, Potentials, and Military Implications.* Strategic Studies Institute, US Army War College. pp. 1–55.
- Cafarella, J.–Dunford, J.–Land, M.–Wallace, B. (2020) *Turkey Commits to Idlib.* – Institute for the Study of War, 18 March 2020.
<https://understandingwar.org/background/turkey-commits-idlib> [Accessed: 03.02.2024].
- Clarity From Above: PwC Global report on the commercial applications of drone technology.
<https://www.pwc.pl/pl/pdf/clarity-from-above-pwc.pdf> [Accessed: 01.02.2024].
- Crino, S.–Dreby, C. (2020) *Drone Attacks Against Critical Infrastructure: A Real and Present Threat.* Atlantic Council, May 1.
<https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/drone-attacks-against-critical-infrastructure-a-real-and-present-threat/> [Accessed: 27.10.2023].
- Deák, V. (2019) *Social engineering alapú információszerzés a kibertérben megvalósuló lélektani műveletek során.* *Hadtudományi Szemle*, Vol. 12. Nr. 3. pp. 95–111.
- Dedrone Official site, Customers
<https://www.dedrone.com/customers/volke> [Accessed: 02.02.2024].
- Dillow, C. (2011) "A DIY UAV That Hacks Wi-Fi Networks, Cracks Passwords, and Poses as a Cell Phone Tower". *Popular Science*, 2011. [Online].
<https://www.popsci.com/technology/article/2011-07/diy-uav-hacks-wi-fi-networks-cracks-passwords-and-poses-cell-phone-tower> [Accessed: 31.01.2024].
- Findings, K. (2018) *Cybersecurity risks Posed by Unmanned Aircraft Systems.* Homeland Security. National Protection and Programs directorate. Office of Cyber and Infrastructure Analysis.
<https://www.hsdl.org/c/view?docid=825402> [Accessed: 03.02.2024].
- Greenwood, J. (2015) *The Phantom Menace – Weaponising a Consumer Drone.*
<https://www.4armed.com/blog/phantom-menace-weaponising-drones/> [Accessed: 03.02.2024].
- Hartmann, K.–Giles, K. (2016) *UAV Exploitation: A New Domain for Cyber Power.* 2016 8th International Conference on Cyber Conflict.
https://www.researchgate.net/publication/305871943_UAV_exploitation_A_new_domain_for_cyber_power [Accessed: 05.01.2024].
- Holland, A. M. (2020) *Unarmed and Dangerous. The Lethal Applications of Non-Weaponized Drones.*
<https://dronecenter.bard.edu/files/2020/03/CSD-Unarmed-and-Dangerous-Web.pdf> [Accessed: 04.02.2024].
- Iordanou, I. (2018) *The Spy Chiefs of Renaissance Venice: Intelligence Leadership in the Early Modern World.* Washington.
- Kosseff, J. (2020) *Hacking Cybersecurity Law.* *University of Illinois Law Review*, 2020.
<https://ssrn.com/abstract=3331350> or <http://dx.doi.org/10.2139/ssrn.3331350> [Accessed: 04.02.2024].
- Kovács, E. (2019) *Pwn2own 2019: researchers Win Tesla After Hacking Its Browser.* *SecuriteWeek.Cybersecurity News, Insights & Analysis.*
<https://www.securityweek.com/pwn2own-2019-researchers-win-tesla-after-hacking-its-browser/> [Accessed: 02.02.2024].
- Kovacs, E. (2020) *Tesla Car Hacked remotely From Drone via Zero-Click Exploit.* *SecuriteWeek.Cybersecurity News, Insights & Analysis.*
<https://www.securityweek.com/tesla-car-hacked-remotely-drone-zero-click-exploit/> [Accessed: 02.02.2024].
- Kumar, M. (2015) *Drones Spying on Cell Phone Users for Advertisers.*

- <https://thehackernews.com/2015/03/drone-cell-phone-spy.html> [Accessed: 31.01.2024].
- Nasheri, H. (2004) *Economic Espionage and Industrial Spying*. Cambridge University Press, Cambridge.
- Nassi B.–Nassi, D.–Netanel, R. B.–Mirsky, Y.–Drokin, O.–Elovici, Y. (2020) *Pahntom of the ADAS: Securing Advanced Driver – Assistance Systems from Split-Second Pahntom Attacks*. ACM SigSac Conference on Computer and Communications Security. 2020. pp. 293–308.
<https://dl.acm.org/doi/10.1145/3372297.3423359>
- Paráda, I.–Tóth, A. (2020) A Metasploit tulajdonságai egy biztonságos FTP démon exploit tükrében. *Hadmérnök*, Vol. 15. Nr. 3. pp. 219–230.
- Pilch, J. (ed 1936) *A hírszerzés és a kémkedés története*. Budapest.
- Prajda, K. (2015) *Justice in the Florentine Trading Community of Late Medieval Buda*. Mélanges de l'École française de Rome-Moyen Âge.
<https://journals.openedition.org/mefrm/2716> [Accessed: 03.02.2024].
- Preto, P. (2016) *I servizi segreti de Venezia. Spionaggio e controspionaggio ai temo delle Serenissima*. Milano.
Protecting against the threat of unmanned aircraft systems (UAS). An Interagency Security Committee Best Practice. U.S. Department of Homeland Security, Cybersecurity and Infrastructure Security Agency, Integrity Security Committee, 2020.
- Rassler, D. (2016) *Remotely Piloted Innovation: Terrorism, Drones and Supportive Technology*. Combating Terrorism Center, No. October.
<https://ctc.westpoint.edu/remotely-piloted-innovation-terrorism-drones-and-supportive-technology/> [Accessed: 05.01.2024].
- Rassler, D. (2018) *Drone Games, Terror Drone Diffusion, and Near-Term Threats*. The Islamic State and Drones: Supply, Scale, and Future Threats.
<https://www.jstor.org/stable/resrep21486.7> [Accessed: 27.10.2023].
- Rassler, D. (2018) *The Islamic State Drones*. Supply, Scale, and Future Threats. Combating Terrorism Center at West Point.
- Report (2022) *FBI Warns of drone risk after detections at Louisiana chemical facilities*. Report Staff. 102industryreport.
<https://www.102industryreport.com/technology/fbi-warns-of-drone-risk-after-detections-at-louisiana-chemical-facilities/> [Accessed: 02.02.2024].
- Rowe, E. A. (2016) *Rats, Traps, and Trade Secrets*. 57 B.C. L. REV 381 (2016)
<https://scholarship.law.ufl.edu/cgi/viewcontent.cgi?article=1769&context=facultypub> [Accessed: 20.03.2024].
- Russel, F. S. (1999) *Information Gathering in Classical Greece*. Michigan.
- Sands, G. (2022) *FBI warns drones pose potential risk to critical infrastructure after some spotter over Louisiana chemical facilities*.
<https://edition.cnn.com/2022/09/30/politics/drones-risk-critical-infrastructure-spotted-louisiana-chemical-facilities/index.html> [Accessed: 02.02.2024].
- Schneider, M.–Lichte, D.–Witte, D.–Gimbel, S.–Brucherseifer, E. (2021) *Scenario Analysis of Threats Posed to Critical Infrastructures by Civilian Drones*. 31st European Safety and Reliability Conference, ESREL pp. 520–527. Research Publishing Services. Proceedings of the 31st European Safety and Reliability Conference (ESREL 2021), Angers, Frankreich. doi: [10.3850/978-981-18-2016-8_234-cd](https://doi.org/10.3850/978-981-18-2016-8_234-cd). ISBN 978-981182016-8.
- Schubert, B. (2019): *Az ipari kémkedés megjelenése a magyar büntetőjogban – a Huawei-ügy tükrében*. Ars Boni.
<https://arsboni.hu/ipari-kemkedes-a-huawei-ugy-tukreben/> [Accessed: 30.01.2024].
- Scott, C. (2021) *Corporate Espionage by Drone: Why Corporations Need Better Physical and Legal Protections*.
<https://ssrn.com/abstract=3772434> or <http://dx.doi.org/10.2139/ssrn.3772434> [Accessed: 04.02.2024].
- Sneath, S. (2021) *Gov. Edwards gets bill that doubles maximum punishment for flying drones over critical infrastructure*. Petrochemical facilities, pipelines, and grain elevators are no-fly zones for drones. Louisiana Illuminator.
<https://lailluminator.com/2021/06/09/gov-edwards-gets-bill-that-doubles-maximum-punishment-for-flying-drones-over-critical-infrastructure/> [Accessed: 02.02.2024].
- Syvänne, I. (2016) *The Eyes and Ears: The Sasanian and Roman Spies ca. AD 222-450*. *Historia I Swiat*, Nr. 5. pp. 107–131.
- Tassey, M.–Perkins, R. (2011) *Wireless Aerial Surveillance Platform*. DEFCON 19, 2011.
<https://www.defcon.org/images/defcon-19/dc-19-presentations/Tassey-Perkins/DEFCON-19-Tassey-Perkins-Wireless-Aerial-Surveillance-Platform.pdf> [Accessed: 31.01.2024].
- Theotokis, G. (2018) *Byzantine Military Tactics in Syria and Mesopotamia in the 10th Century*. Edinburgh.
- Tókei, F. (trans 1995) *Szun-Ce. A hadviselés törvényei*. Budapest.
- WatchMArker Report (2018)
<https://www.marketwatch.com/story/apple-cracks-down-on-drone-pilot-who-shoots-epic-apple-campus-videos-2018-04-16> [Accessed: 02.02.2024].

Wendt, P.–Voltes, A.–Suau-Sanchez, P. (2020)
Estimating the costs for the airport operator and airlines of a drone-related shutdown: an application to Frankfurt International Airport. *Journal of Transportation Security*, Vol.13, Nr. 1, pp. 93–116.

Az üzleti pszichológia és coaching adaptálhatósága a rendvédelmi szervek mindennapi működésébe egy empirikus vizsgálat tükrében

The adaptability of business psychology and coaching in the daily work of law enforcement agencies in the light of an empirical study

DOI: [HTTPS:// DOI.ORG/10.53793/RV.2024.2.6](https://doi.org/10.53793/RV.2024.2.6)

Absztrakt

Jelen tanulmányban az üzleti pszichológia és a coaching értelmezését követően egy empirikus kutatás eredményei kerülnek bemutatásra, melyben gyakorló pszichológusok és coach szakember vélekedését ismerhetjük meg a témával kapcsolatban. A felmérés az üzleti pszichológia és coaching rendvédelmi szervezetekben való beépítésének lehetőségeire fókuszál. A hierarchikus struktúrájú szervezet számos egyéniséggel rendelkezik, akik a rendvédelmet választották hivatásuknak, sok ezren több évtizede a rendszert szolgálják. Egy szervezet működtetése nehéz feladat, amely a rendvédelem vonatkozásában fokozottan igaz, hiszen a parancsuralmi rendszer, a jogszabályi kötelek alapjaiban meghatározzák a szervezet kereteit, a szervezeti kultúrát. Azt viszont nem szabad figyelmen kívül hagyni, hogy a rendszerben dolgozó személyek nem csupán fizikai testek, ők is érző emberek, akiknek a testi egészségük mellett a lelki egészségükre is figyelni kell ahhoz, hogy holisztikus értelemben véve egészségesek tudjanak maradni.

KULCSSZAVAK: ÜZLETI PSZICHOLÓGIA, COACH, RENDVÉDELEM

Abstract

This study presents the results of an empirical study on business psychology and coaching, which includes the views of practicing psychologists and coaches. The survey focuses on the potential of business psychology and coaching in law enforcement organisations. An organisation with a hierarchical structure has a large number of individuals who have chosen law enforcement as their vocation, many thousands of whom have served the system for decades. Running an organisation is a difficult task, and this is particularly true in the case of law enforcement, where the command structure and legal constraints fundamentally determine the framework and culture of the organisation. However, it must not be overlooked that the people working in the system are not just physical bodies, they are also sentient human beings who, in addition to their physical health, also need to be mentally healthy in order to remain healthy in a holistic sense.

KEYWORDS: BUSINESS PSYCHOLOGY, COACH, LAW ENFORCEMENT

Bevezetés

A rendvédelmi szervezetek minden tagja napi szinten szembesül a társadalom komplex és gyakran kihívást jelentő problémáival, melyből kifolyólag működésük pszichológiai aspektusa nem csak a belső dinamikára, azaz a szervezet tagjainak lelki egészségére és egymás közötti interakcióira terjed ki, hanem a külső kapcsolatokra, azaz a közösséggel és az állampolgárokkal való interakciókra is.

A rendvédelmi szervezetekben dolgozók pszichológiai állapota kiemelten fontos, hiszen a munka természete számos stressztényezővel jár, mint például a személyes biztonságuk érzékelhető kockázata, a szakmai döntések gyakran életet és halált jelentő súlya, vagy az érzelmi túlerheltség. A rendvédelmi szervezetek pszichológiai támogatása nemcsak a tagok személyes jólétét és

munkahelyi teljesítményét javítja, hanem azokat a módszereket és eszközöket is megadja számukra, amelyek segítségével képesek hatékonyan kezelni a társadalommal szembeni felelősséget. Ezenkívül a pszichológia szerepe a rendvédelmi szervezetek közösséggel való kapcsolatában is elengedhetetlen. A rendvédelmi szervezetek és az általuk szolgált közösségek közötti bizalom és kommunikáció javítása érdekében a pszichológiai ismeretek alapvető jelentőségűek. Ez a kölcsönhatás kihívásokat és lehetőségeket is rejt magában a rendvédelmi szervezetek számára a jövőbeli fejlődés és adaptáció terén.

Az üzleti pszichológia fogalma és célja

Általánosságban elmondhatjuk, hogy az üzleti pszichológia hozzájárul ahhoz, hogy jobban megértsük, hogyan működnek a szervezetek és hogyan reagálnak a belső és külső változásokra. A pszichológia segítségével jobban megérthetjük a munkahelyi viselkedést, a motivációkat, a teljesítményt, a vezetést és a csapatdinamikát, így hatékonyabban tudunk változtatásokat véghez vinni és a felmerülő problémákat kezelni.

Mint minden más ágazat, a rendvédelem is igyekszik felvenni a versenyt az új kihívásokkal. A társadalmi, politikai, gazdasági helyzet állandó jelleggel befolyásolja a közhangulatot, így az állampolgárok viselkedését is. Nehéz lépést tartani ezen változó, megújuló igényekkel. A rendvédelmi szervekre általánosságban igaz, hogy kevésbé felkészültek a változásokra, szervezeti mobilitásuk lassúbb, mint a piaci szereplőké, amelynek egyrészt oka a szervezet mérete, másrészt maga a hierarchikus struktúrájú szervezeti felépítés, amely mindennemű intézkedést, döntést, újítás lehetőségét jelentősen lelassít (Hídvégi–Balogh 2022).

Farkas (2018) megfogalmazásában „a közszolgálatban dolgozók tudásának pszichológiai ismeretekkel való kiegészülése hozzásegítheti a szervezetben dinamikus zajló folyamatok megértését, amelyen keresztül a szervezet hatékonyabb működése valósulhat meg. A köztisztviselők a feladataik ellátása során olyan döntéseket hoznak – adott esetben közhatalmat gyakorolhatnak, kényszert alkalmazhatnak –, amelyek hatással lehetnek nemcsak az egyén, hanem a társadalom kisebb-nagyobb közösségeinek életére is. Feladataik sokrétűek, döntéseik nagy horderejűek lehetnek, így a társadalmi felelősségvállalásuk az átlagnál magasabb színvonalon érvényesül, ezért az emberi magatartás alapvető törvényszerűségeinek elsajátítása különösen fontosá válik szakmai fejlődésük szempontjából” (Farkas 2018: 13).

Az üzleti pszichológia tudománya nagyon széles területet fed le, az egyén, a csoportok és a szervezetek szintjén egyaránt. Ez a diszciplína azonosítja és elemzi a munkahelyi viselkedést, az emberi teljesítményt és a szervezeti folyamatokat annak érdekében, hogy a szervezetek hatékonyabban működjenek (Arnold et al. 2016). Az ABP (The Association for Business Psychology²⁹) megfogalmazásában az üzleti pszichológia a munkaélet javításának tanulmányozása és gyakorlata. Az emberi viselkedés tudományának megértését ötvözi a munka világának tapasztalatával, hogy hatékony és fenntartható teljesítményt érjen el mind az egyének, mind a szervezetek számára. Az üzleti pszichológia alkalmazott tudomány, amely azt vizsgálja, hogyan lehet

az embereket és a szervezeteket hatékonyabbá tenni. Társadalomtudományi kutatási módszereket használ az emberek, munkahelyek és szervezetek tanulmányozására, hogy jobban összehangolja a sokrétű és néha egymással versengő igényeiket. Célja, hogy egészséges, produktív és kölcsönösen előnyös kapcsolatokat hozzon létre az emberek és a szervezetek között.

A coaching fogalma és célja

A coaching, coach fogalmak mára már beépültek a köztudatba, mindenki hallotta, ismeri ezen kifejezéseket. A business coach angol szót, ha szó szerint fordítjuk magyarrá, üzleti edzőt jelent, olyan edzőt, aki a legjobbat hozza ki az ügyfélből: „A business coach személyre szabott vezetőfejlesztő. Személyre szabott, mert nem egy általános prezentációt mond el, hanem az ügyfél aktuális dilemmáival foglalkozik” (Komócsin 2018: 21).

A coaching egy személyközpontú fejlesztőeszköz, amely támogatja az embereket személyes és szakmai céljaik elérésében. A coaching folyamat során a coach segít az ügyfélnek önmagát jobban megismerni, erősségeit felismerni és fejleszteni, valamint a kihívásokkal és változásokkal hatékonyan megbirkózni. A coaching célja, hogy növelje az ügyfél önreflexióját, önirányítást és problémamegoldó képességét, elősegítve ezzel fejlődését és teljesítményét. A coaching interaktív, személyre szabott fejlesztési folyamat, melynek során a coach segít az ügyfélnek megfogalmazni és elérni személyes és szakmai céljait. Ez a kapcsolat dinamikus és gyakran egyenrangú, ahol a coach elsődlegesen arra összpontosít, hogy aktiválja és fejlessze az ügyfél belső erőforrásait (Whitmore 2009).

A coaching az üzleti pszichológia természetes kiegészítője, amelynek célja, hogy fejlessze az emberek készségeit, hogy a lehető legjobban hozzájáruljanak a szervezet céljainak eléréséhez (Grant 2012). A coaching segít az embereknek megérteni és kiaknázni saját erősségeiket, valamint kezelni gyengeségeiket. Ezenkívül segít az embereknek a személyes és szakmai céljaik elérésében, segít kezelni a stresszt és a konfliktusokat, valamint segít felkészülni a változásokra és az új kihívásokra (Kilburg 1996). A coaching hatékonysága a pszichológiai elveken és a pozitív pszichológián alapul (Seligman–Csikszentmihályi 2000). A pozitív pszichológia célja, hogy elősegítse az emberek jólétét és boldogságát, ill. segítse őket abban, hogy kihozzák magukból a legjobbat. A coaching során használt technikák, mint például a célkitűzés, a visszajelzés, a szelf-reflexió és az erősségekre

kiválóságok alkotják, akik az üzleti tevékenység és a benne dolgozó emberek életének minőségi javítását tűzték ki célul.

²⁹ Az ABP (The Association for Business Psychology) csoportot széles körben elismert szakmai és gyakorlati

összpontosítás mind a pozitív pszichológia elvein alapulnak.

A rendvédelmi szervezetekben dolgozók a társadalom biztonságának fenntartásáért felelnek, ami jelentős stresszt és nyomást jelenthet személyes és szakmai életükben. A coaching szemlélet alkalmazása egy olyan eszköz lehet, amely támogatja a személyes fejlődést, a hatékonyságot, a csapatmunkát és a vezetői fejlődést. Stréhli (2020) doktori értekezésében a közszolgálati vezetőképítés új útjait vizsgálja, az action learning módszerének alkalmazási lehetőségeit a közigazgatásban és rendvédelemben. Összegezett következtetésként a szerző megállapítja, hogy „a vezetőknek szemléletmódváltásra van szükségük és új megoldásokra, ha hatékonyan szeretnének reflektálni a rájuk váró kihívásokra. A vezetőknek praktikus, gyakorlatias, kézzelfogható támogatást kell nyújtani” (Stréhli 2020: 193).

Az üzleti pszichológia és coaching kapcsolata és különbségei

Az üzleti pszichológia és a coaching kapcsolata szoros és komplementer, mivel mindkét diszciplína az emberek és szervezetek fejlődését és hatékonyságának javítását tűzte ki célul. Az üzleti pszichológia az egyének, csapatok és szervezetek munkahelyi viselkedésének megértésére összpontosít, míg a coaching elsősorban a személyes és szakmai fejlődés elősegítését célozza meg (Passmore–Fillery-Travis 2011).

Az üzleti pszichológia és a coaching között azonban vannak jelentős különbségek is. Az üzleti pszichológia általában nagyobb hangsúlyt fektet a kutatásra és a viselkedési minták elemzésére, míg a coaching gyakorlatorientáltabb és inkább az egyének vagy csoportok fejlesztésére fókuszál. Emellett, míg az üzleti pszichológia a munkahelyi viselkedést kontextusban vizsgálja, a coaching folyamat gyakran összpontosít az ügyfél belső perspektívájára és személyes céljaira (Grant 2011). Annak ellenére, hogy vannak különbségek, az üzleti pszichológia és a coaching gyakran egymást kiegészítve használható a szervezeti fejlesztésben. Az üzleti pszichológiai kutatások értékes betekintést nyújthatnak a munkahelyi viselkedésbe és a szervezeti folyamatokba, míg a coaching lehetővé teszi ezeknek az ismereteknek a gyakorlatba történő átültetését, elősegítve az egyének és szervezetek fejlődését (Grant–Palmer 2002).

Az üzleti pszichológia és a coaching együttes alkalmazása erőteljes eszköztárat kínál a szervezeti fejlesztéshez. Mindkettő olyan pszichológiai ismereteket és technikákat használ, amelyek az emberek viselkedésének és gondolkodásának megértésére, illetve fejlesztésére irányulnak (Passmore–Peterson–Freire

2013). Bár mind az üzleti pszichológia, mind a coaching az emberi viselkedés megértésére törekszik, a coaching gyakran specifikusabb és személyre szabottabb megközelítést alkalmaz. A coachingban a központi szerepet a coach és az ügyfél közötti kapcsolat játssza, amelyben a coach segíti az ügyfél autonómiáját és felelősségvállalását a saját fejlődéséért (Rogers 2012). Ugyanakkor, míg a coaching gyakran az egyéni fejlődést és teljesítményt hangsúlyozza, az üzleti pszichológia szélesebb látószögű és magában foglalja az egyének, csapatok és szervezetek szintjét. Az üzleti pszichológia tanulmányozza az emberi viselkedés szervezeti kontextusban való megértését és befolyásolását, beleértve a motivációt, a vezetést, a csapatmunkát, a szervezeti változást és kultúrát (Arnold et al. 2016). Ezért, bár vannak különbségek az üzleti pszichológia és a coaching között, ezeket a különbségeket gyakran előnyként lehet használni a szervezeti fejlesztés során. Az üzleti pszichológiai ismeretek és a coaching módszertanának kombinációja olyan erőteljes eszköztárat kínál, amely elősegíti az egyének, csapatok és szervezetek hatékony működését és fejlődését (Palmer–Whybrow 2007).

Empirikus kutatás a pszichológus és a coach rendvédelemben való jelenlétéről

Kutatási probléma

Empirikus kutatásomban arra a kérdésre kerestem választ, hogy hogyan látják a gyakorló pszichológusok és coachok egymás munkáját, ill. miben látják saját módszereiknek, technikáiknak adaptálhatóságát a rendvédelmi szervek mindennapi működésébe.

Kutatási cél

Kutatásom fő célja volt megállapítani, hogy az üzleti pszichológiában és coachban alkalmazott módszerek és technikák a pszichológusok és coachok szemszögéből vizsgálva hogyan lennének adaptálhatóak a rendvédelmi szervezetek mindennapi munkájába. Fő célom elérése érdekében további alcélokat fogalmaztam meg az alábbiak szerint:

Cél 1. Megállapítani, hogyan látják a pszichológusok és coachok egymás munkáját, mennyire ismerik az azonosságokat és különbségeket munkamódszereikben.

Cél 2. Megállapítani, hogy a megkérdezett szakemberek miben látják az okát annak, hogy a rendvédelmi szerveknél a pszichológia jellemzően az alkalmassági vizsgálat részét képezi, további

adaptálhatóság ellen a hierarchikus szervezetek elzárkóznak.

Cél 3. Megállapítani, hogy milyen hozzáadott értéket tudna adni a rendszernek, ha pszichológusokat és / vagy coachokat vonnának be a rendvédelmi szervezetek mindennapi működésébe, a pszichológiai szűrővizsgálatokon túlmutatva.

Kutatási hipotézisek

A vizsgált témához kapcsolódóan az alábbi hipotéziseket fogalmaztam meg:

Hipotézis 1: Feltételezem, hogy a pszichológusok kevésbé fogadják el a coachokat, a coach képzésen szerzett tudást nem tartják elegendőnek ahhoz, hogy a coaching folyamat eredményes lehessen.

Hipotézis 2: Feltételezem, hogy az interjúban megkérdezettek válaszai alapján elsődlegesen az időhiányban látják az okát annak, hogy nem biztosítanak számukra nagyobb lehetőséget a rendvédelemben, másodlagos okként az adott szervezeti egység vezetőinek téma iránti fókuszhiánya áll fenn.

Hipotézis 3: Feltételezem, hogy az üzleti pszichológiában és a coachingban alkalmazott módszerek közül több módszer is alkalmazható lenne a rendvédelmi szervezetek mindennapjaiban, mely lehetőségeket a rendvédelmi szervek jellemzően nem vesznek igénybe az interjúba bevont releváns tapasztalattal rendelkező kollégák válaszai alapján.

Kutatás módszere és eszköze

A kutatás módszereként a kikérdezés módszerét, azon belül a szóbeli kikérdezést, az interjúkat alkalmaztam. A strukturált interjú során a hipotézisek igazolására vagy elvetésére alkalmas kérdéssor mentén zajlott a beszélgetés 2023 március és június között. Az interjúra való felkérés személyesen vagy telefonon történő felhívással valósult meg. Valamennyi kolléga, aki irányában felkéréssel fordultam, vállalta az interjút. Az interjú lebonyolítása személyesen vagy online beszélgetés formájában valósult meg, több tréner kollégával a földrajzi távolság miatt nem volt lehetőségem a személyes találkozásra. Az interjú időtartamát tekintve átlagosan 1- 1,5 órás beszélgetésekre került sor, amely során a strukturált kérdések mentén részletesen megismerhettem az adott kérdésekkel kapcsolatos véleményüket, tapasztalataikat.

Kutatási minta jellemzői

A felmérésbe összesen hat interjúalany került bevonásra. Az interjúalanyok kiválasztása célzottan történt, melynek alapkritériuma volt, hogy pszichológus és / vagy coach végzettséggel rendelkezzen a személy (1. sz. táblázat).

Interjúalanyok	Interjúalany neme	Interjúalany pszichológus/coach	Interjúalany jelenleg aktív a szakmában	Interjúalany szakmában eltöltött évei száma
V ₁	férfi	pszichológus	igen	26 év
V ₂	férfi	pszichológus	igen	26 év
V ₃	férfi	pszichológus	igen	26 év
V ₄	nő	pszichológus	igen	17 év
V ₅	nő	pszichológus	igen	15 év
V ₆	nő	coach	igen	12 év

1. sz. táblázat: Az interjúalanyok áttekintése

Forrás: Saját szerkesztés

A felmérésben három férfi és három női szakember véleményét ismerhettem meg, ebből öt pszichológus és egy coach. Gyakorlati éveiket tekintve hárman 26 éves szakmai gyakorlattal, egy fő 17 éves praxissal és egy fő 15 éves gyakorlattal rendelkezett, valamint egy fő mögött 12 éves gyakorlat áll.

Kutatási eredmények

Interjúalanyaim több évtizedes tapasztalattal rendelkeznek, melyet a kutatás szempontjából azért tartok kiemelkedőnek, mert olyan gyakorló

szakemberek gondolatait, véleményét sikerült megismernem, akik szakmailag hitelesek és jelentős gyakorlati tudás, tapasztalat áll mögöttük. A válaszadók mindegyike jelenleg is aktívan dolgozik a szakmában.

Első kérdésem a pszichológus és coach, mint szakemberek fogalmának tisztázására irányult. Napjainkban ezen szakemberek gyakran összerosódnak és nagyon sokan pszichológusnak vagy coachnak vallják magukat. Az interjúalanyok véleménye közel azonos ezen kérdésről: a pszichológia egyetemi tanulmányokhoz kötött, melyet követően adott tapasztalati éveknél kell eltelnie és szakvizsgát kell tenni ahhoz, hogy ténylegesen pszichológusnak vallhassa magát a szakember. Az egyik interjúalany meglátása szerint az utóbbi években történtek olyan változások, amelyek ennek a két szakterületnek a lehatárolására törekuszenek, jogszabályban, kódexekben szabályozzák a pszichológus és a coach szakterületét, kompetenciahatárait. Fontos kiemelt különbségként hangzott el az is, hogy a pszichológus patológiás esetekkel is dolgozik az egészséges ügyfelek mellett. Ugyanakkor az egyik pszichológus kollégának a focihoz kötődő hasonlat jutott eszébe a kérdés feltevése kapcsán, amihez már mindenki ért. A coach szakember más szemszögből világított rá erre a kérdésre. Az ő elmondása szerint vannak emberek, akik születésüktől kezdve arra rendeltetnek, hogy segítsenek másoknak, kísérik őket útjaikon. Vannak, akik alap-emberi természetüknél fogva rendelkeznek az ehhez szükséges készségekkel, képességekkel, kompetenciákkal. Ugyanakkor számos olyan szakember is van, akik elvégzik az iskolákat és mégsem lesznek jó szakemberek. Mindkettő szakmához kell az a többlet, kell az a plusz, a belső intuíció, amely képessé teszi az embert a másik segítségére, és ez a plusz nem feltétlenül tanulható – vallja hatodik coach interjúalanyom.

Ezt követően a pszichológus és coach közötti különbségre irányult kérdésem. A pszichológiát valamennyien tudománynak tekintik, mely a lélek mélyebb rétegeivel foglalkozik, a lelki problémákban való elakadásokat tárja fel, bizonyos betegségek lelki háttereire világít rá a klienssel való közös munka során. A szakemberek többsége a coachingot a pszichológia szűkebb területének véli, a coach az, aki mélyebb összefüggéseket nem tár fel, ugyanakkor a kliens személyes képességeit, erősségeit azonosítja és facilitátor szerepben az egyéni fejlődés lehetőségét támogatja. Eszköz- és módszertanban a pszichológia eszközeit és módszereit használja fel. Az egyik pszichológus kolléga nyíltan megfogalmazta, hogy számára a coach nem szakember, úgy véli, szakmai képzettség nélkül avatkozhatnak be pszichológiai folyamatokba. A megkérdezett coach a két szakember leglényegesebb különbségét abban látja, hogy a kliens személyének mely

rétegeivel foglalkoznak. A coach a felszíni jellel foglalkozik, míg a pszichológus a mélyebb rétegekig is lemegy. Amíg a pszichológus a mélyebb okok feltárására fókuszál és a feltárt tényeket elemzi, értelmezi kliensével, addig a coach a jelen kihívásaival dolgozik, fejlődési irányt és mérföldkövet megjelölve az ügyféllel közösen.

Az interjúban résztvevő szakemberek közül egy pszichológus kivételével mindenki foglalkozik egyénnel, családokkal és szervezetekkel egyaránt. Jellemzően munka- és szervezetszociológiáról számoltak be, mint legfőbb szervezeti szinten végzett tevékenység. Négy fő pszichológus a hadseregnél, rendvédelmi szerveknél, önkormányzatoknál, egészségügyi intézményben, gyermekvédelemben lát el pszichológus feladatokat. A coach végzettségű interjúalanyom elsősorban IT cégekkel foglalkozik, de dolgozott már alapítvánnyal, fogorvosi rendelővel is.

Interjúm következő kérdése így hangzott: Meglátása szerint mikor hatékony a pszichológia? Az öt fő pszichológus közül három fő úgy véli akkor, ha a kliens oldalról van fogadókészség, nyitottság, megfelelő együttműködni tudás és megfelelő empátia. Két pszichológus szerint az élet minden pillanatában, az élet minden területén hatékony tud lenni a pszichológia. Hasonlóan vélekedett a coach is a coach hatékonyságára vonatkozóan. Az ügyfél elkötelezettsége és a coach felkészültsége mellett a bizalom a hatékonyság kulcsfogalma.

A pszichológiával kapcsolatos általánosságban való gondolkodásról az egyik elhangzott vélemény a nyitottság volt, hiszen a pszichológia minden tekintetben, mindenhol jelen van az ember életében. Megemlítésre került a tudományosság is, hiszen a tudomány eredményeit, tapasztalatait az élet minden területén fel lehet használni, a lelki működések háttereinek feltárása segítheti megérteni az egészséges lelki működést. Coach interjúalanyom a coaching folyamatát egy csiszolt tükörhöz hasonlította, amelyben lehetőség van torzítás nélkül megvizsgálni az adott egyéni élethelyzetet és szükség szerint lehet változtatni rajta.

A pszichológiai folyamat a megkérdezett pszichológusok mindegyike szerint folyamatlépésekből áll, melyet meghatároz az adott kliens problémája, az a terület, amivel foglalkozni kell. A kezdeti, ismerkedő fázist, melyben a kapcsolatfelvétel, a tájékozódás is megtörténik egy aktív szakasz követi, melyben az intervenció, a közös munka, a segítségnyújtás zajlik. A folyamat végével lezárásra kerül sor, amely a probléma típusától és a kliens személyétől függően 4-5 alkalom, vagy akár egy vagy több éves időtávlatban következik be. Egyik interjúalany véleménye szerint egy elkezdett kliens és pszichológusi együttműködés összetett,

váratlan eseményeket is magában hordoz, ami miatt nehéz mérni, előre jóslatokba bocsátkozni, hogy milyen is lesz ez a folyamat. A coaching folyamata is lépésekből áll, általánosságban 10-12 alkalom szükséges egy megkezdett folyamat végig viteléhez. A közös munka során az egyén személyétől és a problémától függően különböző módszerek állnak a coach rendelkezésére.

Valamennyi pszichológus interjúalanyom egybehangzóan úgy véli, hogy minden olyan élethelyzetben, amikor az egyénnek, szervezetnek problémája van, érdemes szakemberhez fordulni. Az egyik kolléga külön megemlíti, hogy a külföldi kultúrákban a pszichológus szerepe a mindennapi élet természetes része. Egyéni megakadás esetén, ha a laikus, családi, baráti segítség már nem elég és az egyén életminőségére az adott probléma kihatással van, mindenképp érdemes szakemberhez fordulni. Többségük úgy véli, bármilyen élethelyzet indokoltá tudja tenni a pszichológushoz fordulást. Szervezetek esetén a szervezetfejlesztési igénnyel kapcsolták össze az interjúalanyok a pszichológus bevonásának szükségességét, munkavállalók érdekében, konfliktusok kezelésére. A pszichológusok válaszához hasonlóan a coach szakember is úgy véli, hogy az élet minden szakaszában, bármely élethelyzetben helye van a coachnak, mint segítő szakembernek. Hívhatjuk business és life coachingnak a folyamatokat, ugyan úgy emberi, egyéni sorsokkal foglalkozik, hiszen az üzleti világban, a szervezetekben is egyének dolgoznak. Ugyanakkor a coach interjúalany kiemelte, hogy nem akkor érdemes coachhoz fordulni, ha már baj van. Érdemes tudatosan gondolkodnunk önmagunkról, kompetenciáinkról, melyeket, ha megismerünk és fejlesztünk, számos krízishelyzetet meg tudunk előzni, tudatosabban tudjuk kezelni emberi és párkapcsolatainkat.

A kiválasztott interjúalanyok mindegyike rendelkezik oktatói, tréneri tapasztalattal. Egy fő 23 éves felsőoktatási és 6 éves tréneri tapasztalattal bír. Egy másik kolléga 26 éve oktat különböző korosztályban. Egy pszichológus több mint 20 éve foglalkozik gyerekekkel és 5 éves tréneri tapasztalattal is rendelkezik. Az interjúalanyok között az egyik pszichológus 34 éves oktatói és 20 éves tréneri tapasztalattal rendelkezik. A legkevesebb oktatói, tréneri évek száma 6 év volt az interjúban résztvevők körében. A megkérdezett coach szakember 10 éve tanít felsőoktatásban, elsősorban kommunikációt és önismeretet, és 9 évet foglalkozott gyerekekkel és kamaszokkal ugyanezen témákban.

Interjúkérdéseim között következő kérdésként az interjúalanyok rendvédelemmel való kapcsolódását vizsgáltam. A kutatásba bevont öt pszichológus mindegyikének van / volt szakmai kötődése a rendvédelemhez. Négy kolléga a honvédségnél látott el

pszichológusi feladatokat, egy kolléga jelenleg is a honvédség kötelékében szolgál, mint pszichológus. Az évek számát tekintve 10 évet meghaladó tapasztalat van a pszichológusok mögött, egy fő 6 éve dolgozik a rendvédelemben, szakpszichológusként. Az egyik interjúalanyom a rendőrségnél 10 évet dolgozott, ahova a 12 éves honvédségben töltött idő után szerelt át. A coach interjúalanyomnak nincs ilyen jellegű múltja, nem rendelkezik rendvédelmi szervezetek keretében töltött tapasztalattal.

Arra a kérdésre, hogy a megkérdezettek hogyan látják pszichológusként a pszichológus jelenlétét a rendvédelemben, milyen feladatokat lát el vagy tudna ellátni a szervezetnél, hasonló válaszok érkeztek. Az állomány pszichés gondozását többen is fontosnak vélik. Egyik interjúalanyom úgy fogalmazott, hogy fokozottabban lenne szükség az állománnyal való foglalkozásra, míg egy másik kolléga a dinamikus biztonságot látja a pszichológusi munkában. A honvédségben dolgozó kolléga arról számolt be, hogy a honvédségnél a pszichológusnak kiemelt szerepe van. Külön választódnak az alkalmassági vizsgálatban résztvevő kollégák a szervezeti egységben feladatot ellátó pszichológusoktól, segítve ezzel az előítéletmentességet. A coach kolléga tételesen nevezett meg olyan területeket, amelyben a coachingnak látna helyet a rendvédelemben, mint például kommunikáció, önismeret, kiégés megelőzése, stresszkezelés, vezetői kompetenciák fejlesztése és együttműködési készségek fejlesztése.

Interjúkérdéseim között a következőkben arra kértem az interjúalanyokat, hogy fogalmazzák meg, az ő meglátásuk szerint a rendvédelmi szervek vezetői mennyire lennének nyitottak arra, hogy szervezetükbe külsős pszichológusok megbízásokat kapjanak tréningek, képzések tartására, egyéni fejlesztésekre. Két válasz egybehangzóan megfogalmazta a rendszer bürokratikus volta miatti elzárkózást. A rendszer belső ismerete nélkül olyan szakember, aki nem ismeri a rendszert belülről, nem tud érvényesülni. Ugyanakkor egy másik vélemény szerint, mely saját tapasztalatokra hivatkozik, a vezetők nyitottak. Ezen kolléga meglátása szerint a külsős pszichológus választása célravezetőbb, hiszen ő nem képezi a rendszer részét belülről, így a bizalmatlanság kérdése sem merülne fel az ő esetében. Ezen kolléga válaszával ellentétes volt egy másik kolléga meglátása, aki szerint a dolgozók könnyebben fogadják el azokat a pszichológus kollégákat, akik a rendszer sajátosságait ismerik és a rendszerben tevékenykednek. Az ötödik interjúalany a kérdést visszafordítva azt kérdezte meg, hogy vajon a trénernek rendelkeznie-e kellő ismerettel a rendvédelmi szervek zárt világáról, ami a hitelességhez elengedhetetlen. Hatodik interjúalanyom, a coach szakember ugyanúgy a bizalom

kérdését emelte ki elsődleges kritikus pontként a rendvédelemben való külső szakemberek foglalkoztatásának nehézségeként. Ugyanakkor fontosnak tartja a külső szakértői jelenlétet, hiszen a rendszert kívülről szemlélve számos olyan pont beazonosítható, melyet belülről már nem lehet észrevenni és tartós fennállása kifejezetten káros lehet.

Az interjúban résztvevő valamennyi pszichológus válasza megegyezett arra vonatkozóan, hogy a beosztott állománynak szüksége és igénye is lenne pszichológussal való beszélgetésre. Miután gyakorló pszichológusokat kérdeztem, elmondták, hogy jelenleg is járnak hozzájuk rendvédelemben dolgozók, illetve munkájuk során is foglalkoznak egyéni problémákkal, élethelyzetekkel. Egy kolléga külön megjegyezte, hogy bár szükségét érzik az állományi tagok a pszichológussal való beszélgetést, ugyanakkor a bizalmatlanságból fakadó ellenállás jelen van a rendszerben. Nem látja másként a coach sem ezt a kérdés, a coachnak egyértelműen helye van a rendvédelemben dolgozó beosztott állomány segítségével, támogatásában.

Szerettem volna megismerni az interjúalanyok véleményét arról, hogy szerintük az üzleti pszichológia és a coaching mely módszerei és eszközei lehetnének adaptálhatóak a rendvédelmi szervek működésébe. Egy kolléga határozott véleménye, hogy egyik módszer vagy eszköz sem adaptálható a szervezet hierarchikus stukturája miatt, ami nem enged mozgásteret a dolgozóknak, túl központosított a rendszer. Volt kolléga, aki a szervezetfejlesztést és a vezetőknek, beosztottaknak szóló tréningeket látja bevonhatónak. Egy másik kolléga az önmenedzselést és mentorálást nevesítette, mint adaptálható módszert, melyben az alábbi területeket jelölte meg: a beosztottak motiválása, delegálás, problémamegoldás, az emberi kapcsolatok kezelése, értékelés. A kiegészítés-prevenció és kezelés, a kommunikáció és stresszkezelés ugyancsak olyan fókuszterületként hangzott el, melyeknek meglátása szerint helye lenne ezen alá-fölrendelt szervezet mindennapjaiban. Az üzleti pszichológia és a coaching folyamat módszerei és eszközei közül a coach kolléga elsődlegesen a kompetenciafejlesztő tréningeket emelte ki, egyéni és csoportos szinten egyaránt, melyek bevezetését több szinten, strukturáltan tudja elképzelni. A vezető beosztásban dolgozóknál egyéni felmérést követően egyéni fejlesztések vezethetnének eredményhez.

A következő kérdés így hangzott: Milyen érveket tudna felsorolni amellet, hogy Önt, mint pszichológust holnaptól a rendvédelmi szervek foglalkoztassák, illetve milyen plusz hozadékot tudna nyújtani a rendszernek, a benne dolgozó állománynak? A kapott válaszok egybeesnek a korábbi kérdésekben már érintőlegesen feltett kérdésekre adott válaszokkal, kiemelve a segítő,

támogató jelenlétet, az egészséges személyiség szerkezet megőrzését, fejlesztését szolgáló plusz hozadékot. Két kolléga ezen kérdésre nem adott választ. Coach interjúalanyom tudást, elkötelezettséget és inspirációt tudna nyújtani a coaching folyamatok rendvédelemben való bevitelével. A kulturált, magabiztos fellépést az emberekkel, a hatékony kommunikációt a csapatban való összetartást és együttműködést tudja eredményezni, melyhez magabiztos, felkészült vezetőkre van szüksége a rendszernek.

Interjúkérdésem következő blokkja a pszichológus és coach közötti viszonyra, a coach képzés ismeretére irányult. A pszichológusok és coachok közötti viszonyt az öt interjúalany különbözőképpen látja. Vannak, akik kimondottan ellenségesnek, feszültnek vélik a két szakember közötti helyzetet, van, aki alárendeltként jellemzi viszonyukat. Van, aki számára érdektelen a coach, és mindössze egy pszichológus fogalmazott úgy, hogy nem tapasztal negatív megítélést, csak együttműködést és egymásra épülést. A coach interjúalany ugyancsak pozitívan látja a pszichológus és coach közötti viszonyt, egymást kiegészítő szakembereknek véli magukat.

A coach képzést a megkérdezett öt pszichológus közül négyen nem ismerik, egy kolléga meglátása szerint vannak jó és silány képzések. A pszichológusok a coachok elfogadásáról a következőképpen vélekednek: vegyes, nem lehet általánosítani, változó, személyfüggő. Az egyik interjúalany a beszélgetés során kimondta, hogy ő nem fogadja el a coachot, és egy kolléga szerint a pszichológusok elfogadóak, ha a kompetenciahatárokat betartják. A coach szakember hasonlóan vélekedik erről a kérdéssel, mint az egyik pszichológus. A pszichológus és a coach személyétől függ szerinte az elfogadás. Mindkét szakterületen vannak nagyon értékes, magas színvonalon dolgozó szakemberek, és vannak sérült, magukat gyógyítani akaró emberek is.

A következő kérdéssel arra kerestem a választ, hogy a pszichológia és a coaching mennyire választható szét egymástól. Három interjú adó pszichológus válaszában megfogalmazta, hogy pszichológiai alapok nélkül nem lehet a coaching sikeres, így a pszichológiai képzettséget hozzákötnék a coach szakma megszerzéséhez. Két kolléga szerint egyértelműen elhatárolható a két szakma a munka fókusza miatt, és a kompetenciahatárok betartása is ezt a szétválasztást szolgálja, mely véleményt egyértelműen osztja a megkérdezett coach is.

Az interjú utolsó előtti kérdésében a manapság használatos mentor fogalomra kérdeztem rá, hogy mi a különbség a mentor, a coach és a pszichológus között. Hasonlóan vélekednek a szakemberek a mentorról, aki a beilleszkedést támogató ember, más típusú segítő szakember, meghatározott feladatterületen motivál, képez, aki idősebb, tapasztalattal bír és megosztja

tanítványaival a tudást. Céljaként a karrierépítést fogalmazta meg az egyik kolléga, a szakmai és lelki tanácsadás egyik formájaként. A pszichológus elsősorban az élet működésével foglalkozik, egyik interjúalany megfogalmazásában a pszichológus szélesebb ismeretekkel rendelkezik, a másik két szakma szűkebb spektrumot ölel fel. A coach is úgy vélekedik a mentorról, hogy komoly tapasztalattal rendelkezik, aki saját, egyéni módszerekkel dolgozik, egyfajta tanítóként, útmutatóként működve.

Utolsó kérdésként arra kértem az interjúalanyokat, hogy ha van olyan, a témával kapcsolatos releváns gondolat, melyet szívesen megosztanának velem, tegyék meg. Három pszichológus kolléga és a coach szakember az alábbi gondolatokat osztotta meg: az egyik pszichológus jelenleg is dolgozik coach-csal és jó a kapcsolat közöttük, mert mindegyikük betartja a kompetenciahatárokat. A másik kolléga arra vonatkozóan tett kiegészítést, hogy a mentorok képzése fontos, amiben a pszichológusnak kellene részt venni. A harmadik pszichológus annyival szerette volna kiegészíteni a beszélgetést, hogy a segítő szakmában nem a megnevezés a lényegi kérdés, hanem az, hogy az emberben rejlő képességeket elő tudja hozni a megfelelő módszer alkalmazásával, amely az egyén problémájához igazodik. A coach szakember saját cégéről mesélt, ahol 12 szakember dolgozik együtt, köztük pszichológusok, trénerok, terapeuták, mintegy szakmai műhelyként teret adva egyéneknek, csoportoknak, vezetőknek egyaránt. Kooperativitás van náluk, melynek a közepén az egyén áll. Szavait idézve: „mindenki teszi, amit tud és ez épp elég”.

Összegzés

Kutatásom fő célja volt megállapítani, hogy az üzleti pszichológiában és coachban alkalmazott módszereket és technikákat hogyan lehet adaptálni a rendvédelmi szervek mindennapi munkájába a gyakorló pszichológusok és coachok szerint. Az interjúkérdések feldolgozása során a hat interjúalany válaszait szövegesen rögzítettem, majd elemeztem azokat.

A kutatási célokhoz kapcsolódóan felállított három hipotézis igazolása az alábbiak szerint alakult:

Hipotézis 1: Feltételezem, hogy a pszichológusok kevésbé fogadják el a coachokat, a coach képzésen szerzett tudást nem tartják elegendőnek ahhoz, hogy a coaching folyamat eredményes lehessen.

A kapott válaszok alapján *első hipotézisem első része nem nyert igazolást*. A hat interjúalany válaszából egy fő egyértelműen nem fogadja el, egy fő szerint elfogadó a pszichológus a coach irányában, négy fő válasza nem egyértelmű az elfogadás és nem elfogadás között. Az elfogadás függését a szakemberek személyéhez köti

mind a négy válaszadó. A *hipotézis második része nem nyert igazolást* a pszichológusok válaszai alapján. A válaszadók nem ismerik a coach képzést, így ezen feltételezésünk érdemben nem vizsgálható. Ugyanakkor az interjú során több kérdésben is érintettük a coachingoláshoz szükséges képességeket, tudást, melyben a szakemberek egységes állásponton voltak a tekintetben, hogy a pszichológia bizonyos területeinek ismerete nélkülözhetetlen a coach folyamatokba. A pszichológus és a coach az emberi lélek, a személyiség más rétegével dolgozik, így más szintű tudásra van szüksége a két szakterületnek.

Hipotézis 2: Feltételezem, hogy az interjúban megkérdezettek válaszai alapján elsődlegesen az időhiányban látják az okát annak, hogy nem biztosítanak számukra nagyobb lehetőséget a rendvédelemben, másodlagos okként az adott szervezeti egység vezetőinek téma iránti fókuszhiánya áll fenn.

Második hipotézisem nem nyert igazolást a kapott válaszok alapján. A rendvédelem valamennyi interjúalany meglátása szerint sajátos szervezet, mely bürokratikus, zárkózott, speciális szervezeti felépítéssel és szervezeti kultúrával bír. Ezen sajátosságban látják az elsődleges okot, amiért a rendszer nem enged nagyobb teret a külsős szakembereknek. Ebből adódik a bizalmatlanság, aminek következtében nem szívesen engedik be a rendszeren kívüli szakembereket.

Hipotézis 3: Feltételezem, hogy az üzleti pszichológiában és a coachingban alkalmazott módszerek közül több módszer is alkalmazható lenne a rendvédelmi szervezetek mindennapjaiban, mely lehetőségeket a rendvédelmi szervek jellemzően nem vesznek igénybe az interjúba bevont releváns tapasztalattal rendelkező kollégák válaszai alapján.

Harmadik hipotézisem igazolást nyert. A hat interjúalany közül négyen egyértelműen nevesíteni tudták azokat a módszereket, melyek a rendvédelem mindennapi munkájában alkalmazhatóak lennének a beosztotti állomány és a vezetők részére egyaránt.

A kutatómunkámba bevont gyakorló pszichológus és coach szakemberek interjúkérdéseimre adott válaszai alapján egyértelműen megállapíthatjuk a vizsgált téma aktualitásának fontosságát és jelentőségét. Valamennyi szervezetet egyének, emberek sokasága alkotja, az ő közösségük hozza létre az adott szervezet kultúráját, értékeit, akár állami vagy piaci szektorról, akár bürokratikus vagy demokratikus vezetésű szervezetről van szó.

A témával kapcsolatos irodalmak és saját kutatásom alapján úgy vélem, hogy a hierarchikus struktúrájú szervezetek vezetői szintjén egyre sürgetőbben javasolt szemléletváltás a tekintetben, hogy ezen szervezetben dolgozó emberek testi, lelki és pszichés állapotával egyaránt foglalkozni szükséges, mely nem merülhet ki

alkalmassági vizsgálatok elvégzésében. A vezetők merjenek külső szakértőket, szakembereket bevonni vezetői kompetenciafejlesztésre és beosztott állományi tagok fejlesztésére, tréningezésére egyaránt. A rendvédelmi szervek felső- és középszintje a külső szakemberekkel szembeni bizalmatlanság feloldására engedjenek teret a megismerésre, váljanak nyitottá ezen külső szakértők irányába a saját rendszerük jobbá tétele érdekében. Vezetői szinten javasolt átgondolni annak tényét, hogy a különböző szakértelemmel bíró tapasztalt, gyakorlati tudással rendelkező külsős szakemberek (pszichológus, coach, tréner, mentor) szakmai tudásának összeadódása képes olyan erőt és tudást képviselni, amely a rendszer működését hatékonyabbá teheti és meglévő állományát egyéni és szervezeti célok elérésére tudja formálni. Úgy vélem, a külső szakértők bevonása hozzá tudna járulni a szervezet eredményességéhez, mivel az állomány fizikai és mentális egészsége közvetlenül befolyásolja a teljesítményt, és ezzel egyidejűleg hozzájárulna a munkahelyi elégedettséghez és lojalitáshoz, amire a rendvédelmi szerveknek egyre nagyobb figyelmet kellene fordítania.

Irodalomjegyzék

- Arnold, J. et al. (2016) Work psychology: understanding human behaviour in the workplace. Pearson Education.
- Grant, A. M. (2011) Is it time to REGROW the GROW model? Issues related to teaching coaching session structures. *The Coaching Psychologist*, Vol. 7. No. 2. pp. 118-126. [doi: 10.1002/9781119835714.ch3](https://doi.org/10.1002/9781119835714.ch3)
- Farkas, J. (2018) Pszichológia a közszolgálatban. In: Haller, J.–Farkas, J. (szerk.) *Pszichológia a közszolgálatban I.* Budapest, Dialóg Campus Kiadó.
- Hídvégi, P.–Balogh, M. (2022) *Fiatalok képzése és kompetenciafejlesztése tanulmány. Lektorált kutatási beszámoló. 3. Megoldási javaslat tanulmány.* GINOP-5.3.5.-18-2018-00054 Munkaerőpiaci
- alkalmazkodóképesség fejlesztését célzó tematikus projektek: Munkavállalói ismeretek oktatása a közigazgatás, védelem ágazatba belépő fiatalok részére. pp. 26 o.
- Kilburg, R. R. (1996) Toward a conceptual understanding and definition of executive coaching. *Consulting Psychology Journal: Practice and Research*, Vol. 48. No. 2. pp. 134-144. [doi: 10.1037//1061-4087.48.2.134](https://doi.org/10.1037//1061-4087.48.2.134)
- Komócsin, L. (2018) *Módszertani kézikönyv coachoknak és coachingszemléletű vezetőknek I.* Budapest, Manager Könyvkiadó- Business Coach Kft.
- Passmore, J.–Fillery-Travis, A. (2011) A critical review of executive coaching research: a decade of progress and what's to come. *Coaching: An International Journal of Theory, Research and Practice*, Vol. 4. No. 2. pp. 70-88. [doi: 10.1080/17521882.2011.596484](https://doi.org/10.1080/17521882.2011.596484)
- Passmore, J.–Peterson, D.–Freire, T. (2013) The psychology of coaching and mentoring. In: Passmore, J.–Peterson, D.–Freire, T. (eds.) *The Wiley-Blackwell handbook of the psychology of coaching and mentoring.* John Wiley & Sons Ltd. Published, pp. 1-12.
- Palmer, S.–Whybrow, A. (2019). *Handbook of coaching psychology: A guide for practitioners*, New York, Routledge. [doi: 10.4324/9781315820217](https://doi.org/10.4324/9781315820217)
- Rogers, J. (2012) *Coaching skills: A handbook.* Open University Press.
- Seligman, M. E.–Csikszentmihalyi, M. (2000) Positive psychology: An introduction. *American Psychological Association*, Vol. 55. No. 1. pp. 5-14. [doi: 10.1037//0003-066X.55.1.5](https://doi.org/10.1037//0003-066X.55.1.5)
- Stréhlly, G. (2022) Új utakon a közszolgálati vezetőképzés – az action learning módszerének alkalmazási lehetőségei a közigazgatásban és a rendvédelemben. Doktori (PhD) értekezés. Nemzeti Közszolgálati Egyetem, Közigazgatás- tudományi Doktori Iskola, Budapest, 2022.
- Whitmore, J. (2009) *Coaching for Performance: Growing People, Performance and Purpose.* Nicholas Brealey Publishing.

Az igazgatók fejlesztésének lehetőségei és irányai a magyar köznevelés rendszerében - kutatási beszámoló

Opportunities and directions of director development in the Hungarian public education system - research report

DOI: [HTTPS:// DOI.ORG/10.53793/RV.2024.2.7](https://doi.org/10.53793/RV.2024.2.7)

Absztrakt

Ezen kutatási beszámoló célja, hogy összefoglalja a Belügyi Tudományos Tanács által kiírt pályázat keretében megvalósuló gyakornoki program ideje alatt végzett kutatásomat, ismertesse a megállapított részeredményeket, valamint alapját képezze a későbbi doktori munkámnak. A tanulmány elsősorban a már gyakorlott igazgatók további képzési, fejlesztési lehetőségeinek szükségességét, lehetőségeit, irányait vizsgálja és egy erre irányuló koncepció kidolgozásához az alapvetéseket tartalmazza.

KULCSSZAVAK: INTÉZMÉNYVEZETŐI SZAKKÉPZETTSÉG, VEZETŐKÉPZÉS, JELENLEGI KÉPZÉSI GYAKORLAT, TOVÁBBI KÉPZÉS IRÁNTI IGÉNY, KIDOLGOZANDÓ KONCEPCIÓ

Abstract

The purpose of this research report is to summarize my research carried out during the internship program implemented within the framework of the tender announced by the Scientific Council for Internal Affairs, to present the partial results found, and to form the basis of my future doctoral work. The study primarily examines the need, possibilities and directions of further training and development opportunities for already experienced directors and contains the basics for the development of a concept aimed at this.

KEYWORDS: QUALIFICATION OF HEADS OF INSTITUTIONS, LEADERSHIP TRAINING, CURRENT TRAINING PRACTICE, NEED FOR FURTHER TRAINING, CONCEPT TO BE DEVELOPED

Bevezetés

Az Eötvös Loránd Tudományegyetem Pedagógiai és Pszichológia Kar Neveléstudományi Doktori Iskolájának doktorandusz hallgatójaként 2023 májusában kutatási tervet nyújtottam be gyakornoki rendszer keretében tudományos kutatás folytatására kiírt pályázatra a Belügyi Tudományos Tanácshoz. A pályázat elnyerését követően a gyakornoki időszakom 2023 szeptemberétől 2024 februárjáig tartott, a befogadó szerv az Oktatási Hivatal volt.

A pályázat benyújtásának az volt az oka, hogy doktori munkám a Belügyminisztériumhoz tartozó egyik új tudományterülethez, a közneveléshez kapcsolódik. Kutatási témám és egyben disszertációm tárgya az igazgatók felkészítési, fejlesztési lehetőségeinek vizsgálata. A témában való jártasságomat eddigi szakmai tevékenységem alapozza meg: közel 20 éve foglalkozom oktatás-igazgatással, intézményfenntartással, illetve több mint tíz éve tankörvezetőként, szakdolgozati témavezetőként és bírálóként dolgozom a Budapesti Műszaki és

Gazdaságtudományi Egyetem által alapított közoktatási vezető szakirányú továbbképzésben. Hosszú ideje vagyok tehát mindennapi munkakapcsolatban a köznevelési intézmények igazgatóival, és aktívan részese vagyok annak a folyamatnak is, mely pályára való felkészítésüket hivatott szolgálni.

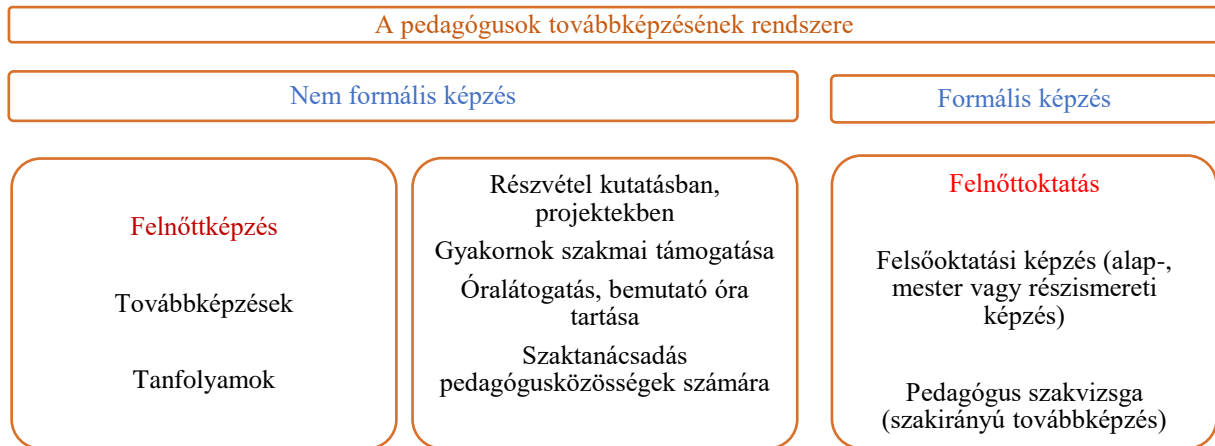
Gyakornoki tevékenységem alatt az volt elsősorban a célom, hogy a téma elméleti háttérének felmérése, valamint a kérdőíves kutatás lebonyolítása és annak elemzése során hozzájuthassak a vonatkozó adatbázisokhoz, szakmai anyagokhoz, könyvtári állományhoz, illetve, hogy a fenti folyamatokhoz szakmai mentorálásban részesülhessek.

A kutatás elméleti kerete

A köznevelési intézmények vezetőinek felkészítése és fejlesztése a pedagógus továbbképzés keretein belül valósul meg. Rendszerét a 277/1997. (XII. 22.) számú kormányrendelet (URL1) akként szabályozza, hogy a

pedagógusnak az 55. életévének betöltéséig kötelező továbbképzésen részt vennie, melynek keretében 7 évenként 120 kreditpontot kell szereznie. A továbbképzési kötelezettség teljesíthető az Oktatási Hivatal által akkreditált és hivatalos jegyzék formájában közzétett továbbképzéssel, felsőoktatási intézmény által kínált pedagógus munkakör betöltésére jogosító

graduális és posztgraduális képzésen való részvétellel, valamint szakirányú továbbképzés keretében pedagógus-szakvizsga letételével, illetve a 120 óra legfeljebb 25%-ának mértékéig kutatással, gyakornok felkészítéssel, óralátogatással vagy szaktanácsadással (Korm. rend. 4-5. §). A továbbképzési rendszer felépítését a lenti, 1. sz. ábra szemlélteti.

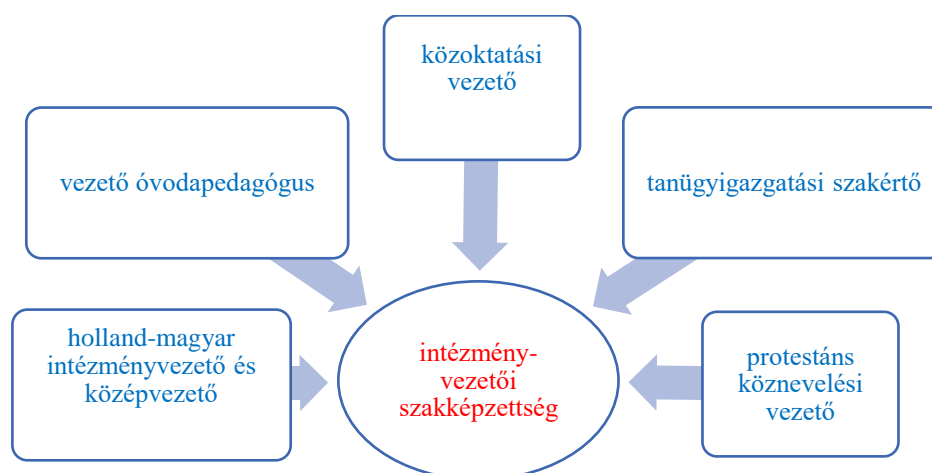


1. sz. ábra: A pedagógusok továbbképzésének rendszere

Forrás: Vereckei 2021

Fejlesztési lehetőségként az igazgatók számára is a fenti kínálat áll rendelkezésre – azzal a fontos különbséggel, hogy 2013 óta az igazgatói munkakör betöltésének feltétele egy felkészítő jellegű képzés

elvégzése, mely intézményvezetői szakképzettséget ad. Ez a szakképzettség felsőoktatási intézmények által kínált, 4 szemeszteres, pedagógus szakvizsgát adó, szakirányú továbbképzés keretében szerezhető meg. A képzések a 2. sz. ábrán kerülnek bemutatásra.



2. sz. ábra: Az intézményvezetői végzettséget biztosító képzések

Forrás: Vereckei 2022

Az igazgatók speciális képzése azonban a vezetőképzés elvégzését követően lezárul, a vezetői életpálya során kötelező folytatása nincs. A fentebb bemutatott, hét évenkénti továbbképzési kötelezettség természetesen vonatkozik rájuk is, de semmi sem szabályozza, hogy azt milyen tartalommal töltsék meg, a további vezetői fejlesztésre előírás nem vonatkozik. Az igazgatók felkészítése, fejlesztése tárgyú kutatásom így egyrészt a kötelező felkészítő képzéselem megfelelőségére, másrészt a vezetői pálya során a további képzési elem szükségességének, lehetséges irányainak felmérésére irányul.

A kutatás lebonyolítása

A kutatás két fő témakörének megfelelően az igazgatói célcsoport véleményének, igényeinek felmérésére kérdőívet állítottam össze, melynek kérdései több csoportba sorolhatóak a következők szerint:

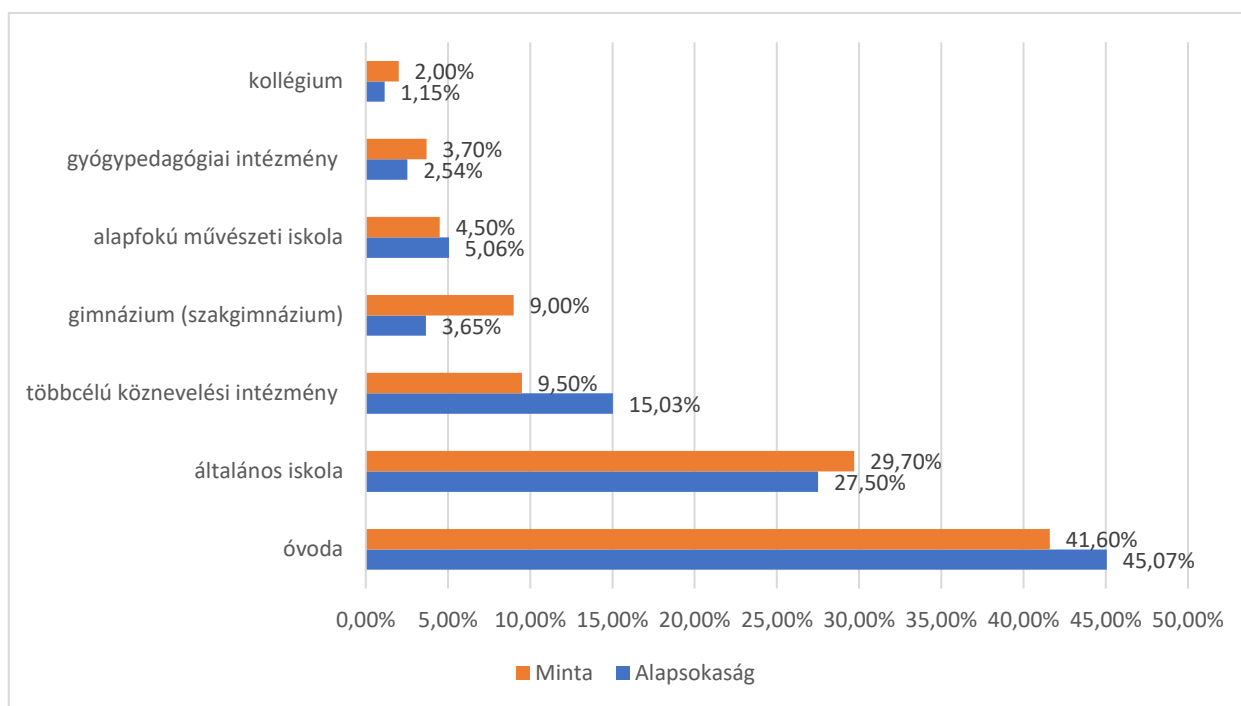
- az I. egység kérdései arra az intézményre vonatkoztak, amelyben az igazgató dolgozik;
- a II. egység kérdései a kérdőív kitöltőjét érintő, ún. demográfiai kérdések (életkor, nem, vezetői feladatellátás jellege, ideje, vezetői munkakör betöltésére vonatkozó jövőbeli szándék stb.);
- a III. egység kérdései az intézményvezetői szakképzettséget biztosító vezetőképzésre vonatkoztak, melyet az igazgatói célcsoport

munkakörének betöltési feltételeként már el kellett végeznie;

- a IV. egység kérdései a vezetőképzés elvégzését követő további képzési szükségletekre, véleményekre, igényekre irányultak.

Gyakornoki tevékenységem fókuszába az utóbbi, a IV. egységhez tartozó kérdések feldolgozása került, így lett jelen beszámolóm témája az intézményvezetői szakképzettséggel már rendelkező igazgatók további fejlesztésének szükségessége és lehetséges irányai.

A kérdőív az Oktatási Hivatal által működtetett, a köznevelés információs rendszere (a továbbiakban: KIR) adatbázisában szereplő minden köznevelési intézmény számára, elektronikus úton megküldésre került (Oktatási Hivatal 2023). A kérdőívet tartalmazó Google űrlapot végül 558 igazgató töltötte ki, mely adat a teljes alapsokaság kb. 10%-ának felel meg. A minta reprezentativitását biztosítja a teljes mértékben véletlenszerű mintavétel, valamint a minta elemszáma, de egyéb adatok is alátámasztják. Kiemelem ezek közül a lenti kimutatást (3. sz. ábra), mely azt szemlélteti, hogy nagyságrendileg megegyezik az intézménytípus szerinti megoszlás az alapsokaságban és a mintában, azzal a kitételrel, hogy a többcélú köznevelési intézmények és gimnáziumok esetében az előfordulási gyakoriság a mintában nagyobb mértékben tér el az alapsokaságra jellemző arányoktól.



3. sz. ábra: Az alapsokaság és a minta intézménytípus szerinti megoszlása (%)

Forrás: Saját szerkesztés

Beszámoló a kutatás (rész)eredményeiről

Az említett IV. kérdéscsoportban az igazgatók összesen 18 kérdésre válaszoltak, melyek elsősorban képzési szokásaikra (milyen jellegű, témájú továbbképzéseken és miért vettek részt, illetve ezek mennyire járultak hozzá vezetői fejlődésükhöz), valamint a folyamatos vezetői képzés iránti igényükre (beleértve a szükségesnek tartott képzés típusa, témája, valamint a képző személye iránti igényt is) és jelenlegi lehetőségeik ismeretére vonatkoztak.

A kérdésekre adott válaszok feldolgozását követően a folyamatos vezetőfejlesztés szükségességéről, annak lehetséges irányairól a lenti részletes megállapítások tehetők.

A jelenlegi továbbképzési gyakorlatról

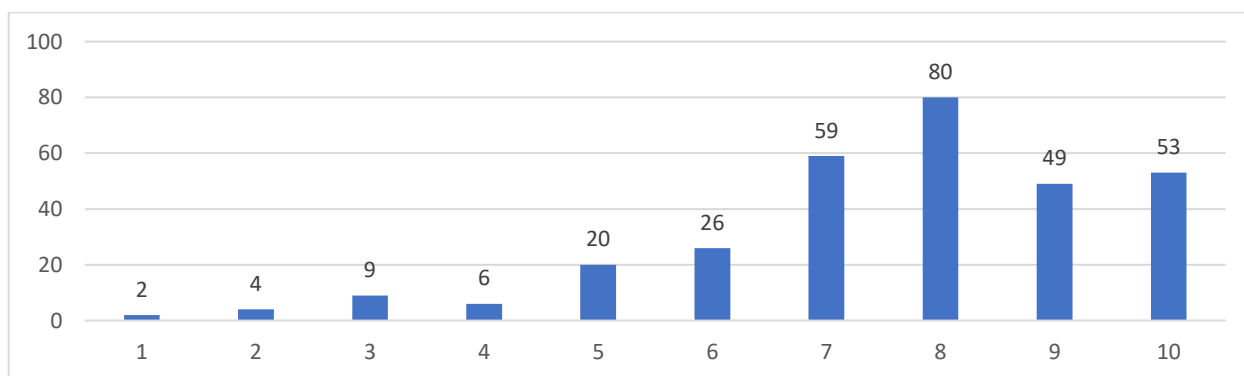
Az igazgatók valamivel több mint fele (55%-a) vett már részt vezetőknek szóló további képzésen, méghozzá gyakorisági sorrendben elsősorban a következő tárgykörökben/témakörökben:

- újabb vezetőképzés
- pedagógus minősítés, tanfelügyelet, önértékelés
- jog, jogszabályi változások
- konfliktuskezelés

- kommunikáció
- szervezetfejlesztés
- pályázatírás
- gazdálkodás
- dokumentumok kezelése és adminisztráció.

A felsorolt tárgykörökben legtöbben, 90%-uk, nem hagyományos pedagógus vagy szakirányú továbbképzést, hanem rövid ideig tartó, nem formális képzést választott: előadást, tréninget vagy egyéb, ehhez hasonló képzéstípust. A fentiek mellett azért még mindig sokan (70%) vettek részt pedagógus továbbképzésen, a hosszabb ideig (több szemeszteren át) tartó, formális képzés elvégzésére vállalkoztak a legkevesebben: a válaszadók ötöde végzett el szakirányú továbbképzést.

Az elvégzett képzés hasznosságát a válaszolók alapvetően pozitívan ítélték meg: az 1-től 10-ig terjedő skálán átlagosan 7,5 pontra értékelték a következők szerint: 78%-uk a 7 és a 10 közötti tartományba tette a hasznosságot, a legtöbben 8-asra minősítették, a 7-es, a 9-es és a 10-es minősítések aránya nagyjából egyező. Közepesen hasznosnak (5-ös, 6-os skálaértékek választása) ítélte meg a képzést a válaszadók 14%-a és mindössze 7%-uk adta valamelyik legalsó, 1-4-es értéket. A részletes kimutatás a 4. sz. ábrán látható.



4. sz. ábra. Hozzájárulás a vezetői fejlődéshez (fő)

Forrás: Saját szerkesztés

Az igazgatók másik fele, a „nem”-mel válaszolók, a képzés elmaradásának a következő leggyakoribb okait jelölték meg:

- időhiány (33%)
- képzési kínálat hiánya (18%)
- anyagi ok (14%).

A válaszadók körülbelül ötöde (pontosan 22%-a) nagyjából egyenlő megoszlásban a lenti négy ok valamelyikét választotta:

- utazás nehézsége a képzés helyszínére
- nem kötelező a részvétel

- nem járul hozzá a képzés a vezetői fejlődéséhez
- a képzésszervezés módja nem megfelelő.

11%-a a válaszolóknak nem tudott a megjelölt lehetőségekből választani, ami arra utal, hogy a felsoroltakon túl egyéb oka is van annak, miért nem vesznek részt az igazgatók további vezetői képzésen.

Összegzésképpen a jelenlegi vezetőfejlesztési gyakorlatról megállapítható, hogy az igazgatóknak körülbelül a fele vett részt a vezetői feladatra felkészítő kötelező képzés elvégzését követően is további vezetői képzésen, és azt munkájuk szempontjából hasznosnak is

tartották. A legtöbben a nem formális, rövidebb idejű képzéseket preferálták, ennek keretében egyaránt választottak olyan tartalmakat, melyek új vagy módosított „operatív” igazgatói/intézményi feladat elvégzéséhez kapcsolódik: jogszabályi változások, minősítés, tanfelügyelet, önértékelés, adminisztráció, gazdálkodás, illetve olyanokat is, mely a mindennapi, általános vezetői munka vagy vezetői funkció szerves részét képezi, mint pl. kommunikáció, konfliktuskezelés, szervezetfejlesztés. A vezetőfejlesztés népszerűsége növelhető lenne a részvételi díj mérséklésével, esetlegesen elengedésével, a képzési kínálat bővítésével, valamint a képzésszervezés módjának változtatásával, elsősorban a kevesebb idő igénybevételének megoldásával és a képzés helyszínére történő eljutási nehézség kiküszöbölésével.

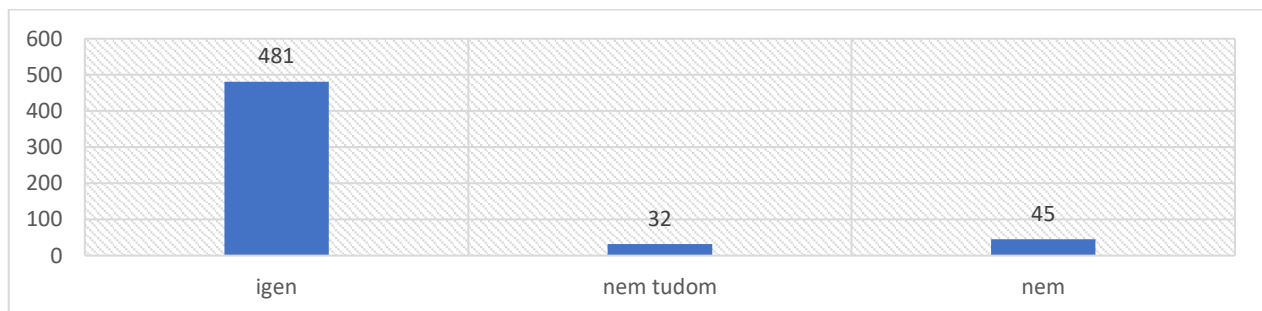
A jelenlegi képzési gyakorlat bemutatásához hozzátartozik az is, hogy a speciálisan vezetőknek szóló képzéseken való részvétel mellett, annál valamivel magasabb arányban, a megkérdezettek 61%-a végzett el az elmúlt három évben nem speciálisan vezetőknek, hanem általánosan pedagógusoknak szóló továbbképzést. A leggyakoribb témák ebben az esetben

az intézmények szakmai munkáját érintő aktualitások voltak:

- digitalizáció, IKT, informatika, robotika
- diabéteszes gyermekek ellátására vonatkozó képzés
- mérés-értékelés
- környezetvédelem, fenntarthatóság
- Komplex Alapprogram
- sajátos nevelési igényű tanulók, fejlesztés, gyógypedagógia
- tehetséggondozás.

A vezetői továbbképzéssel kapcsolatos igazgatói igények, vélemények

Az igazgatói célcsoportot alkotó 558 fő megkérdezett meghatározó része, 86%-a szükségesnek tartja, hogy az intézményvezetői szakképzettség megszerzését követően az igazgatói tevékenység gyakorlása során is folyamatosan részt vegyen vezetőknek szóló képzésen. Ettől mindössze a válaszadók 8%-a zárkózik el, míg 6%-a bizonytalan (5. sz. ábra).



5. sz. ábra. A vezetői szakképzettség megszerzését követő további képzés szükségességének felmérése (fő)

Forrás: Saját szerkesztés

A szöveges válaszokban a folyamatos fejlesztés szükségességét elsősorban a következőkkel indokolták:

- folyamatos a változás a köznevelés területén és szükség van ennek követésére;
- szeretnének mind szakmai, mind vezetéselméleti tudásukban naprakészek maradni/lenni;
- szükség van az állandó megújulásra;
- igény az élethosszig tartó tanulásra, a folyamatos fejlődésre.

A képzés típusára vonatkozóan azok az igazgatók, akik szükségesnek tartják a további képzést, a 8 felkínált lehetőség közül leginkább a kooperatív tanuláson alapuló, csoportos képzési formákat (pl. tréning, esetmegbeszélés, problémamegoldás, team coaching) preferálták, valamint a helyszíni megfigyelést (tanulmányút, intézménylátogatás). E két képzéstípust követik az egyéni képzés, a részvétel előadásokon/konferencián, valamint az önképzés. Legkevésbé jónak a szakvizsgás képzésen való részvételt, a projekt munkát és a tanfolyamot tartották (6. sz. ábra).

Képzéstípus	Értékelési pontszám (1-legkevésbé, 6-leginkább)						Össz- pontszám
	6	5	4	3	2	1	
tanfolyam	89	96	72	84	56	84	1750
szakvizsga	54	53	53	68	80	173	1338
kooperatív tanuláson alapuló, csoportos képzési formák (tréning, esetmegbeszélés, problémamegoldás, team coaching, stb)	205	115	65	45	30	21	2281
előadás(sorozat), konferencia	94	124	87	96	58	22	1958
egyéni képzés, fejlesztés (tanácsadás, mentorálás, coaching stb.)	139	119	99	66	35	23	2116
önképzés, önirányító képzési forma	115	97	91	86	55	37	1853
helyszíni megfigyelés (tanulmányút, intézménylátogatás stb.)	184	115	82	55	29	16	2246
projektmunka	48	80	95	106	74	78	1612

6. sz. ábra. A képzéstípusok értékelése (fő)

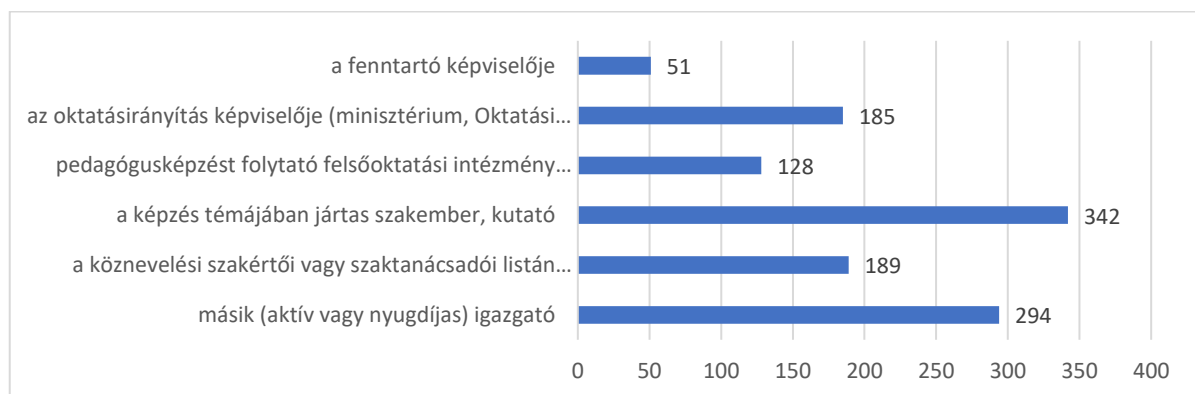
Forrás: Saját szerkesztés

A további képzés témájára vonatkozó igényként leggyakrabban a lenti témaköröket nevezték meg az igazgatók szöveges válaszaikban:

1. jog (alkalmazás, változáskövetés, értelmezés, munkajog, törvényi előírásoknak való megfelelés, aktualitások stb.)
2. változásmenedzsment, változáskezelés, változáskövetés
3. konfliktuskezelés
4. szervezettefejlesztés
5. kommunikáció.

Az igazgatók körülbelül fele arra a kérdésre, hogy kitől tanulnának szívesen a vezetőknek szóló további

képzésen, a „képzés témájában jártas szakember, kutató”, illetve „másik igazgató” válaszokat adták. A következő egységet „a köznevelési szakértői vagy szaktanácsadói listán szereplő szakértő/szaktanácsadó” és az „oktatásirányítás képviselője” (minisztérium, Oktatási Hivatal, kormányhivatal stb.) adja, nagyjából egyforma, 16%-os, illetve 15%-os megoszlással. A legkevesebben pedagógusképzést folytató felsőoktatási intézmény oktatójától (11%), illetve a fenntartó képviselőjétől (4%) tanulnának. Az egyéb válaszok minimális jelenléte (2%) arra utal, hogy a felkínált válaszok közül az érintett csoport tagjai választani tudtak (7. sz. ábra).



7. sz. ábra. Kitől tanulna szívesen a vezetőknek szóló továbbképzésen? (fő)

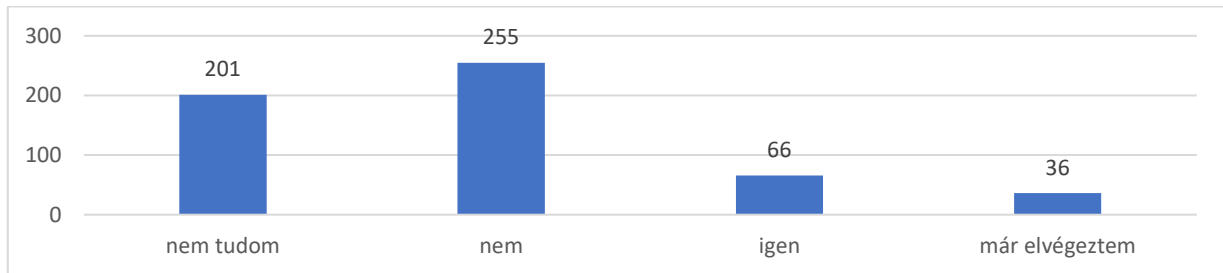
Forrás: Saját szerkesztés

A kérdőív következő kérdései arra irányultak, ismerik-e az igazgatók, milyen lehetőségeik vannak további képzés elvégzésére? Ugyanis jelenleg a

továbbképzési jegyzékben vannak kifejezetten vezetőknek szóló továbbképzések (Waldorf – intézményvezető képzés; Intézményfejlesztés-Változásmenedzsment; Intézményvezetők

megoldásfókuszú és fejlődésszemléletű attitűdjének fejlesztése; Vezetők szendvicshelyzetben) és egy szakirányú továbbképzés, a Köznevelési mestervezető képzés is ezt a célt szolgálja. A válaszok alapján a továbbképzési jegyzék tartalmáról az igazgatók közel felének nincsen tudomása, illetve 16% azt a választ adta, hogy nincs kifejezetten nekik szóló továbbképzés akkreditálva. Az igazgatóknak tehát 35%-a rendelkezik ismerettel a továbbképzési jegyzék kifejezetten vezetőkre vonatkozó tartalmáról. A szakvizsgás képzésről 56%-ának volt, 44%-ának nem volt tudomása.

Az igazgatók közel fele „nem”-mel válaszol arra a kérdésre, hogy szándékában áll-e elvégezni a kifejezetten több éves vezetői tapasztalattal rendelkező hallgatókra épülő köznevelési mestervezető képzést, míg a harmada bizonytalan. Így kevesebb, mint a megkérdezettek ötöde végezte már el vagy áll szándékában elvégezni a gyakorlott vezetők további képzését célzó továbbképzést (megjegyzem, ugyanez az arány mutatkozott meg az előző egységben, az igazgatók továbbképzési gyakorlatát bemutató részben is) (8. sz. ábra).



8. sz. ábra. Mestervezető képzés elvégzésére irányuló szándék (fő)

Forrás: Saját szerkesztés

Összegzésképpen megállapítható, hogy az igazgatók meghatározó része, közel 90%-a igényli a további vezetőképzést, bár annak jelenleg elérhető formáiról kétharmadának nincsenek tudomása. A képzési tartalmat érintően egyaránt van igény a mindennapi intézményi munkát érintő szabályozási környezet változásának követésére, valamint az általános vezetői tevékenység fejlesztésére (kommunikáció, konfliktuskezelés, szervezetfejlesztés, változáskezelés). Egyértelmű az is, hogy nem a hagyományos képzési formát és módszertant preferálják az igazgatók, hanem elsősorban a problémaközpontú, csoportos, kooperatív, a tapasztalatcserén alapuló képzési formákat (pl. tréning, esetmegbeszélés, problémamegoldás, team coaching, tanulmányút, intézménylátogatás), mely során ismereteiket vagy egymástól, vagy a képzés témájában jártas szakembertől, kutatótól, szakértőtől, szaktanácsadótól, illetve az oktatásirányítás képviselőjétől sajátítanak el.

1. Az igazgatók igénylik a vezetői szakképzettség megszerzése után is a további képzést.
2. Az igazgatók nagy része a fejlesztésükre irányuló jelenlegi lehetőségeket nem ismeri.
3. A képzési kínálatot bővíteni szükséges, mert jelenleg négy pedagógus és egy szakirányú továbbképzés biztosítja a speciálisan vezetőkre vonatkozó további képzést.
4. Egy lehetséges koncepció kidolgozása során figyelemmel kell lenni arra, hogy képzésszervezési szempontból elsősorban nem a hagyományos, formális, hanem az idő- és utazási problémákat megoldó, gyakorlatorientált képzéseket szükséges preferálni, mindezt mérsékelt képzési díjért vagy ingyenesen.
5. A képzés témája vonatkozásában egyaránt igény van az aktuális szakmai, valamint az általános vezetőfejlesztési tartalmakra.

A kutatási eredmények összegzése, a további lehetséges irányok meghatározása

A gyakornoki időszak alatt megvalósított kutatási tevékenység részeredményei alapján az igazgatók fejlesztésére irányuló koncepció kidolgozása során a következő tényezőkre mindenképpen figyelemmel kell majd lenni:

Köszönetnyilvánítás

A beszámoló befejezéseként szeretnék köszönetet mondani a gyakornoki tevékenységem támogatásáért mind a Belügyminisztériumnak, mind az Oktatási Hivatalnak, és bízom benne, hogy a teljes koncepció a kutatás többi részterületének kidolgozásával a doktori munkámban megismerhető lesz. Mivel mind a pedagógusok, mind az igazgatók továbbképzési rendszerének megújítása jelenleg aktuális a

köznevelésben, remélem kutatásom eredménye vagy annak egy része adaptálható és felhasználható lesz ebbe a folyamatba is.

Irodalomjegyzék

- Oktatási Hivatal (2023) *Működő köznevelési intézmények listája* (aktuális adatokkal).
<https://dari.oktatas.hu/kirpub/index>
[Letöltve: 2023.06.05.].
- URL: 277/1997. (XII. 22.) Korm. rendelet a pedagógus-továbbképzésről, a pedagógus-szakvizsgáról, valamint a továbbképzésben részt vevők juttatásairól és kedvezményeiről.
<https://net.jogtar.hu/jogszabaly?docid=99700277.kor> [Letöltve: 2024.03.28.].
- Vereckei, J. (2021) A közoktatási vezetőképzés helye a pedagógusok szakmai képzésének rendszerében. In: Karlovitz, J. (szerk.) *Szaktárszertani és pedagógiai tanulmányok a világjárvány idején*. Komárno, International Research Institute s.r.o. pp. 119-130.

Szerzők

dr. Albert Ágota Katalin

mesterséges intelligencia és technológiai jogi szakjogász
adatvédelmi és adatbiztonsági szakjogász

Balogh Mónika

Nemzeti Közszolgálati Egyetem
Katonai Műszaki Doktori Iskola
PhD hallgató

Baranyai András

BAHACO GDPR AI Cloud Kft.
adatvédelmi és workshop felelős

Horváth Illés

Belügyminisztérium

Karsa Róbert

tűzoltó ezredes
Baranya Vármegyei Kataszrófavédelmi Igazgatóság
igazgató-helyettes

Nagy Péter

Debreceni Egyetem
Gazdaságtudományi Kar
Gazdálkodástudományi Intézet
PhD hallgató

Vereckei Judit

Tatabányai Tankerületi Központ
igazgató