TUDOMÁNYOS ÉS ÜZLETI HORIZONTOK KONFERENCIÁJA

BIZALOM AZ AI KORÁBAN

INTERDISZCIPLINÁRIS
NEMZETKÖZI KONFERENCIA

AZ MTA-VEAB GAZDASÁG-, JOG- ÉS TÁRSADALOMTUDOMÁNY SZAKBIZOTTSÁG ÉS A MI ÉS A KREATÍV IPARÁGAK TÁRSADALOMTUDOMÁNYI KUTATÁSA MUNKABIZOTTSÁG

KUTATÓK, ÜZLETI ÉLET SZEREPLŐI ÉS DIÁKKUTATÁSOK

A konferencia elnöke: Garaczi Imre
A szervező bizottság elnöke: Molnárné Dr. László Andrea

MAGYAR • TUDOMÁNYOS • AKADÉMIA •
VEAB

Mind Mate Inspiration

2025.

Bizalom az AI korában

Konferenciakötet

2025. Május

Veszprém

VEAB

MI és a kreatív iparágak társadalomtudományi munkabizottsága

Felelős Kiadó: Molnárné Dr. László Andrea

ISBN 978-615-02-5808-9

Tartalom

Az oktatás, a mesterséges intelligencia és a bizalom metszéspontjai

The Intersection of Education, Artificial Intelligence, and Trust

Molnárné Dr. László Andrea[1]

Absztrakt

A mesterséges intelligencia (MI) megjelenése alapjaiban formálja át az oktatás szerepét, eszközeit és a tanári hivatás értelmezését. Tanulmányunk arra vállalkozik, hogy feltárja az MI oktatásban való alkalmazásának pedagógiai, bizalmi és etikai dimenzióit. Különös figyelmet fordítunk arra, hogy milyen új tanári kompetenciák szükségesek az MI-alapú tanulási környezetek hatékony kezeléséhez, és hogyan tanítható meg a tanulóknak a reflektív bizalom gyakorlata – vagyis az MI-vel szembeni kritikus, mégis együttműködő attitűd. A tanulmány gyakorlati példákon keresztül mutatja be az oktatási szereplők felelősségét az MI használatának pedagógiai integrációjában, kiemelve a tudás újradefiniálásának szükségességét a digitális korszakban.

Abstract

The emergence of artificial intelligence (AI) is fundamentally reshaping the role of education, its tools, and the meaning of the teaching profession. This study explores the pedagogical, ethical, and trust-related dimensions of integrating AI into educational contexts. Special emphasis is placed on the evolving competencies required of teachers to effectively manage AI-driven learning environments, and on how students can be taught reflective trust – a critical yet cooperative stance towards AI technologies. Through practical examples, the study highlights the responsibility of educational stakeholders in the pedagogical integration of AI and underlines the urgent need to redefine knowledge in the digital age.

Keywords: education, artificial intelligence, trust, pedagogical innovation, teacher roles

Bevezetés

A mesterséges intelligencia (MI) térnyerése nemcsak a gazdasági és társadalmi struktúrákat formálja át, hanem az oktatási rendszerek szerepét és működését is alapvetően megkérdőjelezi. E tanulmány célja az MI oktatásban betöltött szerepének, valamint az ezzel összefüggő bizalmi

[1] Molnárné Dr. László Andrea, MyNovaMind alapítója, kutatótanár, tanár-tréner, LLM tananyagfejlesztő molnarne7@gmail.com

kérdéseknek a vizsgálata, különös tekintettel a tanári szerepek változásaira és a tanulói attitűdök átalakulására.

Kutatói és pedagógiai háttér

A szerző oktatási munkája során K–12-es intézményben tanít, így közvetlen tapasztalattal rendelkezik a különböző korosztályú tanulók digitális kompetenciáiról, MI-hoz való viszonyulásáról és az oktatási technológiák iskolai integrációjának kihívásairól. Tudományos tevékenysége nemzetközi együttműködésekre és kutatásokra épül, fő fókuszában az MI pedagógiai alkalmazása és a tanárok digitális transzformációjának támogatása áll.

1. Az MI oktatásra gyakorolt hatása – jelen és jövő

A mesterséges intelligencia (MI) megjelenése és gyors elterjedése alapjaiban kérdőjelezi meg az oktatási rendszerek hagyományos működését. Az „így szoktuk” módszertana, amely hosszú ideig biztosította az oktatási folyamat stabilitását, ma már nem elegendő. Az MI nem csupán egy új technológiai eszköz, hanem egy rendszerszintű változást generáló erő, amely átalakítja a társadalmi működés minden rétegét – így az oktatást is. Az oktatásnak tehát nem passzívan követnie, hanem proaktívan reagálnia kell az MI kihívásaira és lehetőségeire. A tanulás önmagában is rendkívül komplex, többszintű folyamat, amely kognitív, érzelmi, szociális és motivációs tényezők finom egyensúlyán alapul. Ebbe a dinamikus rendszerbe illeszkedik be az MI, amely képes átrendezni a tanulási helyzetek szerkezetét, a tanár-diák viszonyt, a tudáshoz való hozzáférés módját, valamint a tanulói felelősségvállalás és önirányítás szerepét. Az a vélekedés, miszerint az MI nem lesz hatással az oktatásra, nem tartható fenn – épp ellenkezőleg: az MI olyan paradigmaváltást kényszerít ki, amely a tanulási környezetek teljes újragondolását igényli. Az MI már most is átalakítja a munkaerőpiacot, új kompetenciákat helyezve előtérbe: problémamegoldás, rendszerszintű gondolkodás, adatértelmezés, etikus technológiahasználat, digitális együttműködés, és mindenekelőtt – kritikus gondolkodás. E kompetenciák iskolai fejlesztése nem választható lehetőség, hanem szükségszerűség. A pedagógiai gyakorlatnak tudatossá kell tennie azokat a területeket, ahol a tanulók digitális eszközökkel végzett önálló tanulása nemcsak lehetőség, hanem elvárás – és ahol a felhasználói felelősség, az algoritmusok működésének ismerete, valamint az adatokkal kapcsolatos bizalom és kontroll tudatos fejlesztést igényel. Mindezek fényében a tanári szerep újradefiniálása elkerülhetetlen. A klasszikus tudástranszfer – amelynek során a tanár a tudás elsődleges forrásaként jelenik meg – háttérbe szorul, helyét a tanulási folyamat irányítása, a források hitelességének értelmezése,

valamint a tanulói kritikai gondolkodás fejlesztése veszi át. A pedagógus ma már nemcsak információközvetítő, hanem mentorként, facilitátorként, technológiai értelmezőként és etikai iránytűként is funkcionál.

Az MI jelenléte rámutat az oktatás egyik legkritikusabb hiányosságára: a bizalom és a kontroll egyensúlyának hiányára. A tanulók nagy része kontrollálatlanul elfogadja az MI által generált válaszokat – különösen, ha azok formailag meggyőzőek vagy gyorsak. Ezért a tanítás egyik új feladata a „reflektív bizalom" kialakítása: annak megtanítása, hogyan lehet egyszerre bízni a technológia potenciáljában, miközben állandóan tudatában vagyunk annak is, hogy működése nem tévedhetetlen.

Az MI tehát nem csupán egy új eszköz az oktatásban – hanem egy új környezet, amelyben a pedagógiának új nyelvet, új elveket és új célokat kell találnia. A jelen kihívása az, hogy képesek vagyunk-e elengedni a „régi normalitást", és létrehozni egy olyan oktatási rendszert, amely nemcsak reagál az MI jelenlétére, hanem aktívan formálja annak értelmes és etikus felhasználását.

2. A tanulási folyamat és a bizalom összefüggése

A tanulás nem létezhet bizalom nélkül. Amennyiben az MI-t bevonjuk a tanítás-tanulás folyamatába, alapvető kérdés, hogy milyen szintű bizalommal fordulhatnak felé a tanulók. A kritikus gondolkodás és a kontrollált felhasználás kiemelt jelentőséggel bír, különösen fiatalabb korosztályoknál, akik nagyobb mértékben hajlamosak az MI válaszait kontroll nélkül elfogadni.

3. A tudás újradefiniálása a XXI. században

Az információs társadalom kihívásai újfajta tudásértelmezést követelnek. A lexikális ismeretanyag szerepe csökken, a "túlélő tudás" pedig egyre inkább az elemző, szintetizáló, értelmező képességek irányába tolódik. Az oktatás célja nem pusztán ismeretátadás, hanem annak megtanítása, hogyan kezelje a tanuló az MI által közvetített információt, hogyan értékelje annak megbízhatóságát.

A tudás saját értelmezésemben az alábbi módon definiálható:
A 21. század tudása egy olyan problémamegoldás-központú, piaci igényekre érzékeny, szelektív és elemző ismeretkészlet, amely magában foglalja a digitális műveltséget, a különböző tudományágak alapvető ismeretét és megértését, beleértve azok céljait, alapfogalmait, alap összefüggéseit, logikus felépítését, az etikai és jogi vonatkozásokkal bíró

interdiszciplináris szemléletet, valamint a globális referencia keret alkalmazását az információk értelmezésében és integrálásában. A tudás ebben az új megközelítésben áttevődik a "kevésről sokat" paradigmáról a "mindenről kicsit" irányába, ahol a széleskörű alapműveltség és a különböző területek közötti összefüggések felismerése, a diverzitás válik fontossá. Emellett tartalmazza a promptmérnöki tudást, az MI-vel való hatékony kommunikáció alapismereteit, a kommunikációs szabályok alapjait, valamint a fordítástudományi, nyelvtudományi ismereteket, amelyek mind hozzájárulnak a komplex problémák megoldásához a nyelvi modellek megértéséhez, sikeres alkamazásához. A 21. századi tudás magában foglalja annak felismerését, hogy az ismeretek különböző nézőpontokból közelíthetők meg és többféle perspektívából elemezhetők. Az egyén rendelkezik azzal a tudással, amely lehetővé teszi számára, hogy az új információforrásokat azonosítsa és hozzáférjen azokhoz. Tudatában van annak is, hogy az ismeretek dinamikusak, folyamatosan bővülőek, és képes az eltérő nézőpontok értelmezésére és integrálására annak érdekében, hogy a felmerülő problémákat szélesebb, holisztikusabb kontextusban vizsgálja és oldja meg.

Azonban nemcsak a tudást, a bizalmat is újra kell kereteznünk és meg kell tanítani a diákoknak egy újfajta bizalom létét.

4. Az MI-vel kapcsolatos bizalom típusai

A reflektív bizalom fogalmát O'Neil alkotta meg, amely új alapokra helyezi az MI-rendszerekhez fűződő viszonyunkat. A koncepció szerint nem elegendő vakon hinni a technológiai válaszok helyességében: a felhasználónak értenie kell az adott MI működését, fel kell mérnie annak képességeit, és el kell tudnia dönteni, milyen kontextusban alkalmazható megbízhatóan. E bizalom feltételezett, és nem abszolút – dinamikus, folyamatos újraértékelésre épül. Az MI-vel kapcsolatos felhasználói attitűd tehát nemcsak pszichológiai, hanem kognitív és etikai kérdés is. A bizalommal kapcsolatban három kulcstényező jelenik meg: a rendszer megértése, a megfelelő feladatra való alkalmasság, valamint a következetes, elszámoltatható működés. Ezek együtt képezik azt a bizalmi alapstruktúrát, amely nélkülözhetetlen a felelős MI-használathoz. Mindez szembemegy a technológiai determinizmus hagyományos, feltétel nélküli elfogadásával, és önálló értelmezés és a kontroll szükségességét feltételezi. A technológia korábban gyakran automatikusan megbízhatónak számított – például a számológépek vagy bármi más esetében. Az MI azonban prediktív és nem egzakt, így működésének validitása kontextusfüggő, hibalehetőséggel terhelt. A bizalom tehát csak akkor tartható fenn, ha a felhasználó tisztában van a rendszer korlátaival és aktívan figyeli annak működését. Oktatási környezetben ezt a tudatosságot és kritikai szemléletet kell tanítani a diákoknak. Az MI nem feltétlenül hibás – de nem is megkérdőjelezhetetlen.

A tanulók felkészítése erre a típusú bizalomra pedagógiai felelősség. Az oktatónak nem csupán az eszközhasználatot kell megtanítania, hanem az azt övező kritikus gondolkodást és értelmezési keretet is. A reflektív bizalom tehát nem csupán hozzáállás, hanem fejlesztendő kompetencia is.

5. A tanári feladatkörök újradefiniálása

A tanár szerepe az MI által nem csökken, hanem radikálisan átalakul. A pedagógus feladata már nem pusztán a tudás közvetítése, hanem a tanulók gondolkodásának irányítása, az MI által generált tartalmak értelmezésének segítése, valamint a bizalom kritikus értelmezésének tanítása. A tanár többé nem kizárólagos tudásforrás, hanem facilitátor és irányító, aki segít a tanulóknak eligazodni az információs túlkínálatban és a mesterséges intelligencia által létrehozott tartalmak sokféleségében. Ez a szerepváltás nem a tanári autoritás elvesztését jelenti, hanem annak újradefiniálását: a tanár a jövőben a „kritikai szűrő" szerepét tölti be, aki képessé teszi a diákokat a megbízható tudás azonosítására, a manipuláció felismerésére és a digitális környezetben való etikus eligazodásra.

A kritikus gondolkodás tanítása így elsődleges céllá válik. A kérdés – „Bízhatunk-e úgy, hogy nem válunk vak követőkké?" – jól rávilágít a bizalom kettősségére: egyszerre kell nyitottnak lenni az MI lehetőségeire, ugyanakkor képesnek lenni annak reflexív, elemző értelmezésére. Ez a kompetencia nem alakul ki magától, hanem tudatos fejlesztést és következetes tanári támogatást igényel. A tanulók számára az MI-hez kapcsolódó kompetenciák elsajátítása kulcsfontosságúvá válik: az adatértés, az algoritmikus gondolkodás, a digitális etika és az adatbiztonság mind olyan területek, amelyek nélkül a jövőben nem lehet teljes értékű digitális állampolgárságot gyakorolni.

A mai tanári feladatok között ezért egyre hangsúlyosabban jelenik meg az MI működésének megértetése, az adatkezelési alapelvek és a GDPR-nak való megfelelés tanítása, valamint a válaszok megbízhatóságának és az algoritmikus torzításoknak a felismerése. A pedagógusnak nemcsak a tartalmak validálásában van szerepe, hanem abban is, hogy segítse a tanulókat olyan gondolkodási minták kialakításában, amelyek lehetővé teszik a technológia felelős használatát. Az oktatás tehát nem maradhat meg az információátadás szintjén: a gondolkodásra tanítás, a reflexív tudásépítés és a problémamegoldó kompetenciák fejlesztése felé kell elmozdulnia.

A kompetenciák tudatos fejlesztése többlépcsős folyamat. Először is szükséges a megváltozott világ megértése: annak belátása, hogy az MI nem kiegészítő, hanem alapvető tényezővé vált a tanulási és munkafolyamatokban. Másodszor a pedagógusoknak el kell sajátítaniuk azokat az

AI-kompetenciákat, amelyek lehetővé teszik számukra, hogy ne csak saját munkájukban alkalmazzák, hanem a diákok számára is átadható tudásként építsék be a tanítás folyamatába. A módszertani repertoár bővítése kulcsfontosságú: az MI bevonása a tanórák tervezésébe, a differenciálásba, a formatív értékelésbe és a projektalapú tanulásba
mind olyan területek, ahol a tanár közvetlenül megtapasztalhatja, hogy az MI nem helyettesíti, hanem erősíti pedagógiai szerepét.

Ez a folyamat egyben szemléletváltást is kíván: a pedagógus nem az MI-vel szemben határozza meg önmagát, hanem az MI-vel együttműködve alakítja ki új tanári identitását. Ezzel nemcsak a tanulók számára válik példává, hanem hozzájárul egy olyan oktatási kultúra kialakításához is, amelyben az etikus és tudatos technológiahasználat az alapvető értékek közé emelkedik. A tanár tehát a jövő iskolájában egyszerre lesz mentor, kritikai gondolkodásra nevelő, adatértelmező és módszertani újító – olyan szakember, aki képes integrálni a mesterséges intelligenciát az oktatás egészébe, miközben folyamatosan reflektál annak lehetőségeire és veszélyeire.

Az alábbi összefoglalóban áttekinthetjük a tanári AI-kompetenciák és fejlesztési lépések lehetőségét. Értelemszerűen ezek az AI kompetenciákra vonatkoznak. Az egyéb tudatos kompetenciafejlesztés (pl. kreativitás, csapatmunka, önálló tanulásra való készség, felelősségvállalás, stb.) is részét kell, hogy képezze az új pedagógiai munkának.

Kompetencia	Fejlesztési fókusz	Tanári alkalmazás a gyakorlatban
Promptolás és iteráció	Tudatos kérdésfeltevés, pontosítás és újrafogalmazás képessége	A tanár modellezi a tanulóknak, hogyan lehet jó kérdéseket feltenni az MI-nek; tanórán közösen elemzik a különböző promptok eredményeit.
Feladatbontás és tervezés	Összetett problémák lebontása, tanulási folyamat strukturálása	Projektmunkák előkészítése során az MI segítségével ütemtervet készít, majd a diákokkal értékelteti annak reális elemeit.
Eszközválasztás és összehasonlítás	Különböző AI-eszközök ismerete és kritikus kiválasztása	A tanár megmutatja, hogy nem minden MI-eszköz alkalmas ugyanarra, és közösen döntenek arról, melyik illeszkedik legjobban egy adott feladathoz.
Hitelesség és forrásellenőrzés	Információk verifikálása, megbízhatóság vizsgálata	„Hallucinációvadászat” keretében a diákokkal hibákat keresnek az MI válaszaiban, majd forráskritikai beszélgetést folytatnak.

Kompetencia	Fejlesztési fókusz	Tanári alkalmazás a gyakorlatban
Etika és torzítás tudatosítása	Algoritmikus előítéletek és adatvédelmi kérdések felismerése	Konkrét példákon keresztül (pl. sztereotípiák megjelenése) beszélgetést vezet a felelősségteljes MI-használatról.
Adatértés és GDPR	Adatbiztonsági szabályok megértése és alkalmazása	Tanórán tudatosítja a személyes adatok védelmének fontosságát, és felhívja a figyelmet a felelőtlen adatmegosztás következményeire.
Algoritmikus gondolkodás	Problémamegoldás lépésről lépésre, algoritmusok logikájának megértése	Diákokkal közösen „szimulálják” az MI működését: pl. szabályok alapján emberi szereplők játszanak „algoritmust”.
Kritikus bizalom fejlesztése	Nyitottság az MI lehetőségeire, de reflexív használat	Megvitatják a kérdést: „Bízhatunk-e az MI-ben úgy, hogy közben megőrizzük a saját kritikai szűrőinket?”

6. Pedagógiai gyakorlatok a bizalom fejlesztésére

A bizalom oktatása nem pusztán elméleti megközelítés kérdése, hanem pedagógiai gyakorlatban is beépíthető. Az MI-hez kapcsolódó kritikus bizalom kialakítása ugyanis akkor lehet eredményes, ha a tanulók nem csupán absztrakt szinten értik meg a technológia működésének kockázatait és lehetőségeit, hanem közvetlen tapasztalatokon keresztül reflektálnak is azokra. Ennek érdekében a tanórai keretek között olyan feladatok alkalmazhatók, amelyek a technológiai rendszer átláthatóságát és megbízhatóságát teszik vizsgálat tárgyává.

A gyakorlatok – például a prompt-elemzés, a „hallucinációvadászat” vagy a szemléletformáló projektek – nem öncélú tevékenységek, hanem a tanulói autonómia, a kritikai gondolkodás és az etikus technológiahasználat fejlesztésének eszközei. E feladatok keretében a tanulók nemcsak a generált tartalom minőségére reflektálnak, hanem implicit módon saját viszonyukat is alakítják a mesterséges intelligenciához mint tudáskonstrukciós partnerhez.

Különösen fontos, hogy az ilyen tanulási folyamatokban a bizalom nem passzív elfogadásként jelenik meg, hanem aktív, kritikai szűrőn átesett viszonyként. A pedagógiai kontextusban a „bizalom oktatása” tehát nem a technológia kritikátlan idealizálását, hanem éppen annak tudatos használatát és a felelősségteljes döntéshozatal támogatását jelenti. Ily módon a tanórai

integráció nem csupán készségeket fejleszt, hanem hozzájárul a digitális állampolgárság és az oktatási etika alapvető dimenzióinak megerősítéséhez is.

A bizalom oktatása nem elméleti kérdés. Konkrét tanórai gyakorlatokat is be lehet építeni:

- **Prompt-elemzés**: különböző bemenetek hatásának vizsgálata a generált tartalomra;
- **„Hallucinációvadászat"**: MI által generált hibás állítások keresése és kritikája;
- **Szemléletformáló projektek**: például a sztereotípiák detektálása különböző bemenetek alapján.

Mindezek a gyakorlatok jól mutatják, hogy a bizalom tudatos alakítása a mesterséges intelligencia használatában nem mellőzheti a kritikai szempontokat és a reflektív hozzáállást. Ugyanakkor éppen a bizalom pedagógiai fejlesztése világít rá arra is, hogy a technológia alkalmazása számos kockázatot hordoz magában. Az, hogy a tanulók képesek legyenek felismerni és értelmezni e veszélyeket, kulcsfontosságú feltétele annak, hogy ne csupán felhasználói, hanem felelős értelmezői és alakítói legyenek a digitális környezetnek. A következőkben ezért a mesterséges intelligencia oktatási kontextusban való megjelenésének veszélyeit tárgyaljuk, különös tekintettel azok pszichológiai, pedagógiai és társadalmi implikációira.

7. Veszélyek

A technológia iránt érzett bizalom/bizalmatlanság nem mentes a veszélyektől. Az alábbiakban hat olyan terület kerül bemutatásra, amelyeket érdemes szem előtt tartani a pedagógiai folyamatok során.

7.1. Vak engedelmesség

Kutatások szerint az emberek hajlamosabbak elfogadni egy tanácsot, ha azt algoritmushoz kötik – például azonos tanácsot *22%-kal* nagyobb arányban fogadnak el, ha azt mesterséges intelligencia adja. Ez a jelenség az „algoritmus-tekintély" hatására épül, amely a technológia iránti túlzott bizalomból fakad. Oktatási környezetben ez különösen veszélyes, mert a diákok kritikai szűrés nélkül fogadhatják el az MI által generált válaszokat, ami hosszú távon csökkenti az önálló gondolkodást és az információforrások összehasonlításának képességét. A pedagógiai cél itt az, hogy a tanulók megértsék: az MI nem tévedhetetlen, a kapott információt minden esetben ellenőrizni kell.

7.2. Irracionális elutasítás

A másik véglet az, amikor az algoritmus egyszeri hibája után a bizalom jelentősen visszaesik – kutatások szerint akár *37%-kal* is. Ez a „hiba után teljes elutasítás" jelenség szintén káros, mert gátolja a technológia konstruktív alkalmazását. Az oktatásban fontos megtanítani, hogy

egyetlen hiba nem jelenti a rendszer teljes alkalmatlanságát, és a hibák gyakran a helytelen felhasználásból, pontatlan bemenetből vagy kontextushiányból fakadnak. A tudatos felhasználó képes különbséget tenni az egyszeri tévedés és a rendszerszintű megbízhatatlanság között.

7.3. Hallucináció felismerése

Az MI-rendszerek – különösen a generatív nyelvi modellek – hajlamosak úgynevezett „hallucinációkra", amikor valósnak tűnő, de hamis információkat generálnak. Ezek felismerése fejlett forráskritikai és összehasonlító képességet igényel. Oktatásban hatékony gyakorlat lehet több MI által adott válasz összehasonlítása, majd hiteles forrásokkal való ellenőrzése. Így a diákok megtanulják, hogy a meggyőző formátum nem garantálja a tartalmi helytállóságot.

7.4. Félrevezetett félelem

A mesterséges intelligencia fejlődése sokakban technofóbiát vált ki. Ez a félelem gyakran abból fakad, hogy a felhasználók nem értik a technológia működését, és emiatt inkább teljesen elutasítják azt. A fejlődés elutasítása viszont versenyhátrányt eredményezhet, különösen a munkaerőpiacon. Az oktatás szerepe kulcsfontosságú a félelmek oldásában: megfelelő ismeretek és gyakorlati tapasztalat nyújtásával a technológia a fenyegetés helyett lehetőséggé válhat.

7.5. Fordítógép/Képgenerátor-függőség

A fordítógépek, képgenerátorok és egyéb kreatív MI-eszközök rendkívül hasznosak, de túlzott használatuk készségvesztéshez vezethet. Például a diák nyelvtudása romolhat, ha minden fordítási feladatot automatikusan MI-vel végez, vagy a vizuális kreativitása csökkenhet, ha kizárólag képgenerátorokra támaszkodik. A pedagógusnak tudatosítania kell a tanulókban, hogy az MI eszköz, nem pedig helyettesítő – a felhasználás célja a támogatás, nem a teljes helyettesítés.

7.6. Kiszolgáltatottság

Az MI-t értő, használni tudó személyek jelentős versenyelőnybe kerülnek a nem hozzáférőkkel szemben, ami társadalmi szakadékot mélyíthet. Ez a „digitális egyenlőtlenség" komoly oktatáspolitikai kérdés is: ha az iskolák és tanárok nem biztosítanak egyenlő hozzáférést és képzést, akkor a hátrányos helyzetű diákok még nagyobb lemaradásba kerülnek. Az oktatás feladata, hogy minden tanulót felkészítsen az MI-korszak kompetenciáira, függetlenül a kiinduló helyzetüktől.

Ez a hat terület együttesen mutatja, hogy a mesterséges intelligencia használata nem csupán technológiai, hanem komoly pedagógiai, pszichológiai és társadalmi kihívás is. A „reflektív

bizalom” tanítása kulcsfontosságú ahhoz, hogy a diákok elkerüljék mind a túlzott bizalmat, mind az irracionális elutasítást, és tudatos, felelős technológiahasználókká váljanak.

Az űrlap alja

Összegzés

A mesterséges intelligencia oktatásba való integrálása nem pusztán technológiai fejlesztés, hanem mély pedagógiai és társadalmi változás katalizátora. A tanulmány bemutatta, hogy az MI által generált új kihívások – a tudás újradefiniálásától a reflektív bizalom kialakításáig – komplex, egymással összefonódó területeket érintenek. A jövő tanárai nem egyszerűen a tartalomátadásért felelnek, hanem az értelmezés, a kritikai gondolkodás, az etikai mérlegelés és az algoritmikus tudatosság fejlesztéséért is.

Az MI-vel való pedagógiai együttműködés sikere azon múlik, hogy képesek vagyunk-e tudatosan kiegyensúlyozni a bizalom és a kontroll viszonyát, miközben minden tanulót felkészítünk a technológiai környezet aktív és felelős alakítására. Ez nem egyszeri feladat, hanem folyamatos reflexiót és alkalmazkodást igénylő folyamat, amelyben az oktatás szereplői – tanárok, diákok és döntéshozók egyaránt – közösen formálják a jövő tanulási kultúráját. „A bizalom nem abból fakad, hogy mindent elhiszünk – hanem abból, hogy tudjuk, mikor kérdezzünk vissza” (O’Neil, 2002.)

Irodalomjegyzék

AI4K12 & IBM (2024) *IBM AI Education course description*.

Bank for International Settlements (2025) *Governance of AI adoption in central banks*.

Dietvorst, B.J., Simmons, J.P. & Massey, C. (2015) ‘Algorithm aversion’, *Journal of Experimental Psychology*.

European Parliament (2024) *Artificial Intelligence Act – végleges szöveg és idővonal*.

Logg, J.M., Minson, J.A. & Moore, D.A. (2019) ‘Algorithm appreciation’, *Organizational Behavior and Human Decision Processes*, 151, pp. 90–103.

McKinsey & Company (2025) *The state of AI: How organizations are rewiring to capture value*.

MIT RAISE (2024) *Day of AI impact report*. Available at: https://dayofai.org

NIST (2023) *AI Risk Management Framework 1.0*.

O'Neill, O. (2002). *A question of trust: The BBC Reith Lectures 2002*. Cambridge University Press.

Patel, S. & Thurston, D. (2024) *Ethics and AI*. *Leadership Magazine*.

PwC (2024) *2024 US Responsible AI Survey*.

Sokoloff, N. (2025) 'AI is already here' – Connecticut school districts…, *CTInsider*.

University of Helsinki (2023) *Elements of AI reached one million learners*.

Vodafone Group (2024) *Annual Report 2024 – AI Governance Board*.

Az MI lehetőségei az óvodai nevelésben

Potential of AI in Early Childhood Education

Piróth István[2]

Absztrakt

A tanulmány célja az MI alkalmazásának óvodai nevelésben megjelenő lehetőségeinek áttekintése, különös figyelmet fordítva az óvodapedagógusok szemléletére és annak kapcsolódásaira. Az MI eszközök alkalmazása új pedagógiai, etikai és technológiai kérdéseket vet fel, amelyek hatással vannak a gyermekek fejlődésére és a pedagógusok munkájára egyaránt. Az MI segíthet a személyre szabott nevelési környezet kialakításában, de új kihívásokat is hoz a fejlett digitalizáció és személyes interakciók egyensúlyának megtartásában. Az MI fokozatos bevezetése mellett ügyelni kell, hogy a gyermekek kognitív, szociális és érzelmi fejlődése szoros összefüggésben áll az óvodapedagógus felkészültségével és a szülői támogatással. Emellett fontos az etikai és adatvédelmi kérdések megfelelő kezelése is, fókuszban a gyermekek jogaival és méltóságával.

kulcsszavak: óvodapedagógia, mesterséges intelligencia

Abstract

The aim of the study is to review the possibilities of AI application in preschool education, paying special attention to the approach of kindergarten teachers and its effects. The use of AI tools raises new pedagogical, ethical and technological questions that have an impact on children's development and teachers' work. AI can help create personalized educational environments, but it also brings new challenges in balancing digitalisation and face-to-face interactions. In addition to the gradual introduction of AI, it should be borne in mind that the cognitive, social and emotional development of children is closely related to the preparedness of the kindergarten teacher and parental support. It is also important to properly address ethical and data protection issues, taking into account children's rights and dignity.

keywords: early childhood education, artificial intelligence

[2] A szerző kulturális antropológus, informatikus, a Nemzeti Kutatás-fejlesztési és Innovációs Hivatal K+F szakértője. piroth.istvan1@gmail.com

Bevezető

A tanulmány célja, hogy áttekintse az MI (mesterséges intelligencia) alkalmazásának legfontosabb lehetőségeit és kihívásait az óvodai nevelésben[3], különös figyelmet fordítva az óvodapedagógusok különböző szemléleteire és azok hatásaira a MI integrálásának folyamatában. Az óvodapedagógia, mint rendkívül érzékeny és emberközpontú terület, számos kérdést vet fel az MI alkalmazásával kapcsolatban, különösen akkor, amikor figyelembe vesszük a pedagógusok különböző filozófiai és módszertani alapú hozzáállásait. E tanulmányban a *technológiaszkeptikus* és a *technológiaoptimista* szemléletű óvodapedagógusok nézőpontjait is figyelembe veszem, miként befolyásolják azok az MI-eszközök alkalmazásának elfogadását és hatékony használatát az óvodai környezetben (Lindahl és Folkesson 2012, Blackwell et al 2014, Palaiologou 2016).

A tanulmányban arra törekszem, hogy szekunder kutatás alapján bemutassam azokat a szakmai elemeket, amelyek véleményem szerint a címbéli szakterület leírását, kutatását, elemzését, rendszerezést és jobb megértését szolgálják. Ilyen módon iránytűnek tekintem az érdeklődők számára, hogyan kezdjenek hozzá, folytassák, egészítsék ki vagy éppen vitatkozzanak az MI óvodai nevelés univerzumának alább olvasható elemeivel, tartalmával, összefüggéseivel.

[3] A „digitalizáció az óvodákban" és a „mesterséges intelligencia (MI) az óvodákban" két különböző technológiai fejlesztési szintet képvisel az óvodai nevelés területén. Míg a digitalizáció alapvető digitális eszközök integrálását jelenti, addig az MI alkalmazása fejlettebb technológiák bevezetését igényli, amelyek képesek tanulni és alkalmazkodni. A digitalizáció célja a meglévő pedagógiai módszerek támogatása és gazdagítása digitális eszközökkel. Ezzel szemben az MI lehetőséget nyújt a pedagógiai folyamatok személyre szabására és új módszertanok kialakítására. A digitalizác során a fő kihívás a megfelelő eszközök beszerzése és a pedagógusok digitális kompetenciáinak fejlesztése. Az MI integrációja azonban további kihívásokat jelent, mint például az adatvédelem, az etikai kérdések kezelése és a technológia magas költségei.

1.Az MI jelentősége az óvodai nevelésben

A mesterséges intelligencia (MI) rohamos fejlődése az oktatás minden szintjén új lehetőségeket teremt (Baditzné Pálvölgyi és Jakab 2023), beleértve az óvodai nevelést is. Ugyanakkor a technológia integrálása számos kérdést és kihívást vet fel, amelyek megoldása elengedhetetlen a felelős és hatékony használathoz (Lee és Xiong 2024**)**. Ez az fejezet az MI óvodai nevelésben betöltött szerepét, előnyeit és potenciális korlátait tekinti át.

Az óvodai környezetben a mesterséges intelligencia (MI) olyan technológiák és rendszerek gyűjtőfogalma, amelyek képesek tanulni, problémákat megoldani és alkalmazkodni az emberi interakciókhoz, hogy támogassák az oktatási és nevelési folyamatokat. Az MI ebben a kontextusban nem csupán technikai megoldásokra korlátozódik, hanem olyan eszközök és szoftverek formájában jelenik meg, amelyek hozzájárulnak a gyermekek egyéni fejlődésének követéséhez, az oktatási és a nevelési folyamatok személyre szabásához és az óvodapedagógusok mindennapi munkájának megkönnyítéséhez (Honghu et al 2023, Neugnot-Cerioli és Laurenty 2024).

Az *adaptív tanulási rendszerek* olyan intelligens szoftverek vagy platformok, amelyek valós időben elemzik a felhasználók interakcióit, hogy az oktatási és a nevelési tartalmat a gyermekek egyéni igényeihez igazítsák. A *Smartick* egy mesterséges intelligenciára épülő matematikai és készségfejlesztési oktatási program minden gyermek számára egyedi tanulási tervet készít, elemzi a korábbi teljesítményét, hibáit és erősségeit. A rendszer azonnali visszajelzést ad a gyermekeknek a helyes és helytelen válaszokról, hogy azok azonnal korrigálhatók legyenek, így a feladatok fokozatosan válnak nehezebbé, ahogy a gyermek készségei fejlődnek. Az *adaptív tanulási rendszerek* lehetőséget nyújtanak a gyermekek egyéni tanulási ütemének és szükségleteinek megfelelő feladatok kialakítására. Az olyan alkalmazások, mint a *Smartick* vagy a *Khan Academy Kids*, adaptív módon alkalmazkodnak a gyermekek fejlődési szintjéhez, így minden gyermek saját tempójában haladhat, miközben egyéni támogatást kap a szükséges területeken. Az *Epic!* olvasó alkalmazás lehetőséget ad a gyermekeknek, hogy személyre szabott olvasmányokat válasszanak, fejlesztve így a szövegértési és szókincsfejlesztési képességeiket Su et al 2023).

Az *adaptív tanulási játékok* segíthetnek a gyermekeknek olyan szituációk kezelésében, amelyek szociális interakciót igényelnek, például együttműködés, segítségnyújtás, vagy a mások érzelmi állapotának felismerése. Az alkalmazások játékos formában tanítanak az együttműködésről, miközben pozitívan befolyásolják a gyerekek közötti kommunikációt és együttműködést. A *Kimochi* vagy a *Feelings Flashcards* segíthetnek a gyerekeknek, hogy megértsék és kifejezzék érzelmeiket, lehetőséget nyújtanak, hogy a gyerekek játékos módon tanuljanak a különböző

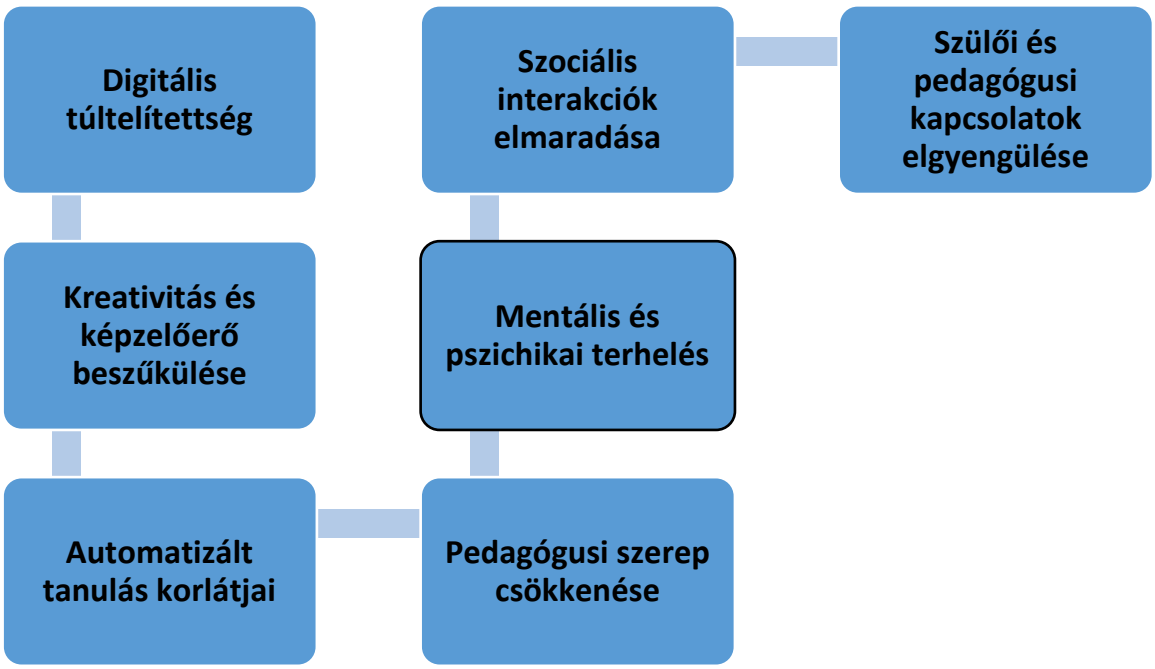
érzelmekről, és hogy felismerjék, hogyan reagáljanak mások érzéseire. A *My Emotional World* támogatja a gyerekeket, hogy szituációkat szimuláljanak, és így megtanulják, hogyan támogassák egymást, valamint hogyan kezeljék a nehéz érzelmi helyzeteket (Su et al 2023).

Az *interaktív robotok*, például a *QTrobot* egy interaktív, humanoid robot, amelyet kifejezetten oktatási és fejlesztési célokra terveztek, különösen a sajátos nevelési igényű gyermekek számára. A robot arckifejezéseket, gesztusokat és érzelmeket jelenít meg, amelyeket a gyermekek utánozhatnak vagy elemezhetnek, ezzel segítve az érzelmi intelligencia és az empátia fejlődését (Su et al 2023).

A *digitális asszisztensek* olyan MI-alapú alkalmazások, amelyek megkönnyítik az óvodapedagógusok számára az adminisztratív munkát, például a gyermekek fejlődési naplóinak vezetését vagy a szülőkkel való kommunikációt. A *ClassDojo* egy digitális asszisztens lehetővé teszi, hogy a pedagógusok közvetlenül kapcsolatba lépjenek a szülőkkel, valós idejű értesítéseket küldjenek a gyermekek napi tevékenységeiről, illetve a fejlődésük során tett előrehaladásról (Su et al 2023).

Az MI-alapú eszközök használatának korlátait a gyermekek fejlődése szempontjából folyamatszerűen is le lehet írni. Az alábbiakban egy ilyen folyamatábrát (1. ábra) adok meg, amely lépésről lépésre követi azokat a potenciális hatásokat, amelyek az MI-eszközök nem megfelelő alkalmazásából eredhetnek.

1. ábra: Az MI-alapú eszközök használatának korlátajai (Limitations of Using AI-based Tools)



forrás: saját szerkesztés

Ezek a folyamatok együttesen mutatják, hogyan vezethet a túlzott MI-használat a gyermekek fejlődésének egyes területein bekövetkező visszaeséshez. A gyermekek számára a legfontosabb a kiegyensúlyozott, interaktív környezet, amely lehetőséget biztosít mind a technológiai eszközök használatára, mind pedig a valós társas kapcsolatokra és a kreativitásra (Honghu et al 2023).

2. Az óvodapedagógusok szempontjai

Az MI alkalmazása az óvodai környezetben új lehetőségeket kínál az óvodapedagógusok számára, segítve őket a gyermekek egyéni igényeihez igazodó nevelési-tanulási környezet kialakításában, a nevelési-tanítási folyamatok hatékonyságának növelésében és az adminisztratív terhek csökkentésében. Az óvodapedagógusok szempontjából az MI nem csupán eszköz, hanem a pedagógiai munkájukat támogató segítő technológiai univerzum, amely egyaránt jelent új kihívásokat és lehetőségeket (Bessenyei 2023). Az alábbi fejezetben bemutatom az MI alkalmazásának előnyeit és kihívásait az óvodapedagógusok munkájában (Su et al 2023, Neugnot-Cerioli és Laurenty 2024, Tölgyes 2023).

Adminisztrációs segédeszközök közé tartozik pl. *a Brightwheel,* amely lehetővé teszi az óvodapedagógusok számára a gyermekek napi tevékenységeinek, fejlődésének és kommunikációjának egyszerű nyomon követését. Az MI segítségével automatikusan generálhatók fejlődési jelentések a szülők számára, és a napi adminisztráció gyorsan és hatékonyan kezelhető. A *Kindergarten 2.0* segít a gyermekek teljesítmények, jelenlét, napi aktivitások, illetve a szociális fejlődéseinek nyilvántartásában, miközben az adatok automatikusan szinkronizálódnak és elemzések is készíthetők.

Fejlődési naplók és adatgyűjtéshez használható a *Seesaw* interaktív platform, amely lehetővé teszi az óvodapedagógusok számára, hogy a gyermekek napi előrehaladását, munkáit és projektszerű feladatait rögzítsék. Az MI alapú analitika segít abban, hogy az adatok alapján személyre szabott fejlesztési tervet készíthessenek az óvodapedagógusok. A *Tiggly* támogatja a gyermekek kognitív és szociális készségeinek fejlesztését, miközben interaktív módon segíti a pedagógusokat a naplózási és elemzési feladatok elvégzésében.

A tanulási tervek automatizálásához használható a *Smartick***,** egy adaptív tanulási platform, amely a fejlődésüknek megfelelő matematikai feladatokat mutat be a gyermekek számára, A *Classcraft* egy játékos alapú tanulási rendszer, amely lehetőséget ad az óvodapedagógusoknak, hogy személyre szabott tanulási élményeket biztosítsanak a gyermekek számára. Kiváló kommunikációs eszköz a *ClassDojo*, amely lehetővé teszi az óvodapedagógusok számára, hogy egyszerűen kommunikáljanak a szülőkkel, oszthassanak meg információkat a gyermekek előrehaladásáról, viselkedéséről, és napi aktivitásaikról.

Az óvodapedagógusoknak képesnek kell lenniük arra, hogy *digitális tananyagokat és adaptív tanulási terveket* készítsenek, amelyek a gyermekek egyéni fejlődési igényeire és tempójára szabottak (Beták 2019, Beták és Szabó-Váradi 2023). Ehhez az alkalmazások és a tanulási platformok mélyebb megértésére van szükségük. Alkalmasnak kell lenniük az *interaktív*

tanulási eszközök (pl. *Osmo Learning*, *Kodable*) használatára, amelyek a gyerekek szociális, érzelmi és kognitív fejlődését támogatják. Rendelkezniük kell a *digitálisan gazdag tanulási környezetet* kialakításának képességével, amelyben a technológia nemcsak eszközként, hanem a tanulási élmény gazdagításaként is szolgál. Tisztában kell lenniük a *digitális állampolgári készségekkel*, amelyek segítik a gyermekeket az online környezetekben való felelős és biztonságos navigálásban (Neugnot-Cerioli és Laurenty 2024).

Az MI eszközök alkalmazásával az óvodapedagógusok szerepe jelentősen átalakulhat (Blackwell et al 2014). Míg korábban az oktatás és a tanulás irányítása alapvetően az óvodapedagógus feladata volt, addig most egyre inkább a *technológiai integrációra* és a *digitális tanulási környezetek kialakítására* helyeződik a hangsúly. Az óvodapedagógusok nemcsak a hagyományos oktatási módszereket kell alkalmazzák, hanem az MI-alapú alkalmazások kezelését és az azok által biztosított személyre szabott tanulási élményeket is koordinálniuk kell (Lindahl és Folkesson 2012).

Bár az MI jelentős támogatást nyújthat az óvodapedagógusok számára, fontos, hogy megőrizzék *szakmai autonómiájukat*. Az automatizálás és a mesterséges intelligencia alkalmazásának egyik lehetséges kockázata, hogy az óvodapedagógusok elveszíthetik a *személyes kapcsolatot* a gyermekekkel, és a tanulási folyamatok felett egyre inkább a gépek irányítanak. Az MI rendszerek által ajánlott megoldások gyakran nem veszik figyelembe a gyermekek egyedi szükségleteit, érzelmi állapotait és szociális fejlődését, ami miatt az óvodapedagógusoknak továbbra is *aktívan be kell avatkozniuk* a tanulási és a nevelésifolyamatokba, hogy ellássák azok megfelelő irányítását (Ertmer és Ottenbreit-Leftwich 2010, Palaiologou 2016).

Az MI bevezetésével új *etikai és szakmai dilemmák* is felmerülhetnek. Az óvodapedagógusoknak meg kell küzdeniük azzal, hogy hogyan biztosíthatják az *etikus alkalmazást* az MI-eszközökkel, figyelembe véve a gyermekek jogait, adatvédelmét és személyes fejlődését. A gyermekek személyes adatait kezelése olyan etikai kérdéseket vethetnek fel, mint a *magánélet védelme* és az *adatbiztonság*, különösen a GDPR-hoz való megfelelés szempontjából (Neugnot-Cerioli és Laurenty 2024).

3. Az óvodás gyermekek fejlődésére gyakorolt hatások

Ez a fejezet azokra a hatásokra összpontosít, amelyeket az MI eszközök gyakorolnak az óvodás gyermekek fejlődésére. A technológia alkalmazása nemcsak a kognitív fejlődést, hanem a szociális és érzelmi készségeket is segítheti, miközben új kihívások elé állítja az óvoda pedagógusokat és a szülőket (Dietz 2020). Az alábbiakban bemutatom, hogyan befolyásolhatják a különböző MI-eszközök a gyermekek fejlődését, és milyen előnyökkel, illetve kihívásokkal járhatnak ezek a technológiai újítások.

Az MI alapú *adaptív tanulási rendszerek* képesek a gyermekek egyéni fejlődési üteme alapján személyre szabott feladatokat adni, ami lehetővé teszi számukra, hogy saját tempójukban fejlődjenek, miközben fokozatosan új kihívásokkal találkoznak. Például a *Smartick* matematikai alkalmazás személyre szabja a feladatokat, figyelembe véve a gyermekek aktuális tudásállapotát, és folyamatosan alkalmazkodik a fejlődésükhöz. Az MI alapú nyelvtanulási eszközök, mint a *Duolingo for Kids*, interaktív játékos formában segítenek a gyermekek szókincsének bővítésében és a nyelvi készségek fejlesztésében.

Az olyan MI-alapú *interaktív mesemondó robotok*, mint az *Emo the Robot*, képesek érzelmi interakciókat generálni a gyermekekben, segítve őket az érzelmi kifejezés és az empátia fejlesztésében. Az érzelmek kifejezésére reagáló robotok ösztönzik a gyerekeket a különböző érzéseik felismerésére és azokra való reagálásra. *Virtuális beszélgető partnerek*, mint az *AI-driven therapy apps*, segíthetnek a gyermekeknek az érzelmeik kezelésében, például a szorongás csökkentésében vagy a stressz kezelésében, így hozzájárulva az érzelmi intelligencia fejlődéséhez.

A *szociális készségeket fejlesztő játékok*, pl. az *Osmo Learning System*, lehetőséget adnak a gyermekeknek, hogy együttműködjenek másokkal, közösen oldjanak meg problémákat, és osztozzanak a sikerélményekben. Az ilyen típusú alkalmazások segíthetnek a gyermekeknek a csapatmunka, a türelem és a problémamegoldás képességeinek fejlesztésében. Az *interaktív csoportos játékok,* amelyek több játékost is bevonhatnak, mint például a *CoSpaces Edu*, elősegíthetik a gyerekek közötti együttműködést és kommunikációt, így erősítve szociális készségeiket.

Az MI technológiák hatása a *gyermekek kreativitására és képzelőerejére* számos szempontból pozitív lehet. Az interaktív, személyre szabott és inspiráló eszközök lehetőséget nyújtanak a gyermekek számára, hogy szabadon alkossanak, új ötleteket valósítsanak meg és kreatív módon oldjanak meg problémákat. Az olyan alkalmazások, mint a *TinkerCAD*, lehetőséget adnak a gyerekeknek, hogy önálló alkotásokat hozzanak létre, kísérletezzenek a digitális térben.

Digitális művészeti alkalmazások, mint például a *Procreate* vagy a *Pixlr*, segítik a gyermekeket, hogy kreatívan fejezzék ki magukat, digitálisan festhessenek, rajzoljanak, fotókat manipuláljanak, anélkül, hogy hagyományos eszközökre lenne szükségük. Az *MI alapú alkotó művészeti eszközök* (pl. *Musicmaker Jam* vagy *Soundation*) segíthetnek a gyermekeknek abban, hogy saját zenét komponáljanak, miközben felfedezik az érzelmi és művészeti önkifejezés új formáit, lehetővé téve számukra, hogy a művészeti jellegű alkotásokon keresztül kommunikáljanak és kifejezzék magukat, ami pozitívan befolyásolja kreativitásukat.

Az MI alapú játékok, mint az *Osmo Creative Studio* vagy a *Kano Computer Kit*, ösztönzik a gyermekeket a problémamegoldásra, a történetmesélésre és az önálló játékra. Az ilyen típusú eszközök játékos formában fejlesztik a kreatív gondolkodást és a képzelőerőt, miközben a gyerekek aktívan hozzájárulnak a játék kimeneteléhez. Az *interaktív történetmesélő eszközök*, mint a *Plotagon* vagy a *Toontastic*, lehetővé teszik a gyerekek számára, hogy saját történeteiket meséljék el, karaktereket alkossanak és új világokat építsenek, ezáltal serkentve a kreatív írást és a vizuális kifejezésmódot.

Az olyan *MI-alapú játékok*, amelyek a szabad alkotáson alapulnak, például a *Minecraft Education Edition*, lehetővé teszik a gyermekek számára, hogy szabadon építsenek, új világokat alkossanak és felfedezzék a lehetőségek határait. Az ilyen típusú alkalmazások segítenek a kreativitásuk kibontakoztatásában, miközben tanulási folyamatokat is magukba építenek. Az alkalmazások figyelembe veszik a gyermekek egyedi igényeit, és olyan kihívásokat adnak, amelyek fejlesztik a problémamegoldó készségeket, miközben ösztönzik a kreatív gondolkodást. Például a *Smart Toys* által kínált eszközök lehetőséget biztosítanak a gyerekeknek, hogy kódolási feladatokon keresztül váljanak kreatív problémamegoldókká.

Ha a gyermekek túl sok időt töltenek digitális eszközök előtt, különösen passzív módon (pl. tévénézés, videojátékok), az csökkentheti a kreatív gondolkodást, mivel nem ösztönzik őket aktív alkotásra vagy problémamegoldásra. A *túlzott képernyőidő és passzív tartalomfogyasztás* könnyen elvonhatja a figyelmet a valós világban való felfedezésről, és a gyerekek függővé válhatnak a képernyőn megjelenő előre megadott tartalmaktól. A passzív tartalomfogyasztás túlzottan redukálhatja a gyermekek szellemi és érzelmi rugalmasságát, ami hosszú távon csökkentheti találékonyságukat. A gyermekek hajlamosak lehetnek a valóság és a digitális világ közötti határvonalat nehezebben megkülönböztetni, így egyre kevésbé képesek önállóan alkotó szellemű gondolatokat generálni (Fáyné Dombi et al 2016).

Az MI-alapú alkalmazások, különösen azok, amelyek strukturált, irányított tevékenységeket kínálnak, csökkenthetik a *kötetlen játék*, az *önálló felfedezés* és a *spontán kreativitás* lehetőségét. A gyerekeknek szükségük van a kötetlen játékkal kapcsolatos tapasztalatokra,

hogy fejlődhessenek a problémamegoldó képességekben, a társas készségekben és az érzelmi intelligenciában. Ha az MI-vezérelt eszközök folyamatosan „túlpörgetik” a gyerekeket, vagy túl erőteljes külső irányítást alkalmaznak, az a gyermekek saját döntéseik és a szabad ötleteléseik folyamatait gátolhatja. Ez a túlságos strukturáltság szorongást és stresszt okozhat, mivel a gyermekek nem élhetik meg a hibázás, kísérletezés örömét (Gyarmathy 2019).

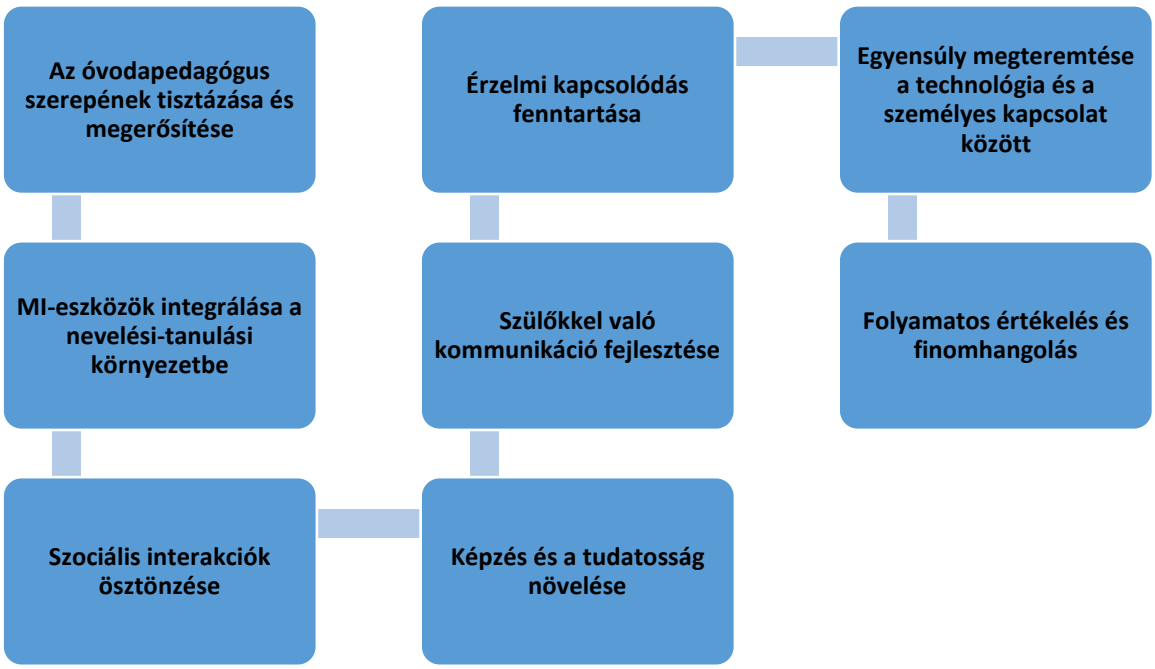
Azok az MI alkalmazások, amelyek szigorú feladatokat és elvárásokat támasztanak (például adaptív tanulási rendszerek), elősegíthetik a gyermekek teljesítményorientáltságát, miközben korlátozzák a *kreatív önkifejezés szabadságát*. Ha az MI-eszközök folyamatosan ellemőrzik és értékelik a gyermekek tevékenységét, ők hajlamosak lehetnek túlságosan megfelelni a rendszer által elvárt „tökéletes” válaszoknak, ezzel gátolva a spontán gondolkodást és az innovatív ötletek születését. Ha a gyermekek úgy érzik, hogy minden lépésüket az MI-eszközökkel összehasonlítják és értékelik, az növelheti a frusztrációt és csökkentheti a kreativitás iránti motivációjukat (Sándor-Schmidt B. é.n).

Az MI-alapú játékok elvonhatják a gyermekek figyelmét a *valódi társas interakcióktól*. Bár az MI képes támogatni a kognitív és kreatív fejlődést, a gyermekek szociális készségei, például a kommunikáció, az empátia és az együttműködés nem fejleszthetők elvárt szinten, ha a gyermekek túlzottan az MI-alapú alkalmazásokra támaszkodnak. A túlzott digitális eszközhasználat, különösen akkor, ha nem kiegészül valódi szociális tapasztalatokkal, elszigeteltséget és szociális készségek elmaradását okozhatja. A gyerekek, akik túl sok időt töltenek egyedül az MI-vezérelt játékokkal, kevésbé lesznek képesek megbirkózni a valós életben felmerülő szociális helyzetekkel, amelyek alkotó szellemű megoldásokat és empátiát igényelnek (Gyarmathy 2019).

A gyermekek *túlzott digitalizációja* elvonhatja a figyelmet a hagyományos, fizikai és szociális interakcióktól, a gyermekek mentális és pszichikai állapotára olyan negatív hatással lehetnek, mint például a fokozott fáradtság, az ingerlékenység vagy a koncentrációs problémák. A *túlzott képernyőidő* káros hatással lehet a gyermekek pszichés jólétére, növelheti az alvászavart és csökkentheti a figyelem fenntartásának képességét (Su és Yang 2022).

Az MI-eszközök használata az *óvodapedagógusok* és a *szülők szerepének* átrendeződéséhez vezethet, mivel az eszközök alkalmanként részben helyettesíthetik az óvodapedagógusokat. Ez csökkentheti a gyermekek számára a szükséges emberi kapcsolódást és a személyre szabott támogatást, mert az MI nem képes helyettesíteni az érzelmi támogatást, amely az egészséges fejlődéshez szükséges (Tóthné 2020). Az MI-eszközök óvodai környezetben történő alkalmazása esetén kulcsfontosságú, hogy azok ne helyettesítsék, hanem kiegészítsék az emberi kapcsolatokat (2. ábra), különösen az óvodapedagógusok és gyermekek közötti interakciókat.

4. ábra: Az emberi kapcsolatok kiegészítésének folyamata (Process of Complementing Human Relationships)



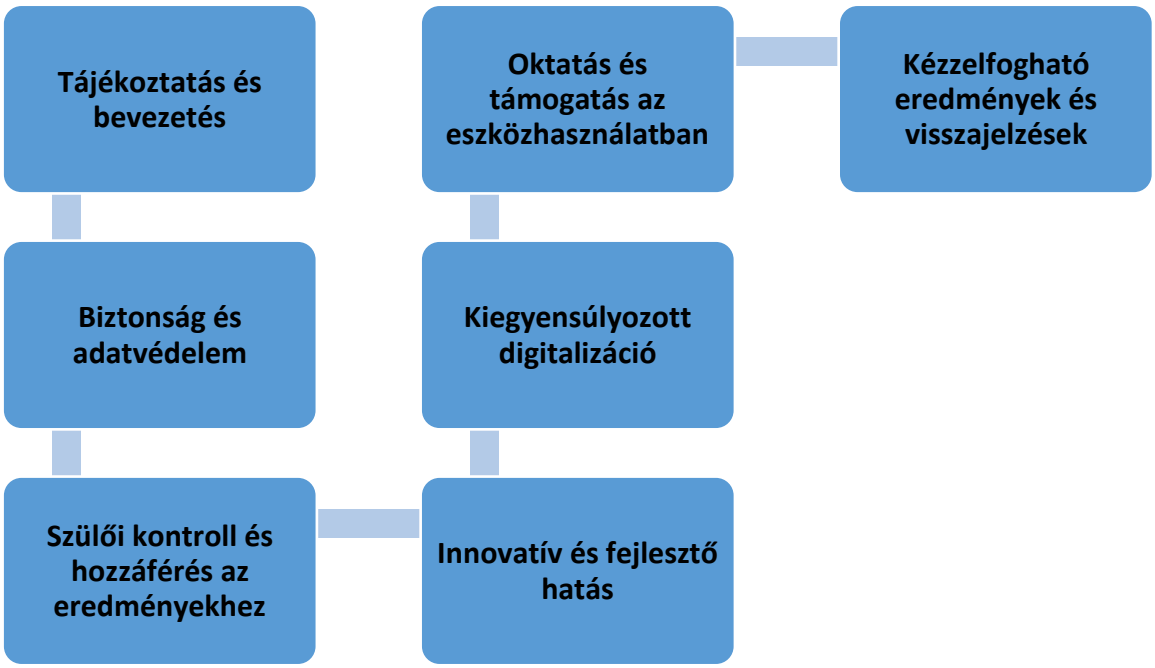
forrás: saját szerkesztés

5. A szülők nézőpontjai

A gyermekek fejlődése szoros összefüggésben áll a szülői támogatással, az otthoni környezetben kialakított nevelési elvekkel és a családi értékekkel. A szülők szerepe nem csupán a gyermekek szociális, érzelmi és kognitív fejlődésének alapja, hanem az óvodai nevelési folyamatokkal való szoros együttműködésük is elengedhetetlen a gyermekek optimális fejlődése érdekében (Plowman et al 2008, Plowman et al 2012). A fejezet célja, hogy feltárja a szülői nézőpontokat az MI alkalmazásával kapcsolatban, figyelembe véve a szülők aggályait, elvárásait és támogatását a technológiai eszközök integrálásában az óvodai nevelésbe.

A szülők elvárásai az MI használatával kapcsolatban az óvodákban változatosak lehetnek, hiszen minden család más háttérrel és igényekkel rendelkezik. Azonban általánosan elmondható, hogy felmerülhetnek az alábbi főbb elvárások (3. ábra). A folyamat megértése és követése segíthet abban, hogy az óvodák és az óvodapedagógusok hatékonyan reagáljanak a szülők elvárásaira, és elérjék, hogy az MI-eszközök alkalmazása minden szempontból a gyermekek javát szolgálja.

6. ábra: A szülők főbb elvárásainak folyamata (Process of Parents' Expectations)



forrás: saját szerkesztés

Az egyik legfontosabb aggály a szülők körében az *adatvédelem és a gyermekek biztonsága* (NETEDUCATIO 2022). Az MI-eszközök gyakran gyűjtenek érzékeny adatokat a gyermekek tevékenységeiről, tanulási előrehaladásukról és viselkedésükről, az adatok tárolása és feldolgozása nemcsak a gyermekek, hanem az óvodák és a szülők számára is komoly biztonsági kockázatokat rejthet.

A szülők gyakran aggódnak a gyermekek *digitális függősége* miatt. A túlzott digitális interakciók helyettesíthetik a valódi, személyes kapcsolatokat, amelyek kulcsfontosságúak a szociális és érzelmi fejlődéshez. Nyugtalankodhatnak amiatt, hogy az MI-alapú eszközök túlzott használata *csökkentheti az emberi kapcsolatokat* az óvodai környezetben. A szülők gyakran úgy érezhetik, hogy az eszközök túlságosan elvonják a figyelmet a óvodapedagógusoktól és a gyermekek közötti társas kapcsolatoktól, ami hosszú távon káros hatással lehet a gyermekek fejlődésére (Zhu és Zhang 2023).

A szülők másik félelme, hogy az MI-eszközök által *használt algoritmusok nem mindig átláthatóak*. Az algoritmusok döntései és ajánlásai hatással lehetnek a gyermekek oktatására és fejlődésére, de a szülők nem mindig tudják, hogyan működnek ezek a rendszerek. Az ilyen elfogultságok torzíthatják a gyermekek tanulási élményét és lehetőségeit (Sun et al. 2024).

A szülők tarthatnak attól is, hogy az MI-alapú eszközök, különösen a túlzottan interaktív vagy játékos rendszerek, nem nyújtanak elegendő *érzelmi támogatást* a gyermekek számára. Az ilyen technológiák nem képesek érzékelni a gyermekek érzelmi állapotát, és nem kínálnak olyan szintű empátiát, mint egy felnőtt. A szülők félhetnek attól, hogy a gyermekek részesülnek olyan mély, emberi interakciókban, amely a szociális és érzelmi fejlődésükhöz szükséges (Zhu és Zhang 2023).

A szülők aggályaiként merülhet fel az is, hogy nem minden gyermek fér hozzá az MI-alapú eszközökhöz otthoni környezetében. Azok a családok, akik nem rendelkeznek megfelelő technológiai eszközökkel, hátrányba kerülhetnek a technológiahasználatban. Ez *digitális egyenlőtlenséghez* vezethet, amely tovább súlyosbíthatja a szociális különbségeket (Sun et al. 2024).

A szülők előítéletei az MI-alapú eszközök alkalmazásával kapcsolatban jogosak, és figyelembe kell venni őket az intézmények és az óvodapedagógusok részéről. Az adatvédelem, a gyermekek digitális jóléte, az emberi kapcsolatok megőrzése és a technológiai egyenlőtlenség mind olyan kérdések, amelyek kezelése elengedhetetlen ahhoz, hogy az MI valóban segíthesse a gyermekek fejlődését anélkül, hogy kárt okozna.

7. Etikai és adatvédelmi kérdések

Mivel az óvodás gyermekek még fejlődésük korai szakaszában járnak, különösen fontos, hogy a technológiai megoldások ne sértsék meg az őket megillető jogokat és méltóságukat. A gyermekek személyes adatai és digitális aktivitásai érzékeny információk, amelyek védelmét el kell látni, és meg kell oldani, hogy az MI rendszerek ne csak a hatékonyságot, hanem az etikai normákat is figyelembe vegyék (Zóka 2019). A fejezet célja, hogy bemutassa azokat a legfontosabb etikai és adatvédelmi kérdéseket, amelyek az MI-eszközök alkalmazásakor felmerülhetnek az óvodai nevelés során.

Az MI-alapú rendszerek az óvodákban különféle adatokat gyűjthetnek, amelyek segítségével a gyermekek tanulási folyamatát és fejlődését személyre szabott módon támogathatják (European Commission 2022). Az adatgyűjtés során legtöbbször az alábbi típusú információk jöhetnek létre: *tanulási adatok (fejlődési nyomon követés, interakciós adatok), valamint magatartási és érzelmi adatok (érzelemfelismerés, szociális interakciók).* Alapvető, hogy a gyermekek személyes adatait csak a *szükséges mértékben* gyűjtsék és dolgozzák fel. Az adatokat csak a tanítási és nevelési célokra szabad felhasználni, és felügyelni kell, azok ne kerüljenek illetéktelen kezekbe. Az adatok *anonimitása és titkosítása* elengedhetetlen, hogy megvédjék a gyermekek személyes információit (Poth 2024).

A szülőknek joguk van ahhoz, hogy tájékoztatást kapjanak a gyermekeik adatainak gyűjtéséről és felhasználásáról. A *beleegyezésük nélkül nem szabad adatokat gyűjteni*. Ezért az intézményeknek világos *adatkezelési irányelveket* kell alkalmazniuk, és biztosítaniuk kell, hogy minden alkalmazott és pedagógus tisztában legyen az adatvédelmi szabályokkal is (New America 2024).

A gyermekek adatainak védelme érdekében *rendszeres auditálás* szükséges, hogy az adatok gyűjtése és feldolgozása megfeleljen az előírásoknak. Az intézményeknek meg kell teremteniük az átláthatóság feltételeit a szülők és pedagógusok számára, így minden érintett fél tudomással bírhat az adatkezelési gyakorlatokban alkalmazott módszerekről és azok hatékonyságáról. Az MI rendszerek által generált *esetleges torzítások vagy hibák kezelése* különösen fontos, mivel ezek befolyásolhatják a gyermekek fejlődésére tett hatásokat, valamint a pedagógiai döntéseket is (New America 2024).

Az etikai irányelvek betartása az MI alkalmazása során kulcsfontosságú, hogy a gyermekek, az óvodapedagógusok és a szülők biztonságban érezzék magukat az eszközök használata során. Az átláthatóság, a gyermekek jogainak tiszteletben tartása, a tisztességes bánásmód, és a folyamatos etikai felülvizsgálat eredménye, hogy az MI hozzájáruljon a gyermekek

fejlődéséhez, miközben védelmet nyújt a potenciális káros hatásokkal szemben (MIT Responsible AI for Social Empowerment and Education 2024).

8. Ajánlások kutatóknak és fejlesztőknek

Az MI integrálása az óvodapedagógiába számos új kutatási lehetőséget kínál társadalomtudományi és módszertani szempontból (Mező 2019). Az alábbiakban néhány elméleti és empirikus kutatási témajavaslatot sorolok fel.

Az MI etikai kérdései az óvodai nevelésben: vizsgálhatja az MI alkalmazásának etikai vonatkozásait az óvodai környezetben, beleértve az adatvédelem, a gyermekek jogai és az óvodapedagógusok szerepének változását.

Az MI társadalmi fogadtatása az óvodapedagógiában: elemezheti, hogy a társadalom és a pedagógusok miként reagálnak az MI bevezetésére az óvodai nevelésben, és milyen félelmek vagy elvárások kapcsolódnak hozzá.

Az óvodapedagógusok felkészítése az MI alkalmazására az óvodai nevelésben: feltárhatja, hogy milyen képzési programok és módszertani támogatás szükséges ahhoz, hogy az óvodapedagógusok hatékonyan integrálják az MI-t a mindennapi nevelési gyakorlatba.

Az MI szerepe a gyermekek fejlődésének nyomon követésében: vizsgálhatja, hogy az MI alapú rendszerek miként segíthetik a gyermekek kognitív, szociális és érzelmi fejlődésének monitorozását és értékelését az óvodai környezetben.

Ezek a kutatási témák hozzájárulhatnak az MI alkalmazásának mélyebb megértéséhez és hatékony integrációjához az óvodapedagógiai gyakorlatban.

Az MI platformokat fejlesztő Ed-Tech cégek számos módon támogathatják az MI hasznosságának fejlődését az óvodákban.

Személyre szabott tanulási platformok fejlesztése: az MI lehetővé teszi a tanulás individualizálását, alkalmazkodva az egyes gyermekek tanulási stílusához és üteméhez. Az adaptív tanulási platformok segítségével az óvodapedagógusok hatékonyabban támogathatják a gyermekek fejlődését.

Automatizált értékelési rendszerek kialakítása: az értékelő eszközök gyors és pontos visszajelzést nyújtanak a gyermekek teljesítményéről, lehetővé téve az óvodapedagógusok számára a fejlődési területek azonosítását és a szükséges beavatkozások megtervezését.

Virtuális asszisztensek és chatbotok integrálása: a virtuális asszisztensek segíthetnek az óvodapedagógusoknak az adminisztratív feladatok kezelésében, valamint támogatást nyújthatnak a gyermekeknek a tanulási folyamat során.
Képzési programok és továbbképzések biztosítása: az Ed-Tech cégek szerepet vállalhatnak az óvodapedagógusok digitális kompetenciáinak fejlesztésében, képzési programok és workshopok szervezésével, amelyek bemutatják az MI eszközök hatékony alkalmazását az óvodai környezetben.
Biztonságos és etikus MI alkalmazások fejlesztése: fontos, hogy a cégek olyan MI megoldásokat kínáljanak, amelyek megfelelnek az adatvédelem és az etikai normák követelményeinek, biztosítva a gyermekek biztonságát és a szülők bizalmát.
Ezekkel a lépésekkel az Ed-Tech cégek hozzájárulhatnak az óvodai nevelés minőségének javításához, támogatva az óvodapedagógusok erőfeszítéseit és elősegítve a gyermekek holisztikus fejlődését.

Összefoglaló

A tanulmány átfogó képet adott az MI alkalmazásának lehetőségeiről és kihívásairól az óvodai nevelésben, figyelmet fordítva a pedagógusok különböző szemléleteire, különösen a tec*hnológiaszkeptikus* és a *technológiaoptimista* megközelítésekre. Az óvodai nevelésben az MI szerepe nem csupán technológiai újítás, hanem pedagógiai és etikai dilemmákat is felvet, amelyek eltérő reakciókat válthatnak ki a különböző szemléletű óvodapedagógusok körében.
Az MI bevezetése az óvodai környezetbe mindkét szemlélet számára számos előnnyel és kihívással jár. A technológiaszkeptikus pedagógusok számára a legnagyobb kihívás az lehet, hogy az új technológiák hogyan illeszkednek a gyermekek szociális, érzelmi és kognitív fejlődéséhez, miközben fenntartják a személyes kapcsolatok és az emberi nevelés hagyományos szerepét. A technológiaoptimista pedagógusok pedig az innovációk előnyeit keresik, de fontos számukra a technológiai eszközök felelős és etikusan alkalmazott integrálása is.
A jövőben elengedhetetlen, hogy az MI alkalmazása az óvodai nevelésben mindkét szemlélet szem előtt tartásával valósuljon meg. Az egyik legfontosabb cél, hogy az új technológiai eszközöket úgy integráljuk, hogy azok kiegészítsék, és ne helyettesítsék az óvodapedagógusok személyes szerepét és a gyermekek számára biztosított emberi kapcsolatokat. Ezen kívül az óvodapedagógusok számára kiemelten fontos, hogy megfelelő képzéseken vegyenek részt az MI-eszközök használatával kapcsolatosan, így biztosítva a technológia hatékony és etikusan alkalmazott bevezetését.

Az MI alkalmazása az óvodákban valóban felvethet bizonyos aggályokat, különösen azok számára, akik a gyermekek fejlődésére és jólétére helyezik a hangsúlyt. Azonban fontos megnyugtatni mind a szülőket, mind az óvodapedagógusokat, hogy az MI-nak nem célja a hagyományos pedagógiai módszerek és emberi kapcsolatok helyettesítése, hanem azok támogatása és kiegészítése.

Köszönetnyilvánítás

Egy új kutatási terület felfedezését és kíváncsiságom, hogy megértsem azt, valamint ösztönzését jelen tanulmány megírásához köszönöm feleségemnek, Diának, az ÓVODAPEDAGÓGUSNAK, aki rávezetett arra, hogy elfogadjam: az óvónők tudásvágya, szorgalma, lelkesedése és elkötelezettsége hegyeket mozgathat meg. A munkaidőt és az óvodai tükröt köszönöm kislányunknak, Lizának, az ÓVODÁSNAK, akinek rám szegezett csillogó okos tekintete folyamatosan elgondolkodtat a jövő intelligenciáit illetően.
Apaként aggódom, kutatóként reménykedem.

Irodalomjegyzék

Baditzné Pálvölgyi K. és Jakab L. 2023, A mesterséges intelligencia alkalmazási lehetőségei a tanításban: kezdeti tapasztalatok és jó gyakorlatok. DOI: 10.21030/anyp.2023.3.3 url: https://www.anyanyelv-pedagogia.hu/cikkek.php?id=1024 (utolsó letöltés: 2024.12.15.)

Bessenyei I. 2023, Mire jó a mesterséges intelligencia? Taní-tani Online. A szabad pedagógiai gondolkodás fóruma. 2023. május 6. url: https://www.tani-tani.info/mire_jo_a_mesterseges_intelligencia (utolsó letöltés: 2024.12.15.)

Beták N. 2019, A robotika, mint a 21. századi készségek fejlesztésének eszköze. (szerk. Kaposi J. és Szőke-Milinte E.: Pázmány Péter Katolikus Egyetem Budapest, Pedagógiai változások – a változás pedagógiája. pp. 491-498. ISBN 978-615-5224-86-7 url:https://publikacio.ppke.hu/id/eprint/1264/1/Pedagogiai_valtozasok_1.pdf (utolsó letöltés: 2024.12.15.)

Beták N. és Szabó-Váradi É. 2023, A digitális óvodapedagógus és digitális eszközhasználat az óvodai csoportszobákban. OXIPO: Interdiszciplinárs e- folyóir, 5 (1). pp. 25-41. ISSN 2676-8771 url: https://real.mtak.hu/164120/1/OxIPO_2023_1_025_Betak_Szabo-Varadi.pdf (utolsó letöltés: 2024.12.15.)

Blackwell, C. K., Lauricella, A. R., és Wartella, E. 2014, Factors influencing digital technology use in early childhood education. Computers & Education, 77, 82–90. url: https://d1wqtxts1xzle7.cloudfront.net/101791210/Blackwell.Lauricella.Wartella.2014.Factors-influencing-digital-tech-use-in-early-education-libre.pdf?1683131479=&response-content-disposition=inline%3B+filename%3DFactors_influencing_digital_technology_u.pdf&Expires=1737366833&Signature=DiEyXWDUDbQBPJW~BipVKFouA1WW3tF5cc~hVPP6JjnnXvCYHwsm6JQOMUB7p~UpNitabx4sJv70jFQLZ2GXaCm5BK3pfQ6Nj-QLf1A3iPpkMrdThTmZyw5asXEJLU9Y0B4BRYt1oBujkkKMeTZHiwncxuW-VWtiapuBK6IfLllGy6SdSMnabe4Rkgr1KGV72lGT9ipylVJlkc1~QqVIw4vXPeVWW2-lilhtsbxCBTrbQvx~WjGVFEQvRD6En0LHq~phJ1zTDQpHcUI6r6DIJ2z1wrLs-0Axb9~2pge7ut8oOVyERbvoy6bh8mWYZQgrRSyJLQDczUHx90ic5rLPnQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA (utolsó letöltés: 2024.12.15.)

Dietz F. 2020, A mesterséges intelligencia az oktatásban: kihívások és lehetőségek. Scientia et Securitas 1: pp. 54–63. url: https://akjournals.com/view/journals/112/1/1/article-p54.xml https://doi.org/10.1556/112.2020.00009 (utolsó letöltés: 2024.12.15.)

Ertmer, P. A., és Ottenbreit-Leftwich, A. T. 2010, Teacher Technology Change: How Knowledge, Confidence, Beliefs, and Culture Intersect. Journal of Research on Technology in Education, 42(3), pp. 255–284. url: https://files.eric.ed.gov/fulltext/EJ882506.pdf (utolsó letöltés: 2024.12.15.)

European Commission. (2022). Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators. url: https://education.ec.europa.eu/news/ethical-guidelines-on-the-use-of-artificial-intelligence-and-data-in-teaching-and-learning-for-educators (utolsó letöltés: 2024.12.15.)

Poth, R. D. (2024). AI and the Law: What Educators Need to Know. url: https://www.edutopia.org/article/laws-ai-education/?utm_source=chatgpt.com (utolsó letöltés: 2024.12.15.)

Fáyné Dombi A:, Hódi Á. és Kiss R. 2016, IKT az óvodában: kihívások és lehetőségek. Magyar Pedagógia. 116. évf. 1. szám 91-117. DOI: 10.17670/MPed.2016.1.91

Gyarmathy É. 2011, A digitális kor és a sajátos nevelési igényű tehetség. url: http://www.diszlexia.hu/GyarmathyFordulopont.pdf. (utolsó letöltés: 2024.12.15.)

Gyarmathy É. 2019, Stabilitás a mobilitásban. In: Fehér Á. és Megyeriné Runyó A. (szerk.): A digitális világ hatása a gyermekekre A BRUNSZVIK TERÉZ SZAKMAI NAPOK keretében szervezett III. Nemzetközi Kisgyermek-nevelési Konferencia kötete Vác pp. 114-121. url:

https://real.mtak.hu/92172/1/A%20digit%C3%A1lis%20vil%C3%A1g%20hat%C3%A1sa%20a%20gyermekekre.pdf (utolsó letöltés: 2024.12.15.)

Honghu, Y., Ting, L., & Gongjin, L. 2023, The Key Artificial Intelligence Technologies in Early Childhood Education: A Review. Artificial Intelligence Review. url: https://arxiv.org/pdf/2401.05403 (utolsó letöltés: 2024.12.15.)

Lee, C., és Xiong, J. 2024, From Keyboard to Chatbot: An AI-powered Integration Platform with Large-Language Models for Teaching Computational Thinking to Young Children. arXiv preprint arXiv:2405.00750 url: https://arxiv.org/pdf/2405.00750 (utolsó letöltés: 2024.12.15.)

Lindahl, M. G., és Folkesson, A. M. 2012, ICT in preschool: Friend or foe? The significance of norms in a changing practice. International Journal of Early Years Education, 20(4), 422–436.

Mező F. 2019, Interdiszciplináris kapcsolódási lehetőségek a mesterséges intelligenciára irányuló cél-, eszköz- és hatásorientált kutatáshoz. Mesterséges intelligencia – interdiszciplináris folyóirat, I. évf. 2019/1. szám. 9–29. doi: 10.35406/MI.2019.1.9 url: https://real.mtak.hu/102842/1/MI_2019_1_009_Mezo_Mezo.pdf (utolsó letöltés: 2024.12.15.)

MIT Responsible AI for Social Empowerment and Education (RAISE). 2024, Securing Student Data in the Age of Generative AI. url: https://ospi.k12.wa.us/sites/default/files/2024-06/ai-guidance_ethics.pdf?utm_source=chatgpt.com (utolsó letöltés: 2024.12.15.)

NETEDUCATIO 2022, A digitális biztonság az óvodában. Letöltés: 2023. 02. 01. url: https://neteducatio.hu/digitalisbiztonsag/(utolsó letöltés: 2024.12.08)

Neugnot-Cerioli, M., és Laurenty, O. M. 2024, The Future of Child Development in the AI Era: Cross-Disciplinary Perspectives Between AI and Child Development Experts. arXiv preprint arXiv:2405.19275. url: https://arxiv.org/pdf/2405.19275 (utolsó letöltés: 2024.12.15.)

New America. (2024). Artificial Intelligence in Schools: Privacy and Security Considerations. url: https://www.newamerica.org/oti/blog/artificial-intelligence-in-schools-privacy-and-security-considerations/ (utolsó letöltés: 2024.12.15.)

Palaiologou, I. 2016, Teachers' dispositions towards the role of digital devices in play-based pedagogy in early childhood education. Early Years, 36(3), pp. 305–321.

Plowman L., Stevenson O. és Stephen Ch. 2012, Preschool children's learning with technology at home. Computers & Education. 59(1). pp. 30–37. url: https://www.research.ed.ac.uk/files/10854135/Plowman_et_al_preschool_children_and_technology_Computers_Education.pdf (utolsó letöltés: 2024.12.15.)

Plowman, L, McPake J. és Stephen Ch. 2008, Just picking it up? Young children learning with technology at home. Cambridge Journal of Education 2008. 38(3). pp. 303–319. url: https://www.research.ed.ac.uk/files/10854734/Plowman_et_al_2008_Just_picking_it_up_.pdf (utolsó letöltés: 2024.12.15.)

Sándor-Schmidt B. é.n., Új utak az óvodapedagógiában. A többszörös intelligenciák koncepció elméleti és gyakorlati keretei In: Karlovitz János Tibor (szerk.): Tanulás és fejlődés, p 75-80. ISBN 978-80-89691-31-9 url:https://www.academia.edu/22540485/%C3%9Aj_utak_az_%C3%B3vodapedag%C3%B3gi%C3%A1ban_A_t%C3%B6bbsz%C3%B6r%C3%B6s_intelligenci%C3%A1k_koncepci%C3%B3_elm%C3%A9leti_%C3%A9s_gyakorlati_keretei (utolsó letöltés: 2024.12.15.)

Su, J. és Yang, W. 2022, Artificial intelligence in early childhood education: A scoping review. In: Computers and Education: Artificial Intelligence 3. url: https://www.sciencedirect.com/science/article/pii/S2666920X22000042 (utolsó letöltés: 2024.12.15.)

Sun, Y., Liu, J., Yao, B., Chen, J., Wang, D., Ma, X., Lu, Y., Xu, Y., & He, L. (2024). Exploring Parents' Needs for Children-Centered AI to Support Preschoolers' Storytelling and Reading Activities. url: https://arxiv.org/abs/2401.13804?utm_source=chatgpt.com (utolsó letöltés: 2024.12.15.)

Tóthné Parázsló L 2020, Digitális intelligencia – készségek a sikeres digitális élethez. In: A kultúraváltás hatása az egyéni kompetenciákra: a digitális kompetencia modelljei. Eszterházy Károly Egyetem Líceum Kiadó, Eger, pp. 106-118. ISBN 978-963-496-160-4 url: https://real.mtak.hu/120274/1/106_118_T%C3%B3thn%C3%A9.pdf (utolsó letöltés: 2024.12.15.)

Tölgyes L. A. 2023, A mesterséges intelligencia integrálása az oktatásba. ICTGlobal. 2023. május 23. url: https://ictglobal.hu/iparagi-megoldasok/a-mesterseges-intelligencia-integralasa-az-oktatasba/ (utolsó letöltés: 2024.12.15.)

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. Computers and Education: Artificial Intelligence, 4, 100124. url: https://www.sciencedirect.com/science/article/pii/S2666920X23000036 (utolsó letöltés: 2024.12.15.)

Zhu, K., & Zhang, M. (2023). Kindergarten Parents' Perceptions of the Use of AI Technologies and AI Literacy Education: Positive Views but Practical Concerns. url:

https://www.researchgate.net/publication/379954587_Kindergarten_parents%27_perceptions_of_the_use_of_AI_technologies_and_AI_literacy_education_Positive_views_but_practical_concerns?utm_source=chatgpt.com (utolsó letöltés: 2024.12.15.)

Zóka K. 2019, A világ befogadásának változásai kisgyermekkorban. In: Fehér Á. és Megyeriné Runyó A. (szerk.): A digitális világ hatása a gyermekekre A BRUNSZVIK TERÉZ SZAKMAI NAPOK keretében szervezett III. Nemzetközi Kisgyermek-nevelési Konferencia kötete Vác. pp. 159-173. url: https://real.mtak.hu/92172/1/A%20digit%C3%A1lis%20vil%C3%A1g%20hat%C3%A1sa%20a%20gyermekekre.pdf (utolsó letöltés: 2024.12.15.)

Oktatási chatbotokkal kapcsolatos tanulói bizalmat meghatározó tényezők

Factors determining student trust in educational chatbots

Ollé János[4]

Absztrakt

A mesterséges intelligenciára épülő oktatási chatbotok egyre gyakrabban jelennek meg a digitális tanulási környezetekben, különösen a felsőoktatásban és a felnőttképzésben. A velük szembeni tanulói bizalom azonban alapvetően befolyásolja, hogy a hallgatók milyen mértékben fogadják el és használják ezeket az eszközöket tanulástámogatásra. A tanulmány egy feltáró jellegű pilotvizsgálat eredményeit mutatja be, amely alapján egy kérdőíves adatgyűjtés alapján megállapítható, hogy a bizalom feltételekhez kötött: a válaszadók számára kiemelten fontos a chatbot válaszainak szakmai pontossága, a működés átláthatósága, valamint az emberi jelenlét – különösen az oktatói kontroll – biztosítása. A résztvevők a chatbotot elsősorban kiegészítő, nem pedig helyettesítő tanulási eszközként fogadják el, leginkább elméleti tudás elsajátításában és tanulásszervezési támogatásban használnák. A tanulmány a tanulói bizalom főbb dimenzióit mutatja be, és javaslatokat fogalmaz meg az oktatási célú chatbotok tudatos és hiteles fejlesztéséhez és alkalmazásához.

Kulcsfogalmak: oktatási chatbot, tanulói bizalom, mesterséges intelligencia, felsőoktatás, felnőttképzés, tanulástámogatás.

Abstract

Educational chatbots powered by artificial intelligence are increasingly present in digital learning environments, particularly in higher and adult education. However, learners' trust in these systems fundamentally influences the extent to which they accept and use them for learning support. This study presents the results of an exploratory pilot survey, which reveals that trust in educational chatbots is conditional. Based on questionnaire data, respondents

[4] PhD. habil. Pannon Egyetem, Humántudományi Kar, Digitális Módszertani Intézet, tudományos főmunkatárs, olle.janos@htk.uni-pannon.hu

emphasized the importance of the chatbot's professional accuracy, operational transparency, and the assurance of human presence—especially instructional oversight. Participants primarily view chatbots as supplementary rather than substitutive tools in learning, and would mostly use them for acquiring theoretical knowledge and receiving organizational support. The study identifies key dimensions of learner trust and offers recommendations for the conscious and credible development and application of educational chatbots.
Keywords: educational chatbot, learner trust, artificial intelligence, higher education, adult learning, learning support.

Bevezető

A mesterséges intelligencián alapuló oktatási chatbotok megjelenése új dimenziót nyitott az online tanulási környezetekben. Az azonnali üzenetváltásra és természetes nyelvi párbeszédre képes rendszerek személyre szabott támogatást nyújtanak, azonnali visszajelzést adnak, és fokozhatják az interaktivitást a tanulás során. Különösen a ChatGPT 2022. végi megjelenését követően tapasztalható robbanásszerű érdeklődés az oktatási célú chatbotok iránt, amelyet a felsőoktatásban és a felnőttképzésben is egyre szélesebb körű kísérletezés követett. Az oktatási chatbotok sikeressége nem csak technológiai fejlettségükön vagy funkcionalitásukon múlik. A tanulók részéről megnyilvánuló bizalom kulcsfontosságú tényező: ha a hallgatók nem bíznak a chatbotban, nem veszik igénybe annak segítségét, nem használják fel a tanulásban, vagy akár el is utasítják az így biztosított tanulástámogatási lehetőséget. A bizalom mértéke nemcsak az elfogadásra van hatással, hanem visszahat a tanulási eredményekre is, befolyásolhatja a teljesítményt, a tanulási környezettel kapcsolatos elégedettséget, vagy a tanulási élményt.

A bizalom ebben a kontextusban is többdimenziós jelenség. Egyrészt technológiai jellegű (bízik-e a tanuló a válaszok pontosságában és relevanciájában), másrészt pszichológiai (képes-e komfortosan és biztonságérzettel használni a rendszert), harmadrészt pedig az oktatási kimenetekhez kapcsolódik (érzi-e, hogy a chatbot támogatja az előrehaladását). A tanulói bizalom egy komplex, multifaktoriális konstrukció, amely technológiai, pszichológiai és pedagógiai elemek kölcsönhatásából épül fel (Schei et al. 2024).

Ez a tanulmány egy feltáró jellegű empirikus pilot kutatás keretében vizsgálja és elemzi, hogy felsőoktatásban, posztgraduális képzésen tanuló felnőtt hallgatókból álló csoport hogyan viszonyul a chatbotokkal támogatott oktatáshoz és tanulási környezethez. Jelen tanulmány célja, hogy empirikus és elméleti alapokon feltárja a tanulói bizalmat meghatározó főbb

tényezőket, hozzájárulva ezzel az oktatási technológiák megalapozottabb alkalmazásához, valamint a bizalomépítés stratégiáinak tudományos megalapozásához.

1.A chatbotokkal kapcsolatos bizalom szakirodalmi háttere

Az oktatási chatbotok tanulási környezetekbe történő integrálása egy olyan transzformatív eszközt vezetett be, amely javíthatja az oktatási eredményeket, személyre szabott támogatást nyújthat és fokozhatja az elköteleződést. Azonban e chatbotok sikeressége nagymértékben a tanulók bizalmától függ, amelyet különféle tényezők befolyásolnak.

1.1.Technológiai pontosság és bizalom

A technológiai pontosság a chatbotok által szolgáltatott információk megbízhatóságát, helyességét és következetességét jelenti. Ez a tényező jelentős mértékben befolyásolja a tanulók bizalmát, mivel a pontatlan vagy megbízhatatlan válaszok szkepticizmust és a bizalom csökkenését válthatják ki. A tanulók hasznos eszköznek tartják a chatbotokat olyan tanulmányi feladatokhoz, mint az írás, programozás vagy problémamegoldás, de gyakoriak az aggodalmak a válaszok pontosságával és megbízhatóságával kapcsolatban (Schei et al. 2024; Bettayeb et al. 2024). A válaszok becsült pontosságának átláthatóvá tétele növelheti a felhasználói bizalmat, de az átláthatóság önmagában nem mindig elegendő a bizalom kialakításához, mivel a felhasználók eltérően értelmezhetik az átláthatósági jeleket. Ez hangsúlyozza annak fontosságát, hogy a chatbotokat egyértelmű és intuitív átláthatósági funkciókkal tervezzék meg a bizalom növelése érdekében. A szakértői pozicionálás – azaz a chatbot bemutatása egy adott szakterülethez kapcsolódó eszközként – szintén hatással lehet a bizalomra. Ha a chatbotot egy adott terület (pl. építőipari feladatok) szakértőjeként keretezik, az növeli a felhasználók észlelt bizalmát és a chatbot iránti támaszkodást (Wienrich és Obremski 2024). Ez arra utal, hogy a chatbot képességeinek és korlátainak egyértelmű kommunikálása növelheti annak hitelességét és megbízhatóságát.

1.2. Pszichológiai észlelés és bizalom

A pszichológiai észlelés arra utal, hogy a tanulók hogyan tapasztalják meg és értelmezik a chatbotokkal való interakciójukat. Ez magában foglalja az érzelmi kötődéseket, az önhatékonyságot és a személyiségjegyeket, amelyek jelentős szerepet játszanak a bizalom kialakulásában. A chatbotoknak az a képessége, hogy érzelmi kapcsolatot alakítsanak ki a tanulókkal, jelentősen befolyásolhatja a bizalmat. Az érzelmi kifejezés és az empátia beépítése a chatbot-interakciókba növelheti a tanulók érzelmi biztonságérzetét és bizalmát A ChatGPT

használatát vizsgáló kutatások is hangsúlyozzák az érzelmi támogatás fontosságát, hogy a tanulók a chatbotot megbízható és hasznos eszköznek érzékeljék (Bettayeb et al. 2024).
Az önhatékonyság – az a meggyőződés, hogy a tanuló képes sikeresen elvégezni egy feladatot – hatással van a chatbotokkal szembeni bizalomra. Egy kutatás azt találta, hogy azok a tanulók, akik magasabb önhatékonysággal rendelkeznek, nagyobb valószínűséggel bíznak meg a chatbotokban és használják őket tanulmányi feladatokhoz. A chatbotok növelhetik a bizalmat azáltal, hogy erősítik a tanulók kontroll- és kompetenciaérzetét. A személyiségjegyek, például az érzelmi instabilitás és az újdonságokra való nyitottság szintén hatással lehetnek a chatbotokkal szembeni bizalomra. Az alacsonyabb neuroticizmussal rendelkeznek (azaz kevésbé hajlamosak a szorongásra és stresszre), nagyobb valószínűséggel fogadják el és bíznak meg a chatbotokban (Raiche et al. 2023).

2. Tanulási eredményesség és hatékonyság hatása a bizalomra

A tanulmányi teljesítmény, a tanulással való elégedettség és a viselkedési szándékok – szintén összefüggésben állnak a tanulók chatbotokba vetett bizalmával. A bizalom és az oktatási eredmények közötti kapcsolat kétirányú: a pozitív kimenetek erősíthetik a bizalmat, míg a bizalom elősegítheti a jobb tanulási eredményeket. A chatbotokba vetett bizalom pozitív korrelációt mutat a tanulmányi teljesítménnyel. Azok a tanulók, akik bíztak a chatbot megbízhatóságában és pontosságában, magasabb osztályzatokat és jobb tananyagmegértést tapasztaltak (Uppal és Hajian 2024). Hasonló eredmények szerint a chatbotok képesek javítani a tanulási eredményeket a személyre szabott támogatás és az azonnali visszajelzés alapján (Bettayeb et al. 2024).
A tanulói bizalom az oktatási chatbotokkal kapcsolatban egy összetett, többtényezős konstrukció, amelyet technológiai pontosság, pszichológiai észlelés és oktatási kimenetek befolyásolnak. A technológiai pontosság – beleértve a válaszok megbízhatóságát és az átláthatóságot – alapvető a bizalom kiépítéséhez. A pszichológiai tényezők, mint az érzelmi kapcsolódás, önhatékonyság és személyiségvonások, szintén kulcsszerepet játszanak. Az oktatási eredmények – ideértve a teljesítményt, a viselkedési szándékokat és az elégedettséget – egyszerre befolyásolják és megerősítik a bizalmat. Ahhoz, hogy az oktatási chatbotok valóban betölthessék tanulástámogató szerepüket, a fejlesztőknek és oktatóknak olyan chatbotokat kell tervezniük, amelyek nemcsak pontosak és átláthatóak, hanem érzelmileg is támogatóak, és illeszkednek a tanulók igényeihez és preferenciáihoz. E tényezők figyelembevételével az

oktatási chatbotok megbízható eszközökké válhatnak, amelyek gazdagítják a tanulási élményeket és javítják a tulmányi eredményeket.

Módszertan

A tanulmány empirikus alapját egy kérdőíves felmérés képezte, amelyet a digitális tanulási környezetekkel aktívan foglalkozó felnőtt tanulói csoport körében végeztünk. A kutatás célcsoportját az e-learning fejlesztő szakirányú továbbképzés hallgatói és végzett hallgatói alkották (N=45 fő), akik a felsőoktatás posztgraduális képzési rendszerében vesznek (illetve vettek) részt. A minta nem reprezentatív, és speciális szakterületi kontextusban értelmezendő; ennek megfelelően a vizsgálat feltáró, pilot jellegű. A kutatásnak ezt a jellemzőjét figyelembe kell vennünk a kutatási eredmények értelmezésnél és az eredmények korlátainál is. Fontos megjegyezni, hogy a célcsoportra a képzésük témája alapján nem jellemzők az alapvető technikai ellenérzések, illetve számukra az AI-alapú tanulási technológiák, különösen az oktatási chatbotok témája nem volt ismeretlen. Ezáltal kizárhatóak azok a tipikus torzító hatások (mint például a technológiai szorongás vagy az információhiány és tapasztalatlanság), amelyek egy tágabban értelemezett vizsgált minta esetén gyakran jelentkeznek. Egy átfogó kérdőív részeként a chatbotokkal támogatott tanulási helyzetekhez kapcsolódó 9 kérdés közül 3 Likert-skálás item rögzítette a válaszadók attitűdjeit és vélekedéseit, illetve 6 nyílt végű, kvalitatív jellegű kérdés is szerepelt, amelyek a hallgatók szubjektív élményeinek, értelmezéseinek és elvárásainak feltárását célozták. A nyílt kérdések lehetőséget adtak a tanulói bizalom árnyaltabb, kontextusba ágyazott megközelítésére és feltárására is. A nyílt végű válaszokat kvalitatív tartalomelemzés segítségével dolgoztuk fel. A válaszok elsődleges, nyílt kódolása után tematikus kategóriákat alakítottunk ki, amelyek a bizalom különböző dimenziói (pl. technológiai megbízhatóság, oktatói kontroll, érzelmi viszonyulás) szerint rendeződtek. Az elemzés során induktív kategóriaképzést alkalmaztunk, majd a kapott témákat kvantifikáltuk a válaszok előfordulási gyakorisága alapján, hogy a kvalitatív eredmények strukturáltan beépíthetők legyenek a kvantitatív eredmények értelmezésébe. Az adatgyűjtés online formában történt: a kérdőív egy webalapú űrlapon keresztül volt elérhető, és a kitöltés önkéntes, anonim módon zajlott. A válaszadásra rendelkezésre álló időszak két hét volt. Az így gyűjtött adatok lehetőséget biztosítottak arra, hogy az elméleti háttérben azonosított bizalmi dimenziókat a valós oktatási gyakorlat szempontjából is értelmezzük.

Eredmények

A tanulói bizalom típusai: eltérő bizalmi attitűdök megoszlása

A kérdőíves vizsgálat egyik kulcskérdése arra irányult, hogy a válaszadók milyen mértékben bíznának meg egy olyan képzés során alkalmazott oktatási chatbotban, amely tanulástámogató szerepet tölt be. A nyílt kérdésre adott válaszok alapján négy jól elkülöníthető bizalmi attitűdöt azonosítottunk, amelyek kvalitatív értelmezéssel, majd kvantitatív osztályozással kerültek besorolásra. A válaszadók többsége (51%) feltételes bizalmat fogalmazott meg az oktatási chatbotokkal kapcsolatban. Ezek a kitöltők alapvetően nem zárkóznak el a chatbot alkalmazásától, ám bizalmuk előfeltételekhez kötött. Kiemelték például, hogy a chatbotnak validált, szakszerűen előkészített tananyagra kell épülnie, működését emberi kontrollnak kell kísérnie, valamint elvárják a célzott oktatási feladathoz való illeszkedést is. A hallgatók véleménye a technológiai pontosság és a transzparencia kulcsszerepét emeli ki a bizalom kialakulásában.

A válaszadók 20%-a bizalmatlanságot vagy szkepszist fejezett ki. Az ebbe a csoportba tartozók elsősorban a chatbot megbízhatóságát, hibázási lehetőségét és az emberi kapcsolódás hiányát említették aggályosnak. Többen egyenesen veszélyesnek tartják a chatbotok oktatási célú alkalmazását, mivel szerintük ezek nem alkalmasak a mélyebb pedagógiai támogatásra. Ez a csoport a pszichológiai komfortérzet és a kontroll hiányát tekinti a bizalom legfőbb akadályának.

Szintén 20%-ot képviseltek azok a válaszadók, akik elvi bizalomról vagy elfogadásról számoltak be. Ők alapvetően pozitívan viszonyulnak az oktatási célú MI-alkalmazásokhoz, és vagy magában a technológiában, vagy az adott képzési környezetben (pl. a tanári vagy intézményi döntéshozók szakértelmében) bíznak. Ez a csoport különösen nyitott az innovációkra, és az újdonság elfogadását nem feltételezi a személyes tapasztalat meglétéhez.

A válaszadók legkisebb csoportja (9%) tapasztalat-alapú bizalmat fogalmazott meg. Ők konkrét pozitív élményekre (pl. ChatGPT-vel végzett sikeres tanulási helyzetekre) hivatkozva jelezték, hogy már most is megbíznak a chatbotokban. Ez a csoport a bizalmat közvetlen interakciók és jól működő használati tapasztalatok alapján építette fel, ami alátámasztja azt az elméleti állítást, hogy a tanulói bizalom dinamikusan fejlődő konstrukció, amelyet a tanulási folyamat során szerzett élmények is formálnak.

Az eredmények összességében arra utalnak, hogy a válaszadók többsége nem elutasító az oktatási chatbotokkal szemben, azonban a bizalom mértéke és típusa jelentős eltéréseket mutat.

A feltételes bizalom dominanciája azt sugallja, hogy a bizalom nem automatikusan adott, hanem tudatos fejlesztést, pedagógiai és technológiai garanciákat igényel. A skála két végpontján a szkeptikus és az élményalapú elfogadó csoportok állnak, amelyek a leginkább kritikus, illetve a leginkább motivált felhasználói rétegeket képviselik.

A tanulói bizalmat növelő tényezők: preferált feltételek és elvárások

A válaszadók körében feltett nyitott kérdés arra irányult, hogy milyen körülmények, jellemzők vagy működési feltételek növelnék a bizalmukat egy oktatási chatbot iránt. A válaszok kvalitatív kódolása és kategorizálása négy fő bizalomnövelő dimenziót eredményezett. A válaszadók 36%-a a szakmai hitelességet és a forrásmegjelölést emelte ki legfontosabb bizalomépítő tényezőként. Ezek az attitűdök elsősorban a chatbot által adott válaszok megalapozottságára, szakirodalmi vagy tananyagbeli hivatkozásaira, illetve a válaszok mögötti tudástartalom átláthatóságára fókuszálnak. Több válaszadó külön is hangsúlyozta, hogy akkor érezné megbízhatónak a chatbotot, ha a forrásait feltüntetné, és nyíltan jelezné, ha nem rendelkezik kellő információval. Ez az elvárás közvetlen kapcsolatban áll a technológiai pontosság és transzparencia korábbi kutatásokban is megjelenő elméleti dimenzióival.

A válaszadók 23%-a az emberi kontroll és tanári jelenlét biztosítását nevezte meg kulcsfontosságú tényezőként. Ebben a csoportban többen úgy vélték, hogy a chatbot csak akkor tud bizalomgerjesztő módon működni, ha mellette elérhető marad valamilyen emberi támogatás – például a tanári ellenőrzés, egy oktató által adott visszajelzés vagy az azonnali segítségkérés lehetősége. Egy válaszadó így fogalmazott: „Ha pl. a kérdésemre nem találja a választ az algoritmusok között, akkor beinvitál a chatbe egy valós oktatót.”Ez a nézőpont megerősíti, hogy a tanulók bizalma nem kizárólag a technológiára, hanem az azt körülvevő pedagógiai környezetre is irányul.”

A válaszadók 23%-a az átláthatóságot és a működési transzparenciát tekintette kulcsfontosságúnak. Ide azok az észrevételek tartoztak, amelyek a chatbot működési logikájának, adatforrásainak, válaszképzési mechanizmusának és tanítási módjának átláthatóságát hiányolták. A válaszadók ebben a kategóriában fontosnak tartották, hogy tudják, „miből és hogyan születik egy-egy válasz”, és ezzel kapcsolatban világos visszajelzést kapjanak a rendszer működéséről. Ezzel a dimenzióval a felhasználói kontrollérzet és az információs autonómia jelentősége kerül előtérbe. A válaszok 18%-a sorolható egyéb bizalmi tényezők kategóriájába, amelyek között megjelentek a személyre szabhatóság, az adatvédelem

biztosítása, a tanulói visszajelzések figyelembevétele, valamint a pozitív közösségi értékelés (pl. más tanulók ajánlása). Ezek a válaszok egyfajta bizalom-ökoszisztémára utalnak, amelyben a technológiai, pszichológiai és társas tényezők egyaránt szerepet játszanak.

Összességében az eredmények azt mutatják, hogy a tanulói bizalom erősítéséhez nem elegendő csupán a technológia funkcionalitása. A válaszadók a bizalom alapjául világos működési kereteket, emberi támogatási lehetőségeket, és szakmailag hiteles tartalmat várnak el. Ezek az elvárások megerősítik a szakirodalomban is megjelenő álláspontokat, miszerint a tanulói bizalom multidimenzionális jelenség, és tudatos fejlesztési stratégiákat igényel az oktatási chatbotokat alkalmazó rendszerekben.

A tanulói bizalmat csökkentő tényezők: aggályok és fenntartások

A válaszadók arra a kérdésre is reflektáltak, hogy milyen jellemzők vagy tapasztalatok esetén csökkenne a bizalmuk egy oktatási chatbot iránt. A nyílt végű válaszok kvalitatív tartalomelemzése és kódolása alapján öt fő bizalomgyengítő tényező rajzolódott ki. A legtöbben, a válaszadók 42%-a, a hallucinációt vagy téves választ nevezte meg elsődleges aggodalomként. Ők olyan helyzetekre utaltak, amikor a chatbot bizonyos témákban magabiztos, de tartalmilag hibás választ ad, illetve nem ismeri be, ha nem tud biztos információval szolgálni. A „hallucináció” jelensége – azaz a hihetőnek tűnő, de valótlan állítás generálása – a válaszok alapján a legnagyobb kockázati tényezőnek számít a tanulói bizalom szempontjából, különösen abban az esetben, ha a hallgatók nem rendelkeznek megfelelő tudással a válasz ellenőrzéséhez.

A második leggyakoribb aggály, a pontatlan vagy hiányos tudás (22%) szorosan kapcsolódik az előzőhöz. Ide azok a megjegyzések kerültek, amelyek a chatbot válaszainak felületességét, részletességének hiányát vagy konkrét kérdések megválaszolására való képtelenségét említették. A válaszadók ebben a kategóriában gyakran említették, hogy a chatbot nem tudott releváns példát hozni, „nem tantárgyspecifikus” módon reagált, vagy a válaszai túl általánosak maradtak.

A válaszok 18%-a egyéb tényezőket jelölt meg a bizalom csökkentésének forrásaként. Ezek között szerepelt például az adatvédelem hiányossága, a tanulói adatok feletti kontroll elvesztése, vagy a rendszer túlzott általánosítása. Itt jelenik meg az az elvárás, hogy egy oktatási chatbot ne csak nyelvi képességeiben legyen meggyőző, hanem strukturálisan is illeszkedjen az adott tantárgyhoz, tananyagtartalomhoz és tanulói helyzethez. Kisebb arányban (9%) ugyan, de jelen volt az ismétlés és automatizmus miatti bizalomcsökkenés. E válaszadók szerint a

chatbot gyakran ismétli magát, sablonos válaszokat ad, vagy túl mechanikusan közelíti meg a problémákat, ami hosszú távon csökkenti a hitelességet és a tanulói érdeklődést.
Ugyancsak 9% említette a bizalomcsökkentés okaként az emberi kapcsolat hiányát. Ezek a hallgatók hangsúlyozták, hogy számukra elengedhetetlen a tanári visszajelzés, a valós idejű segítség vagy az emberi interakció lehetősége – akár érzelmi, akár didaktikai szempontból.E szempont különösen fontos lehet a személyre szabott tanulás és az önirányított tanulási folyamatok támogatása szempontjából.
Az eredmények világosan mutatják, hogy a tanulói bizalom elvesztése leginkább tartalmi és működésbeli megbízhatatlanságból, illetve a technológia korlátainak felismeréséből ered. A bizalom nemcsak akkor sérül, ha a chatbot téved, hanem akkor is, ha nem képes érzékenyen és személyre szabottan kezelni a tanulók szükségleteit. A válaszadók aggályai azt jelzik, hogy az oktatási chatbotok tervezésekor nem elegendő a technológiai teljesítményre koncentrálni; legalább ilyen fontos az átláthatóság, a tantárgyspecifikus hitelesség és az emberi jelenlét lehetőségének biztosítása is.

A chatbotok lehetséges szerepe a képzésekben: tanulói preferenciák és elutasítási küszöbök

A kérdőív egyik zárt kérdéssora azt vizsgálta, hogy a válaszadók milyen mértékben tartják elfogadhatónak vagy elutasítandónak a chatbotok különféle oktatási alkalmazási lehetőségeit. A három feltett kérdésben a válaszadók egy 5 fokú Likert-skálán értékelték, hogy szerintük mennyire jó vagy rossz megoldás lenne az adott forma.
A legnagyobb elfogadottsággal az a lehetőség rendelkezik, hogy a chatbot a kurzusokban tanulási feladatokkal kapcsolatban legyen használható. A válaszok átlaga (-2 és 2 közötti korrigált skálán 1,36) ebben az esetben pozitív irányba tolódott, ami arra utal, hogy a hallgatók kifejezetten támogatják a chatbot alkalmazását akkor, ha az a gyakorlati feladatmegoldásban, például értelmezésben, példák keresésében vagy ötletgenerálásban segíti őket. Ez összhangban áll a korábbi kutatásokkal, amelyek szerint a tanulók különösen hasznosnak ítélik a chatbotokat azonnali visszajelzést vagy támogatást igénylő helyzetekben.
Szintén pozitív, bár valamivel visszafogottabb elfogadottságot mutatott az a lehetőség, hogy a chatbot a tananyaggal kapcsolatban legyen használható. A válaszadók többsége ebben az esetben is támogató volt, de a válaszok átlaga már közelebb állt a semleges tartományhoz (-2 és 2 közötti korrigált skálán 1,16). Ez arra utalhat, hogy a tananyagközpontú interakciók során

a válaszadók magasabb elvárásokat támasztanak a chatbot tartalmi pontosságával, fogalmi mélységével és hitelességével szemben.

A legerősebb elutasítás azokban a válaszokban jelent meg, amelyek egy teljes képzési program chatbotra épülő kiváltását vetették fel. Ebben az esetben a válaszok átlaga negatív tartományban helyezkedett el, jelezve, hogy a hallgatók többsége nem támogatja a teljesen automatizált, embermentes oktatás koncepcióját (-2 és 2 közötti korrigált skálán -0,85). A visszautasítás mögött vélhetően nemcsak a tartalmi és technológiai aggályok, hanem a tanári jelenléttel, személyességgel és támogatottsággal kapcsolatos elvárások is meghúzódnak.

Összességében az eredmények azt jelzik, hogy a válaszadók nyitottak a chatbotok kiegészítő, támogató szerepére a tanulási folyamatban – különösen a feladatokkal kapcsolatos interakciók során –, ugyanakkor erőteljes fenntartásaik vannak azokkal a megoldásokkal szemben, amelyek a tanulás teljes folyamatát gépesítenék. Ez a preferenciarendszer megerősíti azt a korábbi megállapítást, hogy a tanulói bizalom nem csupán a technológia teljesítményén, hanem annak pedagógiai integrációján és arányosságán is múlik.

A tanulói elfogadás határai: kizáró tényezők és kilépési küszöbök

A kérdőív záró szakaszában a válaszadók arra a hipotetikus helyzetre reflektáltak, hogy milyen körülmény esetén nem lennének hajlandóak egy képzésben részt venni, ha az oktatás jelentős részét oktatási chatbot támogatná. A kérdés célja az volt, hogy feltárja azokat a határokat, amelyeken túl a bizalom már nem tartható fenn, és amelyek adott esetben a képzésből való kilépéshez is vezethetnek.

A legtöbben, a válaszadók 36%-a azt jelezte, hogy számukra a chatbot kizárólagossága és a tanári kapcsolat teljes hiánya elfogadhatatlan lenne. Az ide tartozó hallgatók a válaszaikban hangsúlyozták, hogy a személyes visszajelzés, a pedagógiai irányítás és az emberi jelenlét elengedhetetlen a tanulás során. Több válaszadó szerint egy olyan rendszer, amelyben „nincs lehetőség kérdezni egy oktatótól” vagy „teljesen a chatbot dönt”, nemcsak motiválatlanná tenné őket, hanem akár a képzés megszakítását is indokolttá tenné.

A második legnagyobb csoport (25%) a nem megfelelő válaszminőséget nevezte meg kilépési okként. Ők elsősorban a sablonos, felületes, félreértelmezett vagy irreleváns válaszokat tartják kizárónak. A chatbot használhatósága szerintük akkor sérül, ha az nem képes komplex, tantárgyspecifikus vagy elmélyült kérdésekre érdemben reagálni. Ezek a vélemények megerősítik az eredmények korábbi szakaszában is megjelenő aggodalmat a „hallucináció” és tartalmi megbízhatatlanság kapcsán.

A válaszadók 16%-a elfogadó hozzáállást tanúsított, vagyis azt jelezte, hogy számukra nincs olyan kizáró tényező, amely önmagában a képzés megszakítását indokolná. Ők többnyire nyitottabbak a technológiai újításokra, illetve feltételes elfogadással viszonyulnak a chatbot-alapú oktatási környezetekhez. 14%-uk technikai problémákat és a rendszer megbízhatatlanságát említette olyan küszöbtényezőként, amely tartós fennállás esetén akár a képzésből való kilépést is kiválthatná. Ezek a válaszadók főként az instabil működés, a hibás válaszadás, a rendszerleállás vagy az adatvesztés veszélyeit hangsúlyozták. Végül 9% a válaszában egyéb szempontokat említett, például adatvédelmi aggályokat, a chatbot tanítási módszereinek idegenségét, vagy a személytelenség miatti elidegenedést.

Összességében az eredmények azt mutatják, hogy a tanulói bizalomnak és elfogadásnak határai vannak. A hallgatók kétharmada szerint az emberi kapcsolat nélküli, kizárólag chatbotra épülő tanulási helyzet vagy a tartós hibázási és pontatlansági problémák elfogadhatatlanok. Ez a megállapítás különösen fontos lehet az oktatástechnológiai rendszerek tervezésében: a chatbotokat nem önállóan, hanem integrált, emberközpontú oktatási ökoszisztéma részeként érdemes alkalmazni.

Elfogadás feltételei: milyen körülmények között lenne vállalható a chatbot-alapú tanulás?

A kutatás során arra is kerestük a választ, hogy milyen feltételek mellett fogadnák el a válaszadók azt a helyzetet, ha a tanulásban egy oktatási chatbot alkalmazása elvárásként jelenne meg. A cél annak feltárása volt, hogy milyen bizalmi és működési garanciákat várnak el a hallgatók egy ilyen tanulási környezettől.

A leggyakrabban megnevezett feltétel a megbízhatóság és szakmai pontosság volt, amelyet a válaszadók 36%-a emelt ki. Ez a csoport azt hangsúlyozta, hogy a chatbot csak akkor válhat a tanulás elfogadott eszközévé, ha a válaszai szakszerűek, pontosak, konzisztens logikájúak és hibamentesek. Több válaszadó konkrétan elvárta, hogy a chatbot ne „hazudjon”, ne találjon ki válaszokat, illetve legyen képes korrekt módon jelezni, ha nem tud válaszolni. Ez a feltétel összhangban áll az empirikus eredmények többi részében is megjelenő bizalomépítő elvárásokkal. A válaszadók 23%-a a korlátozott szerepet és az oktatói jelenlét szükségességét nevezte meg feltételként. Szerintük a chatbot csak kiegészítő eszközként elfogadható, de nem helyettesítheti az oktatót. Fontosnak tartják, hogy legyen lehetőség emberi konzultációra, visszajelzésre, illetve hogy a chatbot használata ne legyen kötelező jellegű. Egyes válaszok

kiemelték, hogy a tanulóknak legyen döntési lehetőségük abban, mikor és milyen célra használják az eszközt.

A transzparencia és tájékoztatás a működésről szintén központi elemként jelent meg a válaszok 14%-ában. Ezek a válaszadók elvárják, hogy a rendszer egyértelműen és közérthetően ismertesse saját működési logikáját, korlátait, tanítási módját. Többen megemlítették, hogy elfogadhatóbb lenne a chatbot használata, ha tudnák, milyen forrásokra épít, mikor válik bizonytalanná a működése, és hogyan kezeli az esetleges válaszhibákat. Végül a válaszadók 11%-a technikai és tanulásszervezési támogatást várna el a chatbottól. Ebben a csoportban gyakran említették, hogy hasznos lenne, ha a chatbot segítene eligazodni a tanulási felületeken, figyelmeztetne határidőkre, vagy adminisztratív terheket venne le a tanulóról. Ez a funkcionális támogatás ugyan nem közvetlenül a bizalomhoz kötődik, de hozzájárulhat a pozitív felhasználói élmény kialakulásához, ami visszahat a bizalmi megítélésre is.

Az eredmények egyértelműen azt mutatják, hogy az oktatási chatbotok elfogadása feltételhez kötött, és a tanulók tudatosan megfogalmazott elvárások mentén hajlandók beépíteni ezeket a tanulási folyamataikba. A bizalom nem adottságként, hanem pedagógiai és technológiai feltételrendszerként értelmezhető. A sikeres integrációhoz megbízhatóság, emberi jelenlét, transzparencia és tanulást támogató funkciók együttesen szükségesek.

Összefoglaló

A kutatás célja annak feltárása volt, hogy milyen tényezők befolyásolják a tanulói bizalmat az oktatási célú chatbotok alkalmazásával kapcsolatban. A tanulmány a mesterséges intelligencia oktatási környezetbe való integrációjának egyik kulcskérdésére fókuszált: milyen feltételek mellett hajlandók a felsőoktatásban és posztgraduális képzésekben részt vevő hallgatók elfogadni a chatbotok tanulástámogató szerepét, és milyen körülmények vezethetnek bizalomvesztéshez vagy elutasításhoz.

A vizsgálat alapján a hallgatók többsége nem utasítja el kategorikusan az oktatási chatbotok használatát, azonban bizalmuk feltételes, és számos szempont teljesüléséhez kötött. A legfontosabb bizalomépítő tényezők közé tartozik a szakmai hitelesség, a forrásmegjelölés, valamint a válaszok pontosságának és megalapozottságának biztosítása. Ugyanilyen jelentőséggel rendelkezik a chatbot működési korlátainak nyílt és átlátható kommunikációja, különösen annak beismerése, ha a rendszer nem képes válaszolni egy kérdésre. A válaszadók számára a „hallucináció” (téves, de magabiztos válasz generálása) mellett az is a bizalmat

csökkentő tényezők közé tartozik, ha a chatbot felületessége vagy visszajelzés-nélkülisége akadályozza a tanulást.

A hallgatók elvárják, hogy az oktatási chatbot ne működjön elszigetelten, és ne váltsa ki teljes egészében az emberi oktatói jelenlétet. A válaszok többsége hangsúlyozza a tanári kontroll, konzultáció és visszajelzés lehetőségének megőrzését, amely a tanulói bizalom egyik legfontosabb alapja. Ez összefügg azzal a megállapítással is, hogy a chatbot csak kiegészítő, támogató szerepet tölthet be a képzésben. A hallgatók elutasítják azt az elképzelést, hogy egy teljes képzés kizárólag chatbotra épüljön, ugyanakkor pozitívan értékelik a chatbot tananyagtámogató és tanulási feladatorientált funkcióit.

A kutatás rávilágít arra is, hogy a tanulói bizalom nemcsak a technológiai működés, hanem a kurzuskörnyezet egészének jellemzői alapján alakul ki. A válaszadók számára lényeges, hogy a chatbot alkalmazása humán közegbe ágyazott, oktatóval kiegészített környezetben történjen, amely lehetőséget ad az egyéni visszacsatolásra, értelmezésre és korrekcióra. A chatbotot a hallgatók leginkább elméleti tudás közvetítésére, kisebb mértékben gyakorlati alkalmazások támogatására, és tanulásszervezési segítségként fogadják el.

Összegzésként elmondható, hogy a tanulói bizalom szempontjából a chatbotok alkalmazása nem önmagában, hanem pedagógiai kontextusba ágyazott módon, világos működési keretekkel, transzparens kommunikációval és emberi jelenléttel együtt lehet sikeres. A feltételes bizalom megerősítése érdekében a fejlesztőknek és oktatóknak egyaránt törekedniük kell arra, hogy a chatbot ne helyettesítő, hanem kiegészítő eszközként támogassa a tanulási folyamatokat.

Jelen pilotkutatás korlátai közé tartozik a viszonylag kis mintanagyság, a nem reprezentatív, szakmailag specializált minta, valamint a kérdőíves mérési forma korlátozott mélysége. Az eredmények ezért nem általánosíthatók a teljes hallgatói populációra, ugyanakkor értékes betekintést nyújtanak egy technológiailag affinis, célzott tanulói csoport bizalmi attitűdjeibe.

A továbblépés érdekében javasolt egy kiterjesztett, több intézményre kiterjedő longitudinális vizsgálat, amely képes feltárni a bizalom időbeli alakulását és a tanulói attitűdök változását a chatbot-használat során. Emellett mélyinterjúkon vagy fókuszcsoportokon alapuló kvalitatív kutatások is indokoltak, amelyek a tanulói értelmezések finomabb rétegeit és a tapasztalatok érzelmi aspektusait is képesek feltárni. A tanulói bizalom összetett, dinamikusan változó konstrukció, amelynek megértéséhez többszintű és több módszertant ötvöző kutatási megközelítés szükséges.

Irodalomjegyzék

Bettayeb, A. M., Talib, M. A., Altayasinah, A. Z. S., & Dakalbab, F. 2024, *Exploring the impact of ChatGPT: conversational AI in education.Frontiers in Education*. https://doi.org/10.3389/feduc.2024.1379796

Raiche, A., Luca, G. D., Rivest-Henault, D., Vaughan, T., Proulx, C., & Guay, J.-P. 2023, *Factors influencing acceptance and trust of chatbots in juvenile offenders' risk assessment training.Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2023.1184016

Schei, O. M., Møgelvang, A., & Ludvigsen, K. 2024, *Perceptions and Use of AI Chatbots among Students in Higher Education: A Scoping Review of Empirical Studies.Education Sciences.* https://doi.org/10.3390/educsci14080922

Uppal, K., & Hajian, S. 2024, *Students' Perceptions of ChatGPT in Higher Education: A Study of Academic Enhancement, Procrastination, and Ethical Concerns.European Journal of Educational Research*. https://doi.org/10.12973/eu-jer.14.1.199

Wienrich, C., & Obremski, D. 2024, *The Impact of Transparency and Expert-Framing on Trust in Conversational Ai*. https://doi.org/10.2139/ssrn.4715193

AI-EdTech és bizalom a jövőtudatos oktatási folyamatban

AI-EdTech and Trust

in the Future-Conscious Educational Process

Dr. Zuti Pál[5]

Absztrakt

A **prezentációs előadásom célja** egyrészt az AI-EdTech, a jövőtudatosság, a bizalom dimenzióinak vizsgálata, másrészt a változások, kihívások, hatások, lehetőségek kezelése és a jelenben megfogalmazott, de a jövőbe mutató kitörési pontok és válaszok keresése

A **digitális kultúra** egy olyan fogalom, amely leírja, hogy a digitális technológia (IKT), az AI-EdTech, az AI és az IoT hogyan alakítja azt, ahogyan emberként kölcsönhatásba lépünk egymással (kommunikáció – technikai kommunikáció – designkommunikáció). Ez az a mód, ahogyan viselkedünk, gondolkodunk és kommunikálunk a társadalomban, melyben kitüntetett szerepe van a **digitális kultúra 7 dimenziójának**.

A **jövőorientált készségek és a jövőtudatos gondolkodás fejlesztése** a digitális korban, hatása van az intézményi közösségformáló szemléletre és célrendszerre is. A **jövőtudatos gondolkodásban** nagyon fontos szerepe van a **jövőtudatosság célrendszerének (SMART célok)** és az **adatnak**. Az intézményi keretek között a **bizalom fenntartása és a biztonságunk megtartása** az intézményi elfogadottságunk és támogatottságunk megélése, fejlesztés nélkül nem lehetséges.

Az oktatás meglévő stratégiáját elengedhetetlenül összhangba kell hozni a digitalizációról (AI-EdTech) , az innovációról (felelősségteljes innováció) és az oktatásról (jövőtudatos oktatás) szóló tudományok adataival.

Kulcsszavak: digitalizáció, innováció, AI-EdTech, jövőtudatosság, bizalom.

---

[5] Mind Mate Inspiráció Egyesület felnőttképzési és digitális kultúra fejlesztési vezető, mentor, matematika – műszaki ismeretek – informatika – pedagógia – oktatástechnológia tanár, Európai Digitális Oktatási Központ (European Digital Education Hub) tag, UNISPACE Oktatási Műhold Munkacsoport alapító tag, Magyar Pedagógiai Társaság Békés Vármegyei tagozat társelnök, Magyar Pedagógiai Társaság Projektpedagógiai Szakosztály alapító tag, Független Akkreditált Vizsgaközpontok szakképzési szakértő, EdTech Koalíció tag, dr.zuti@gmail.com

Abstract

The aim of my presentation is, on the one hand, to examine the dimensions of AI-EdTech, future consciousness, and trust, and on the other hand, to manage changes, challenges, impacts, and opportunities while seeking breakthrough points and answers that are formulated in the present but point toward the future.

Digital culture is a concept that describes how digital technology (ICT), AI-EdTech, AI, and IoT shape the way we interact with each other as human beings (communication – technical communication – design communication). This is the way we behave, think, and communicate in society, in which **the 7 dimensions of digital culture** play a prominent role.

The development of future-oriented skills and future-conscious thinking in the digital age also has an impact on institutional community-building perspectives and goal systems. **Future-conscious thinking** places great importance on **the goal system of future consciousness (SMART goals)** and **data**. Within institutional frameworks, **maintaining trust and preserving our security** is impossible without experiencing and developing institutional acceptance and support.

The existing strategy of education must inevitably be aligned with the data from sciences about digitalization (AI-EdTech), innovation (responsible innovation), and education (future-conscious education).

Keywords: digitalization, innovation, AI-EdTech, future consciousness, trust

Bevezetés

Az oktatás a 21. században a digitalizáció, a mesterséges intelligencia (AI) és az oktatástechnológia (EdTech) hatására radikális átalakuláson megy keresztül. E változások nem pusztán technikai eszközök és módszerek bevezetését jelentik, hanem az oktatási folyamat minden szereplőjének – tanulók, pedagógusok, intézmények, családok – kultúráját, szemléletét, kapcsolati hálóit és együttműködési mintázatait is formálják. (Bostrom 2014)

A digitális kor jövőtudatos oktatásában kiemelt jelentőségű a bizalom, amely a társadalmi-gazdasági kihívások, az adatbiztonság, az etikai kérdések, valamint az intézményi elfogadottság és támogatottság szempontjából is nélkülözhetetlen.

Jelen tanulmány célja, hogy a legújabb tudományos eredmények és gyakorlati tapasztalatok fényében átfogó képet adjon az AI-EdTech fókuszú, jövőtudatos oktatási folyamat kulcsszereplőiről, a kapcsolódó kihívásokról és lehetőségekről, valamint a fenntartható, bizalmi alapú intézményi kultúra építésének módjairól.

1. Fogalmak és elméleti keretek

1.1. Digitalizáció és AI-EdTech

A **digitalizáció** az információs és kommunikációs technológiák (IKT) adaptációját jelenti az oktatás minden szintjén, ideértve a hálózati, platformalapú és automatizált megoldásokat is. Az EdTech, vagyis az oktatástechnológia, az oktatási folyamatok fejlesztésének céljával integrálja a digitális eszközöket, míg az AI-EdTech a mesterséges intelligencia-alapú rendszerekkel (tanulóanalitika, adaptív tanulás, intelligens tutorrendszerek stb.) egészíti ki az oktatási technológiát.

Az **AI-EdTech** térnyerése a differenciált, személyre szabott tanulás, a tanulási eredményesség, az adminisztrációs folyamatok hatékonyabb menedzselése felé mutat, ugyanakkor újfajta etikai, adatvédelmi és szakmai dilemmákat is felvet. Az AI-EdTech lehetőséget kínál az adaptív tanulásra, kiemeli a tanulói önállóságot, és a tanítási- tanulási folyamat minden szintjén eszközrendszert biztosít a jövőorientált készségek fejlesztéséhez. Ugyanakkor új kockázatok és dilemmák is felmerülnek, például adatvédelem, etikus algoritmikus döntéshozatal, intézményi felelősségvállalás és a digitális szakadék problematikája. Ennek megfelelően a tanulmány célja a fogalmi keretek tisztázása, a digitális pedagógiai újítások bemutatása és az AI-EdTech, illetve a bizalom dimenzióinak részletes elemzése. (Kozma, T 2019; Zuti, P 2024d)

A digitalizáció az oktatásban azt jelenti, hogy a tanulási-tanítási folyamat résztvevői (tanulók, tanárok, intézmények) digitális eszközök, platformok, tartalmak, kommunikációs és adatgyűjtési eljárások integrációjával érik el a tanulási célokat. Az EdTech – oktatástechnológia

– tágabb fogalom: minden olyan eszköz, szoftver, szolgáltatás ide sorolható, amely a tanulást, tanítást, a pedagógiai innovációkat támogatja. Az AI-EdTech fókuszban a mesterséges intelligencia (gépi tanulás, természetes nyelvfeldolgozás, predikciós modellezés stb.) szerepe áll: személyre szabott tanulási tartalom, motivációs visszacsatolás, tanulóanalitika, intelligens tutor- vagy értékelő rendszerek váltak elérhetővé. (Stuart, R & Peter, N 2020; Stone, James V 2019)

A digitális kultúra korában a digitális megoldás – amely a technológiára támaszkodik az összetett problémák megoldásában – ígéretté és buktatóvá is válik egyszerre. Míg a digitális platformok demokratizálják az információkhoz való hozzáférést, elmélyíthetik a felügyeletet, az adatkivonást és az algoritmikus torzítást is. Ugyanakkor ezek a platformok a polarizáció, a félretájékoztatás és az áruvá alakítás színterei is. Ahogy eligazodunk a digitális kultúrák összetettségei között, egy tanulság kiemelkedik: **nincsenek egyedi megoldások vagy narratívák.** A digitális kultúrák sokfélesége tükrözi az emberi tapasztalatok sokszínűségét, megköveteli, hogy alázattal, kíváncsisággal, **bizalommal** és a méltányosság iránti elkötelezettséggel közelítsünk a **digitális technológiához.**

1.2. Jövőtudatosság, jövőorientált készségek

A jövőtudatosság nem csupán technológiai vagy infrastrukturális felkészültséget jelent, hanem a jövőorientált, adaptív gondolkodás fejlesztését az egyéni, közösségi és intézményi szinteken is. Kulcsfontosságúak a 21. századi kompetenciák: algoritmikus és kritikus gondolkodás, kreativitás, kollaboráció, digitális írástudás, valamint az önálló tanulás képessége. (Zuti 2024b) Ennek részét képezik az ún. **„future literacy**", azaz a ***jövő változásainak* értelmezése**, **tervezése és irányítása, valamint a digitális és AI eszközökkel való reflexív bánásmód is**. A ***jövőtudatosság* („futures literacy") azon képesség, hogy a lehetséges jövőképeket céltudatosan, kreatívan, együttműködően használjuk a jelen döntéseinek meghozatalához**. Kulcskompetenciák ennek mentén: kritikus gondolkodás, kreativitás, digitális írástudás, kommunikáció, kollaboráció, alkalmazkodóképesség, problémamegoldás életvezetési, önálló tanulási készségek. A jövőtudatos oktatás jellemzője a SMART célrendszer alkalmazása: specifikus – mérhető – elérhető – releváns – időhöz kötött célok kijelölése, ezek rendszeres monitorozása és reflektív fejlesztése. (Pintér, G & Nemes, G 2024; Zuti, P 2024d)

1.3. A bizalom pedagógiai és technológiai dimenziói

A bizalom olyan alapvető tényező, amely egyszerre köt össze embereket, intézményeket és rendszereket. A digitális környezetben a bizalom kettős: egyaránt szól az emberek (diákok, tanárok) közti személyes viszonyról, valamint a digitális eszközök, AI-megoldások, platformok iránti technológiai bizalomról is. A bizalom pedagógiai értelemben az inkluzív,

tanulóközpontú, partneri kapcsolatokban, illetve az autonómia és kompetenciaérzés erősödésében nyilvánul meg. Technológiai szempontból a transzparens, etikus fejlesztés, az adatbiztonság és az AI működésének érthetősége a legfőbb mozzanatok. Az AI-EdTech fejlesztése és alkalmazása csak akkor vezet a tanulási eredmények, az intézményi kultúra és a társadalmi elfogadottság pozitív változásához, ha mindkét bizalmi dimenzió kiemelt figyelmet kap, és transzparens, etikus eljárások mentén valósul meg. (Kováts 2018)

2. Az oktatási szereplők és a digitális átalakulás

2.1. Tanulók digitális kultúrája

A tanulók digitális kultúrája a digitális eszközhasználatból, szoftverek, platformok használatából, média- és információfogyasztási szokásokból, valamint az online kommunikációs mintázatokból áll össze. Az algoritmizált tartalomfogyasztás újraértékeli az önálló tanulás, a kritikus médiahasználat és az információs önrendelkezés jelentőségét. Egyre nagyobb kihívást jelent a képernyőidő, a figyelem menedzsmentje, az online identitás és etikus digitális jelenlét. Az AI-alapú rendszerek képesek személyre szabott tananyagokat ajánlani, támogatni a differenciált haladást, de ezzel együtt szükség van a tudatos, reflektált médiapedagógiára. (Zuti és Bóczi 2023)

A tanulók digitális szokásainak, motivációinak, készségeinek alapos ismerete nélkül lehetetlen a korszerű pedagógia sikeres megvalósítása. Az AI-vezérelt tartalomszolgáltatók (pl. TikTok, YouTube) algoritmusai gyakran fokozzák a figyelem-ingert, addikciós mintákat erősítenek, és uniformizálják a tanulók előtti információs teret. (Zuti 2024d) Ez szükségessé teszi a tudatos, reflektív médiapedagógiát és a tanulói digitális önrendelkezés fejlesztését. (Zuti, P 2023c; Zuti, P & Sióréti, G 2024)

2.3 Tanári szerepek és tanulásirányítás

A tanári szerep a digitális korban differenciálódik: egyszerre igényel technológiai kompetenciákat, AI-műveltséget, pedagógiai-informatikai ismereteket és az együttműködés újraértelmezett formáit. Fontossá válik a támogató, facilitátori, mentoráló szerep, amelyben a tanár a tanulást irányító, a digitális térben eligazodást segítő szakember. Megjelenik az adatelemzés, a tanulói haladás mérésének, értékelésének és visszacsatolásának digitális dimenziója. (Prensky 2012)

Kiemelendő a tanári autonómia, digitális önképzés és szakmai hálózatosodás, amely révén folyamatosan naprakészen tartható a digitális kultúra és AI-EdTech eszköztára. A tanári önképzés, szakmai hálózatosodás, a digitális portfóliók, reflektív gyakorlatok pedig a szakmai fejlődés zálogai. (Szűts 2020)

A **digitális korban** a tanár naprakész ismerete hatványozottabban érvényesül. Ha nem rendelkezünk naprakész ismeretekkel, akkor sérülnek a legfontosabb érdekek, a **tanulói érdekek**! A **tanár-diák kapcsolatnál**, mindig a tanár kell, hogy viszonyuljon a tanulóhoz, az ő életkorához. Sohasem tőle várom el - nem is várhatom el - a viszonyulást. **Ez a tanár lehetősége és felelőssége!** Az **algoritmikus gondolkodás és szemlélet** a pedagógiai tevékenység meghatározó mozgató ereje. A **Digitalizáció** – az **Algoritmikus gondolkodás** – a **Feladatmegoldás** – és a **Felelősségteljes innováció** iránymutatásainak megfelelően a személyiségfejlesztés, a hatékonyság és a versenyképesség fejlesztése történik. (Zuti, P & Gulyás, Zs 2022; Zuti, P & Harangozó, H 2023) Ennek legfontosabb jellemzői: minőség; megbízhatóság; adaptivitás; tanulóközpontúság. Az oktatásban a versenyképesség lényegében a minőség, a minőségirányítás szinonimája. A versenyképesség és a vonzóképesség viszonylatában **legfontosabb az emberi tényező szerepe**. Ezt sohasem szabad figyelmen kívül hagyni! Továbbá nagyon fontos megemlíteni, hogy a versenyképességben az innováció és az oktatás a kulcskérdés. A **felelősségteljes innovatív pedagógiai tevékenységnek megfelelően** a tanulási-tanítási **szituációk** minőségi javítása érdekében, olyan pedagógiai módszertani kultúra kialakítására lehet törekedni, amely lehetővé és szükségessé teszi az **önálló ismeretelsajátítást,** a közös megbeszéléseket, ezek általános érvényű tanulságainak megfogalmazását, rögzítését. **Nagyon fontos,** hogy a módszer és tartalom nem szakadhat el egymástól. **A megformálatlan tartalom ugyanis nem tartalom!**

Napjainkban a tanárnak naprakész szakmai ismeretekkel, tapasztalattal, szakmai, intellektuális, szociális, módszertani, digitális kompetenciákkal és megfelelő felelősségtudattal, felelősségteljes innovációval kell rendelkeznie ahhoz, hogy a digitalizáció világában hatékonyan eligazodni képes innovatív pedagógus legyen, az Ipar 4.0 filozófiájának megfelelően. Ez az a szemlélet és gondolkodásmód, amely egyértelműen különbséget tesz a **problémamegoldás és a feladatmegoldás** között. Ugyanis a feladatmegoldás az a problémamegoldás, amely algoritmizálható, vagyis digitalizálható! Ez összhangban van azzal, hogy **a 21. század a kompetencia, a kompetens tudás birtoklásának az évszázada.** A **tanárok folyamatosan támogatják** a tanulókat digitális technológiák (AI) és eszközök segítségével. Online konzultációk, csoportos projektek és virtuális órák biztosítják, hogy minden tanuló a lehető legjobban kihasználja a tanulásra fordított időt.

Kompetenciafejlesztés az oktatásban azt jelenti, hogy új módszereket kel alkalmazni, új szemléletmódot kell kialakítani:

- Ma már megkerülhetetlen a **modern technológia, a digitális tananyagok** alkalmazása, az **e-learning eszközök** adta lehetőségek minél jobb kihasználása

- Az oktatásnak életszerűnek, **élményszerűnek** kell lennie: interaktív feladatok, AI applikációk, szimulációk, 3D modellezés, videók stb. adta lehetőségek kihasználásával

A digitális kompetenciák jellemzői:

A digitális kompetencia az információs társadalom technológiáinak (**ITT**) magabiztos és kritikus használatára való képesség a munkában, a szabadidőben és a kommunikációban. **Ezek a kompetenciák** az algoritmikus, kreatív és kritikai (**AKK**) gondolkodással, a magas szintű információkezelési készségekkel és a fejlett kommunikációs készségekkel állnak kapcsolatban. A**z információs és kommunikációs technológiák (IKT)** alkalmazásához kapcsolódó készségek a legalapvetőbb szinten a multimédia technológiájú információk keresését, értékelését, tárolását, létrehozását, bemutatását és átadását, valamint az internetes kommunikációt, a webes felületeken való részvétel képességét ölelik fel. (Rousseau et al. 1998)

Digitális kompetenciateszt területei

1. Információ- és adatkezelési jártasság
2. Kommunikáció és együttműködés
3. Digitális tartalomgyártás
4. Biztonság
5. Problémamegoldás (Feladatmegoldás)

Az **AI szolgáltatásokat** a társadalmi és egyéni igényeknek, a piaci trendeknek megfelelően, továbbá az érintettek visszajelzései alapján **folyamatosan kell fejleszteni.** Az oktatásban rendkívül ébernek kell lennünk, hogy figyelemmel kísérjük az AI fejlődését és világunkban betöltött általános szerepét. Figyelnünk és ellenőriznünk kell az adott technológia kimenetelét, hogy teljes mértékben megértsük a viselkedését, és megbizonyosodjunk arról, hogy nem sérti (emberi) erkölcsi értékeinket*,* továbbá az AI izgalmas lehetőségeiről az oktatásban. **Ezekre a speciális kompetenciára, az AI miatt nagy igény lehet a közeljövőben:**

1. Adatelemzés és értelmezés
2. Technikai készségek
3. Projektmenedzsment
4. Alkalmazkodóképesség
5. Etikus döntéshozatal

Az AI olyan új kompetenciákat igényel, amelyek (valószínűleg) egyrészt inkább a kognitív területet érinti az emberiség fejlődésében.

2.2. Intézményi közösség, szülői együttműködés

Az oktatási intézmény közösségformáló ereje a digitális, hálózati működésben új dimenziókat kap. (Zuti, P & Hegedűs, I 2023; Zuti, P 2024a; Zuti, P & Menyhárt, E 2024a) A közös célok (pl. jövőtudatosság, fenntarthatóság), a SMART célrendszer alkalmazása, a közösségi tanulás és innováció a szervezeti kultúra lényeges elemei. A szülők digitális bevonása, informálása, s a családi környezetben jelentkező digitális kihívások (pl. algoritmikus tartalomszűrés, adatvédelem) közös kezelése nélkülözhetetlen a sikeres digitális átálláshoz.

Az online tér kihívásainak közös kezelése nemcsak a digitális biztonságot, hanem a szociális bizalmi tőkét is építi. Tudatosulnia kell annak, hogy a digitalizáció nem cél, hanem eszköz. Azonban amíg ez a gondolat nem hatja át a vezetői mentalitást, az intézmények nem lesznek képesek az oktatási folyamatok előnyeit kihasználni. A tanulási környezete világos, nyitott és inspiráló, vagyis modern tanulási környezet**,** ahol a jövő: az innovációs intelligencia. (Zuti 2024d)

Nagyon fontos a szülőknél is, hogy tudatosodjon: a digitalizáció nem cél, hanem eszköz. Meghatározó a családi digitális házirend kialakítása, az aktív részvétel: a szülő ne csak szabályokat állítson fel, hanem vegyen részt a gyermek online tevékenységében, mutassa meg, hogyan működnek a szűrők, mire kell figyelni. Abban az esetben hiteles minta: ha a szülő saját digitális szokásai is példát mutatnak.

A párbeszéd és bizalom alapja a közös internetezés és a nyílt kommunikáció, amely erősíti a bizalmat, csökkenti a veszélyeket, és segít abban, hogy a gyermek kevesebb időt töltsön a képernyő előtt. Szülői példamutatás meghatározó a biztonság és tudatosság (adatvédelem) terén. Az algoritmusok szerepére is ki kell térnünk: az algoritmusok a digitális világban – például a közösségi média platformokon – automatikusan ajánlanak tartalmakat a felhasználók viselkedése alapján.

Ezek a rendszerek képesek gyorsan felismerni, mi köti le a gyermek figyelmét, és egyre hasonlóbb tartalmakat kínálnak neki, ezzel növelve a képernyőidőt és befolyásolva a médiafogyasztási szokásokat. Az algoritmusok tehát részben „pótszülőként” is működhetnek, hiszen ők szabják meg, milyen tartalmakkal találkozik a gyermek, ha a szülő nem felügyeli aktívan a digitális jelenlétet. (Zuti 2023a)

3. Algoritmusok, adat és etikai kihívások

3.1. Az algoritmusok hatása a tanulói életvilágokra

Az algoritmusok – különösen a közösségi médiában és digitális tanulási platformokon – rövid idő alatt radikálisan befolyásolták a tanulók médiafogyasztását, figyelemmenedzsmentjét és tanulási stratégiáit. Mivel ezek a rendszerek a felhasználói viselkedésből tanulnak, növelik a tartalmak személyre szabhatóságát, de ezzel együtt az online szokások uniformizálódását és akár veszélyes addikciós mintázatokat is generálhatnak.

A pedagógia feladata ennek tudatosítása, a tanulók kritikai médiatudatosságának fejlesztése, a digitális önvédelem, a személyes adatok felelős kezelése és a digitális jóllét hangsúlyozása. Az automatikus tartalomajánlók a tanulók érdeklődésének megfelelő, de gyakran elfogult, információs buborékot eredményező tartalmakat kínálnak. Az „algoritmikus pótszülő" szerep veszélye, hogy a tanulói médiafogyasztás elszakad a kritikus reflexiótól és a pedagógiai kontrolltól. (Zuti 2024d) Ennek kapcsán kiemelten fontos a médiatudatosság fejlesztése, a tanulók képessé tétele arra, hogy felismerjék a manipuláció lehetséges formáit.

3.2. Adatvédelem és kiberbiztonság

A digitális oktatási környezetben az adatok kezelése – legyen az tanulói teljesítmény, részvételi statisztika, szociális vagy egészségügyi adat – kiemelt érzékenységű kérdés. Az adatvédelmi tudatosság növelése, a GDPR-nak való megfelelés, a biztonságos adatkezelési gyakorlatok bevezetése, a kibertérből eredő veszélyek elleni prevenció és gyors reagálás alapvető fontosságú. Ennek része a tantervi szintű kiberbiztonsági oktatás és az intézményi adatkezelési protokollok szigorú kidolgozása. (European Commission 2022; Digital Education Action Plan 2021-2027)

3.3. AI és EdTech etikai keretei

Az AI-EdTech alkalmazása során a legfontosabb etikai kérdések az átláthatóság, a méltányosság, az előítélet-mentesség, a biztonság, a megbízhatóság és az AI-algoritmusok döntéseinek nyomon követhetősége. Meghatározó, hogy az AI-rendszerek fejlesztői és alkalmazói – legyenek azok fejlesztő cégek vagy oktatási intézmények – törekedjenek az etikus, felelősségteljes innovációra, a tanulók adatainak óvására, a tanulók védelmére és az oktatók példamutató magatartására. Mind a fejlesztők, mind a pedagógiai gyakorlat felelőssége, hogy ennek megfelelő megoldásokat alakítsanak ki, és az AI-alapú rendszerek fejlesztésében a pedagógiai, pszichológiai, etikai és informatikai szakterületek szoros együttműködése valósuljon meg.

4. Innováció, versenyképesség és minőség az oktatásban

4.1. Intézményi innováció és stratégiai célkitűzések

Az innováció nem csupán a technológiai fejlesztésekről szól, hanem az egész intézményi kultúra, pedagógiai program, vezetési szemlélet és a partnerekhez való viszony újragondolását is megköveteli. A SMART (Specifikus, Mérhető, Elérhető, Releváns, Időhöz kötött) célrendszer alkalmazása – egyéni és intézményi szinten – a jövőtudatos működés alapja. (Pintér és Nemes 2024) Az innováció nem csupán eszközhasználatot, hanem az oktatás célszerű átszervezését, új pedagógiai filozófiák bevezetését, támogató intézményi kultúra kialakítását is igényli. A SMART célrendszer elterjesztése – konkrét, mérhető, motiváló, releváns és időhöz kötött célok kitűzése – nélkülözhetetlen a tervezett fejlődéshez. Az AI-EdTech integrációja partnerséget, folyamatos értékelést, az eredmények monitorozását és az illeszkedést segítő flexibilitást kíván meg. (Zuti, P 2023b; Zuti, P & Gulyás, Zs Megjelenés alatt)

4.2. Oktatási minőség, adaptivitás, tanulóközpontúság

Az oktatás minősége a digitális korban az adaptivitásban, a tanulói differenciálásban, az egyénhez illeszkedő tanulási útvonalak biztosításában és a tanulóközpontúságban is mérhető. Az AI-megoldások adaptív rendszerként támogatni tudják a lemaradók felzárkózását, a tehetségek fejlődését, az egyenlő esélyek megteremtését, miközben újfajta pedagógiai kihívásokat is eredményeznek (pl. tanulói önállóság, digitális motiváció, tanári visszacsatolás minősége). Az AI-rendszerek lehetőséget adnak arra, hogy az egyéni tanulási tempót, felzárkóztatást vagy tehetséggondozást és a felzárkóztatást személyre szabottan valósítsuk meg, miközben megmarad a közösségi tanulási élmény is.

4.3. Tehetséggondozás és esélynövelő programok

A tehetséggondozás a jövőorientált oktatás egyik legfontosabb eleme, amelyben a digitális eszközök, AI-analitika és differenciált pedagógia kulcsszerepet kap. Ugyanilyen fontos a hátrányos helyzetű tanulók támogatása, a felzárkóztatás, amely nem a tanulási képességek, hanem a társadalmi-kulturális tőke különbségeiben keresendő. A tehetséggondozás és a hátrányos helyzetű tanulók támogatása a digitális korszakban differenciált pedagógiai válaszokat igényel: digitális mentorrendszerek, személyes visszacsatolás, online kompetenciafejlesztő modulok mind hozzájárulnak az egyenlő esélyek biztosításához.

5. Az intézményi kultúra megújítása

5.1. Jövőtudatos vezetési szemlélet

A jövőtudatos intézményi vezetés a közös célok mentén állandó reflexióra, együttműködésre és a változás menedzselésére épül. Előtérbe kerül a mentori támogatás, a tanárok közös fejlesztése, az AI-EdTech rendszerek kipróbálása, a pilot projektek indítása és tapasztalatainak szervezett megosztása. Fontos az önértékelési rendszerek, a tanár-diák-szülő visszacsatolások

beépítése, a nyílt tudásmegosztás és a folyamatos szakmai fejlődés támogatása. (Zuti 2024c) A fenntartható, jövőtudatos intézményi kultúra kialakítása folyamatos reflexiós, értékelési, visszacsatolási és fejlesztési folyamat. Ennek motorja a vezetés: a tanárok mentorálása, az AI-rendszerek kipróbálása és a pilot projektek tapasztalatainak megosztása. Nyitott tudásmegosztásra, konstruktív partnerekre és egyéni autonómiára éppúgy szükség van, mint a digitális biztonság és etika garantálására. (Selwyn 2019)

5.2. Operacionalizált bizalom

A bizalom nemcsak deklarált elv, hanem operacionalizált, a mindennapi működésben érvényesülő szervezeti norma is. Ez magában foglalja az átlátható döntéshozatali mechanizmusokat, az adatkezelési eljárásokat, a tanulási folyamatok követhetőségét és a felelős, etikus szakmai magatartás normáit. A bizalom operacionalizálása azt jelenti, hogy szervezeti normává válik az átlátható döntéshozatal, az adatkezelési szabályozottság, a felelős információáramlás és az etikus szakmai magatartás. Ezek nélkül az AI-EdTech fejlesztések nem vezetnek eredményes tanulási- tanítási folyamatokhoz, sőt, akár bizalmi válságot is generálhatnak mind a tanulók, mind a szülők körében. (Németh 2022)

5.3. Felelősség, együttműködés, szakmai fejlődés

Az AI-EdTech és digitális transzformáció közepette a felelős cselekvés, a tanári és tanulói autonómia támogatása, a szakmai közösségek, tudásmegosztó műhelyek, valamint a nemzetközi hálózatosodás (pl. European Digital Education Hub) döntő jelentőségű. Az oktatási szféra együttműködése más ágazatokkal erősíti a rendszerszintű innovációt, és hozzájárul a versenyképes, inkluzív, fenntartható tudástársadalom megalapozásához.

6. Kihívások és lehetőségek az AI-EdTech alkalmazásában

6.1. Tanulói és tanári kompetenciák fejlesztése

A tanulók és tanárok digitális és mesterséges intelligencia alapú kompetenciáinak fejlesztése folyamatos, élethossziglan tartó tanulási feladat. Szükséges a digitális írástudás kiterjesztése, a tanulók probléma- és feladatmegoldó, kritikus gondolkodási és együttműködési készségeinek fejlesztése, a tanári AI-transzformációs készségek erősítése. (Lynch 2018)

6.2. Digitalizáció határai és a pedagógiai értékek megtartása

A digitális korszak oktatásának határa a „humanizált technológia”, vagyis a tanulói szükségleteket, egyéni fejlesztési lehetőségeket, érzelmi és kapcsolati igényeket tiszteletben tartó, etikus, méltányos pedagógiai gyakorlat. Az AI-EdTech innovációk mellett meg kell őrizni a személyességet, a kapcsolati tanulást és méltóságot, az önreflexiót és közös tanulási élményeket. (Prensky 2012)

A digitalizáció, bármilyen előnyökkel is jár, nem helyettesítheti az emberi kapcsolatokat, a személyességet, a tanulói önreflexiót és a kapcsolati tanulást. Az emberi tényező, a tanári jelenlét, empátia, személyes visszacsatolás és közös tanulási élmény továbbra is a pedagógiai értékrend központi elemei maradnak

6.3. Fenntartható digitális pedagógia

A fenntartható digitális pedagógia a technológiai megújulás, az oktatási minőség, a társadalmi egyenlőség és ökológiai szempontok összehangolását célozza. Kiemelten fontos a környezettudatos digitális eszközhasználat, a digitális karbonlábnyom csökkentése, az adatközpontok energiahatékony működtetése, valamint a fenntarthatóság tantervi integrációja is. (Zuti, P & Menyhárt, E 2024a; Zuti, P & Menyhárt, E 2024b)

Összegzés

A digitális kor, az AI-EdTech és a jövőtudatos oktatás nem csupán technológiai racionalizációt, hanem mély, rendszerszintű szemléletváltást hoz az oktatás minden szintjén. A bizalom, az együttműködés, az etikus adatkezelés, a tanulói, tanári és intézményi autonómia alapvető feltétele a jövőorientált, inkluzív és versenyképes tanulási-tanítási kultúra megteremtésének. (Molnár, Gy 2021; Zuti, P & Szakály, Z 2024)

Az AI és az EdTech gyors fejlődése révén egyre fontosabbá válik a szakmai megújulás, a digitalizáció kritikus reflexiója, a fenntartható intézményi fejlődés. A siker kulcsa a pedagógiai értékek és a modern technológiák összehangolásában, a közös célok kijelölésében, a bizalom operacionalizálásában, valamint a tanulók és tanárok aktív részvételében rejlik.

Az AI-EdTech jövőtudatos és értékelvű alkalmazása csak a résztvevők együttműködésére, bizalomépítésére, folyamatos reflektív gondolkodására és az értékek megőrzésére alapozható. A jövő tudatos alakítása, a kompetens, felelősségteljes, együttműködő szereplők közös tanulása elengedhetetlen nemcsak a digitális kultúra szempontjából. (Zuti 2025)

Az AI-EdTech-alapú oktatás jövője nem csupán technológia kérdése. Az emberi tényező, a bizalom kiépítése, a tanári támogatás és a tanulói aktív részvétel legalább ilyen fontos. Az intézményi kultúra átalakulása, az átlátható, felelős, etikus döntéshozatal és adatkezelés, a felhasználók digitális kompetenciáinak fejlesztése, valamint a fenntarthatóság biztosítása lehetnek a siker indikátorai. Az AI-EdTech jövőtudatos és értékelvű alkalmazása csak a résztvevők együttműködésére, bizalomépítésére, folyamatos reflektív gondolkodására és az értékek megőrzésére alapozható. (Zuti 2024d)

Irodalomjegyzék

Bostrom, N. 2014, *Superintelligence – Paths, Dangers, Strategies*. Oxford University Press.

European Commission. 2022, Digital Education Action Plan 2021-2027.

Kováts, G. 2018, 'A sokszínű bizalom' *Educatio*, Vol. 27. 4. pp. 531–547.

Kozma, T. 2019, 'Jövőorientált tanulás és tanulói kompetenciák' *Magyar Pedagógia*, Vol. 119. 4. pp. 355-372.

Lynch, M. 2018, How Artificial Intelligence Is Already Transforming Education.

Molnár, Gy. 2021, *Pedagógia, innováció, technológia, digitális kultúra – a digitalizáció új irányai*. Typotex.

Németh, B. 2022, 'A pedagógiai bizalom stratégiái' *Educatio*.

Pintér, G & Nemes, G. 2024, *Jövőtudatos gondolkodás*. Mind Mate Inspiration.

Prensky, M. 2012, *From Digital Natives to Digital Wisdom: Hopeful Essays for 21st Century Learning.* Corwin.

Rousseau, D M, Sitkin, S B, Burt, R S & Camerer, C. 1998, 'Not so different after all: A cross-discipline view of trust' *Academy of Management Review*, Vol. 23. 3. pp. 393–404.

Selwyn, N. 2019, *Should Robots Replace Teachers? AI and the Future of Education*. Polity Press.

Stone, James V. 2019, *Artificial Intelligence Engines: A Tutorial Introduction to the Mathematics of Deep Learning*. Sebtel Press.

Stuart, R & Peter, N. 2020, *Artificial Intelligence: A Modern Approach*. Pearson, 4. kiadás.

Szűts, Z. 2020, *A digitális pedagógia elmélete*. Akadémiai Kiadó.

Zuti, P & Bóczi, Zs. 2023, *A digitális kultúra és az alfa generáció*. Digitális Témahét. 2023. március 27-31. Budapest. https://www.erzsebetvarosiiskola.hu/?p=3448

Zuti, P & Gulyás, Zs. 2022, *A digitalizáció és felelősségteljes innováció a jövőtudatos képzési folyamatban – Hatása a versenyképességre és a vonzóképességre*. 28th Multimedia in Education Online Conference XXVII. Multimédia az oktatásban online nemzetközi konferencia. Conference Proceedings. 15-20.o. DOI: 10.26801/MMO.2022.1.028. 2022. július 6-7. Eszék. Horvátország. Zuti, P & Gulyás, Zs. Megjelenés alatt, *Digitization and responsible innovation in the future-oriented training process – Its impact on competitiveness and attractiveness.* Journal of Applied Multimedia. ISSN 1789-6967.

Zuti, P & Harangozó, H. 2023, *A digitalizáció és felelősségteljes innováció*. Digitális Témahét. 2023. március 27-31. Budapest. https://www.erzsebetvarosiiskola.hu/?p=3448

Zuti, P & Hegedűs, I. 2023, *A digitális kultúra 7 dimenziója a Technika és tervezés tantárgy tanításában (Best practices).* Digitális Témahét. 2023. március 27-31. Budapest. https://www.erzsebetvarosiiskola.hu/?p=3448

Zuti, P & Menyhárt, E. 2024a, *A digitalizáció, a jövőtudatosság és a digitális kultúra dimenziói a Budapesti Műszaki SZC Neumann János Informatikai Technikumban*. 30. JUBILEUMI Multimédia az oktatásban nemzetközi konferencia. MTA Humán Tudományok Kutatóháza.

Zuti, P & Menyhárt, E. 2024b, *Digitális világ: next level a szakképzésben*. XVI. TANÍ-TANI nemzetközi konferencia. Miskolci Egyetem Bölcsészet- és Társadalomtudományi Kar - Magyar Pedagógiai Társaság.

Zuti, P & Sióréti, G. 2024, *Felnőttképzés 4.0 - Tananyagfejlesztés a digitális korban*. 20. Nemzeti és Nemzetközi Lifelong Learning Konferencia. Óbuda Egyetem.

Zuti, P & Szakály, Z. 2024, *Digitális technológiák és módszerek a jövőtudatos felnőttképzésben*. Felnőtt tanulás hete. Orosháza.

Zuti, P. 2023a, *Az adatvédelmi tudatosság fejlesztése. IT- és információbiztonság*. Digitális Témahét. 2023. március 27-31. Budapest. https://www.erzsebetvarosiiskola.hu/?p=3448

Zuti, P. 2023b, *SMART tudásplatform fejlesztése. A digitális Óperencián túl, vagy mégsem*. Multimedia in Education Online Conference XXIX. Multimédia az oktatásban online nemzetközi konferencia. Conference Proceedings. .1.029. 2023. július 6-7. Szeged.

Zuti, P. 2023c, *Adult Education and Training 4.0*. EPALE. Community Story. 2023. október 6. https://epale.ec.europa.eu/en/blog/pal-zuti-adult-education-and-training-40

Zuti, P. 2024a, *A digitális kultúra dimenziói a jövőtudatos képzési folyamatban*. XV. Taní-tani Online Nemzetközi Tudományos Konferencia. Miskolci Egyetem Bölcsészet- és Társadalomtudományi Kar - Magyar Pedagógiai Társaság.

Zuti, P. 2024b, *Felnőttképzés 4.0 - A jövőtudatos felnőttképzés a digitális korszak sodrásában.* XV. Taní-tani Online Nemzetközi Tudományos Konferencia. Miskolci Egyetem Bölcsészet- és Társadalomtudományi Kar - Magyar Pedagógiai Társaság.

Zuti, P. 2024c, *Tudásplatform fejlesztése a digitális korban*. „Tavaszi szél vizet áraszt" konferencia. Gál Ferenc Egyetem.

Zuti, P. 2024d, AI-EdTech és bizalom a jövőtudatos oktatási folyamatban. (Prezentációs absztrakt, kézirat)

Zuti, P. 2025, *Álomiskolám a digitális korban*. XVI. TANÍ-TANI nemzetközi konferencia. Miskolci Egyetem Bölcsészet- és Társadalomtudományi Kar - Magyar Pedagógiai Társaság.

Kontextus a humán-MI kooperációban a szoftverfejlesztés szemszögéből

A nagy nyelvi modelleken alapuló szoftverek fejlesztésének kihívásai az emberi interakció felhasználói elvárásaira figyelemmel

Context in human-AI cooperation from a software development perspective

Challenges in developing software based on large language models with respect to user expectations of human interaction

Dr.[6] Ország-Krisz Axel[7]

Absztrakt

A nagy nyelvi modellek számos hibát vétenek a humán-MI kooperáció során, melyek jelentős része a kontextus téves kezelésén alapul. Ez a dolgozat bemutat kilenc olyan gyakori hibajelenséget, amelyek bármelyikét megtapasztalhatja a felhasználó, ha nagy nyelvi modellen alapuló megoldást használ. A tévedések generált képek segítségével kerülnek bemutatásra a könnyeb megértés segítése végett, de a hibák minden szoftverben hasonló természetűek. Ezt mutatja az is, hogy noha a jelenséggel foglalkozó kutatások csupán néhány éves múltra tekintenek vissza és más-más területet vizsgálnak, megállapításaik mégis egységesek. Erre figyelemmel a dolgozat több megoldást mutat a felhasználók számára a tévedések kezelése érdekében és irányokat is megfogalmaz a fejlesztői oldal felé a kontextus kezeléséhez kapcsolódó anomáliák megszüntetése céljából.

Kulcsfogalmak: nagy nyelvi modell, tévedés a kontextus kezelésében, humán-MI kooperáció, szoftverfejlesztés, prompt készítés

Abstract

Large language models make numerous mistakes during human-AI cooperation. Most of these errors are based on incorrect context management or recognition. This paper presents nine common failures that users may encounter when using large language model (LLM) based solutions. The mistakes are illustrated using generated images to ease understanding. However, all of these issues appear in each software that applies large language models in their workflow. Findings of the related fields of science are consistent even though research on the phenomenon is only a few years old and examines different areas and software. This proves the presence and importance of the demonstrated failures. Therefore, the paper provides several solutions for

[6] Nem tudományos fokozat, jogász doktor

[7] mesterséges intelligencia fejlesztő és szakértő, adattudós, email@drorszagkriszaxel.hu

users to manage these errors and also gives suggestions for developers to eliminate context-related anomalies effectively.
Keywords: large language model, issue in handling context, human-AI cooperation, software development, prompt engineering

Bevezető

Az emberi intelligencia mesterséges, gépesített megvalósítása már az ókorban foglalkoztatta a gondolkodókat, ám a 20. századig kellett várni rá, hogy a mesterséges intelligencia kilépjen a fantáziák és mítoszok világából. A mai értelemben vett MI megalkotását a programozható digitális számítógépek kifejlesztése alapozta meg, mely technológia már az 1940-es években rendelkezésre állt. A folyamat innentől kezdetben igencsak felgyorsult, hiszen a Turing teszt már 1948-ban felvetette a mesterséges intelligencia egy lehetséges határát és 1956-ban, tehát mindössze 16 évvel később, az USA-ban található Dartmouth College főiskolán már el is kezdődött a mesterségesen létrehozott intelligencia megvalósíthatóságának kutatása (Newquist 1994). Azonban a gyakorlati szoftverkészítés még sokáig váratott magára, mivel az elméletek évtizedeken keresztül messze a technikai lehetőségek előtt jártak. 2005-ben, a „big data" fogalmának megjelenésével indult el az a folyamat, mely a mai értelemben vett mesterséges intelligencia megoldások kifejlesztéséhez vezetett (Lecun et al. 2015). A jelen keretek közt vizsgált területhez, a nagy nyelvi modellek hétköznapi alkalmazásához viszont csak a ChatGPT-3 modell 2020-as megjelenése vezetett. Az LLM fejlesztésének ugyanis ez volt az első olyan szintje, amely, az általános célú használhatóság tekintetében áttörést ígért, illetve jelentett (Hariri 2023).

Számos tudományos cikk, dolgozat foglalkozik a mesterséges intelligencia fejlesztésének technológiai korlátaival vagy éppen az MI-nek a világra gyakorolt hatásával ideértve a társadalmi, etikai és gazdasági kérdéseket is. Ugyanakkor a mesterséges intelligenciát, mint szoftverfejlesztési folyamatot a fejlesztő személyek oldaláról vizsgáló kutatás már kevesebb. Ez nyilvánvalóan egyfelől annak tudható be, hogy a terület még nagyon új, másfelől pedig nagyon gyakorlatorientált.

Tekintettel arra, hogy e pillanatban még minden mesterséges intelligencia modell végső soron olyan kód, amelyet emberi kéz alkotott, az emberi faktor és az MI fejlesztés problematikájának emberi megértése kulcsfontosságú ahhoz, hogy olyan megoldásokat fejlesszünk, amelyek valóban azt is úgy végzik el, ahogyan azt az eredeti cél megköveteli.

1.A kontextus, mint probléma

A fogalom kérdéskörének áttekintése

A kontextus az ember számára olyan hétköznapi és természetes, hogy definiálni sem egyszerű. A fogalom meghatározásának kérdésével a nyelvészet és a kommunikációhoz kapcsolódó egyéb tudományok foglalkoznak behatóan és ezek mindegyike területén elmondható, hogy a fogalmak történeti és elméleti szinten is jelentős változatosságot mutatnak. A szerző nem kívánja bemutatni egyik kontextus-fogalmat sem és nem is tesz kísérletet ezek összevetésére, mivel ez kívül esik jelen dolgozat fókuszán. Ugyanakkor azt mindenképpen érdemes megjegyezni, hogy a kontextus, mint a szövegkörnyezet leírója, olyan hatással van a humán-MI interakcióra, mely a jelenleg tárgyalt kérdés alapproblémája.

A kontextus, a fentebb említett komplex meghatározástól elkülönülten, az informatikában is létező entitás, méghozzá egy kifejezetten gyakorlati fogalom. Épp ez utóbbi aspektusára figyelemmel nagyon is konkrét jelentéssel bír, ám az alkalmazás tekintetében már akadnak különbségek az egyes programnyelvekben. A kontextus minden esetben olyan valami, amely a munkakörnyezet egy részét kiemelten írja el. A legtöbb programozási nyelvben a kontextus csak az adatáramlás kezelésére korlátozódik. A Python például a változók életciklusa nyomon követésének egyszerűsítésére alkalmazza a kontextusmenedzsert, mely a használata elején megnyitja, a végén pedig lezárja az érintett adatfolyamot (Van Rossum & Coghlan 2005). Ezzel szemben azok a programnyelvek, amelyek teljesen elszigetelt környezetben futnak, egyfajta kontextust biztosítanak az alaphardver elérésére, mely a korábban bemutatott alkalmazással ellentétben állandóan jelen van, de az egyes felhasználási területekre figyelemmel folyamatosan változó, korlátozott formában érhető el. A kontextust így kezelő környezetre példa az Android vagy az iOS operációs rendszerek és az eszeken használt Java, Kotlin és Swift nyelvek.

A szoftverfejlesztés történetében már viszonylag korán felmerült olyan kód készítése, mely alkalmas arra, hogy az ember szabadszavas módszerrel, ha úgy tetszik „természetes kommunikációval” léphessen kapcsolatba a számítógéppel. Az első ilyen program az 1964 és 1967 között a Massachusetts Institute of Technology (MIT) intézetben fejlesztett Eliza nevű alkalmazás volt, amelyet pszichoterápiás eszközként is sokáig használtak. Ezt a programot tekintik a természetes nyelvfeldolgozás úttörőjének is. Maga a megoldás teljes egészében az angol nyelv merev mondatalkotási szabályainak kihasználására épült, de évtizedeken keresztül volt képes ámulatba ejteni az embereket az értelmesen kommunikáló gép illúziójával. Ez nem is csoda, hiszen, mint azt az Alan Turing at 100 (2012) is kiemeli, Eliza eredeti célja a Turing teszt teljesítése volt.

A tényleges természetes nyelvi kommunikáció megvalósulását azonban csak a nagy nyelvi modellek hozták el és a nyelvészeti értelemben vett kontextus problémája pedig először akkor került elő, amikor 2020-ban a már fentebb említett ChatGPT-3 megjelent. Ez a szolgáltatás volt ugyanis az, amely áttörte az informatika határait és a nagy nyelvi modelleket eljuttatta a felhasználók széles köréhez, olyanokhoz is, akik számára a szoftverek informatikai paraméterei és korlátai nem relevánsak és nem is ismertek. Ebből fakadóan jelentősen kitágult az a nyelvi kontextus, amelyben a szoftvernek, mint szolgáltatásnak helyt kellett állnia (Wang et al. 2024). Figyelemmel arra, hogy maga a probléma is csak néhány éve került fókuszba, az e területet vizsgáló kutatások még nagyon frissek és a megállapítások finomítása, ellenőrzése egy kezdeti fázisban lévő folyamat. Ugyanakkor a kérdés megismerése, elemzése és megértése nagyon is fontos, hiszen a nagy nyelvi modellek által kiszolgált megoldások, épp az egyszerű humán-szoftver interakció ígérete miatt, egyre inkább beköltőnek hétköznapjainkba.

A kontextus szoftvers leképezésének megértése, illetve a leképezés javítása mindkét kommunikációs fél számára nélkülözhetetlen. Az ember, azaz a felhasználó a kontextus-kezelés megértésével hatékonyabban tud kommunikálni a nagy nyelvi modellekkel, míg a fejlesztők a jövőben olyan megoldásokat tudnak majd előállítani, amelyek hatékonyabb válaszokat lesznek képesek adni olyan interakciók esetén is, ahol az ember teljes egészében saját, természetes kommunikációs közegében marad, és a gép minden tekintetben, maradéktalanul alkalmazkodik a felhasználó kontextusához.

2. A kontextus-probléma mibenléte

A kontextus problematikáját gyakorlati módon lehet leginkább meghatározni, és messziről tekintve egy nagyon egyszerű jelenségben foglalható össze; a nagy nyelvi modell és az ember közötti kommunikáció során a szoftver már alapjaiban sem azt a választ adja, amire az ember számít. E megfogalmazás kellőképpen laikusnak hat, hogy figyelemfelkeltő legyen, ugyanakkor minden egyszerűsége ellenére viszonylag pontos is, mivel tartalmazza azt a kritériumot, hogy megalapozottan számít a kommunikáló valamiféle válaszra, tehát minden olyan helyzetet kizár, amelyben a „hiba” a szoftver egyéb sajátosságából vagy a futtató környezet helytelen működéséből, illetve a felhasználói szándékból fakad.

Ahhoz, hogy a fenti állítás értelmezése helyes legyen, érdemes tisztázni, hogyan is működik felhasználói szinten a nagy nyelvi modell munkafolyamat. A felhasználó megfogalmaz valamit a szoftver számára, ezt promptnak nevezzük. A szoftver belső szerkezeti felépítését tekintve egy ún. „pipeline”, amely voltaképp azt jelenti, hogy a szolgáltatás egymásra épülő szerkezeti egységekből épül fel, amelyek együttesen járulnak hozzá a végeredmény kialakításához (Ciu

et.al. 2024). A folyamat lényegének leírására a szállítószalag a legképszerűbb megoldás, hiszen az adat egyik állomásról a másikra halad tovább, miközben átalakul, így a végén elkészül a termék, amely voltaképp a válasz, illetve a reakció az emberi kommunikációra. A prompt tehát bekerül ebbe a folyamatba, előzetes feldolgozáson megy keresztül, amely során a szöveg ún. tokenekké alakul. Ez egy olyan kódolási folyamat, amely a modellek számára értelmezhető formára alakítja az emberi kommunikációt. Ezt követően, mint arra például Li 2024-es munkája is rámutat, nagyon különböző módokon folytatódik az adatok feldolgozása, hiszen a modellek értelmezést és keresést is végeznek, illetve a folyamat második felében generatív ágensek segítségével alakítanak ki egy tokenekből álló választ, amely az utófeldolgozás során válik olyan szöveggé, amit a felhasználó megért. A kontextushoz köthető hibák felmerülhetnek minden olyan fent nevezett fázisban, amelyek tényleges modell akciót tartalmaznak, tehát az elő- és utófeldolgozás kivételével bárhol.

A szerző elsődleges célja a hibás kontextusértelmezés jelenségének bemutatása. Épp ezért e dolgozat keretei közt nem kerülnek tételesen bemutatásra és ütköztetésre a témával foglalkozó tudósok tanulmányai. Ugyanakkor a mélyebben érdeklődő olvasó számára hasznos lehet, ha megismeri Neuman 2024-es, Spector és munkatársai 2022-es, Zhu és munkatársai 2024-es, Patil és Gudivada 2024-es, Kaddour és munkatársai 2023-as, illetve Deng és munkatársai 2024-es munkáit, mert az alábbi felsorolás, illetve a következő fejezet alfejezetei egy kivételével olyan kontextuskezelési hibákat mutatnak be, amelyek valamennyi szerző felismer és jelen dolgozat kereteit messze meghaladóan ismertet is.

A fent említett kutatók összességére is figyelemmel és a szerző tapasztalatai alapján a kontextus-hibák leginkább jellemző területei, melyek bemutatásra is kerülnek:

- Koherencia vagy kohézió hiánya
- Az általános jelentés ismeretének hiánya
- Anafora-feloldás hibája
- Többértelmű kontextus
- Túltanulás
- Szakmai nyelv megértésének hiánya
- A világ korlátozott ismerete

A szándék felismerésének hiánya

A felsoroltakon túlmenően bemutatásra kerül egy további jelenség is, amelynek lényege a válaszkényszerből adódó hiba. A mindenáron válaszadás szükségszerűsége olyan mélyen alapvető tulajdonságnak számít a nagy nyelvi modellek, sőt általában a mesterséges

intelligencia alkalmazási területein, hogy önálló hibaként fel sem merül, pedig a hétköznapivá váló szolgáltatások okán hasznos lenne ezzel a kérdéssel is foglalkozni, hiszen a felhasználók – akaratlanul vagy épp szándékosan – olyan kérdések, kérések, kihívások elé is állítják a szoftvert olykor, amelyekre nincs valós megoldás. Ugyanakkor az MI-vel szemben elvárássá vált, hogy „mindenre tudjon válaszolni". Márpedig, ha nincs „értelmes", illetve valós válasz egy adott promptra, azt a szerző álláspontja szerint célszerű lenne közölni a felhasználóval.

3. A kontextus problémák megoldásának irányai

A fentebb vázolt problémák feloldására két jellemző választ ad az elméleti és a gyakorlati informatika egyaránt. Az egyik az ún. „Retrieval-Augmented Generation", röviden RAG, magyarra frodítva „feltáró kereséssel bővített alkotás", a másik pedig az ún. „kis nyelvi modellek" megalkotása, amelyek csupán egy-egy részterületre fókuszálva teljesítenek kiemelkedően. A Retrieval-Augmented Generation olyan megoldást jelent, amely a betanított adatokon túlmenően a feldolgozással összekötött keresések segítségével igyekszik az általános nagy nyelvi modelleknél pontosabb választ adni (Gupta, Ranjan & Singh 2024). Ezzel szemben a kis nyelvi modellek nem a szolgáltatás részeként, hanem önálló szolgáltatásként működnek már előre meghatározva azt a kontextust, amelyben jól fognak teljesíteni (Heaven 2025).
Mindkét megoldási módra igaz, hogy bőven van még hova tökéletesedniük, illetve az is, hogy jelentős eltéréseket mutatnak a szolgáltatók azon ígéreteivel szemben, mely szerint küszöbön az általános célú mesterséges intelligencia. Ez utóbbi tekintetében fontos azt megjegyezni, hogy egy ilyen ígéretű szolgáltatás minden bizonnyal RAG vagy sok kis nyelvi modell együttesén, esetleg a kettő kombinációján alapul, tehát összességében a belső felépítése mindenképpen távol áll attól, amit jellemzően elvárnánk az általános gondolkodástól és sokkal inkább lesz egy olyan szoftver-láncolat, amely a tudomány nagyon röviden itt is ismertetett állása szerinti építőkövekből áll össze.

4. A kontextus-probléma természete a gyakorlatban

Jelen dolgozat a kontextus kezelésének leggyakoribb problémáit egy egyszerű eszközzel, képgenerálással kívánja bemutatni. A bemutatási folyamat során az első kivételével valamennyi prompt angol nyelven került megfogalmazásra annak érdekében, hogy a vizsgálatból zavaró tényezőként kikerüljön a nyelvek közötti fordítás, mint potenciális hibaforrás.
Fontos megjegyezni, hogy a szerző a felhasznált generatív mesterséges intelligencia modellnek nem fejlesztője vagy üzemeltetője, illetve sem a részletes forráskódot, sem a betanító adatokat

nem ismeri. Ugyanakkor az alább látható jelenségek a nagy nyelvi modelleken alapuló szolgáltatások általános jellemzőit tökéletesen mutatják be, ezen felül az egyes képek küllemére figyelemmel alappal tételezhetjük fel, hogy a kiválasztott szolgáltatás nagy mértékben azokra a szabad felhasználású képi anyagokra támaszkodva került tréningezésre, amelyekre az összes többi modell betanítása is alapul.

5. Koherencia vagy kohézió hiánya

Ez a hiba azt jelenti, hogy a modell nem képes a szavak vagy mondatok, szövegrészek közötti kapcsolatot, összefüggést feltárni. A jelenség a *„bor és seprője”* prompt megadásával kerül bemutatásra.

1. ábra "bor és seprője"

forrás: Magic Studio Art Generator

Mint azt az 1. ábra jól mutatja, a modellnek gondja akadt a „seprő” szó értelmezésével. Maga a szó kifejezetten a bor készítése során képződő salakanyagot, lerakódást jelenti, ugyanakkor az eljárás során a folyamat valamely pontja nem vette figyelembe a szót megelőző „bor” szót, így rendszer a hasonló hangzású és írásmódú „seprű” szót alkalmazta.

A koherencia hiányából adódódó kontextus-hibát nehéz a gyakorlatban megkerülni, mivel a körülírás jellemzően csak ront a helyzeten, hiszen újabb szavak hozzáadásával csak komplikáltabbá válik a generálandó kép, így nagyobb lesz az esély a nem megfelelő tartalom kialakulására.

6, Az általános jelentés ismeretének hiánya

Ez a hiba olyankor jelenik meg, amikor a modell valamilyen oknál fogva nincs tisztában egy szó vagy szóösszetétel hétköznapi jelentésével, illetve e tekintetben elmaradásban van. A jelenség most a *„cell examination”* prompt segítségével tárul fel.

2. ábra "cell examination"

forrás: Magic Studio Art Generator

A fentebb látható 2. ábra jó példája annak, milyen az, amikor a modell betanító adatai a felhasználáskor nem időszerűek. A prompt cella ellenőrzést és sejtvizsgálatot egyaránt jelent. A szerző vélelmezi, hogy hétköznapi értelenben az orvostudományi fogalom gyakoribb beszédtéma, mint a büntetésvégrehajtási jelentés. Ugyanakkor az is nyilvánvaló, hogy a sejtvizsgálat általánossá válásához szükség volt az orvostudomány és a biológia fejlődésére, mely az utóbbi évtizedben egyéb tudományok mellett a mesterséges intelligenciának is volt köszönhető. Főként szerzői jogi okok miatt a betanító adatok nagyrésze ugyanakkor a régmúltból származik, ahol a sejtvizsgálat még szóba sem jöhetett.

Tisztán a képgenerálás területén ez a hiányosság is nehezen oldható fel a felhasználó prompt alakításával, más nagy nyelvi modellek esetében pedig a körülírás miatt fennáll a fókuszvesztés veszélye. Az előző fejezet végén említett Retrieval-Augmented Generation megoldások épp ezeknek az eseteknek a feloldására hivatottak.

7, Anafora-feloldás hibája

Az anafora irodalmi és nyelvészeti fogalom is. Ez utóbbi értelemben az egyik rész másikra utalását jelenti. A nagy nyelvi modellek esetében azt a jelenséget illetik ezzel a névvel, amikor

a modell nem ismeri fel, hogy a szöveg egy része egy korábbira utal. A hiba vizsgálatához a *„pentatonic scale”* prompt került felhasználásra.

3. ábra "pentatonic scale"

forrás: Magic Studio Art Generator

A generált képet a 3. ábra tartalmazza. Az alkalmazott prompt magyar jelentése pentaton skála. A pentaton skála egy zenei fogalom, mely például a magyar népdalokra is jellemző. Önmagában a „pentatonic” szónak képesnek kellene lennie a modellt rávenni arra, hogy zenei tárgykörben alkosson, ugyanakkor a nagy nyelvi modellekre jellemző figyelem központú működés miatt a „scale” szó uralta a kép létrehozását és az előző szóhoz fűződő, az ember számára egyértelmű kapcsolat teljes egészében eltűnt. Az angol scale skálát és mérleget is jelent egyszerre. A mérleg a jelek szerint annyira általános kifejezés, hogy egészen egyértelmű kontextusokat is képes elnyomni.

Mivel az e probléma szemléltetésére kitalált feladat alapvetően kreatív jellegű, a felhasználó hasonló esetben minden bizonnyal valamilyen énekléshez kötődő képet generálna, ezért ez így megoldható. Kommunikációs helyzetben a modell a pentaton skála fogalom előzetes megkérdezésével nagy eséllyel olyan állapotba hozható, ahol a zenei kontextus immár számára is egyértelmű.

7. Többértelmű kontextus

Ez a hiba akkor jelenik meg, amikor egy adott kifejezésnek több, körülbelül azonos súlyú értelmezése van. Itt nem a szakkifejezések ismeretének hiányáról van szó, hanem olyan esetekről, amikor minden különálló jelentés azonos mértékben gyakori és hétköznapi jellegű. A prompt, amely a jelenséget bemutatja a *„rock collection”*.

4. ábra "rock collection"

forrás: Magic Studio Art Generator

A kép, amelyet a 4. ábra tartalmaz, az egyik gyakori jelentést valósítja meg, amely szerint a *„rock collection”* kövek gyűjteménye. Korosztálytól és érdeklődési területtől függően ugyanilyen eséllyel lehetne szó rock zenei stílusú albumok esetleg relikviák összességéről is. A legtöbb modell esetében előzetes kontextus hiányában nagy eséllyel mindkét jelentés előfordul. Ezt a modellekbe épített véletlen alkalmazása biztosítja, mely az ilyen szoftverek kommunikációját teszi emberszerűbbé és kreatívabbá.

Akár kép készítés, akár más nagy nyelvi modell feladat esetében a felhasználó ezt a hibát könnyen ki tudja küszöbölni megfelelő megfogalmazással. A probléma jellemzően inkább abból szokott adódni, az egyes ember számára az általa használt jelentés annyira magától értetődik, hogy eszébe sem jut a másik jelentés létezése, míg nem szembesül a modell „hibájával”.

8. Túltanulás

Túltanulásról akkor beszélünk a mesterséges intelligencia tekintetében, ha egy adott probléma lehetséges megoldásai közül csak bizonyosakat preferál, illetve csak bizonyos helyzeteket képes érdemben vizsgálni és megoldani. A nagy nyelvi modellek területén ez olyankor fordulhat elő, ha a betanító adatokat egy-egy területre fókuszálva kiemelten kezelik. A *„welcoming wave”* prompt egy ilyen jelenséget mutat.

5. ábra "welcoming wave"

forrás: Magic Studio Art Generator

A fenti 5. ábra tartalma kifejezetten meglepő, hiszen a prompt magyar jelentése üdvözlő integetés. Ráadásul ez a szókapcsolat aligha értelmezhető másképpen. A modell esetében itt nagy valószínűséggel arról van szó, hogy a hullám egy vízi, természeti kontextust teremtett.

Fontos megjegyezni, hogy a túltanulás a gyakorlatban semmilyen hasonlóságot sem mutat a az anafora jelenséggel, noha jelen dolgozat keretei között hasonlónak tűnhetnek, hiszen egy-egy jelző szerkezet téves értelmezésén alapulnak. Ugyanakkor az anafora esetében egy a kontextust orientáló, szakterületi értelemmel bíró első szóról van szó, míg jelen esetben egy teljesen általános értelemmel bíró szerkezet értelmezése volt teljes egészében sikertelen, téves.

A túltanulás ellen a felhasználó nem tud védekezni, ez tipikusan olyan terület, ahol a fejlesztőnek kell bölcsen mérlegelnie, az adott modellt mire és milyen mértékben tanítja, illetve hogyan ismeri fel a modellben az esetlegesen előforduló túltanulást.

9. Szakmai nyelv megértésének hiánya

Ez a hiba akkor jelenik meg, ha kifejezetten szakmai jellegű feladatot adunk egy nagy nyelvi modellnek, de az egyes terminus technicusokat nem ismeri fel a szoftver, nem figyel rájuk. A hiba bemutatásához a *„root path”* prompt került alkalmazásra.

6. ábra "root path"

forrás: Magic Studio Art Generator

A 6. ábra tartalma egy kissé szűrreális alkotás. A képen egy a gyökereit a föld felszínén tartó fa van, amelyhez egy út vezet. A prompt magyar tükörfordítása gyökér útvonal. A jelek szerint a modell is így értelmezte. Ugyanakkor ennek a szavak szintjén megragadó meghatározásnak valójában semmi értelme sincs. Az informatikában viszont ez a szókapcsolat egy adott fájlrendszerben található legfelsőbb elérési pontot jelenti. A fenti prompt esetében indokolt lett volna ezt valamilyen módon ábrázolni, vagy legalább valamilyen informatikai kontextusra jellemző rajzot alkotni.

A modell ismereteinek hiányosságain alapuló hibák kezelése felhasználói szinten voltaképp lehetetlen. Meg lehet próbálkozni ugyan körülírással, melynek hátulütői itt is ugyanazok, amelyek korábban már említésre kerültek.

10. A világ korlátozott ismerete

A nagy nyelvi modellek jellemzően azért ismerik korlátozottan a világot, mert tanulási folyamataik valamilyen szempontból kötöttek. Itt két nagy eset fordulhat elő, a rendelkezésre álló számítási kapacitás vagy az adat képezi a korlátot. Jellemző, hogy többnyire a felhasználható adat fogy el. A *„networking”* prompt elég jól demonstrálja a hibát.

7. ábra "networking"

forrás: Magic Studio Art Generator

A kép, amelyet a 7. ábra tartalmaz elfogadható jó megoldásnak, hiszen a networking hálózatépítést jelent. Ugyanakkor a környezet semmiképpen sem modern, ami azért értelmezhető hibaként, mivel a prompt semmi olyan szöveget nem tartalmazott, amely a historikus jelleget indokolná.

Ha a felhasználó a világ korlátozott ismeretével találkozik egy-egy nagy nyelvi modell esetében, minden kockázat ellenére is érdemes lehet a körülírással kísérletezni, mivel például a korszak olyan tényező, amit prioritásként kezel a modell és alakítja jelen esetben a kép megjelenését. A példán túlmutatva érdemes megjegyezni, hogy más szolgáltatások esetében akár egészen más jellemzők is működhetnek így. A modell korlátozott ismeretét olyan jellemző, amelyet a felhasználónak érdemes kiismernie, hogy saját munkáját megkönnyítse, eredményesebbé tegye.

11. A szándék felismerésének hiánya

Ez a jelenség alapvetően azért fordul elő, mert a mesterséges intelligencia alapú rendszerek nem képesek felismerni és értelmezni a szándékot, épp ezért ezt a hiátust az adott helyzetben általánosnak tekintett szándék vélelmezésével orvosolja a szoftver. A hiba a *„creative model”* prompt segítségével kerül bemutatásra.

8. ábra „creative model”

forrás: Magic Studio Art Generator

Az itt látható 8. ábra egy olyan kép, amelyen a *„model”* szót a nőalak, míg a *„creative”* szót a kezei környékén található tárgyak jelenítik meg. Figyelemmel a dolgozat tárgyára a szándék itt a kreatív modellek ábrázolási kísérletére irányult, mivel maga a prompt magyar fordítása is ez. Akárcsak számos korábbi esetben, itt is a prompt körülírása lehet az a megoldás, amely esetlegesen a felhasználó segítségére lehet a nagy nyelvi modellek alkalmazása során, hogy végeredményt kapjon, amely igazodik elvárásaihoz.

12. A „mindenáron megoldás”, mint hibaforrás

A mesterséges intelligenciát tartalmazó megoldásoknak már a kezdetektől fogva egy sajátos jellemzője, hogy akkor is adnak valamiféle végeredményt, ha a valóságában értelmezési tartományon kívül van számukra a megoldandó feladat. Ez a nagy nyelvi modellek esetében sem téves értelmezés, vagy „hallucináció”. A probléma abból ered, hogy a szoftverek láncolata mindenképp ad „értelmesként” osztályozott végeredményt, mivel egy „nincs valós megoldás” végpont matematikai értelemben nehezen illeszthető a mélytanuló modellekbe. Annak érdekében, hogy ez megmutatható legyen, az összes többi esetben is használt képgeneráló modell ezúttal a *„context”*, azaz kontextus promptot kapta.

9. ábra „context”

forrás: Magic Studio Art Generator

A 9. ábra az utolsó rajz, de tekinthető akár az utolsó utáninak is, hiszen a felhasználó célja jelen esetben az volt, hogy a szoftvert olyan feladat elé állítsa, aminek nincs valós megoldása. Ehhez képest egy folyóparti ház lett a modell válasza.

13. A kontextus-probléma bemutatásának tanulságai

A probléma demonstrálására választott szolgáltatástípus, azaz a képek generálása elég jól használható a kontextushoz kötődő hibák bemutatására, mivel a vizuális megjelenés a szöveges tartalommal szemben alapvetően nyelvfüggetlen. A jelen esetben választott szoftver pedig elég jól megmutatja, a feltárni kívánt hibák miként jelennek meg a gyakorlatban.

Általános tanulságként levonható, hogy a modell bizonytalanság esetén természetet és embert ábrázoló képet hoz létre, mely jellemzően a 20. század első felére jellemző hangulatot áraszt. Az ábrázolás témáit magyarázza, hogy a fent említett témák a leggyakrabban kért tartalmak közé tartoznak (Bie et.al. 2024). Az évszázaddal ezelőtti látványvilágot a szerzői jog határozza meg, mivel a műalkotások jellemzően az alkotók halálát követő 70 év elteltével válnak köztulajdonná.

14. Javaslatok a kontextus-probléma feloldására

A kontextushoz köthető hibákat hosszú távon a szoftverfejlesztőknek kell kiküszöbölniük, ha valóban olyan megoldást kívánnak fejleszteni, amely az embertől nem vár semmilyen előképzettséget, tudást. Ugyanakkor, figyelemmel különösen arra, hogy a nagy nyelvi modelleken alapuló szolgáltatások már most is igen elterjedtek, a felhasználóknak is célszerű

magukat felvértezni olyan ismeretekkel, amelyek a mesterséges intelligencia számára megfelelő nyelvezetet teremtenek. Ez utóbbi már csak azért is fontos, mert ma még nem tudható, mennyi ideig tart, míg ténylegesen kifejlesztésre kerül egy olyan megoldás, amely képes a fent nevesített, illetve az összes többi hasonló hibát tökéletesen, vagy legalább nagyobbrészt kezelni.

15. Mit tehet a felhasználó?

A felhasználó elsődleges feladata ma, hogy ne élő szöveget, hanem promptot fogalmazzon meg. Ez a gyakorlatban azt jelenti, hogy a kommunikáció során csak annyire érdemes szofisztikáltan fogalmazni, amennyire feltétlenül szükséges, hiszen a megfogalmazásban rejlő részletek a gyakorlatban elvesznek, mivel valamelyik fentebb bemutatott kontextus-probléma jó eséllyel „elnyeli” a részleteket is. Ezen felül az sem elhanyagolható, hogy egy emberi értelemben véve alaposan megfogalmazott gondolat bőven tartalmaz olyan logikai és tartalmi fordulópontokat, amely a modellt teljes egészében félreviszi.
A felhasználó számára hasznos tudás az is, ha az általa elérni kívánt hatást olyan szavakkal fejezi ki, amely az adott cél érdekében leginkább általánosan használatos. A terminus technicusokat érdemes hétköznapi fogalmakkal helyettesíteni és azok szakmai sajátosságait, a hétköznapitól való letérését pedig tételesen leírni. Ez különösen képek, videók egyéb tartalmak generálása esetén igaz.

16. Mit tegyen a fejlesztő?

A fejlesztéseknek minél több kontextust fel kell ismerniük és alkalmazkodni kell hozzájuk. Ehhez a már meglévő megoldások tökéletesítésén, de akár teljesen új megközelítéseken keresztül is vezethet út, hiszen a mesterséges intelligencia és a nagy nyelvi modellek fejlesztése még teljes egészében abban a szakaszban van, amikor könnyen elképzelhető, hogy az iparágat és a tudományterületet alapvetően megváltoztató felfedezést tesznek a kísérletezők, a kutatók.

Összefoglalás

A nagy nyelvi modellek a humán interakciók során kapott emberi kommunikációt nem minden esetben értelmezik úgy, ahogyan azt a felhasználók elvárják. Ezen helyzetek jelentős hányadát teszik ki azok az alkalmak, amikor a megértés hibája a kontextus téves értelmezésén alapul. Jelen dolgozat arra tett kísérletet, hogy a kontextus értelmezésének legjellemzőbb hibáit egy kép generáló modell segítségével mutassa be. Az ábrákon található kilenc kép mindegyike egy-egy tipikus hibát jelenít meg.

A szakirodalmi áttekintés során megállapítható, hogy a nagy nyelvi modellek korlátait feltérképező kutatás igencsak friss, ám eredményei mégis azonos irányokba mutatnak. Ez a tény azt is jelenti, hogy az a körülmény mely szerint az eltérő munkafolyamatokat alkalmazó modellek hasonló hibákkal küzdenek arra enged következtetni, a tévedések fennállása a mesterséges intelligencia és a nagy nyelvi modellek alapvető matematikai sajátosságaira vetíthetők vissza. Ez jó eséllyel azt is jelenti, hogy a fejlesztői oldal akkor fog a kérdésekre megoldást találni, ha olyan koncepcióval áll elő, amely a létező matematikai korlátokat meghaladja, illetve a gátakat áttöri.

Tekintettel arra, hogy a nagy nyelvi modellek használata már most is mindennaposnak mondható és a feltárt hibák minden olyan szolgáltatásban jelen vannak, amelyek ilyen megoldásokat alkalmaznak, a felhasználók nem engedhetik meg maguknak, hogy megvárják, míg a fejlesztői oldal és a tudomány megoldja a kérdést. Épp ezért arra is nagy mértékben szükség van, hogy a használók megismerjék e szoftverek hibáit és igyekezzenek hozzá alkalmazkodni, megkerülni azokat mindaddig, míg a mainál sokkal jobban működő megoldások nem kerülnek a piacra.

A mesterséges intelligenciát emberek fejlesztik emberek számára, épp ezért a humán-MI kooperáció közvetetten ember-ember kooperációt is jelent. Ezt mindkét oldalnak figyelembe kell vennie a nagy nyelvi modellek és a velük való munkavégzés sikeréhez egyaránt.

Irodalomjegyzék

Alan Turing at 100, 2012, Harvard Gazette. 13 September 2012.

Bie, F, Yang, Y, Zhou, Zh, Adam Ghanem, A, Zhang, M, Yao, Zh, Wu, X, Holmes, C, Golnari, P, Clifton, D.A, He, Y, Dacheng Tao, D, Song, Sh. L, 2024, RenAIssance: A Survey Into AI Text-to-Image Generation in the Era of Large Model, IEEE, ISSN: 1939-3539

Cui, J, Zhao, Y, Yu, Ch, Huang, J, Wu, Y, Zhao, Y, 2024, Code Comprehension: Review and Large Language Models Exploration, IEEE, ISBN:979-8-3503-7434-6

Deng, J, Shen, Zh, Wang, B, Su, L, Cheng, S, Nie, Y, Wang, J, Yin, D, Ma, J, 2024, FltLM: An Intergrated Long-Context Large Language Model for Effective Context Filtering and Understanding, elérhető itt: https://arxiv.org/abs/2410.06886

Gupta, Sh, Ranjan, R, Singh, S.N, 2024, A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions, elérhető itt: https://doi.org/10.48550/arXiv.2410.12837

Heaven, W.D, 2025, Small language models: 10 Breakthrough Technologies 2025, elérhető itt: https://www.technologyreview.com/2025/01/03/1108800/small-language-models-ai-breakthrough-technologies-2025/

Hariri, W. 2023, Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing, elérhető itt: https://doi.org/10.48550/arXiv.2304.02017

Kaddour, J, Harris, J, Mozes, M, Bradley, H, Raileanu, R, McHardy, R, 2023, Challenges and Applications of Large Language Models, elérhető itt: https://arxiv.org/abs/2307.10169

Lecun Y, Bengio Y & Hinton G, 2015 „Deep learning" Nature, 2015, 521 (7553), pp.436-444.

Li, X, 2024, A Survey on LLM-Based Agentic Workflows and LLM-Profiled Components, elérhető itt: https://doi.org/10.48550/arXiv.2406.05804

Magic Studio Art Generator, 2024, elérhető itt: https://magicstudio.com/ai-art-generator/

Neuman, Y, 2024, AI for Understanding Context, Springer Cham, ISBN: 978-3-031-64209-8

Newquist, H.P. 1994, The brain makers, 1. kiadás, Sams Publisher, Indianapolis, IN, USA, ISBN: 0672304120

Patil, R, Gudivada, V, 2024, A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs), MDPI Applied Science, elérhető itt: https://doi.org/10.3390/app14052074

Spector, A. Z, Norvig, P, Wiggins, C, & Wing, J. M, 2022, Data science in context: Foundations, challenges, opportunities, Cambridge University Press.

Van Rossum G, Coghlan A. 2005. PEP 343 – The "with" Statement, elérhető itt: https://peps.python.org/pep-0343/

Wang, Z, Chu, Z, Doan, T.V. et al., 2024, History, development, and principles of large language models: an introductory survey, AI Ethics (2024), elérhető itt: https://doi.org/10.1007/s43681-024-00583-7

Zhu, Y, Antony, Moniz, J.R.A, Bhargava, Sh, Lu, J, Piraviperumal, D, Li, S, Zhang, Y, Yu, H, Tseng, B, 2024, Can Large Language Models Understand Context?, elérhető itt: https://arxiv.org/abs/2402.00858

A kreativitás matematikája: a mesterséges alkotás határai

The Mathematics of Creativity: The Limits of Artificial Creation

Vécsey Richárd Ádám[8]

Absztrakt

A tanulmány a generatív mesterséges intelligencia kreativitását és annak matematikai mérhetőségét vizsgálja. A kutatás célja, hogy feltárja az MI által létrehozott alkotások egyediségének határait, összehasonlítva az emberi kreativitással. A kreatív teljesítmény elemzésére különböző mérőszámok és benchmarkok szolgálnak, beleértve a Kullback-Leibler divergenciát, a Shannon-entrópiát, valamint az MI-képességeket vizsgáló egységes adatbázisokat és tesztrendszereket. A tanítás során alkalmazott veszteségfüggvények és az adatbázis-alapú értékelési módszerek különböző megközelítéseket kínálnak az alkotások kreativitási szintjének mérésére. Bár az MI képes új tartalmak generálására, alkotásai szigorúan az előzetesen tanult minták extrapolációján alapulnak. A mesterséges kreativitás különösen az emberi és gépi együttműködésben bontakozik ki, ahol az MI támogató eszközként működik. A kutatás megállapítja, hogy bár az MI technikailag kifinomult alkotásokat hozhat létre, az emberi kreativitás spontaneitása és intuíciója továbbra is egyedülálló marad. A jövőbeli kutatásoknak arra kell összpontosítaniuk, hogy milyen módszerekkel lehet az MI kreativitását mélyebb szinteken modellezni, valamint hogyan fejleszthetők a generatív rendszerek az emberi gondolkodásmódhoz közelebb álló alkotási folyamatok irányába.

Kulcsfogalmak: generatív mesterséges intelligencia, kreativitás, matematikai modellezés, veszteségfüggvény, teljesítménymérés

Abstract

The study examines the creativity of generative artificial intelligence and its mathematical measurability. The research aims to explore the boundaries of uniqueness in AI-generated creations, comparing them to human creativity. Various metrics and benchmarks are used to analyze creative performance, including Kullback-Leibler divergence, Shannon entropy, and unified databases and test systems assessing AI capabilities. The loss functions applied during training and database-based evaluation methods offer different approaches to measuring the

[8] dr. Vécsey Richárd Ádám jogász, adattudós, mesterséges intelligencia fejlesztő és tanácsadó, vecsey.richard@gmail.com

creativity level of generated works. While AI is capable of generating new content, its creations are strictly based on the extrapolation of previously learned patterns. Artificial creativity particularly flourishes in human-machine collaboration, where AI works as a supportive tool. The research concludes that while AI can produce technically sophisticated creations, the spontaneity and intuition of human creativity remain unique. Future research should focus on methods for modeling AI creativity at deeper levels and on how generative systems can be developed to align more closely with human thought processes in the creative domain.
Keywords: generative artificial intelligence, creativity, mathematical modeling, loss function, benchmarking

Bevezető

A kutatás célja annak feltérképezése, hogy jelenleg hol tart a generatív mesterséges intelligencia modellekkel (GAN) történő alkotási folyamat, ennek milyen határai vannak, illetve milyen módon mérhető az újonnan elkészült alkotások egyedisége. Ez segít megalapozni olyan kutatásokat, amelyek a lehetséges továbbfejlesztési irányokat vizsgálják. A kutatások során számbavételre kerülnek a kreativitás méréséhez használt adatbázisok és képletek is.
Terjedelmi korlátok miatt célként kizárólag egy összegző, a helyzetet nagyvonalakban áttekintő tanulmány létrehozása vállalható. További célként jelenik meg a humán kreativitással való összevetés, a párhuzamok és esetleges kollízió meghatározása.
A kutatásnak nem célja új tudományos módszertanok kidolgozása, így nem célja új betanítási módszertanok kidolgozása vagy új modell struktúrák létrehozása. Nem célja továbbá olyan új és egyedi matematikai képleteket létrehozása sem, amelyek segítségével a modellek által létrehozott kijövetek kreativitási faktora vagy újdonság hatása mérhető lenne.

Módszer

A téma interdiszciplináris jellege miatt nagyon nehezen átfogható egyetlen kutatásban, mert teljesen eltérő tudományágak összekapcsolására lenne szükség a matematikától az informatikán, a biológián és a pszichológián át, a művészettudományon keresztül egészen a filozófiáig. Ezen tudományterületek mindegyike másként közelíti meg a tudat, a kreativitás és az alkotás fogalmát. Annak érdekében, hogy egy digitális alkotás újdonságát vizsgálhassuk és összevethessük az emberi és mesterséges alkotási folyamatokat, szükség lenne a humán kreativitás minél teljeskörűbb megértésére biológiai és pszichológiai szempontból. Amíg nem vagyunk tisztában az emberi alkotási tevékenység mögöttes folyamataival, nehéz pontos képletet alkotni arra, mi számít egyedinek és kreatívnak.

A kutatás során abból indulok ki, hogy a mesterséges intelligencia modellek, így a generatív modellek hatékony működéséhez is szükséges a megfelelő betanítás. Ehhez szükség van adatokra, melyek az elérni kívánt eredmény szerint lehetnek szöveges, képi vagy hang adatok. Az adatokon túlmenően szükséges a megfelelő hibaszámítási algoritmus meghatározása.
A kutatás nem vizsgálja a különböző modellek struktúrájában rejlő különbségeket, ahogyan a betanítás során használt paraméterek, mint például a tanulási ráta (learning rate) vagy a figyelmen kívül hagyás aránya (dropout rate) sem képezi részét a vizsgálatnak. Nem része a vizsgálatnak a betanító adatok összeállítási technikáinak elemzése sem.
A kutatás tehát két nézőpont alapján közelíti meg a vizsgálat tárgyát. Elsősorban a betanítás során használt hibafüggvény alapján, hiszen ez bármilyen mesterséges intelligencia vagy gépi tanulásos modell esetén használatos. Másodsorban pedig a betanítás során használt adatbázisokra, illetve az azokban lévő betanító adatokra fókuszálva. Ez utóbbi megközelítés alkalmazható egy nagy nyelvi modell (LLM) esetében.

Eredmények

1.Fogalom

Az első lépés a hosszú úton a kreativitás meghatározása. Ez nem könnyű feladat, mert személyenként, illetve művészeti és tudományos területenként is eltérő választ adhatunk. Nagy és társai (Nagy et al. 2022) szerint a kreativitás fogalma egy olyan folyamat, amely során újszerű, hasznos az adott feladatnak és körülményeknek megfelelő ötletek, gondolatok és produktumok keletkeznek. Ez a meghatározás meglehetősen tág, ezért nehezen alkalmazható a GAN-ek esetében. A keletkezéshez nem szükséges emberi alkotó akarat, egy modell is tud eredményt keletkeztetni. Az újszerű és hasznos fogalmak mérése meglehetősen körülményes. Szerencsére az adott körülményeknek való megfelelés pontosan mérhető a veszteségfüggvények és előre elkészített teszt adatbázisok segítségével. Fontos megjegyezni, attól, hogy valami kreatív, nem feltétlenül lesz szép vagy esztétikus a megfigyelő számára.

2.Kreativitás tulajdonságai

A fogalom meghatározása után meg kell vizsgálni a kreativitás minőségét és tulajdonságait annak érdekében, hogy megpróbáljuk mérhetővé tenni, illetve megvizsgálhassuk, hogy a létező mérőszámok ténylegesen alkalmasak-e a kreativitás mérésére. A tulajdonságok vizsgálatára a legjobb kiindulási pont az, ha a különböző művészeti területekből indulunk ki. Amennyiben van olyan tulajdonság, amely átível az eltérő művészeti ágakon, az a vizsgálat során úgyis kiderül. Ez a megközelítés azért is célszerű, mert a mesterséges intelligencia modellek is

hasonló módon tagozódnak. Vannak közöttük olyanok, amelyek szöveget generálnak, mások képet vagy egyéb vizuális tartalmat készítenek, míg megint mások zenét komponálnak. A modellek között nincs átjárás, az egyik művészeti ágra történő rátanítás nem teszi lehetővé számukra, hogy más típusú művet alkossanak.

Szövegalkotás terén a kreativitás többféle módon is megközelíthető. Az első a lexikai diverzitás, mely a szövegben használt szavak változatosságát méri. Minél többféle szóból áll egy szöveg, vagyis a szókincs minél változatosabb, annál kreatívabbak tekinthető a szöveg. Ennek informatikai módon történő mérése egyszerű, csupán a szavakat vagy fogalmakat szükséges megszámolni. A magas lexikai diverzitású szöveg érdekesebb már a különböző szavak számosságánál fogva is. Azonban előfordulhat, különösen tudományos szakszövegeknél, hogy nincs helye bizonyos fogalmak szinonimákkal való helyettesítésének. A szöveg tartalma attól még lehet kreatív, hogy ez a használt fogalmakban nem tükröződik.

Egy lépcsővel feljebb lépve a szavaktól eljutunk a mondatokig. A szintaktikai komplexitás a mondatok szerkezetének összetettségét méri. Azt könnyű belátni, hogy a komplexebb mondatok sokkal kreatívabbak lehetnek. Gondoljunk csak arra, hogy az „A macska alszik.” mondatot sokkal kevesebb lehetséges variációban lehet újraalkotni, mint a „A jóllakott macska a kályha melegénél alszik a gazdája pulóverén.” mondatot. Az összetettebb szövegek segítségével az alkotó könnyebben kifejezheti a gondolatait, több hely és idő áll ezen gondolatok leírására és megértésére. Ebből kifolyólag a szintaktikailag komplex szövegek kreativitása is könnyebben mérhető a kevésbé komplex szövegekéhez képest.

Irodalmi alkotásoknál a szemantikai újdonság vizsgálatának is lehet helye azáltal, hogy a használt fogalmak jelentésének újszerűségét mérjük. A szemantikai újdonság mérése nehéz, mivel a jelentések kontextusfüggőek és változhatnak az idővel. Viszont a versekben használt költői képek esetén a fogalmak új jelentéssel való feltöltése a képzettársítások révén valóban jó mérőszáma a kreativitásnak.

A szöveg egészének megítélésekor nem csak az egyes szavakból, hanem azok kapcsolataiból is kiindulhatunk. A stilisztikai eredetiség az elkészült szöveg stílusának egyediségét tudja mérni. Amennyiben egy szöveget képesek vagyunk új, egyedi stílusban elkészíteni, az utal a kreativitásra is. Természetesen kérdésként merül fel, hogy ez a kreativitás mennyire kötött az alkotóhoz és mennyire érhető tetten a szövegben. Abban az esetben, ha csak egy dokumentum áll a vizsgáló rendelkezésére, abból elég nehéz eldönteni, hogy csupán a stílus alapján mennyire kreatív az írott tartalom. Ha például valaki például népies stílusban ír novellákat, de egy jogi ügy miatt el kell készítenie egy jogi szaknyelven és az adott szakmára jellemző fordulatokkal és felépítéssel ellátott szöveget, a feladat abszolválása nagy kreativitást mutat annak ellenére,

hogy ez nem derülne ki akkor, ha nem egy nagyobb kontextust, a szerző korábbi tevékenységét vizsgálnánk. Az viszont kétségtelen, hogy a stílus hasznos lehet akkor, amikor egy mű szerzőjét akarjuk meghatározni.

A narratív koherencia révén mérhetők a szövegben lévő logikai összefüggések. Egy koherens narratíva a kreativitás fontos összetevője lehet. Nagynyelvi modellek esetén a logikai érvelésre külön tesztet is használnak. Talán az is közrejátszik egy magasabb narratív koherenciájú szöveg kreatívnak történő elfogadásában, hogy az ilyen írásos tartalmak könnyebben érthetők és követhetők, a meggyőző erejük sokkal nagyobb.

Vizuális tartalmak esetén már egészen másféle mutatószámokat használhatunk. Az első a kompozíció újdonsága, egyedisége. Ez egy informatikailag meglehetősen nehezen meghatározható mérőszám, hiszen nem az egyedi képpontokat, hanem azok összességét is szükséges vizsgálni. Ráadásul a képzőművészetben vannak tipikusan használt kompozíciók, beállítások, arányok, így ezen mérőszám megítélése jelentősen függ a témától, stílustól és a megfigyelő szubjektív véleményétől is.

A színhasználat eredetisége és harmóniája szintén lehet egy faktor az alkotások kreativitásának mérésekor. Elemzéskor fókuszálhatunk a színek telítettségére, az árnylatokra, az intenzitásra, de arra is, mennyire illeszkedik a használt szín az adott tárgyhoz, stílushoz, a kompozícióhoz, a környezethez és az alkotás témájához.

A témaválasztás önmagában is megfelelő mérőszám a kreativitáshoz, amennyiben a téma újszerű és egyedi. E faktor megítélése nagyon szubjektív, algoritmizálása ebből kifolyólag nehézkes.

A technikai kivitelezés során a képen lévő objektumok minőségét, eredetiségét és részletgazdagságát vagy éppen egyszerűségét vizsgálhatjuk. A textúrákon túl mérhető az élesség, a megfelelő fény-árnyék hatás érzékeltetése, illetve egy adott objektum helyes elhelyezése a térben. A technikai kivitelezés kiemelt szerepet kap, amikor egy kép minőségét akarjuk megítélni, legyen szó akár szintetikus, akár eredeti alkotásról.

A képek által kiváltott érzelmi hatás is jelezhet kreativitást. Ebben az esetben a kiváltott érzelmek intenzitását, változatosságát és minőségét lehet meghatározni. E mérőszám jelentősen függ a befogadó személyétől, így algoritmizálni meglehetősen nehéz. Könnyebb a helyzet akkor, amikor szöveges tartalmak érzelmi töltését kívánjuk meghatározni, mert ott a szavak tényleges jelentéséből és érzelmi töltöttségéből leehet kiindulni. Vizuális vagy hang alapú tartalomnál ez nehéz, mert egy-egy objektum teljesen eltérő érzelmi hatást válthat ki a különböző megfigyelőkből.

Zenei tartalom esetén a dallam újszerűsége és egyedisége, vagyis a melodikus újdonság nagyon jól jellemzi az alkotásban rejlő kreativitást. Természetes korlátot jelent, hogy a különböző zenei irányzatoknak vannak tipikus mintázatai és szerkezeti sajátosságai, amelytől nem, vagy csak kis mértékben lehet eltérni.

A hangok közötti összefüggések vizsgálatával mérhető a harmonikus komplexitás. Minél bonyolultabb, összetett és egyedibb egy harmónia, annál magasabb kreativitási faktort társíthatunk hozzá. Természetesen lehetnek olyan anomáliák is, amikor pont az egyszerűséghez kell nagyon kreatív alkotónak lenni. Ugyanakkor az esetek többségében a komplexebb, zeneileg teltebb harmónia kreatívabb alkotási folyamatot feltételez. A harmonikus komplexitás elemzése során figyelembe vehető a különböző hangok közötti kapcsolat, a hangközök használata és az akkordok variációi, amelyek mind hozzájárulnak az alkotás egyediségéhez.

A zenei művek egyik mindenki által legkönnyebben felismerhető tulajdonsága a ritmus. Ennek változása, összetettsége és eredetisége a kreativitás szintjét is jelzi. A ritmus az egyik legfontosabb elem, amely meghatározza a zene dinamikáját és energia szintjét. Az egyedülálló ritmikai struktúrák és az innovatív ritmusváltások új zenei élményeket hozhatnak létre, amelyek elragadják és meghökkentik a hallgatóságot. A ritmus elemzéséhez használhatunk különböző mutatószámokat, mint például a ritmikai sűrűség, a szinkopáció és a poliritmika mértéke.

A ritmus mellett a hangszerek kombinációja, számossága és változatossága az alkotó kreativitásának jó mércéje. Vannak bizonyos stílusok, melyeknél a felhasználható hangszerek köre szűkebb, de kevés hangszert is lehet kreatívan használni egy zenei műben. Az egyedi hangszerelési megoldások és a különböző hangszerek közötti interakciók új hangzásvilágot hozhatnak létre. A hangszerelés során figyelhetünk a hangszerek dinamikai változásaira, a hangszínek keveredésére és az egyes hangszerek játékstílusának variációira, amelyek mind hozzájárulnak a zenei mű egyediségéhez és eredetiségéhez.

A struktúra és a forma, vagyis a zene felépítése szintén használható a kreativitás mércéjeként. A zeneszerzők különböző formákat és szerkezeti elemeket alkalmazhatnak, hogy létrehozzanak egyedi zenei narratívákat. Egy forma lehet egyszerű vagy összetett, lineáris vagy ciklikus, amely befolyásolja a hallgatók élményét és a zenei mű érzelmi hatását. Az újszerű szerkezeti megoldások és az egyedi formák hozzájárulnak a zenei művek kreativitásának és innovációjának méréséhez.

Generatív modellek esetén a kreativitás vonatkozhat arra is, hogy az adott modell mennyire tér el a számára megadott szöveges utasítástól. Bizonyos vizuális tartalmakat készítő modellek képesek arra, hogy nem csak szöveges, hanem képi utasítást is elfogadnak. Itt a kreativitás szintje tovább mélyül, hiszen nem csak a prompt és a kép közötti arány megtalálása is

befolyásolja a modell kreativitási faktorát, hanem az így létrejövő együttes parancsból szintetizált utasítástól való eltéré szintje is.

3. Adatbázisok

Léteznek olyan adathalmazok, amelyek segítségével a különböző generatív hálózatok képességeinek mérésére szolgálnak. Ezek egy része valamely tudományterületen való jártasságot mér, amely elsődlegesen nem tekinthető kreatív folyamatnak, hiszen a tárgyi tudás meglétét vizsgálja. Ugyanakkor a kérdés-felelet jellegű mérésben benne rejlik a kreativitás is, hiszen a modellnek meg kell érteni a kérdést, majd arra nem csak nyelvtanilag, hanem tudományos értelemben is helyes választ kell adnia. A legismertebb adatbázisok többsége az LLM-ekhez lettek kifejlesztve, de léteznek képi és zenei alkotások mérésére használtak is.

Az MMLU az egyik legismertebb benchmark a nyelvi modellek értékelésére. 57 különböző tárgykört fog át, beleértve a matematikát, informatikát, jogot, fizikát, történelmet, és a közgazdaságtant is. A feladatok többsége feleletválasztós tesztkérdésekből áll, amelyek emberi tesztelések alapján lettek kialakítva. Az értékelési módszer magában foglalja a zero-shot és few-shot teszteket, így vizsgálva a modellek képességét a kontextus nélküli és minimális kontextusból való tanulásra. A teszt célja annak felmérése, hogy egy LLM mennyire képes valódi emberi tudást szimulálni. A benchmark különösen erős abban, hogy széles körű tudást igényel a sikeres teljesítéshez. (Hendrycks et al. 2020)

A GLUE (General Language Understanding Evaluation) egy szintén széles körben használt benchmark, amely az LLM-ek általános nyelvértési képességeit méri, beleértve a szövegértelmezést, következtetéseket, szemantikai hasonlóságot és párbeszédértést is. Ehhez különféle természetes nyelvi feldolgozási (NLP) feladatokat használ. A tesztet 2018 óta használják. A teszt összesen kilenc különböző NLP feladatból áll, amelyek változatos nehézségi szinteket képviselnek, és különböző típusú nyelvértési képességeket mérnek:

1. CoLA (Corpus of Linguistic Acceptability) – Az angol mondatok nyelvtani helyességét értékeli. A Matthew-korrelációs együtthatót (MCC) használja az értékeléshez, ami -1 és 1 között mozog.
2. SST-2 (Stanford Sentiment Treebank) – Egyértelmű szentiment-analízis, ahol a modelleknek pozitív vagy negatív érzelmi töltetet kell azonosítaniuk filmkritikák mondataiban. A pontosságot (accuracy) használja az adott mondat érzelmének (pozitív/negatív) meghatározására.
3. MRPC (Microsoft Research Paraphrase Corpus) – A mondatok közötti hasonlóságot vizsgálja, azaz hogy két mondat jelentése azonos vagy eltérő. Mind az pontosságot. mind az

F1-pontszámot (F1 score) használja, mert a kategóriák eloszlása egyenlőtlen. A minták 68%-a pozitív.

4. QQP (Quora Question Pairs) – Azonos jelentésű kérdéseket kell felismerni egy nagy adatbázisból. Mind a pontosságot, mind az F1-pontszámot használja, mert a kategóriák eloszlása egyenlőtlen. A minták 63%-a negatív.

5. STS-B (Semantic Textual Similarity Benchmark) – Két mondat közötti szemantikai hasonlóság mérését végzi egy skálán. Pearson és Spearman korrelációs együtthatókat használ az 1-től 5-ig terjedő hasonlósági pontszámok előrejelzésére. Az alap pontszámok emberek által annotáltak.

6. MNLI (Multi-Genre Natural Language Inference) – Különböző szövegtípusokon végzett természetes nyelvi következtetés, amelynek eredménye lehet előfeltétel (entailment), ellentmondás (contradiction) és semleges (neutral) kategória. A pontosságot használja az értékeléshez.

7. QNLI (Question Natural Language Inference) – A kérdés-válasz értelmezésének helyességét méri. A pontosságot használja annak meghatározására, hogy a kontextusban szereplő mondat tartalmazza-e a kérdésre adott választ.

8. RTE (Recognizing Textual Entailment) – Természetes nyelvi következtetés kisebb, nehezebb adatkészleten. A pontosságot használja az előfeltevés értékelésére.

9. WNLI (Winograd Schema Challenge) – Szövegértési teszt, amely a névmások, mint referenciák értelmezését vizsgálja. A pontosságot használja annak előrejelzésére, hogy a mondat alanyának kicserélése után valóban helyes-e.

A teszt a megalkotásakor nagyon erős abban, hogy felmérje a modellek megértési képességét. (Wang et al. 2018)

A GLUE teszt viszonylag gyorsan "könnyűvé" vált az új generációs modellek számára, ezért a kutatók létrehozták a SuperGLUE benchmarkot, amely már sokkal nehezebb feladatokat tartalmaz, hogy tovább tesztelje a modellek magas szintű nyelvi megértését és logikai következtetési képességeit. A teszt nyolc összetett NLP feladatot tartalmaz, amelyek mindegyike nehezebb a GLUE megfelelőihez képest:

1. BoolQ (Boolean Questions) – A modellnek meg kell határoznia, hogy az eldöntendő kérdésre a szövegben található információk alapján lehet-e válaszolni.

2. CB (CommitmentBank) – Természetes nyelvi következtetés (NLI) feladat, amelyben a modellnek meg kell állapítania, hogy egy állítás milyen valószínűséggel igaz egy adott szöveg alapján.

3. COPA (Choice of Plausible Alternatives) – Oksági következtetési feladat, amelyben a modellnek ki kell választania két lehetőség közül azt, amelyik logikusabban következik az adott mondatból, vagy amelyikből az adott mondat logikusabban következik.
4. MultiRC (Multi-Sentence Reading Comprehension) – Többmondatos szövegértési feladat, amelyben egy kérdéshez adott válaszokat kell kiértékelni és eldönteni, hogy melyik helyes. Egy kérdéshez több helyes válasz is tartozhat az adatbázisban.
5. ReCoRD (Reading Comprehension with Commonsense Reasoning Dataset) – Gépi olvasásértési feladat, amely az általános világismeret és következtetési képességek tesztelésére épül. A modellnek egy szöveg alapján kell pótolnia hiányzó kulcsszavakat. A nehézsége, hogy mindegyik lehetséges megoldás helyesnek tűnik.
6. RTE (Recognizing Textual Entailment) – A GLUE-ban megismert teszt bővített változata.
7. WiC (Word-in-Context) – Egy adott szó két különböző szövegkörnyezetben való jelentését kell összehasonlítani és eldönteni, hogy ugyanazt jelenti-e vagy sem.
8. WSC (Winograd Schema Challenge) – A modelleknek meg kell határozniuk, hogy melyik névmás illik az adott szövegkörnyezetbe. A sikeres feladatmegoldáshoz nélkülözhetetlen az alapfokú tudás és érvelési képesség megléte.
A SuperGLUE jóval nehezebb, mint a GLUE, mivel a feladatai több emberi logikát, háttértudást és összetettebb érvelést igényelnek. (Wang et al. 2019)
Az adatbázisokon végzett elemzéseknek már betanított modellek esetén van létjogosultsága. Tudományos szempontból érdekes lehet félkész modelleket is tesztelni, de üzleti szempontból ezt sosem éri meg megcsinálni, hiszen időt és pénzt emészt fel. Üzletileg akkor éri meg ezeket a teszteket elvégezni, amikor van néhány lehetségesen jól betanított modell és azok közül akarjuk kiválasztani a számunkra legjobbat. Fontos hangsúlyozni, hogy egy nagyobb adathalmazon vagy tovább tanított modell sem teljesít feltétlenül jobban az elődeinél a teszteken. Ez látható az 1. ábrán lévő átlagolt pontértékeken is, melyeket az OpenAI GPT 4o modellje ért el.

modell	globális	érvelés	kódolás	matematika	elemzés
gpt-4o-2024-	55,33	53,92	51,44	49,54	60,91

08-06					
gpt-4o-2024-11-20	52,19	55,75	46,08	42,87	56,15

Forrás: https://livebench.ai/#/?organization=OpenAI (megnézve: 2025. 01. 20.) alapján saját szerkesztéssel.

Az újabb modell bár érvelésben jobb volt, kódolásban, elemzésben és a matematikai problémák megoldásában is alulmaradt a három hónapnál korábbi verzióhoz képest.

Képgeneráló modellek esetén is léteznek benchmark adatbázisok. A generált képek stílusukban és tartalmukban elég sokfélék lehetnek a valós adatokat tartalmazó adatbázisokhoz képest. A JourneyDB négy mérőszám mérésére alkalmas a generált képek tartalmi és stílusbeli értelmezési teljesítményének mérésére, úgy, mint a prompt inverzió, stílus elemzés, képaláírás és vizuális kérdések megválaszolása. Ezen túlmenően a teszt elvégzésekor természetesen a modell teljesítménye is mérhető. A négy mérőszám segítségével kifejezhető, mennyire képesek a modellek megérteni a szintetikusan generált képeket. (Sun et al. 2023)

Ha nem a megértés a lényeg, hanem inkább a generált és valós tartalmak megkülönböztetése, arra a GenImage adatbázis használható. Ez a teszt rengeteg hamis képet tartalmaz a nagy képgeneráló modellektől az eredeti képek mellett. Természetesen egy ilyen adatbázissal nem csak olyan modell fejleszthető, amely megkülönbözteti a valós és a szintetikus képet, hanem a generált tartalmak minősége is javítható, hogy hatékonyabban átverjék a detektáló modelleket. (Zhu et al. 2023)

Generált zenei alkotások esetén is megjelent az igény a benchmark adatbázisokra. Ezek mennyisége és minősége a tanulmány írásakor jelentősen elmarad a másik két témakör adatbázisaihoz képest. Solak és társai (Solak et al. 2024) például a kutatásukhoz saját adatbázist készítettek 6000 szintetikusan készített zeneszámból.

4. Képletek

A modellek betanítása során azt vizsgáljuk, hogy a gép által generált tartalom és a célként elérendő tartalom között mekkora a távolság, vagyis mekkora a hiba. Ehhez többféle függvényt használhatunk attól függően, hogy milyen a modellünk és milyen tartalmat akarunk készíteni.

A Cross-Entropy Loss (keresztentrópia-veszteség) veszteségfüggvényt széles körben alkalmazzák szekvenciális adatoknál, például nyelvi modelleknél és képalapú modelleknél. Segít minimalizálni a generált adat és a valós adat közötti különbséget, ezáltal növelve a generált szöveg vagy kép valóságszerűségét.

A Mean Squared Error (átlagos négyzetes hiba) veszteségfüggvényt gyakran alkalmazzák képgenerálásban és rekonstrukciós feladatokban, mint például autoenkóderek és GAN-ek esetében. Az MSE minimalizálása segít abban, hogy a generált képek minél jobban hasonlítsanak az eredeti képekre.

A Perceptual Loss (érzékelési veszteség) veszteségfüggvényt használják képek és videók generálásánál, hogy javítsák a generált tartalom vizuális minőségét. Az érzékelési veszteség a magasabb szintű jellemzők közötti különbséget méri, amelyeket egy előtanított hálózat, például egy ResNet vagy VGG-hálózat készít. (Lin, S, Yang, X. 2024)

A KL-Divergence (Kullback-Leibler divergencia) veszteségfüggvényt a Variational Autoencoder (VAE) modellek esetében használják. Ez segít a generált eloszlás és a valós eloszlás közötti különbségek minimalizálásában, vagyis általános értelemben véve a modell ügyesebb, robosztusabb lesz, miközben az egyedileg előállított képek minősége is magas marad. (Asperti, A, Trentin, M. 2020)

A megfelelő veszteségfüggvény megtalálása általában a betanítást megelőző tesztelési fázis feladata. Az előbb felsoroltaknál jóval többféle létezik, melyek akár vázlatos tárgyalása is túlfeszítené jelen tanulmány kereteit. Hovatovább, a klasszikus veszteségfüggvények mellett a fejlesztők használnak egyedileg programozott funkciókat, hogy az egyedi mérőszámok révén még hatékonyabb eredményt érjenek el. (Ebert-Uphoff et al. 2021)

A korábban bemutatott mérőszámok jelentősen eltérnek a fejlesztés során alkalmazott veszteségfüggvényektől. Ennek oka, hogy informatikailag nagyon nehéz a kreativitás, mint fogalmat megragadni és lekódolni. A betanítás során ugyanis az első és legfontosabb cél, hogy élvezhető eredményt kapjunk. Amíg nincs szemmel látható és könnyen felismerhető eredménye egy modellnek, addig azt nehéz értékelni. Amennyiben azt kérjük egy képgeneráló alkotástól, hogy rajzoljon egy kutyát, de csak színes zajt kapunk, akkor lehet akármilyen szép is a zaj, az utasításunkat a modell nem hajtott végre. Elsődlegesen tehát a végeredménynek igazodnia kell

a parancshoz. Ezután az eredménynek jó minőségűnek kell lennie. Harmadlagos célként jelenhet meg a kreativitás. A kreativitást ugyanakkor jelentősen befolyásolja az, hogy a modell a betanítás során milyen adatokkal találkozott. Fontos megjegyezni, hogy bár a GAN működési elvéből következik, hogy a véletlenszerűségnek rendkívüli hatása van az eredményre, ahogyan azt Zhu és társai (Zhu et al. 2024) elemezték, a véletlenszerűen generált zaj ellenére a modellek képtelenek olyan dolgot készíteni, amivel még nem találkoztak. Szemléletes példa, amikor egy modellt úgy tanítunk be, hogy nem látott egyetlen képet sem kutyákról, majd megkérjük, hogy rajzoljon egy kutyát. Végeredményt fogunk kapni, de az általa készített kutya reprezentáció is olyan alapsémákból áll majd össze, amelyeket a betanítás során látott.

Az LLM modellek esetén az egyik legkönnyebben alkalmazható mérőszám, a Shannon-féle entrópiafüggvény (Shannon 1948).

$$H(X) = -\sum_{x \in X} p(x_i)(x_i)$$

*Ez a képlet e*gy véletlenszerű változó vagy egy üzenet bizonytalanságának, információtartalmának a mérésére szolgál. Minél nagyobb egy rendszer entrópiája, annál nagyobb a bizonytalanság vagy az információtartalom. A nagyobb entrópia jelenthet magasabb kreativitási szintet. Tegyük fel, hogy két szöveget vizsgálunk. Az első szövegben a szavak gyakorisága egyenletes eloszlású, míg a második szövegben néhány szó sokkal gyakrabban fordul elő, mint a többi. Ebben az esetben az első szöveg entrópiája nagyobb lesz, mint a második szövegé. Ez azt jelenti, hogy az első szövegben nagyobb a rendezetlenség vagy bizonytalanság, ami a kreativitás jele lehet.

Összefoglaló

A tanulmány célja a generatív mesterséges intelligencia kreatív képességeinek vizsgálata, különös tekintettel az alkotási folyamat matematikai és algoritmikus megközelítésére. A kutatás rámutat arra, hogy bár az MI képes új tartalmak előállítására, ez elsősorban a tanult minták extrapolációján alapul, és nem egy önálló kreatív gondolkodási folyamat eredménye. Az emberi kreativitás olyan elemeket foglal magában, mint az intuíció, a kontextusfüggő gondolkodás és az érzelmekkel átszőtt alkotási folyamatok, amelyek jelenlegi mesterséges rendszerekkel nehezen reprodukálhatók. Ezen túlmenően nehéz alkotni olyan veszteségfüggvényt, amely pontosan képes leképezni a kreativitás fogalmát a matematika nyelvére úgy, hogy abból nem csak egyértelműen megmondható legyen, egy alkotási folyamat végeredménye kreatív-e vagy sem, hanem a betanítási folyamat során a modell képes legyen ebből úgy tanulni, hogy az egyes

tanulási ciklusok mind kreatívabb generálási folyamatot tegyen lehetővé. Az MI kreativitásának vizsgálata során számos mérőszám alkalmazható, többek között az entrópia, és a Kullback-Leibler divergencia, amelyek a generált tartalmak újdonságát és összetettségét hivatottak mérni. A kutatás foglalkozik a benchmarkok szerepével is, amelyek lehetővé teszik az MI kreatív teljesítményének objektív kiértékelését. Bemutatásra kerültek olyan tesztrendszerek, mint a GLUE és a SuperGLUE, amelyek nyelvi modellek képességeit mérik. Ezen túlmenően megemlítésre kerültek a vizuális és zenei tartalomgenerálás területén használt adatbázisok is.
Az MI által generált tartalmak minősége jelentősen függ a tanítási adatok sokszínűségétől és a tanulási módszerektől. A tanulmány implicit módon rámutat arra, hogy az MI kreativitása korlátozott abból a szempontból, hogy nem rendelkezik valódi inspirációval vagy eredeti szándékkal, hanem csupán statisztikai mintákon alapuló manipulációt végez. Az egyik legfontosabb megállapítás az, hogy a kreativitás mérése továbbra is nyitott kérdés, mivel a meglévő mérőszámok nem tudják teljes mértékben visszaadni az emberi alkotás komplexitását. Az emberi és gépi kreativitás közötti különbségek különösen élesek a művészeti területeken, ahol az alkotások jelentése és érzelmi töltete meghatározó szerepet játszik. Az MI gyakran képes technikailag tökéletes kompozíciókat létrehozni, de ezek sok esetben nélkülözik azt az egyedi perspektívát, amely az emberi alkotások sajátja.
Az MI és az emberi kreativitás közötti szinergia azonban új lehetőségeket nyithat meg, különösen a kreatív együttműködés terén, ahol az MI támogató eszközként funkcionálhat az emberi alkotók számára. A kutatás megmutatta, hogy az MI kreativitása leginkább akkor bontakozik ki, ha az emberi döntéshozatal és az algoritmikus optimalizálás egyensúlyban marad. Bár az MI egyre összetettebb és kifinomultabb alkotásokat képes létrehozni, az emberi kreativitás egyedisége és spontaneitása továbbra is pótolhatatlan marad. A betanításhoz, amennyiben a jelenlegi a módszerek nem változnak, mindig szükség lesz új, friss és kreatív, emberek által létrehozott tartalmakra.
A jövőben a kutatásoknak arra kell összpontosítaniuk, hogy milyen módszerekkel lehet az MI kreativitását mélyebb szinteken modellezni, illetve hogyan lehet a gépi alkotásokat olyan irányba fejleszteni, amely jobban közelít az emberi gondolkodásmódhoz. A tanulmány végső megállapítása szerint a mesterséges kreativitás mérése és fejlesztése nemcsak technikai, hanem filozófiai és pszichológiai kérdés is, amely további interdiszciplináris vizsgálatokat igényel.

Irodalomjegyzék

Asperti, A, Trentin, M. ’Balancing reconstruction error and Kullback-Leibler divergence in Variational Autoencoders’ arXiv:2002.07514, 2020

Ebert-Uphoff, I, Lagerquist, R, Hilburn, K, Lee, Y, Haynes, K, Stock, J, Kumler, C, & Stewart, J Q. ' CIRA Guide to Custom Loss Functions for Neural Networks in Environmental Sciences' arXiv:2106.09757, 2021

Hendrycks, D, Burns, C, Basart, S, Zou, A, Mazeika, M, Song, D & Steinhardt, J. 'Measuring massive multitask language understanding,' arXiv:2009.03300, 2020

Lin, S, Yang, X. ' Diffusion Model with Perceptual Loss' arXiv:2401.00110, 2024

Nagy, B, Csizmadia, P, Kovács, A J, Czigler, I, Gaál, Zs A. 'A kreativitás és a kreatív teljesítményt befolyásoló tényezők pszichometriai vizsgálata fiatal és idős felnőtt populáción' Alkalmazott Pszichológia Vol. 2022, 22.2 pp. 55–89

Shannon, C E. 'A Mathematical Theory of Communication', korrektúrával újranyomott The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

Solak A, Grötschla, F, Lanzendörfer, L A & Wattenhofer R. ' Benchmarking Music Generation Models and Metrics via Human Preference Studies', NeurIPS 2024 Workshop, 2024

Sun, K, Pan, J, Ge, Y, Li, H, Duan, H, Wu, Y, Zhang, R, Zhou, A, Qin, Z, Wang, Y, Dai, J, Qiao, Wang, L & Li, H. 'JourneyDB: A Benchmark for Generative Image Understanding' arXiv:2307.00716, 2023

Wang, A, Singh, A, Michael, J, Hill, F, Levy, O & Bowman, S R. ' Glue: A multi-task benchmark and analysis platform for natural language understanding' arXiv:1804.07461, 2018.

Wang, A, Pruksachatkun, Y, Nangia, N, Singh, A, Michael, J, Hill, F, Levy, O & Bowman, S R. 'SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems' arXiv:1905.00537, 2019

Zhu, M, Chen, H, Yan, Q, Huang, X, Lin, G, Li, W, Tu, Z, Hu, H, Hu, J & Wang Y. ' GenImage: A Million-Scale Benchmark for Detecting AI-Generated Image' arXiv:2306.08571, 2023

Zhu, Z, Xu, T, Li, L & Wang, Y. 'Noise Dimension of GAN: An Image Compression Perspective' arXiv:2403.09196, 2024

Bizalom az AI korában – Trustworthy AI és a blokklánc technológia integrációja

Trust in the Age of AI – The Integration of Trustworthy AI and Blockchain Technology

dr. Halász Rita[9]

Absztrakt:

A mesterséges intelligencia (MI) rendszerek társadalmi elfogadottságának egyik kulcsa a bizalom, aminek pedig előfeltétele a technológiai megbízhatóság. A tanulmány célja, hogy bemutassa, miként építhető be a bizalom strukturális eleme az MI életciklusába az átláthatóság, az etikai megfelelés, az adatbiztonság és az emberi kontroll elveinek érvényesítésével. A tanulmány különös figyelmet fordít az Európai Bizottság által 2018. júniusában létrehozott mesterséges intelligenciával foglalkozó magas szintű független szakértői csoportjának iránymutatásában meghatározott elvekre, a megbízható mesterséges intelligencia "Trustworthy AI" gyakorlati megvalósításának módjaira, valamint a blokklánc-technológia támogató szerepére a megbízhatóság technológiai biztosításában. Az érvelést konkrét nemzetközi ajánlások is alátámasztják. A tanulmány egyik fő megállapítása, hogy a bizalom a jövőben nem utólagos legitimáció, hanem a rendszertervezés kezdetétől integrált, igazolható tulajdonság kell legyen.

Kulcsszavak: megbízható mesterséges intelligencia, dokumentáció, átláthatóság, magyarázhatóság, etikai megfelelés, szabályozás, blockchain

Abstract

One of the key factors for the societal acceptance of artificial intelligence (AI) systems is trust, which in turn relies on technological reliability. The aim of this study is to explore how trust can be embedded as a structural component throughout the AI lifecycle by enforcing the principles of transparency, ethical compliance, data security, and human oversight. Special attention is given to the principles outlined by the High-Level Expert Group on Artificial Intelligence, established by the European Commission in June 2018, to the practical implementation of the concept of “Trustworthy AI,” and to the supportive role of blockchain technology in ensuring technological trustworthiness. The argument is supported by

[9] dr. Halász Rita: jogász-közgazdász, a Blockchain Magyarország Egyesület alelnöke és a DigitalTech EDIH üzletfejlesztési és compliance tanácsadója rita.halasz.dr@blockchainhungary.org

international recommendations. One of the main conclusions of the study is that trust should not be treated as a post hoc justification, but rather as a verifiable and integrated attribute from the very beginning of system design.
Keywords Trustworthy AI, documentation, transparency, explainability, ethical compliance, regulation, blockchain

Bevezetés

A tanulmány alapját a Magyar Tudományos Akadémia Veszprémi Területi Bizottsága (MTA-VEAB) Gazdaság-, Jog- és Társadalomtudományi Szakbizottsága, valamint a „MI és Kreatív Iparágak Társadalomtudományi Kutatása" Munkabizottság által szervezett nemzetközi konferencia keretében, 2025. május 20-án Veszprémben elhangzott előadás képezi. Az előadás „Bizalom az AI korában – Trustworthy AI és a blokklánc technológia integrációja" címmel a megbízható mesterséges intelligencia gyakorlati érvényesülésének, szabályozási megfelelőségi szempontjainak és a technológiai bizalom újradefiniálásának kérdéseit tárgyalta.

A prezentáció központi kérdése az volt, hogy miként értelmezhető és igazolható a bizalom a digitális technológiák, különösen a mesterséges intelligencia kontextusában. Olyan dilemmák kerültek vizsgálat alá, mint például: mit is jelent a „megbízható" MI, és mennyiben bízhatunk meg egy rendszer által közölt információban, ha nem látjuk és nem ellenőrizhetjük az azt megalapozó adatforrásokat vagy döntési folyamatokat. A tanulmány e kérdések mentén vizsgálja, lehet-e a blokklánc technológia a bizalom technológiai biztosítéka, és milyen szerepet tölthet be az MI átláthatóságának és hitelességének erősítésében.

A publikáció célja, hogy elméleti és gyakorlati szempontból egyaránt bemutassa a mesterséges intelligencia és a blokklánc technológia integrációjából származó lehetőségeket. Emellett hangsúlyosan fel kívánja hívni a figyelmet arra, hogy a megbízható MI nem pusztán technikai, hanem társadalmi, jogi és etikai keretrendszert is igényel.

1.A bizalom normatív újraértelmezése a digitális technológiák korában

A digitális korszakban a bizalom többé nem csupán interperszonális viszony, hanem technológiai, jogi és etikai kérdés is. Onora O'Neill[10] szerint a bizalom racionális döntés, amelyet csak akkor adunk meg egy embernek, intézménynek vagy algoritmusnak, ha

[10] O'Neill, Onora. *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge: Cambridge University Press, 2002.

igazolhatóan kompetens, őszinte és felelősséget vállal a következményekért. E három feltétel teremti meg a megbízhatóságot, amelyet átlátható folyamatok, erős adatvédelem, technikai stabilitás, diszkriminációmentesség és elszámoltathatóság támaszt alá. A mesterséges intelligencia korában így a bizalom nem előzetes feltétele, hanem eredménye annak, hogy a rendszer teljesíti-e ezeket a tudatosan kialakított normákat.

2.A Trustworthy AI keretrendszere, avagy a bizalom új alapjai az Európai Bizottság szakértői csoportjának iránymutatása alapján

Az Európai Bizottság mesterséges intelligenciával foglalkozó szakértői csoportja (AI HLEG) által kidolgozott iránymutatás[11], mérföldkőnek tekinthető a megbízható mesterséges intelligencia normatív és gyakorlati alapjainak lefektetésében. A dokumentum célja, hogy irányt mutasson az etikai elvek és jogszabályi követelmények rendszerszintű alkalmazására, különös tekintettel az MI rendszerek fejlesztésére, bevezetésére és felügyeletére.

Az AI HLEG által megalkotott Trustworthy AI keretrendszer nem pusztán elméleti iránymutatás, hanem egy olyan strukturált szempontrendszer, amely hozzájárul az MI társadalmi elfogadásához, a szabályozási megfeleléshez, és a digitális technológiák etikai beágyazásához. Alapját képezi az Európai Unió mesterséges intelligenciára vonatkozó szabályozási törekvéseinek, így különösen az Európai Parlament és a Tanács (EU) a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról szóló 2024/1689 rendeletének (AI Act)[12], és lehetővé teszi, hogy a bizalom ne ígéret, hanem dokumentálható működési minőség legyen.

A keretrendszer[13] négy fő komponensből áll:

1. a Trustworthy AI fogalmi elemei;
2. az etikai alapelvek;
3. a hét fő követelmény; valamint
4. az értékelési szempontok kialakítása.

2.1. A Trustworthy AI fogalmi elemei

Az iránymutatás értelmében egy mesterséges intelligencia rendszer akkor tekinthető megbízhatónak, ha teljes életciklusa során egyszerre három alapvető követelménynek is

[11] European Commission, High-Level Expert Group on Artificial Intelligence. 2019. *Ethics Guidelines for Trustworthy AI* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[12] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.
[13] European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai p.10

megfelel. Egyrészt működésének összhangban kell állnia a hatályos nemzeti és uniós jogszabályokkal, tehát jogszabályi megfelelést kell biztosítania (jogszerű). Másrészt elengedhetetlen, hogy tiszteletben tartsa az alapvető jogokat és az emberi méltóságot, vagyis etikai szempontból is megfelelő legyen (etikus). Harmadrészt az MI rendszernek technikai és társadalmi szempontból is stabilnak kell lennie: biztonságosan, robusztusan kell működnie, miközben képes minimalizálni a nem szándékos károkozás kockázatát is (stabil).
Azt is látni kell, hogy ezek a feltételek önmagukban nem elegendőek, együttes meglétük szükséges ahhoz, hogy egy adott rendszert valóban megbízható mesterséges intelligenciaként értékelhessünk.

2.2. Etikai alapelvek

Az AI HLEG által megfogalmazott etikai alapelvek azt részletezik, miként illeszthetők a mesterséges intelligencia rendszerek egy emberközpontú és igazságos társadalomba. Elsődleges követelmény, hogy az emberi autonómia érvényesüljön: a technológia nem veheti át az emberi döntéshozatal szabadságát (*az emberi autonómia tiszteletben tartásának elve*), hanem olyan környezetet kell teremteni, amelyben az emberi felügyelet, az önrendelkezés és a kiegészítő jelleg harmonikusan találkozik, miközben az MI nem manipulálhat, és nem gyakorolhat megtévesztő befolyást a felhasználókra. Ugyancsak alapvető elv, hogy a rendszerek sem közvetlenül, sem közvetve nem okozhatnak kárt a felhasználóknak, a társadalomnak vagy a környezetnek (*a kár megelőzés elve*); ide tartozik a kiszolgáltatott csoportok fokozott védelme, a biztonsági garanciák biztosítása és a rosszindulatú felhasználás elleni védelem is. A *méltányosság elve* a diszkriminációmentes működést és az erőforrásokhoz, szolgáltatásokhoz, lehetőségekhez való igazságos hozzáférést írja elő, amely mind az előnyök és hátrányok elosztására, mind a jogorvoslati lehetőségekre kiterjed. Végül, a *megmagyarázhatóság elve* megköveteli, hogy az MI céljai, funkciói és döntési logikái átláthatók legyenek; a felhasználóknak érteniük kell, milyen szempontok alapján születik egy döntés, különösen akkor, ha „feketedoboz" jellegű rendszerről[14] van szó. Ilyen esetekben alternatív magyarázhatósági megoldások – például az ellenőrizhetőség és a nyomon követhetőség – alkalmazása válik szükségessé.
Elvi ellentmondások kezelése

[14] Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. 2017. "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models." arXiv preprint arXiv:1708.08296. https://arxiv.org/abs/1708.08296.

Az AI HLEG rámutatott arra, hogy a négy etikai elv között adott esetben ellentmondás alakulhat ki – például amikor a közbiztonság érdekében végzett megfigyelés ütközik az emberek magánéletéhez vagy önrendelkezéséhez fűződő jogaival. Az ilyen helyzeteket – álláspontjuk szerint - nem lehet szabályozással megoldani, hanem olyan döntéshozatali folyamatokra van szükség, amelyek átláthatók, és biztosítják a demokratikus elszámoltathatóságot. Ezen döntéseknek jogilag és erkölcsileg is megalapozottnak kell lenniük.

2.3. A hét fő követelmény

A megbízható mesterséges intelligencia nemcsak elvi, hanem gyakorlati szempontból is számos konkrét követelményt támaszt a fejlesztőkkel és az alkalmazókkal szemben. Az alábbi hét fő követelmény az AI HLEG által kidolgozott keretrendszer alapján részletezi, hogyan valósítható meg az etikai és jogi elvek gyakorlati integrációja.

a) Emberi cselekvőképesség és felügyelet

Az MI rendszerek nem sérthetik az egyének autonómiáját, ezért biztosítani kell, hogy a felhasználók önálló, tájékozott döntéseket hozhassanak az MI használatáról. Ennek érdekében hatásvizsgálatokra, visszacsatolási mechanizmusokra és emberi beavatkozási lehetőségekre („human-on-the-loop")[15] van szükség. Az automatizált döntések nem válhatnak kizárólagossá, az emberi felügyelet gyakorlását is garantálni kell. Minél kevesebb a közvetlen emberi kontroll, annál szigorúbb tesztelés és ellenőrzés szükséges egy adott MI rendszer fejlesztési folyamatában.

[15] European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai p.19

b) Műszaki stabilitás és biztonság

Az MI rendszereknek ellenállónak kell lenniük mind belső hibák, mind külső támadások (pl. adatszennyezés, manipuláció) ellen. Elengedhetetlen a vészhelyzeti protokollok, monitoring eszközök és kockázatfüggő biztonsági intézkedések beépítése a rendszerbe. Ezen túlmenően a követelményen belül érvényesülő pontossági kritérium alapján, a rendszernek képesnek kell lennie megbízható előrejelzések és döntések meghozatalára. Szintén ezen követelmény kritériumi eleme a fejlesztési és tanítási folyamat reprodukálhatósága (pl. reprodukciós fájlokkal), amely biztosítja a transzparens működést és a hibák visszakereshetőségét.

c) Adatvédelem és adatkezelés

Az MI rendszereknek garantálniuk kell az adatminőséget, az adatfeldolgozás jogszerűségét és a személyes adatok védelmét. Az adatkészleteknek pontosnak, elfogulatlannak és torzításmentesnek kell lenniük. Fontos szabály, hogy az adatkészletet a betanítás megkezdése előtt ellenőrizni kell. A teljes MI életciklus során dokumentálni szükséges az adatforrásokat és a kezelési lépéseket. Minden szervezetnek, amely MI rendszereket fejleszt vagy üzemeltet, világos adatkezelési protokollokat kell kialakítania, amelyben meghatározzák, hogy ki férhet hozzá az adatokhoz, milyen feltételek mellett, és milyen céllal.

d) Átláthatóság

Az MI rendszerek működésének minden szakaszát – az adatgyűjtéstől az algoritmushasználatig – dokumentálni kell. Az átláthatóság magába foglalja a rendszer döntéshozatalának érthetőségét, nyomon követhetőségét és a felhasználók megfelelő tájékoztatását is, ideértve annak tisztázását, hogy a rendszer gépi döntéseket hoz. A rendszer korlátairól, pontosságáról és működési logikájáról való tájékoztatás is alapvető követelmény.

e) Sokféleség, diszkriminációmentesség és méltányosság

A tanítóadatok torzításának elkerülése kulcsfontosságú az MI rendszerek esetében. Ennek érdekében már az adatgyűjtési fázisban biztosítani kell a társadalmi sokféleség képviseletét. Szükség van olyan felügyeleti mechanizmusokra, amelyek monitorozzák az esetleges diszkriminációs hatásokat. Az iránymutatás alapvető elvként rögzíti, hogy támogatni kell az

esélyegyenlőséget[16] az oktatáshoz, termékekhez, szolgáltatásokhoz, továbbá a technológiához való hozzáférés terén, amelyet biztosítani kell minden társadalmi csoport számára.

f) Társadalmi és környezeti jólét

Az MI rendszerek fejlesztése és működtetése során törekedni kell a fenntarthatóságra, a környezeti ártalmak minimalizálására, valamint a társadalmi kohézió és demokratikus értékek támogatására. Ezeket a szempontokat minden fejlesztési szakaszban értékelni szükséges.

g) Elszámoltathatóság

Az MI rendszerek működéséért egyértelmű felelősségi köröket kell kijelölni. A döntéseknek átláthatónak és auditálhatónak kell lenniük, az alkalmazott algoritmusok és adatok értékelését pedig belső és külső ellenőrzések révén kell biztosítani. Különösen fontos, hogy a jogsérelmek esetén jogorvoslati lehetőség is rendelkezésre álljon.

2.4.A megbízhatóság értékelése[17]

A megbízható MI keretrendszer egyik fontos gyakorlati eszköze a megbízhatósági kérdéslista, amely az MI rendszerek etikai, jogi és műszaki szempontú önértékelését segíti. Ennek alkalmazásakor elengedhetetlen figyelembe venni a rendszer konkrét működési környezetét, célját és azt, hogy milyen kockázatokat jelenthet az érintettekre nézve. A kérdések azt vizsgálják például, hogy biztosított-e az emberi felügyelet az MI által hozott döntések felett, rendelkezik-e a rendszer hiba esetén működésbe lépő biztonsági mechanizmussal, illetve megfelel-e az adatbiztonsági és GDPR[18] követelményeknek. Ugyanilyen lényeges annak ellenőrzése, hogy a döntési folyamatok dokumentáltak és értelmezhetőek-e, valamint, hogy a tanítóadatok sokféleségét és esetleges torzításait megvizsgálták-e. A kérdéslista kiterjed arra is, történt-e társadalmi és környezeti hatásvizsgálat, valamint, hogy vannak-e kijelölt felelősök, és működnek-e visszacsatolási vagy panaszkezelési mechanizmusok a szervezeten belül. Ezek a szempontok nemcsak a megbízhatóság értékelését teszik lehetővé, hanem hozzájárulnak a rendszerek folyamatos fejlesztéséhez és a felhasználói bizalom erősítéséhez is.

[16] European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai p.15

[17] European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai p.32

[18] European Union. (2016). *General Data Protection Regulation (GDPR)* (EU 2016/679). https://gdpr.eu/

2.5.Műszaki és nem műszaki megfelelési tényezők a megbízható mesterséges intelligencia érdekében

A megbízható mesterséges intelligencia gyakorlati kialakítása egyszerre igényel műszaki és nem műszaki megközelítéseket.

A technológiai oldalon elsődleges, hogy maga a rendszerarchitektúra stabil és biztonságos legyen, és már a tervezése során figyelembe vegyék az etikai és jogállamisági elveket, ne pedig utólag illesszék hozzá. A magyarázható MI (explainable AI, XAI) módszerei biztosítják, hogy az algoritmusok döntései átláthatóak és ellenőrizhetőek maradjanak, míg a fejlesztés minden szakaszára kiterjedő, kontextusérzékeny tesztelés gondoskodik arról, hogy a várt és a tényleges működés közötti eltérések feltárhatók legyenek. A működést támogató szolgáltatási mutatók – például a funkcionalitás, teljesítmény, megbízhatóság, védelem és karbantarthatóság – további bizonyítékot nyújthatnak a biztonsági és műszaki szabványoknak való megfelelésre.

Ezzel párhuzamosan a nem műszaki eszközök teremtenek keretrendszert a felelős alkalmazáshoz. A szabályozások, mint az AI Act és a termékbiztonsági előírások, rögzítik a működési feltételeket és a felelősségi rendet; a magatartási kódexek az etikus szervezeti kultúrát erősítik; a szabványosítás egységes fejlesztési és tesztelési módszertant garantál; a tanúsítás hitelesíti a biztonsági megfelelést, de nem helyettesíti a szervezeti felelősségvállalást, amit jogorvoslati és felülvizsgálati mechanizmusokkal kell biztosítani. Az elszámoltathatóságot szilárd irányítási keretekbe kell ágyazni, az oktatás és a tájékoztatás pedig elősegíti az etikus gondolkodást. Az érintettek bevonása és a folyamatos társadalmi párbeszéd növeli az elfogadottságot, míg a sokszínű, inkluzív fejlesztőcsapatok érzékenyebbé teszik a rendszert a különböző csoportok igényeire.

A szervezeteknek érdemes etikai felelőst kinevezniük vagy független etikai tanácsadó testületet létrehozniuk, amely a fejlesztés és az alkalmazás során felügyeli az elvárt normák érvényesülését; e testület a jogi és adatvédelmi szerepköröket nem helyettesíti, hanem kiegészíti. A megbízható MI előnyeit mindenki számára hozzáférhetővé kell tenni, ezért a vonatkozó követelményeket már az ötletelési és tervezési fázisba integrálni kell, és a rendszert a használat teljes időtartama alatt a beépített adatvédelem és biztonság elvei szerint kell működtetni. Lényeges elvárás[19], hogy ezek az alapelvek és követelmények a teljes életciklus során folyamatosan érvényesüljenek, mert csak így garantálható a valódi, és fenntartható bizalom.

[19] European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. p.18

3. A hét fő követelmény gyakorlati érvényesíthetősége

A mesterséges intelligenciával szembeni bizalom egyik alappillére, hogy az MI rendszerek fejlesztése és működtetése a gyakorlatban is megfeleljen azoknak az alapelveknek és megvalósíthatósági kritériumoknak, amelyeket a nemzetközi és uniós szakpolitikai irányelvek (így például az EU AI HLEG vagy az OECD) rögzítenek. A technológiai fejlődés gyors üteme miatt azonban a gyakorlati alkalmazás és az elméleti elvárások között gyakran jelentős szakadék tapasztalható. Egy az OECD által 2025-ben publikált jelentés[20] szerint bár az európai vállalatok 61 százaléka rendelkezik valamilyen MI szabályzattal, de csak 28 százalékuk képes ténylegesen nyomon követni, hogy algoritmusaik milyen adatforrásokra és döntési logikákra épülnek, ami jelentős szervezeti (stratégiai) kockázatot hordoz. Ez a megállapítás rávilágít arra, hogy az MI működésének átláthatósága és visszakövethetősége még nem megfelelően érvényesül a gyakorlatban. Az MI rendszerek biztonságos és etikus alkalmazásának legnagyobb kockázata azonban elsősorban nem technikai természetű, hanem szervezeti: a nem kellően dokumentált, ellenőrizetlen működés; a kritikus adatminőség és az adatbiztonság, amelyeket a NIS2-irányelv[21] is kiemelten kezel. A hibás vagy manipulált adat nemcsak a stratégiai döntések megalapozottságát veszélyezteti, hanem a teljes rendszer integritását is.

E fejezet célja, hogy bemutassa, a már fentiekben ismertetett hét fő követelmény azon tartalmi kritériumait, és azok gyakorlati érvényesíthetőségét, amelyek megvalósításával biztosítható a megbízható MI keretrendszerének való megfelelés.

3.1. A hét fő követelmény tartalmi kritériumai: átláthatóság, megmagyarázhatóság, nyomon követhetőség, ellenőrizhetőség, elszámoltathatóság, pontosság és reprodukálhatóság

Az átláthatóság az MI rendszerek egyik alapvető követelménye, amely biztosítja, hogy azok működése, döntési logikája, céljai és társadalmi hatásai érthetőek és hozzáférhetők legyenek az érintettek számára. Ez a társadalmi kontroll előfeltétele, és csak akkor valósulhat meg, ha az adatkezelés, a rendszerfelépítés és az üzleti modell is világosan dokumentált. A dokumentáció nem csupán adminisztratív feladat, hanem minden további követelmény – így az ellenőrzés, a nyomon követés, és a megmagyarázhatóság – alapja. Rögzíteni kell többek között az adatgyűjtés és adatcímkézés folyamatait, az adatkészleteket, az alkalmazott algoritmusokat, valamint a fejlesztés során meghozott emberi döntéseket, például az MI rendszer felhasználási

[20] OECD, Boston Consulting Group & INSEAD. 2025. *The Adoption of Artificial Intelligence in Firms*. Paris: OECD Publishing. (oecd.org)

[21] European Union. 2022. *Directive (EU) 2022/2555 on Measures for a High Common Level of Cybersecurity across the Union (NIS 2 Directive)*. *Official Journal of the European Union* L 333. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2555.

területének kijelölését érintőt. A magyarázhatóság (XAI)[22] kritériumából következik, hogy az MI rendszer képes a döntéseit emberi (végső soron felhasználói) szempontból értelmezhető módon megindokolni. Ez különösen fontos érzékeny területeken, például az egészségügyben, ahol a döntések közvetlen és súlyos következményekkel járhatnak.

A nyomon követhetőség elve azt kívánja meg, hogy az MI rendszer működésének minden aspektusa – beleértve az adatbemeneteket, az algoritmikus változásokat és a kimeneteket – részletesen dokumentált legyen. Ez szorosan összefügg az ellenőrizhetőséggel, amely lehetőséget biztosít független auditokra, illetve az algoritmikus döntések visszafejtésére, hogy szükség esetén rekonstruálható legyen, hogyan és miért született egy adott döntés.

Az elszámoltathatóság elve mentén célszerű például egy szervezet működési szabályzatában egyértelműen meghatározni, hogy ki a felelős az MI rendszer egyes működési szakaszaiért és döntéseiért. Ez alapfeltétele annak, hogy a rendszert be lehessen illeszteni jogilag és szervezetileg is értelmezhető keretek közé.

A pontosság és a reprodukálhatóság szintén kulcsfontosságú tényezők. Egy jól működő MI rendszer azonos feltételek mellett ismételten is ugyanazt az eredményt adja, különösen biztonságkritikus környezetekben. Ehhez szükséges, hogy a rendszer helyesen kategorizáljon adatokat, pontos döntéseket és előrejelzéseket hozzon, valamint rendelkezzen olyan fájlokkal, amelyek lehetővé teszik az egész fejlesztési folyamat újraalkotását.

4. Emberi felügyelet biztosítása: human-in/on/out-of-the-loop megközelítés

Az emberi jelenlét és kontroll az MI rendszerek működése során három, egymással fokozatosan csökkenő emberi beavatkozási lehetőséget kínáló modell mentén írható le. A „human-in-the-loop" megközelítésben az ember minden döntési ciklusba aktívan bekapcsolódik, így például egy orvosi diagnosztikai rendszer esetében a szakember minden egyes kimenetet felülvizsgál, a katonai célpont-azonosításnál pedig az operátor adja meg a végső engedélyt; hasonló a helyzet pénzügyi jóváhagyásoknál is. A „human-on-the-loop" szintnél az ember már inkább felügyelő szerepet tölt be: folyamatosan monitorozza az autonóm rendszert, és csak rendellenesség vagy kritikus helyzet esetén avatkozik közbe. Ez a modell teremti meg az autonóm működés és az emberi kontroll egyensúlyát például önvezető járműflották, ipari robotcellák vagy

[22] Samek, Wojciech, Thomas Wiegand, and Klaus-Robert Müller. 2017. "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models." arXiv preprint arXiv:1708.08296. https://arxiv.org/abs/1708.08296.

döntéstámogató rendszerek esetében. Végül a „human-out-of-the-loop” működésnél az ember teljesen kimarad a döntéshozatalból; ezt kizárólag szűk, alacsony kockázatú vagy speciális területeken szabad alkalmazni, mint amilyen a zeneajánló algoritmusoké vagy bizonyos játék alkalmazásoké.

Az, hogy melyik modellt választjuk, döntő hatással van az etikai kontroll, a jogi felelősség és a biztonsági beavatkozás lehetőségére. Magas kockázatú rendszerek esetén az AI Act kifejezetten előírja a megfelelő emberi felügyeletet, mert az elszámoltathatóság és a kritikus helyzetekben történő felülbírálás emberi döntésképességet kíván. Általános tervezési elv, hogy minél kisebb a közvetlen emberi kontroll, annál alaposabb tesztelésre és szigorúbb irányítási keretrendszerre van szükség a megbízható MI rendszer megvalósításához.

5. A fejlesztési folyamat életciklusa: kutatás-adatgyűjtés-betanítás-tervezés-fejlesztés-tesztelés-használat-eredmény

Egy MI projekt esetében a szervezeteknek már a projekt első fázisától úgy kell felépíteniük a munkafolyamatot, hogy a megbízhatóság minden lépést áthasson. Ennek részeként a kutatási fázisban egyértelműen meg kell határozni a megoldandó problémát, fel kell mérni a szóba jövő technológiákat és át kell tekinteni a vonatkozó jogszabályi környezetet. Az adatgyűjtés során kizárólag torzításmentes és jogszerű adatállományokat szabad felhasználni, amelyek pontos dokumentálásával később igazolható a tanítási adatok tisztasága.

A betanítási fázisban gondoskodni kell a teljes fejlesztés átlátható nyomon követéséről: minden adatmozgatást, algoritmus-változatot és emberi döntést rögzíteni kell, hogy a folyamat később auditálható legyen. A tervezés szintjén ki kell alakítani az algoritmikus logikát, a rendszerarchitektúrát és a mérési kritériumokat, miközben biztosítani kell az emberi beavatkozás lehetőségét is. A tényleges fejlesztés után a prototípust nemcsak funkcionális, hanem stressz- és etikai teszteknek is alá kell vetni, majd az élesítéskor folyamatos felügyeletet, visszacsatolást és hatásmérést kell alkalmazni, hogy a tapasztalatok alapján azonnal javítható legyen a rendszer.

Fontos szabály, hogy ha a megoldás alapvető emberi jogokat érint, például arcfelismerést vagy kórházi diagnosztikát, független külső szakértőket is be kell vonni a teljes folyamat objektív ellenőrzésére.

5.1. A hét fő követelmény gyakorlati érvényesülésének eszközei: hatásvizsgálat, dokumentálás, értékelés, belső elszámoltathatóság, érintettek bevonása, jogorvoslat

Egy működő, megbízható MI rendszer csak akkor születhet meg, ha a fejlesztés minden szakaszát következetesen szabályozott gyakorlat kíséri. Már a tervezéskor kötelező felmérni a társadalmi, környezeti és gazdasági hatásokat, különösen ha az alkalmazás alapvető jogokat – például a magánélet védelmét – érintheti. A teljes folyamatot részletesen dokumentálni kell: rögzíteni az adatforrásokat, a modellparamétereket és minden döntési lépést, majd ezt belső és független auditokkal rendszeresen ellenőrizni, hogy kiderüljön, az algoritmus éles környezetben is a kijelölt kockázati keretek között működik-e. A felhasználókat hiteles tájékoztatás illeti meg, és végig meg kell maradnia az emberi beavatkozás lehetőségének. A szervezetnek egyértelmű etikai, jogi és adatvédelmi felelősöket kell kijelölnie, a működést pedig átlátható elszámoltathatósági és hatékony jogorvoslati mechanizmusokkal kell alátámasztani. A rendszer csak akkor marad hosszú távon fenntartható, ha az érintettek folyamatos visszajelzése és a nyílt ellenőrzés beépül a működés mindennapjaiba.

6. Adatkezelés és adatminőség: adatvédelem, adatkészlet minősége, adattorzítás elkerülése

A mesterséges intelligencia rendszerek megbízhatósága nagymértékben függ az adatkezelés minőségétől. Ez megköveteli a jogszabályok – különösen a GDPR – betartását, az adatminimalizálás (szükséges minimumra való törekvés), anonimizálás és célhoz kötött felhasználás elveinek alkalmazásával. Emellett világos hozzáférési szabályokra és biztonságos adatkezelési gyakorlatra van szükség. A fejlesztés során folyamatosan biztosítani kell az adatkészletek pontosságát, reprezentativitását és torzításmentességét. A sokféleségre való törekvés és az emberi felügyelet segít elkerülni a rendszerszintű torzulásokat, amelyek sérthetik az alapvető jogokat. A társadalmilag hasznos és etikus működéshez olyan átfogó adatstratégiára van szükség, amely az adatvédelem, minőség és etika egyensúlyát végig fenntartja. A jó minőségű adatkészlet pontosságot, reprezentativitást és konzisztenciát jelent, amit a fejlesztési életciklus során folyamatosan vizsgálni és dokumentálni kell, mert csak így garantálható, hogy a betanítás előtt, alatt és után is sértetlen marad az adatállomány.

7. Műszaki biztonság: üzembiztos leállítási terv, készenléti terv

A technikai biztonság nem pusztán informatikai feladat, hanem az etikai megbízhatóság alapfeltétele. A mesterséges intelligenciát úgy kell megtervezni, hogy az ember bármikor biztonságosan leállíthassa, majd előre rögzített eljárással újraindíthassa a rendszert. Emellett fel kell készíteni váratlan eseményekre: a szoftvernek támadás vagy hiba esetén automatikusan készenléti módba kell váltania, amit részletes vészforgatókönyvek és beépített biztonsági

elemek tesznek lehetővé. A hibatűrés része, hogy a rendszer észleli és kezelni tudja a meghibásodásokat anélkül, hogy a működés teljesen leállna.

8. A blokklánc-technológia szerepe a megbízható mesterséges intelligencia támogatásában

A mesterséges intelligencia rendszerek megbízhatóságának megteremtése nem csupán etikai és jogi kérdés, hanem technológiai kihívás is. A blokklánc-technológia – technikai sajátosságaiból fakadóan – hatékony eszközként szolgálhat a Trustworthy AI alapelveinek gyakorlati megvalósításához. Az olyan jellemzők, mint a változtathatatlanság, a decentralizáltság, az átláthatóság, valamint a kriptográfiai védelem révén a blokklánc ideális környezetet kínál az MI rendszerek biztonságos, elszámoltatható és auditálható működéséhez.

A blokklánc nem csupán egy adatkezelési eszköz, hanem a digitális bizalom technológiai infrastruktúrájaként is értelmezhető. Alkalmazása révén a mesterséges intelligencia nem csupán technológiailag lesz hatékony, hanem működése – dokumentált és visszakövethető módon – megfelelhet a társadalmi, etikai és jogi normáknak is. A blokklánc-architektúrába illesztett MI rendszerek így képesek lesznek biztosítani az átláthatóságot, az adatintegritást, valamint a reprodukálhatóságot.

9. A blokklánc kulcsjellemzői adatbizalmi szempontból

A blokklánc-technológia sajátos működéséből fakadóan több olyan jellemzőt is hordoz, amelyek alapvetően segítik a mesterséges intelligencia megbízhatósági követelményeinek teljesülését, különösen az adatbizalom területén. A felhasználók azonosítása nem közvetlen személyes adatokon, hanem kriptográfiai azonosítókon alapul, így a rendszer pszeudonimitást biztosít. Emellett a Zero-Knowledge Proof[23] típusú megoldások lehetővé teszik, hogy a felhasználók bizonyos információkat igazoljanak anélkül, hogy azok tényleges tartalmát felfednék, ami elősegíti az adatminimalizálást és az adatvédelmi szabályoknak való megfelelést.

[23] A Zero-Knowledge Proof olyan korszerű kriptográfiai protokoll, amely lehetővé teszi, hogy a bizonyító fél (prover) igazolja egy állítás igazságát a hitelesítő félnek (verifier) anélkül, hogy a bizonyítás tárgyát – például jelszót, privát kulcsot vagy bizalmas adatot – ténylegesen feltárná. Ennek köszönhetően ZKP-k különösen alkalmasak GDPR-kompatibilis azonosító- és engedélyezési rendszerek, illetve blokklánc-alapú tranzakciók adatvédelmi megerősítésére, ahol az érzékeny adatok kiszivárgása elfogadhatatlan kockázatot jelentene. Goldwasser, Shafi, Silvio Micali, és Charles Rackoff. "The Knowledge Complexity of Interactive Proof-Systems." https://epubs.siam.org/doi/10.1137/0218012

A blokkláncra rögzített adatok digitális aláírással és hash-értékekkel vannak ellátva, így biztosított az adatok hitelessége és integritása, ami elengedhetetlen például az MI rendszerek tanítóadatainak vagy döntési naplóinak hiteles dokumentálásához. Mivel minden adatmozgás időbélyeggel ellátva, visszakereshető módon kerül rögzítésre, a technológia a teljes fejlesztési és működési folyamat átláthatóságát is támogatja – ez pedig kulcsfontosságú az auditálhatóság és a társadalmi bizalom szempontjából.

A nyilvános blokklánc-hálózatok révén a rendszerben történő műveletek technikailag ellenőrizhetők és nyomon követhetők, így erősítve az átláthatóságot. Továbbá, az okosszerződések használata lehetővé teszi, hogy bizonyos döntések automatikusan, előre programozott szabályok szerint történjenek, amely nemcsak a működés hatékonyságát növeli, hanem hozzájárul az etikai és jogi megfelelőség biztosításához is.

Összességében a blokklánc-technológia nemcsak technikai támogatást nyújt az MI rendszerek szabályozott és megbízható működéséhez, hanem alapot teremt a transzparens és etikus digitális környezet kialakításához is. Az így létrehozott strukturált, hitelesen dokumentált és visszakövethető működés lehetővé teszi, hogy az MI rendszerek ne csupán hatékonyak, hanem hosszú távon is igazolhatóan megbízhatóak és társadalmilag elfogadottak legyenek.

10. Az INATBA álláspont[24]

Az INATBA (International Association for Trusted Blockchain Applications) 2023-as jelentése szerint a mesterséges intelligencia és a blokklánc-technológia összekapcsolása kulcsfontosságú lehet az átlátható és etikus digitális ökoszisztéma kialakításában. A blokklánc biztosítja az MI rendszerek döntéseinek naplózását, visszakereshetőségét és auditálhatóságát, ezzel növelve az elszámoltathatóságot. Az okosszerződések segíthetik a jogi és etikai megfelelést, míg a decentralizált identitáskezelés lehetővé teszi, hogy a felhasználók maguk rendelkezzenek adataik felett. A tokenizáció nyomon követhetővé és jogilag rendezetté teszi az adat- és algoritmusvagyont.

A jelentés kiemeli a megfelelőségi dokumentáció, a blokklánc-alapú auditmegoldások bevezetése, valamint az adatvédelmi és MI etikai képzések szükségességét. A blokklánc nem helyettesíti az emberi döntéseket, de biztosítja azok nyomon követhetőségét és hitelességét, így támogatja a hosszú távon is megbízható, etikus és társadalmilag elfogadott MI rendszerek kialakítását.

11. Etikai elvekből jogi kötelezettség: a Trustworthy AI szabályozási útja

[24] INATBA PAPER by the AI & Blockchain Convergences task force 2024. *AI Regulation and Blockchain: Bridging Ethics and Governance*

A Trustworthy AI gyakorlati megvalósításának egyik leglényegesebb jogi eszköze az AI Act, amely a mesterséges intelligenciára vonatkozó uniós szabályozás alapját képezi. Az Európai Bizottság szakértői által korábban kidolgozott etikai és technikai elvek ebben a rendeletben nyernek kötelező érvényű, jogi formát. Az AI Act tehát nem csupán egy szabályrendszer, hanem a biztonságos és megbízható mesterséges intelligencia feltételeinek intézményesített garanciája is. Az AI Act nem új elveket vezet be, hanem az eddigiek szabályozási és végrehajtási keretrendszerét alkotja meg, különös tekintettel a magas kockázatú MI rendszerekre. E dokumentum a fejlesztési életciklus minden szakaszában konkrét megfelelési követelményeket fogalmaz meg: kötelező dokumentációt, kockázatelemzést, emberi felügyelet biztosítását, magyarázhatóságot, valamint jogorvoslati és felelősségi mechanizmusokat ír elő.
A szabályozás tehát kettős célt szolgál:

- Bizalomerősítő funkciót, amely a társadalmi elfogadottságot és felhasználói jogokat garantálja;
- Versenyképességi és innovációs keretet, amely jogbiztonságot és iránymutatást nyújt a fejlesztők és szolgáltatók számára.

Az AI Act hatályba lépése – az eltérő fejezeti ütemezés mellett – 2026. augusztus 2-án válik teljes körűen kötelezővé az Európai Unióban. A jogszabály jogi megalapozottságot biztosít a Trustworthy AI elveinek, ezáltal hidat képez az etikai normák és a technológiai gyakorlat között. A blokklánc és MI rendszerek technológiai összhangja pedig e szabályozás technikai megvalósításához is hozzájárulhat, biztosítva az auditálhatóságot, a visszakövethetőséget és az átlátható működést.

Összegzés

A mesterséges intelligencia rendszerek megbízhatósága korunk egyik legösszetettebb technológiai és társadalmi kihívása. A „Trustworthy AI” nem csupán technikai teljesítmény, hanem mélyen normatív elvárás, amely az átláthatóság, az elszámoltathatóság, az adatvédelem és az emberi autonómia tiszteletének együttes érvényesítését követeli meg.
E követelmények teljesülése azonban nem garantálható kizárólag szabályozási vagy jogi eszközökkel, hanem szükség van technológiai támogatásra, társadalmi diskurzusra és intézményi felelősségvállalásra is. Ebben a kontextusban a blokklánc-technológia nem csupán adatkezelési megoldás, hanem a digitális bizalom technikai infrastruktúrájának egyik meghatározó eleme lehet, hiszen lehetővé teszi egy olyan környezet kialakítását, ahol a

megbízhatóság nem elvont ígéret, hanem technikailag visszakövethető és bizonyítható képesség.

A mesterséges intelligencia és a blokklánc technológiák integrációja olyan új digitális ökoszisztémák alapjául szolgálhat, amelyek az emberközpontúság, etikai megfelelés és átláthatóság mentén szerveződnek. A jövő digitális társadalma csak olyan infrastruktúrára épülhet, amelyben a bizalom nem vélelmezett, hanem technikailag alátámasztott és jogilag szabályozott. Ez a szemlélet képezi az etikus, jogszerű és fenntartható MI rendszerek fejlesztésének alapját is.

Irodalomjegyzék

European Commission High-Level Expert Group on Artificial Intelligence 2019, *Ethics guidelines for trustworthy AI*, European Commission, Brussels, available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Union 2016, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, *Official Journal of the European Union*, L 119, available at: https://gdpr.eu/

European Union 2022, *Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS 2 Directive)*, *Official Journal of the European Union*, L 333, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2555

European Union 2024, *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence (AI Act)*

Goldwasser, S., Micali, S. & Rackoff, C. 1989, 'The knowledge complexity of interactive proof-systems', *SIAM Journal on Computing*, vol. 18, no. 1, pp. 186–208, available at: https://epubs.siam.org/doi/10.1137/0218012

INATBA 2024, *AI regulation and blockchain: bridging ethics and governance*, AI & Blockchain Convergences Task Force white paper, International Association for Trusted Blockchain Applications, Brussels.

O'Neill, O. 2002, *A question of trust: The BBC Reith Lectures 2002*, Cambridge University Press, Cambridge, available at: https://www.cambridge.org/9780521529969

OECD, Boston Consulting Group & INSEAD 2025, *The adoption of artificial intelligence in firms*, OECD Publishing, Paris.

Samek, W., Wiegand, T. & Müller, K-R. 2017, 'Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models', *arXiv pre-print*, arXiv:1708.08296, available at: https://arxiv.org/abs/1708.08296

Megbízható teljesítmény, ígéretek, elvárások és bizalom

Az AI startup világ kihívásai

Reliable performance, promises, expectations and trust

The challenges of the AI startup world

Dr.[25] Ország-Krisz Axel[26]

Absztrakt

Az AI startupok sajátos feszültségben működnek, amely a tényleges felhasználók által elvárt megbízhatóság, a vállalkozók tett ígéretek, illetve a befektetők és a piac által támasztott elvárások közt gerjednek. Jelen írás e feszültséget vizsgálja az érintettek, azaz a fejlesztők, az alapítók, a befektetők, a felhasználók, a piac, a szabályozók és az állam szempontjából két rövid példán keresztül. Az egyik egy egyszerű, innovatív AI-megoldás piaci bevezetése, míg a másik egy meglévő szoftvertermék AI-funkcióval történő frissítése. A cikk keretei közt gépi tanulás, a felelős MI, a termékfejlesztés és az irányítás irodalmát szintetizáljuk, bemutatva, hogy a mai MI korlátai, így különösen az adateloszlás-eltolódás, a törékenység, a kalibrációs hibák, a hallucinációk, az adverzárius kockázatok és az átláthatatlan eredet miként alakítják az állításokat, az értékelést, az árazást, a megfelelést és a bizalmat. A dolgozat a végén gyakorlati útmutatást igyekszik adni a tekintetben, hogyan lehet összehangolni a teljesítményt, az ígéreteket, az elvárásokat és a bizalmat egy startup teljes életciklusa folyamán.

Abstract

AI startups face a distinctive tension between the reliability that real users require and the expectations and promises that energize entrepreneurs, investors and markets. This article examines how that tension plays out across stakeholders; developers, founders, investors, users, markets, regulators and the state. There are two short examples: the market launch of a simple, innovative AI solution; and an upgrade to an established software product with AI features. We synthesize insights from machine learning research, responsible AI, product development and governance to show how limitations of contemporary AI, like data shift, brittleness, calibration errors, hallucinations, adversarial risk, and opaque provenance, shape product claims,

[25] Nem tudományos fokozat, jogász doktor

[26] mesterséges intelligencia fejlesztő és szakértő, adattudós, email@drorszagkriszaxel.hu

evaluation, pricing, compliance and trust. We close with pragmatic guidance for aligning performance, promises, expectations and trust throughout the startup lifecycle.

Bevezetés

Napjainkra az MI rendszerek már réges-rég túljutottak a laboratóriumi bemutatók korszakán, hiszen mindennapjaink részéve vált a mesterséges intelligenciát tartalmazó, vagy épp teljes egészében azon alapuló szolgáltatások használata. Noha az MI használata ilyen hétköznapi, az ilyen jellegű szoftverek alapvető problémáinak kiküszöbölése még mindig általános, az egész iparágat kihívások elé állító feladat. A mesterséges intelligencia megoldások teljesítőképességét több tényező is korlátozza. Ezek közül a legjellemzőbbek az érzékenység a betanító adatokra és a kontextusra, a „rövid utak" megtanulása (Geirhos et al., 2020), a bizonyosság mértékének téves kalibrációja (Guo et al., 2017), a generatív modellekre jellemző hallucinációk (Ji et al., 2023), a kétoldalú sérülékenységek (Biggio és Roli, 2018) és az adatszivárgási kockázatok (Carlini et al., 2021). Termékeik és szolgáltatásaik tervezése, fejlesztése és értékesítése során a startupoknak e realitásokat, azaz a fent említett korlátok gyakorlati hatásait kell összeegyeztetniük az ún. „hype" ciklusokkal, azaz az olyan időszakokkal, amikor alkalmazásának egy adott területe a figyelem középpontjába kerül (Fenn és Raskino, 2008). Az ilyen ciklikus hatásokon túlmenően minden technológiai vállalkozásra jellemző egy terjedési dinamika is, mely hatást a növekedésre éhes tőke jeletősen képes befolyásolni (Gompers és Lerner, 2004). Mindezeken túlmenően az egyes piacok, mint például az USA vagy EU szabályozási környezete is lehetőségeket vagy épp korlátokat jelentenek a vállalkozások számára.

Egy sikeres vállalkozás a fent említett összes tényező között képes olyan utat, vagy utakat találni, amelyek a piacra kerülését és a piacon maradását eredményezik. Ugyanakkor tisztán felhasználói szemmel nézve a működésnek pusztán a megbízhatósága fontos, amely e kontextusban első sorban a kiszámítható, holnap is elérhető terméket jelenti.

A dolgozat keretei közt két rövid példán keresztül vizsgáljuk meg, egy adott helyzetben egy vállalkozás hogyan képes a fent leírt tényezők sokszor ellentétes hatását összeegyeztetni. Az első egy egyszerű AI-termék piaci bevezetését követi, míg a második egy AI-centrikus frissítést mutat be egy a piacon már bejáratott termékben. A bemutatás során mindkét esetben a mai technikai korlátokat és a hiteles vállalásokat meghatározó hatásaikat emeljük ki.

1.A fogalmi keretek meghatározása

E fejezetben röviden bemutatásra kerülnek azok a fogalmak, amelyek a dolgozat vizsgálatának tárgyát képezik.

Megbízhatóság:

A megbízhatóság fogalma alatt annak azt értjük jelen írás keretei közt, hogy egy rendszer milyen mértékben felel meg következetesen a teljesítménycéloknak meghatározott feltételek mellett, beleértve a pontosságot, a késleltetést, az üzemidőt és a biztonságot. Ez a meghatározás alapvetően Breck et al., 2017 és Sculley et al., 2015 gondolatain alapul.

Itt fontos megjegyezni, hogy a megbízhatóság felhasználói értelmezése a gyakorlatban nem szükségszerűen tükrözi ezt a gondolatot. A fogalom jelentéstartalmának minden érintett félre kiterjedő pontos meghatározása messze túlmutat ezen írás keretein. Ugyanakkor álláspontunk szerinte egy olyan kérdésről van szó, amely a mesterséges intelligencia, illetve a nagy nyelvi modellek esetében különösen fontos, mivel e területen sokkal nehezebb egzakt validációs technikákat alkalmazni az eredmények megbízhatósága tekintetében.

Ígéretek

Ígéretek alatt jelen írás keretei közt Ries, 2011 alapján az alapítók, csapatok és szállítók által a marketing, az értékesítés, az adománygyűjtés és a sajtó területén tett explicit és implicit állítások értendők, amelyek az adott termékhez szolgáltatáshoz közvetlenül kapcsolódnak. A mesterséges intelligencia és nagy nyelvi modellek előtérbe helyezése okán ezen cikk keretei közt első sorban azokkal ígértetekkel foglalkozunk, amelyek tartalmi jellegűek, így a megjelenéssel és a teljesítménnyel, mint például sebesség, kapcsolatos állításokkal nem foglalkozunk. Azt is érdemes meg kiemelni, hogy az ígéreteknek, épp a témakör miatt, két területe fontos igazán; a marginális helyzetek, illetve egyediség kezelése, valamint általános megértés, azaz a lényeglátás pontossága.

Elvárások

Az írás keretei közt elverásokként kezeljük az érdekelt felek mentális modelljeit arról, hogy mit tud nyújtani a rendszer. Egyes szerzők, mint például Fenn és Raskino, 2008 szerint olyan tényezekről van szó, amelyeket a hype ciklusok, a korábbi termékek és a társadalmi narratívák alakítanak. Az írás keretei között az elvárások időbeli változása kevésbé jelenik meg, illetve azok eltérő természete is csak részben nyer teret a két példa életciklusából fakadó különbözősége folytán. Ennek oka igen egyszerű. A cikk a kihívásokra, illetve azon kihívásokra koncentrál, amelyek a mesterséges intelligencia alkalmazásához, illetve az ehhez kapcsolódó

bizalomhoz követlenül kötődik. Épp ezért az időbeli változás vizsgálata nem fér bele az írás kereteibe.

Bizalom

A fentiekkel ellentétben a bizalom esetében nehéz az egyes fogalommeghatározásokat egységes platformra hozni a szoftver termékek és szolgáltatások esetében. Ez a szakirodalommal kapcsolatos megállapítás két szempontból is érdekes. Egyfelől a bizalom egy kvázi alapfogalomnak tekinthető az életünkben, hiszen olyasvalamiről van szó, ami szinte születésünktől velünk fejlődik, mégis – vagy talán épp ezért – egyben olyan természetű is, amelyet nehéz megragadni. Másfelől a bizalom megítélése egy szoftver esetében annak absztrakt természete miatt talán még nehezebb. Ez utóbbi megállapítás további érdekes gondolatokhoz vezet, mivel a mesterséges intelligencia alkalmazások esetén a bizalom kérdése még fontosabb, hiszen az eredmény nem csak a megírt kódon, hanem a betanító adaton is múlik. Ezen még inkább túlmutatnak a nagy nyelvi modellt használó alkalmazások, mivel az ilyen rendszerek hitelesítése egzakt értelemben lehetetlen, hiszen a leginkább elfogadott mérési módszer az, ha az egyik nyelvi modellt a másikkal validáljuk, ami azért különös megoldás, mivel nem tartalmaz külső, teljesen obejktív tényezőt.

A fentiekre figyelemmel meglepő, hogy az MI és a bizalom kérdése mérsékelten kutatott kérdéskör, illetve a kutatások ellentmondásoktól sem mentesek. Egy egyszerű kiindulásként megjegyzendő, hogy a szoftverekhez kötődő bizalom Mayer et al., 1995 szerint a rendszerre való támaszkodás hajlandóságában mutatkozik bizonytalan körülmények között, amelyet az észlelt kompetencia, az integritás és a jóindulat befolyásolnak.

Általános jellemzők

Minden egyes startup, illetve fejlesztés tekintetében az alább szereplők különíthetők el:

- Ötletgazda
- Szoftvermérnök (fejlesztő)
- Csapat (vállalkozás)
- Kockázatitőke-befektető
- Felhasználó
- Versenytársak
- Szabályozó és felügyelő szervek (hatóságok, állam)

Az alábbiakban röviden bemutatom az egyes szereplők általános helyzetét személyes tapasztalatiam alapján.

Ötletgazda

Az ötletgazda az a személy, akivel a szakirodalom külön nem foglalkozik, őt – szerzőtől függően – a fejlesztő vagy a vállalkozó személyével azonosítják – pedig a valóságban egyáltalán nem ritka, hogy egy-egy ötletgazda külön entitásként van jelen egy startup életében még akkor is, ha adott esetben a menedzsment része, tehát, mint ilyen, részben a vállalkozó szerepkörét gyakorolja. Az ötletgazda célja, hogy a termék vagy szolgáltatás valósuljon meg. Az ő esetében a motivációt három mondattal leírhatjuk:

- *"Legyek jelen."*
- *"Legyek sikeres."*
- *"Legyek átütő."*

Ebből következően az ötletgazda feladata elmagyarázni a terméket, azaz megmondani minden más szereplőnek voltaképp miről is van szó. Első körben az ő feladata az is, hogy elmondja, hogyan lesz az ötletből termék. E ponton fontosnak tartjuk megjegyezni, hogy az ötletgazda nem szükségszerűen egy személy, de mindenképpen olyan személyek egysége, akik egy entitást alkotva dolgoznak az ötlet alapvető megvalósításán, azaz termékké alakításán. Az ötletgazda feladata még ideális esetben a termék bemutatása és képviselete is, akár a teljes életciklus során. A különösen innovatív megoldások elképzelhetetlenek ötletgazdáik személyes története nélkül. Az ötletgazda két legfontosabb dilemmája az értékajánlat és a tényleges működés köré csoportosulnak. Az első esetben az álmokat és a vágyakat, a második esetben pedig az elvárásokat és az erőforrásokat kell összehangolnia a valósággal.

Szoftvermérnök (fejlesztő)

A fejlesztőt egyetlen cél mozgatja. A kívánt megoldás működjön. A terméket illetően ez a cél az alábbi három mondatban összefoglalható:

- *"Tudja, amit kell."*
- *"Legyen stabil."*
- *"Legyen skálázható."*

A fejlesztő első feladata a szoftver megtervezése, ezt a megvalósítás, végül a tesztelés követi. A valóságban ez a három feladat folyamatosan ismétlődik. Ezt az ismétlődést iterációs ciklusnak is nevezik, hiszen a legtöbb modern szoftver fejlesztése iteratív módon történik. Bár nem kimondottan fejlesztés, a szoftvermérnöki csapat feladata az üzemeltetés is. Ez azért különül el, mert itt új funkciók hozzáadása nem lehet cél, ám a megbízhatóság, a stabilitás növelése ezt mégis megkívánhatja olykor-olykor.

A fejlesztő dilemmái abból adódnak, hogy a fenntarthatóságat és az elvárásoknak való minél teljesebb megfelelést egyszerre kell fenntartania. A fenntarthatóság jegyében optimális megoldásokat kellene fejlesszen oly módon, hogy azok jövőállók legyenek. Ez a kettős már

akár önmagában is hordozhat nehezen összeegyeztethető aspektusokat, de az elvárások mindenképp ellentétes élt adnak, hiszen a gyors fejlesztés, a változatok mielőbbi kiadása rendszerint nem segíti a tartós előrelátó tervezést már pusztán az idő szűkössége miatt sem.

Csapat (vállalkozás)

A vállalkozást, amely a terméket vagy szolgáltatást kínálja, voltaképp egyetlen cél mozgatja, hogy a termék, vagy szolgáltatás, azaz a cég létezzen. Ezt három alapvető feltétel biztosítja:

- *"Legyen felhasználó."*
- *"Legyen fenntartható."*
- *"Ne legyen baj."*

Annak érdekében, hogy mindez megvalósuljon, a vállalkozásnak értékesíteni kell tudni a terméket, működtetnie kell a céget, illetve folyamatosan őrködnie kell a kockázatok csökkentésén legyen szó akár működési, akár szabályozói kockázatokról.

A vállalkozás dilemmái két fő szempont köré összpontosulnak, az egyik az erőforrás-menedzsmenthez kapcsolódik, míg a másik oldal az elvárásoknak való maradéktalan megfelelés, illetve a kellő figyelem biztosítása minden egyes feladatra. Mivel az erőforrás a legvégesebb és a legkorlátosabb a valóságban, a dilemmák dőlési iránya egyértelmű, épp ezért a vállalkozás művészete voltaképp az erőforrások megfelelő allokációja.

Minden, a vállalkozói működéssel kapcsolatos kérdés a jelen írásnál sokkal mélyebben tárgyalt más szerzők által, épp ezért itt nem célunk ezt mélyebben ismertetni.

Kockázatitőke-befektető

Minden kockázati tőkebefektető végső célja egyetlen dolog, a profit. Ennek érdekében sok munkát kell neki is elvégeznie. A termék életszakaszait tekintve három fő lépés különíthető el:

- *"Valósuljon meg."*
- *"Ne bukjon el."*
- *"Termeljen profitot."*

Bár talán hajlamosak vagyunk a befektetőre úgy gondolni, mint aki csak a pénzt adja, a valóságban ez a feladatkör sokkal többet rejt magában. Természetesen a befektető feladata biztosítani a tőkét, de ezen túlmenően ő végzi a megvalósító csapat értékelését, ő biztosítja a megfelelő embereket, illetve képes kiegészíteni egy hiányos csapatot, valamint szakmai értelemben gondozza a megvalósítókat, hogy a kellő tudás birtokában nézhessenek szembe a kihívásokkal. Ezen felül folyamatos kontrollt gyakorol a folyamatok felett, hogy a valóság minél inkább közelítsen az elképzelésekhez, elvárásokhoz – első sorban a menedzsment, a pénzügyek és az előrehaladás tekintetében.

A befektető dilemmái a minőséggel kapcsolatosak, azaz a megfelelő ötletet a megfelelő csapattal valósítják-e meg.

Felhasználó

A felhasználót egyetlen cél mozgatja egy új termék esetében. Legyen érdemes használni, vagy másként fogalmazva érje meg a befektetett pénzt, illetve a megtanulásába vetett energiát. Ez a gondolatmenet három alapvető szemponttal is leírható:

- *"Legyen hasznos."*
- *"Legyen jó."*
- *"Legyen megbízható."*

A felhasználónak egyetlen feladata van, használja, illetve élvezze a terméket vagy a szolgáltatást. Egyes szoftverek esetében lehetőség van érdemi visszajelzések küldésére, illetve közreműködésre is, de ezek alapvetően opcionálisak, nem általános szükségletek.

A felhasználónak egyetlen központi dilemmája lehet a termék vásárlása előtt, hogy érdemes e használni. Használat közben pedig folyamatosan azon gondolkozhat, megfelel-e az adott termék, illetve szolgáltatás az általa elvártaknak. Ez utóbbi eldöntése a gyakorlatban a legritkább esetben könnyű, hiszen egészen biztosan sosem pontosan olyan egy termék vagy szolgáltatás, mint amilyennek elképzeltük. Sőt, a gyakorlatban rendszerint van, amiben jobb, miközben van, amiben rosszabb, esetleg egészen máshogy működik, mint gondoltok, így el sem tudjuk dönteni, jobb vagy rosszabb e valójában.

Versenytársak

Egy-egy versenytársnak a valóságban nincs aktív célja egy konkurencia fejlesztése kapcsán, ám a gyakorlatban a saját tevékenységére mindenképp lehetnek alapvető elvárásai, amelynek központi eleme, hogy az új versenytárs ne legyen ártalmas számára. Ez a gyakorlatban három egyszerű gondolattal is kifejezhető:

- *"Ne legyen."*
- *"Ne vigyen el ügyfelet."*
- *"Tartsa be a szabályokat."*

A fenti sorrend egyfajta kívánságlistaként is felfogható, mely tekintetében jóhiszemű és jogszabályokat betartó keretek között az elsőre csak egy felvásárlás révén lehet ráhatása. A másik kettőre viszont reagálhat a cég, így különösen figyelhető a konkurenciát, hogy betart-e minden szabályt, illetve fejlesztéseket eszközölhet, hogy megtartsa vagy akár növelje ügyfélkörét. E tekintetben egy régóta a piacon lévő cég akár inspirációt is nyerhet, tanulhat is a versenytárstól.

A versenytársak tekintetében előálló dilemmákról számos közgazdaságtani forrás részletes és alapos elemzést közöl, melyek messze túlmutatnak ezen írás keretein. Annyi mindenképpen elmondható, hogy a másik piaci fél valós helyzetének és valós értékajánlatának minél pontosabb ismerete nagyon sokat segít a megfelelő piaci pozicionálásban.

Szabályozó és felügyelő szervek (hatóságok, állam)

Bár vállalkozóként nem mindennap gondolja ezt az ember, valójában a szabályozó hatóság, illetve az állam elemi érdeke is, hogy egy-egy cég létezzen, hiszen a cégek a gazdaság motorja. Ugyanakkor a létezésnek számos peremfeltétele van. A nagy kép szintjén három alapvető faktor megléte mindenképpen elsőrendű:

- *"Működjön együtt."*
- *"Legyen jogkövető."*
- *"Legyen hasznos."*

Az állam, illetve szerveinek működése, feladatai és dilemmái tekintetében a szakirodalom tetemes. Ezek felszínes bemutatása sem témája ezen cikknek. Ugyanakkor piaci szinten a legfontosabb feladatnak talán az tekinthető, hogy a szabályozás megfelelő és egyenlő működési környezetet biztosítson, ahol a cégek megfelelő figyelmet kapnak és a hatóságok időben reagálnak a cégek szükségleteire, vagy az esetleges szabályszegésekre egyaránt.

A továbbiakban két, fentebb már felvázolt két példán keresztül azt mutatom be, ugyanezen szereplők milyen problémákkal szembesülnek az adott helyzetben. Annak érdekében, hogy ne csak a saját tapasztalataimat adjam közre az írás keretei közt, az egyes példák esetében konkrét forrásokra is hivatkozok a szereplők kihívásait illetően.

Kihívások egy új, mesterséges intelligencia termék bevezetésénél

Az első példa keretében egy kis csapat egy MI-alapú értekezlet-összefoglalót készít, amely hangot rögzít, beszédet ír le, és tömör összefoglalókat és teendőket készít. A termék kész beszédfelismerést és egy általános célú nyelvi modellt használ enyhe finomhangolással. E példa azért érdekes és gyakori, mert a legtöbb startup valójában olyan terméket vagy szolgáltatást kínál, amely már létező termékek, nyílt forráskódú programok újracsomagolása, újragondolása, vagy ezen szoftverek új összehangolása révén jön létre. Bár elsőre ez a megközelítés nem tűnik túl innovatívnak, a valóságban egy egészen új termék jön létre, ha jól gazdálkodnak a szoftverek adottságaival.

2. Fejlesztői nézőpont

Az előző fejezetben bemutatott általános kérdéseken túlmenően a fejlesztő egy MI alapú megoldás esetében mindig találkozik az adat- és modellkorlátok kérdéskörével. Így például a beszéd megértésének pontossága az akcentusoktól, az adott szakmai terület szókincstől és a mikrofon minőségétől függően változik. A szolgáltatás másik pillére, az összefoglaló sem mentes a hiba lehetőségétől, itt a modell hallucinálhat teendőket, vagy rosszul ismerheti fel a beszélőket (Ji et al., 2023).

A fejlesztés részeként a rendszert folyamatosan kalibrálni kell, ehhez azonban egy értékelési megoldásra is szükség van. Az offline WER (szóhibaarány) és a ROUGE/BERTScore nem elegendőek a felhasználó által észlelt hasznosság mérésére, mely az értékelés egyik alapeszköze. Ehelyett forgatókönyv-alapú értékelésekre, hibatipológiákra és emberi beavatkozáson alapuló értékelésekre van szükség (Breck et al., 2017). Mivel a mai megoldások távol állnak a tökéletestől szükség van arra, hogy a felhasználó is állítani tudja a megbízhatósági küszöbértéket a szövegfelismerés tekintetében. Ennek a kommunikációja segít a termék megfelelő használatában (Guo et al., 2017).

Tekintettel arra, hogy a feldolgozott hanganyagok minden esetben érzékeny adatokat tartalmazhatnak, az adatbiztonság alapkövetelmény. Épp ezért a tárolásnak és feldolgozásnak adatminimalizálást, titkosítást, hozzáférés-vezérlést és megőrzési szabályzatokat kell alkalmaznia az adott felhasználási hely szerinti – például GDPR – szabályokkal és az adatvédelmi normákkal összhangban.

3. Vállalkozói nézőpont

A vállalkozóra, azaz a céget működtető csapatra hárul az egyensúlyozás feladata. Egyfelől a terméket hiteles állításokkal kell bevezetni, amelyek a tipikus körülmények között várható működést mutatják meg. A „Minden jó lesz.” jellegű állítások rendszerint nem működnek. Fontos feladat az is, hogy megteremtődjön az egyensúly a modellek képességei, azaz a modellkártyák és az ügyfél által elvárt képességek, azaz a „képességkártyák” között (Mitchell et al., 2019; Gebru et al., 2021).

A piacra lépés és az életszakaszok menedzsmentje szempontjából a leghasznosabb piaci tapasztalat, hogy a korai alkalmazók tolerálják az extrém hibákat is, ha cserébe korábban nem felelhető értéket vélnek felfedezni a termékben. Később már sokkal fontosabb, hogy a termék produktivitása valós, valamilyen módon mérhető bizonyítékokat tudjon felmutatni.

4. Befektetői nézőpont

A befektetők jó eséllyel nagyon sok mutatós demonstrációt látnak pályafutásuk során. Épp ezért az ő gondolkodásuk messze túl kell mutasson ezen. A lehetőségek mellett fel kell mérniük a kockázatokat is. E folyamat során a legfontosabb tényezők Kaplan et al., 2020 és Hoffmann et al., 2022 szerint is a valós, az adatokból következő előny, a termék tényleges piaci illeszkedése, az irányítás minősége, a belső értékelési folyamatok minősége és költség-haszon görbe.

5. Felhasználói és piaci nézőpont

Egy új termék esetében kiemelten kell figyelni arra, hogy a felhasználók akár kiemelt elvárásokat is megfogalmaznak a mesterséges intelligenciát tartalmazó termékek esetében. Ha nagy az eltérés az elvárás és a valóság között, az túlzott bizalmat vagy bizalmatlanságot okoz (Lee és See, 2004). Épp ezért a bizalmi trendek alakulása kiemelt fontosságú egy-egy fejlesztés jelenét és jövőjét tekintve egyaránt.

6. Szabályozói és állami nézőpont

Bár egy ilyen eszköz jellemzően alacsony kockázatú, az adatvédelmi, fogyasztóvédelmi és platformfeltételek szerinti kötelezettségek azért erre a termékre is érvényesek. Így például a NIST mesterséges intelligencia szoftverekre kidolgozott RMF és az ISO/IEC 23894 szabvány kockázatkezelési eszközeinek alkalmazása erősen javasolt. Ezen felül számos állam fokozottan figyeli azt is, a termék vagy szolgáltatás marketingje nem tartalmaz-e félrevezető állítást az MI-t illetően.

Mint azt már azáltalános részben is említettem, az államok hangsúlyozzák a kkv-k termelékenységét, miközben biztosítják az adatvédelmet és a tisztességes versenyt, épp ezért egy jó szabályozói környezet ösztönözheti a tervezésénél fogva megbízható termékeket (OECD, 2019).

Kihívások meglévő termék MI-t tartalmazó továbbfejlesztésénél

A második példa egy tipikus esetet mutat be. Egy kiforrott SaaS, azaz szoftver-szolgáltatás egy olyan funkcióval bővül, amely rutinszerű ügyfél-e-maileket és összefoglalókat készít a meglévő dokumentumokból. Az alaptermék már most is ügyfeleket szolgál ki, és létrehozott SLA-kkal és auditnaplókkal rendelkezik.

7. Fejlesztői nézőpont

A fejlesztőcspat számára a legfontosabb szempont, hogy az eredeti, determinisztikus programrészeket teljes mértékben el tudja különíteni a generatív modellek által is befolyásolt folyamatoktól, mivel a két megoldás eltérő üzemeltetési és mérési módszereket követel meg. Ezzel párhuzamosan olyan védelmet is be kell iktatni, amely csökkenti a modell hallucinációit. Ezt érdemes a vállalat, illetve a termék már meglévő tartalomi tapasztalataira alapozni (Ji et al., 2023). Az újonnan beépülő funkciók esetében a dokumentálásnak és az adatutaknak is világosnak kell lenniük (Mitchell et al., 2019). Ez utóbbi biztonsági és adatvédelmi kritérium is egyben.

Mint azt már fentebb is említettük, az MI használata az értékelési gyakorlatot is megváltoztatja. Ez utóbbi nem minden esetben csupán egyszerű praktikus szempont, olykor a szabályozói környezet is előírja, így egyes területeken a NIST megfelelés is megköveteli.

8. Vállalkozói nézőpont

Egy vállalkozás számára kiemelten fontos, hogy amennyiben egy szoftvert fejlesztenek, az MI integrációja ne egy szükségszerű frissítésnek, hanem valódi értékteremtő beavatkozásnak tűnjön, amely a rendszer teljesítményének érdemi növekedésével jár valamilyen értelemben. Az új szolgáltatásoknak meg kell jelenniük a felhasználói szerződésben is, amelyet így módosítani kell.

9. Befektetői nézőpont

Egy már üzemelő szoftver esetében a befektetői oldal számára minden más piaci körülménnyel szemben a legfontosabb kérdés az, mit nyer és mit veszít a változással, azaz a gyakorlatban hány új ügyfél érkezik és mennyi morzsolódik le, illetve az ügyfelek összes költése, azaz a tényleges bevétel miként változik (Sculley et al., 2015).

10. Felhasználói és piaci nézőpont

Mint arra Lee és See, 2004 már jóval az MI kora előtt rámutat, minden szoftver továbbfejlesztése során alapvető szempont figyelemmel kísérni a felhasználó elvárásokat és bizalmat a változásokat illetően. Ezen felül a mesterséges intelligencia esetében a technológiai

komplexitást úgy kell megvalósítani, hogy azzal a szolgáltatás hozzáférhetősége semmilyen értelemben sem szenvedjen csorbát (Barocas et al., 2019).

11. Szabályozó és állami nézőpont

Egy termék továbbfejlesztése esetén a szabályozói és állami oldal érdekeltségében akkor áll be változás, ha az adott területen önmagában a mesterséges intelligencia alkalmazása magas kockázatúvá teszi a terméket. Tipikusan ilyen területek a pénzügy, az egészségügy és a közügyek. Ilyenkor különösen fontos szem előtt tartani, hogy az algoritmikus auditálási gyakorlatok és hatásvizsgálatok segítenek a biztosítéki réseket megszüntetni (Raji et al., 2020). Egy mesterséges intelligenciát is tartalmazó szoftverbővítés esetén általános tanulság, hogy egy bevált termék esetében a korlátozó tényezők nem a nyers modellképességből fakadnak, hanem a fegyelmezett integráció lehetőségei által meghatározottak. Egy ilyen folyamatnak minden esetben tiszteletben kell tartania a meglévő megbízhatósági szerződéseket és megfelelési rendszereket.

Gyakorlati útmutató mesterséges intelligencia startupok számára

Néhány technológiai módszerrel minden startup sokat tehet a fenntartható elvárások érdekében:

- Eloszlás-eltolódás és törékenység: Minden betanító adat esetében figyelembe kell venni, hogy a teljesítmény romlik a tanítási eloszlásokon kívül, épp ezért a folyamatos értékelés és adatfrissítés elkerülhetetlen és szükséges gyakorlat (Geirhos et al., 2020).
- Megfelelő kalibráció és szelektív alkalmazás: A modellek tévesen kalibrált bizonyossága félrevezetheti a felhasználót, épp ezért tematikus és egyéb határokat, megbízhatósági küszöbértékeket kell bevezetni, hogy neveljük a biztonságot (Guo et al., 2017; Lee és See, 2004).
- Hallucinációk és a valóság: A generatív rendszerek tényeket gyárthatnak. Ezt a problémát, illetve kockázatot a válaszok visszakereshetősége, a korlátozott dekódolás és utólagos összehasonlítás mérsékli ugyan, de teljes mértékben soha nem szünteti meg (Ji et al., 2023).
- Kibervédelem: Minden szolgáltatást védeni kell a felhasználó rosszindulatú viselkedésével szemben. Ennek eszköze az izoláció, a be- és kimeneti szűrés és a házirendek implementálása (Biggio és Roli, 2018).
- Adatvédelem: A betanító adatok és az érzékeny adatok szivárgása folyamatos kockázat, épp ezért az adatbiztonságnak tervezési szemponttá kell válnia, amelynek keretében

minden adatáramlással és az adatpontok eredetével is tisztában kell lenni (Carlini et al., 2021; Gebru et al., 2021).

- A Goodhart-törvény alkalmazása: Nem szabad hagyni, hogy a mérőszámok hajszolása torz fejlesztési irányokat indukáljon, ennek érdekében folyamatosan változatos metrikákat kell alkalmazni és ezeket kvalitatív felülvizsgálattal kell ellenőrizni.
- Költség–teljesítmény kompromisszumok: Noha elméleti síkon minden rendszer egyszerű módszerekkel is könnyen skálázható, a gyakorlatban a számítási hatékonyságra optimalizált tanítás és a körültekintő modellválasztás hatékonyabb módszer az egyszerű skálázásnél (Kaplan et al., 2020; Hoffmann et al., 2022).

Néhány működési módszerrel minden startup sokat tehet a jó belső működés érdekében:

- Dokumentáció és átláthatóság: A modellkártyák, adatlapok, értékelési protokollok és változásnaplók támogatják a tájékozott használatot és a szabályozói felkészültséget (Mitchell et al., 2019; Gebru et al., 2021).
- Kockázatkezelési keretrendszerek: A NIST AI RMF és az ISO/IEC 23894 megkönnyítik és egyszerűsítik a veszélyek azonosítását, illetve mérési, veszélycsökkentő és monitorozási gyakorlatokat biztosítanak.
- Elszámoltathatóság és auditálás: A belső auditok, harmadik féltől származó értékelések és hatásvizsgálatok megmutatják, mennyire képes a működés és az előrehaladás megfelelni az elvárásoknak (Raji et al., 2020; European Union, 2024).
- Emberközpontú tervezés: Az olyan egységes felület, amely segíti az ellenőrzést, a javítást és implementálja a fellebbezés lehetőségét is növelik a felek közt szükséges bizalmat (Lee és See, 2004).

Néhány termékfejlesztési módszerrel minden startup sokat tehet a siker érdekében:

- A megbízhatósági tartomány meghatározása: Rögzíteni kell az üzemi feltételeket, bemeneteket, várható hibamódokat.
- Értékelési keret: Kombinálni kell az offline metrikákat emberi értékeléssel, helyzet alapú tesztekkel és ún. „red team” tesztekkel.
- Megfigyelhető működés: Monitorozni kell az adatminőséget, teljesítményt, biztonsági eseményeket és költségeket. Implementálni kell a visszakereshetőséget és az incidenskezelést.
- Ígéret-bizonyíték tandem: Képességkártyákat kell létrehozni mért szcenáriókhoz kötve. Az ún. SLA-kat a fenntartható működéshez kell igazítani.

- Adatkezelés és a magánszféra védelme: Kezelni kell az adateredetet, megőrzést és hozzáférést. Minimalizálni kell az érzékeny adatok használatát, és be kell vezetni az adatok felhasználó szintű ellenőrzését.
- Irányítási és működési sztenderdek alkalmazása: Fel kell térképezni a megfelelő iparágban használt szabványokat, így például NIST vagy ISO és ezeknek megfelelően kell implementálni a működést, irányítást.
- Felhasználói visszajelzés és kontroll: Teret kell adni kontrollnak és magyarázatoknak az egyes területeknek megfelelően. Implementálni kell a felhasználói visszajelzések fogadását és kezelését.
- Technikai függőségek diverzifikálása: Amennyire lehetséges, kerülni kel az egy szolgáltatóhoz kötöttséget; modell-útválasztás és caching alkalmazása a költség, késleltetés és minőség egyensúlyára.
- Fejlődési ösztönzők beépítése: Fontos, hogy a növekedési célok a megbízható kimenetekhez kötődjenek, ne hiúságmetrikákhoz; figyelni kell az ún. „Goodhart-hatásokat”.

Összefoglalás

Jelen írás célja bemutatni azt a szűk mezsgyét, amely az MI-t használó startupok működését tartósan jellemzi. A vállalkozások által épített rendszerek valós tudása és a piaci elvárások között rendszerint jelentős különbség van, mely folyamatos feszültségek forrása. Fegyelmezett mérnöki munkával és valós, minél objektívebb módszereket alkalmazó értékeléssel, valamit az ígéretek és az elvárások körültekintő menedzsmentjével, illetve az előrelátó terméktervezéssel a helyzet a legtöbb esetben kezelhető. Ugyanakkor a fentebb felvázolt két példa megmutatja, hogy akár új eszközt indítunk, akár érett terméket frissítünk, a tartós bizalom akkor születik, ha az érintettek ellenőrizhetik a teljesítményállításokat, értik a korlátokat, és számíthatnak olyan mechanizmusokra, amelyek a rendszert a megbízhatósági tartományban tartják. Ám azt végső üzenetként fontosnak tartjuk kihangsúlyozni, a megbízható és hiteles működés korántsem azonos a szolgáltatósok tartalmának feltétel nélküli elfogadhatóságával, mely utóbbi vizsgálata kiemelt és kulcsfontosságú minden egyes alkalmazás esetében.

Irodalomjegyzék

Barocas, S., Hardt, M. and Narayanan, A. (2019) Fairness and Machine Learning. Elérhető itt: https://books.google.hu/books?hl=hu&lr=&id=HuGwEAAAQBAJ&oi=fnd&pg=PR9&dq=Barocas,+S.,+Hardt,+M.+and+Narayanan,+A.+(2019)+Fairness+and+Machine+Learning.&ots=Q

0geIta4ma&sig=kdKwNCTDT-OIyS5dHqf5QJiJjNc&redir_esc=y#v=onepage&q=Barocas%2C%20S.%2C%20Hardt%2C%20M.%20and%20Narayanan%2C%20A.%20(2019)%20Fairness%20and%20Machine%20Learning.&f=false

Bender, E.M., Gebru, T., McMillan-Major, A. and Mitchell, S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT), pp. 610–623. Elérhető itt: https://dl.acm.org/doi/abs/10.1145/3442188.3445922

Biggio, B. and Roli, F. (2018) Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition. Elérhető itt: https://doi.org/10.1016/j.patcog.2018.07.023

Breck, E., Cai, S., Nielsen, E., Salib, M. and Sculley, D. (2017) The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction. Elérhető itt: https://ieeexplore.ieee.org/abstract/document/8258038

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U., Oprea, A. and Raffel, C. (2021) Extracting Training Data from Large Language Models. 30th USENIX Security Symposium, pp. 2633–2650. Elérhető itt: https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting

Christiano, P., Leike, J., Brown, T., Martic, M., Legg, S. and Amodei, D. (2017) Deep Reinforcement Learning from Human Preferences. Advances in Neural Information Processing Systems 30. Elérhető itt: https://proceedings.neurips.cc/paper_files/paper/2017

European Union (2024) Artificial Intelligence Act (EU AI Act). Elérhető itt: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/hun

Fenn, J. and Raskino, M. (2008) Mastering the Hype Cycle: How to Choose the Right Innovation at the Right Time. Harvard Business Review Press. Elérhető itt: https://books.google.hu/books?hl=hu&lr=&id=ApeSoRHCcGwC&oi=fnd&pg=PR6&dq=Fenn,+J.+and+Raskino,+M.+(2008)+Mastering+the+Hype+Cycle:+How+to+Choose+the+Right+Innovation+at+the+Right+Time.+Harvard+Business+Review+Press.&ots=d9OF6C3jaa&sig=ypukAHoEPY1Mu6auUQhHZK8G4UM&redir_esc=y#v=onepage&q=Fenn%2C%20J.%20and%20Raskino%2C%20M.%20(2008)%20Mastering%20the%20Hype%20Cycle%3A%20How%20to%20Choose%20the%20Right%20Innovation%20at%20the%20Right%20Time.%20Harvard%20Business%20Review%20Press.&f=false

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H. and Crawford, K. (2021) Datasheets for Datasets. Communications of the ACM, 64(12), pp. 86–92. Elérhető itt: https://dl.acm.org/doi/fullHtml/10.1145/3458723

Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M. and Wichmann, F.A. (2020) Shortcut learning in deep neural networks. Nature Machine Intelligence, 2, pp. 665–673. Elérhető itt: https://doi.org/10.1038/s42256-020-00257-z

Gompers, P. and Lerner, J. (2004) The Venture Capital Cycle, 2nd ed. MIT Press. Elérhető itt: https://books.google.hu/books?hl=hu&lr=&id=yEAcswbX1fEC&oi=fnd&pg=PR7&dq=Gompers,+P.+and+Lerner,+J.+(2004)+The+Venture+Capital+Cycle,+2nd+ed.+MIT+Press.&ots=uhwSNv6Cuc&sig=xEUKUq98xQMs1C8ujvGyxjREdN0&redir_esc=y#v=onepage&q&f=false

Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q. (2017) On Calibration of Modern Neural Networks. Elérhető itt: https://proceedings.mlr.press/v70/guo17a.html

Hoffmann, J., Borgeaud, S., Mensch, A., et al. (2022) Training Compute-Optimal Large Language Models. Elérhető itt: https://arxiv.org/abs/2203.15556

ISO/IEC (2023) ISO/IEC 23894:2023 Information technology — Artificial intelligence — Risk management. International Organization for Standardization.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A. and Fung, P. (2023) Survey of Hallucination in Natural Language Generation. ACM Computing Surveys (in press). Elérhető itt: https://dl.acm.org/doi/full/10.1145/3571730

Kaplan, J., McCandlish, S., Henighan, T., et al. (2020) Scaling Laws for Neural Language Models. Elérhető itt: https://arxiv.org/pdf/2001.08361/1000

Lee, J.D. and See, K.A. (2004) Trust in Automation: Designing for Appropriate Reliance. Human Factors. Elérhető itt: https://doi.org/10.1518/hfes.46.1.50_30392

Mitchell, M., Wu, S., Zaldivar, A., et al. (2019) Model Cards for Model Reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*). Elérhető itt: https://dl.acm.org/doi/abs/10.1145/3287560.3287596

NIST (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology. Elérhető itt: https://www.nist.gov/itl/ai-risk-management-framework

OECD (2019) Recommendation of the Council on Artificial Intelligence. Elérhető itt: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Ouyang, L., Wu, J., Jiang, X., et al. (2022) Training Language Models to Follow Instructions with Human Feedback. Erléhető itt: https://arxiv.org/abs/2203.02155

Raji, I.D., Smart, A., White, R., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Ledford, J., Theron, D. and Barnes, P. (2020) Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. Proceedings of the 2020 Conference on Fairness,

Accountability, and Transparency (FAT*), pp. 33–44. Elérhető itt: https://doi.org/10.1145/3351095.3372873

Ries, E. (2011) The Lean Startup. Crown Business.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F. and Dennison, D. (2015) Hidden Technical Debt in Machine Learning Systems. Advances in Neural Information Processing Systems 28. Elérhető itt: https://proceedings.neurips.cc/paper_files/paper/2015

The Reliability of Smart Homes and People's Attitude Toward Them

Az okosotthonok megbízhatósága és az emberek bizalma

Zila Zsófia Noémi

Absztrakt

Ez az esszé az okosotthon technológiák megbízhatóságát és az emberek ezekkel kapcsolatos véleményét vizsgálja. Az írás amellett érvel, hogy az okosotthonok megbízhatóbbak, mint azt sokan gondolják, feltéve, hogy a felhasználók tisztában vannak a lehetséges kockázatokkal, és tudják hogyan kezeljék ezen rendszereket. A kutatás vegyes módszertan alkalmazásával történt, beleértve egy 40 fős kérdőívet és egy interjút Balogh Imrével, a ProGuard alapítójával, azon céllal, hogy feltérképezze az okosotthon technológia jelenlegi helyzetét, tapasztalatokat, valamint a leggyakoribb aggályokat, például a biztonságot, a költségeket és a felhasználhatóságot. Az esszé különböző tanulmányok bevonásával szélesebb körű rálátást kínál a témára. Az eredmények azt mutatják, hogy bár a legtöbben hallottak már az okosotthonokról, és sokan használnak okoseszközöket, többen továbbra is bizonytalanok azok megbízhatóságát illetően. Ezek a kétségek elsősorban biztonsági aggodalmakból fakadnak, illetve abból, hogy a felhasználók nem teljesen értik, hogy működnek ezen rendszerek. A tanulmány kitér továbbá a költséghatékonyságra, az energia-megtakarítási lehetőségekre, valamint a felhasználóbarát kialakítás fontosságára is. Kiemeli, hogy az okosotthonok valós előnyöket kínálhatnak (különösen a különleges igényű emberek számára). Ugyanakkor ahhoz, hogy ezek a rendszerek szélesebb körben megbízhatónak minősüljenek, elengedhetetlen a felhasználók jobb tájékoztatása/tájékozottsága és az iparág nagyobb fókusza a megbízhatóságra és a fenntarthatóságra.

Abstract

This essay looks into how reliable smart homes really are and how people feel about them. It aims to show that smart homes are more trustworthy than many believe, as long as users are aware of the possible risks and know how to deal with them. Through a mixed-methods approach (including a 40-participant questionnaire and an interview with Imre Balogh, founder of ProGuard) the research investigates the current state of smart home technology, user experiences, and prevailing concerns such as security, cost, and usability. The essay also includes information from various articles and studies to give a broader view of the topic. The results show that while most people have heard of smart homes and even use some smart

devices, many are still unsure about their reliability, mostly due to safety concerns and not fully understanding how these systems work. The study also discusses the cost, energy-saving features, and the importance of user-friendly designs. It highlights that smart homes can offer real benefits, especially for people with disabilities or special needs, but for these systems to be fully trusted, users need better information and companies should focus more on reliability and sustainability.

I. **Introduction**

Have you ever wondered about what houses will look like in the future and how technology will change them? Technological advancements are changing the world we live in, whether we like it or not. This change is apparent when it comes to houses, and the smart home technology many people are using. In the media, we can see contradicting news about the reliability of smart homes and the security they can offer, some write about scandals relating to smart homes (Kelly, S M 2024; Riley, A 2020), while others write about how good these new inventions are (Cericola 2022). With all this flowing information in the news, it is hard to decide which ones to believe. In my research, I present information about smart homes and how reliable they are. In my research I will not only focus on the reliability of current smart home technologies but also on what people's attitudes are toward these smart home systems. The research aims to prove that smart homes are actually more reliable than people believe them to be, however it is vital that the users know about the possible risks and how to resolve them. This topic will be observed by contrasting people's current opinion on the reliability of smart homes with previously done research about the reliability of smart homes.

II**. Methodology**

There is no univocal definition for *smart item* and *smart home*, however, in the context of this article the terminology *smart item* refers to devices which have the three following characteristics: autonomy, context-awareness, and connectivity. (Silverio-Fernández et al. 2018) In this paradigm, autonomy means that the devices can process information on their own, and they can complete a number of tasks autonomously, without the user having to interact. Connectivity is also key, smart devices are a part of the Internet of Things (IoT), which means that for devices to be considered smart, they have to be able to connect to one another and communicate with each other. Context-awareness refers to a device being able to anticipate the next action to take based on context and based on the users' previous preferences. (Kabir et al.

2015) In the context of this article the term *Smart home* refers to properties equipped with multiple smart appliances/devices connected by a hub, which can be controlled either from personal smart devices (phones, tablets, laptops) or from a smart remote control. I have sent out a questionnaire to find out currently how many people have smart homes, how many people plan on investing in a smart home system and those who do not like smart homes, why they do not like these technologies, and what would make them more appealing. 40 people of different ages participated in the questionnaire, which I sent out online, to the Milestone mailing list, and to my Facebook and Instagram acquaintances, however, I do not know specifically who filled it out (the questions can be found at the end of the document, at appendix 1). Everyone could participate in this research, without any previous knowledge, and most of the participants stated that they do not have substantial knowledge about smart homes. The ages of the participants vary between 14 and 64, so the generations Z (1997-2012) and Y (1981-1996) (Beresford Research 2025), which are the ones of the digital age are represented, as well as the generations born before the internet, which became more commonly used as the 1980s emerged (National Science and Media Museum 2020). Mostly I dealt with quantitative data in my research, which I analyzed by looking at the statistical average for each answer, the median, modus and also by handling the answers of non-smart home owners and smart home owners separately. Afterward I contrasted the analyzed data from the questionnaire with information from previously published studies and articles relating to the origin of smart homes, how reliable they are and how many people have them (the bibliography can be found at the end of the document). These articles that I used for my secondary sources are mostly less than 10 years old, and they all deal with different questions related to smart homes, which are the basis of this article. I have also done an interview with Imre Balogh, the founder and director of ProGuard, and partner company of Ajax Security Systems, focusing on smart security items, as well as other smart home items (interview questions can be found at appendix 2). Although Imre might be biased on the topic, I believe the opinion of someone who profoundly understands how smart homes work is crucial to this research. In my interview I have asked him the most polemical questions about smart homes that people currently have. At the end I will draw a conclusion about the reliability of smart homes and whether they can really be trusted in their current state, or what improvements can be made.

III. **The History of Smart Homes and Their Current State**

The first smart homes began developing in the 1970s, when the first programmable logic controllers were manufactured, which were simple smart home devices used in, for instance,

washing machines or elevators. Ever since then, more and more household objects have a "smart" version of themselves, and there is no sign of this trend stopping (Smartest Home 2024).

In the 1990s the EIB (European Installation Bus), now a part of the KNX (konnex) association was founded, which helped form the smart home market by making the companies in the association produce devices which were able to connect to each other, even if they were produced by different brands (Dzierzek 2013). KNX is a Belgian non-profit-oriented company which is present in 190 countries all over the world (KNX Association n.d.). In the beginning, the devices mainly consisted of lighting, ventilation and temperature controlling items, along with other items in the home.

In the early 2000s the first smart houses with central control were built, and ever since more and more smart houses have been built (Statista n.d.). Currently, smart home items can be controlled by a hub unit, which is responsible for communicating with all the smart devices separately, so it is enough for users to only give instructions to the hub, and the hub can forward these instructions to the desired smart devices (Alzahrani et al. 2020). Most of the time users can communicate with the hub via a phone application, though some hubs have their own remote control panels.

When asked in my questionnaire, 12 out of the 40 participants were neutral to the topic of smart homes, 11 found it interesting, 17 of them were disinterested in the topic, but all of them knew of smart homes. 57,5% of the participants stated that they have some knowledge about smart homes, mostly gained from the internet, such as Facebook, Tiktok or Youtube, or they have heard some things about smart homes from their friends. 29 out of the 40 participants stated that they have some smart items at home, the most common are smart vacuum cleaners, smart TVs and smart lamps. 9 out of the 40 participants defined the home they live in as a smart home (figure 1) but all of them stated that they do not have a fully equipped smart home ecosystem (meaning that not all of their appliances and devices are smart). This shows that even though smart devices have been around for some time, not everyone has them or knows how to use them, and it is important to assess why some people prefer regular homes and how much an average person knows about smart homes.

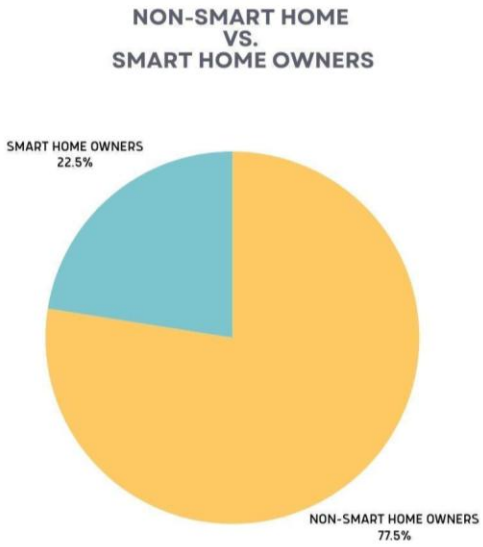


figure 1/1. ábra

Ratio of Smart Home and Non-Smart Home Owners

Okosotthonban és hagyományos otthonban élők aránya

Source/Forrás: edited by the writer/saját szerkesztésű

IV. User Friendliness and Reliability

When it comes to smart homes, there are different kinds of purposes they can serve, based on the devices in them and how the residents utilize these devices and whether or not they can serve their full purpose. In a previously done research (Gøthesen et al. 2023) they state that some homes are simply smart because it is more comfortable for its residents to have a personalized system set up for them, without having to always look for different kinds of remotes, having to vacuum for themselves etc. But there are devices which make houses liveable for people with disabilities, or the elderly, and while devices which offer personalization have a bigger target audience these healthcare smart devices might have a bigger impact on our society in the long run. For example there are cameras, which one can set up inside or outside the house, and they have a built-in fall sensor, which immediately calls for help when someone in the frame falls. The use of touchscreens instead of buttons can also help people living with disabilities, as it can be hard for some people to firmly push buttons, but they can easily touch some part of a screen. Smart home devices can also help parents monitor their kids while tending to errands in the house such as doing laundry or loading the dishwasher. It is cardinal for smart home devices to be easily operable, so that everyone can learn how to use

them, in order for these devices to actually be comfortable to its users. The overwhelming majority (93,5%) of the participants believe that they would easily be able to learn how to use smart home devices, with only two of them disagreeing with this fact.

When asked in the questionnaire whether or not participants think that smart homes are more reliable than regular homes 11 out of the 40 people remained neutral, which was the most commonly occurring answer (figure 2) Because of these functions mentioned above, it is crucial for smart homes to be reliable in terms of consistent performance even though for most people only their comfort depends on it, but for some, it could be their wellbeing or even their life.

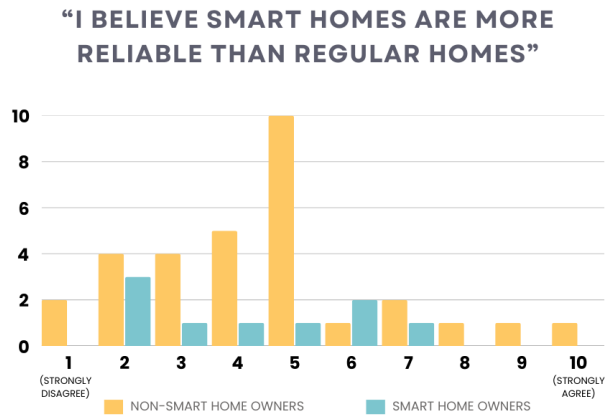


Figure 2/2. ábra

Reliability of Homes

Az otthonok megbízhatósága

Source/Forrás: edited by the writer/saját szerkesztésű

No participant in the questionnaire has completely disagreed with the fact that they have heard negative things about the reliability of smart homes, and 50% of the people said that they do not think that smart homes are more reliable than regular homes, while 27,5% of them remained neutral. So, smart homes are still deemed rather unreliable by many people.

When I asked Imre Balogh about his experience with the reliability of smart homes devices, and the different things that factor into how trustable they are, he said that it is essential to buy smart homes devices from trustable brands, and that solves most of the issues regarding the unreliability of smart home devices. Before buying smart home items it is important to contrast the device we want to other devices and read trustable reviews about them. It is also imperative to note that some devices might only be available for governments and large companies, and not for the civil society. When asked, the participants with a smart home system have stated

that they do some research before buying new devices, however, on average, they did not agree to the fact that they buy products only from brands that they already trust, although most of them agreed that they do thorough research before buying a new device to make sure that they get the most reliable one (figure 3). Buying devices from brands one is unfamiliar with is not necessarily a problem, but it seems to be vital to do thorough research about new brands or devices one wants to buy.

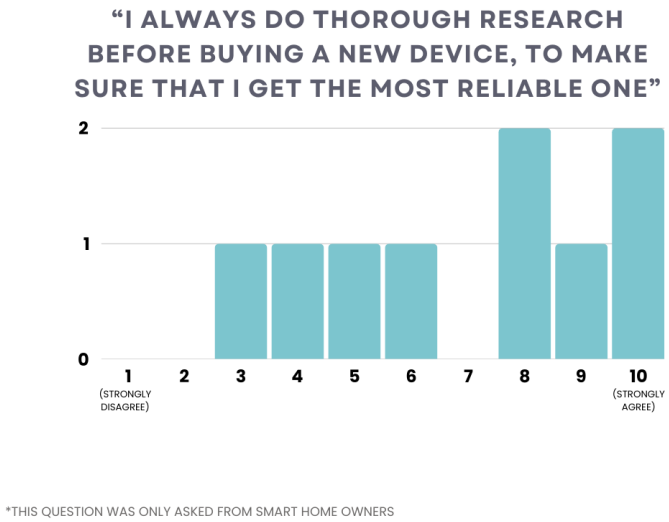


Figure 3/3. ábra

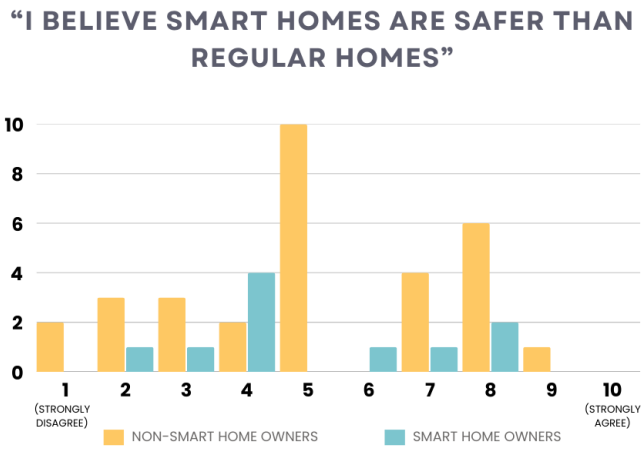
Decision-making process

Döntéshozói folyamat

Source/Forrás: edited by the writer/saját szerkesztésű

V. Security

Home is a place where everyone should feel safe and comfortable according to an article about smart home automation security (Jose and Malekian, 2015). The main concerns about smart homes that we can hear



about in the media are regarding the problem of security. Smart devices record information about the users and they store them as data, which can be a privacy concern in the house (Guhr et al. 2020). Another source points out how high the risk is for personal data being leaked, and that the best way to prevent this from happening is the users being aware of the potential risks and them trying to prevent these (Zielonka et al. 2021). On the other hand, they claim that smart security cameras, smart locks and other smart security devices can be good additions to protecting someone's home if the residents feel comfortable with them. 14 out of the 40 participants remained neutral when they were asked if regular homes or smart homes were safer, the rest of them were torn between the two statements (figure 4).

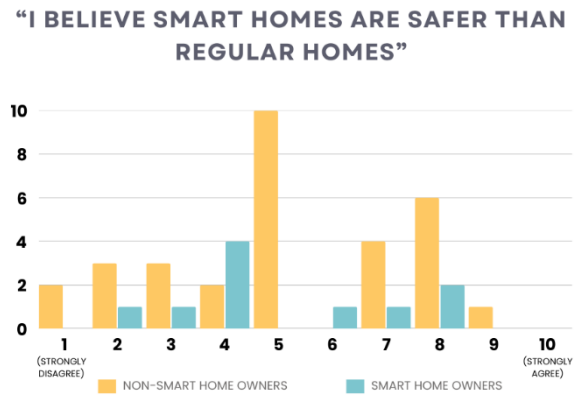


Figure 4/4. ábra

Safety of Homes

Az otthonok biztonsága

Source/Forrás: edited by the writer/saját szerkesztésű

There are ways to make these smart appliances safer, like encryption standards and other protocols that are in use, but that still does not make these devices completely safe (Jose and Malekian 2015). Imre Balogh said that the main privacy concern with smart homes is the fact that many devices communicate through WiFi, which means that anyone who has access to the WiFi, or can hack the WiFi password can see all the information these smart home devices know, and they can access these devices without users even noticing. Instead, devices should communicate in their own encrypted language, which should be independent from WiFi or Bluetooth, so they are immensely less likely to be hacked. This does not mean that such devices do not have internet connection, but they only use WiFi to communicate with the user's phone,

and they do not use WiFi to communicate with each other. These devices can be controlled in different ways from the phone, like an independent control panel if there is a problem with the WiFi, so they can be activated and deactivated unrelated to WiFi connection. When asked about security concerns in the smart homes, none of the participants listed this as an issue that they know about or are concerned about.

The same problem occurs with robotic vacuum cleaners, which know way too much information about one's home for them to be completely considered safe, especially if they are working on the home's WiFi. In order for smart vacuum cleaners to work, they need to have a complete layout of one's home, including doors and windows, which poses a threat for burglary. The European Union has the General Data Protection Regulation (GDPR) to help control the information we can use and store about other people (Guhr et al. 2020) but these regulations often lack updates, and are usually not specific enough. In a research done on the GDPR compliance of smart homes experts argue about whether smart home items actually meet the GDPR standards, (Piasecki 2023) by which users have to receive disclosure of the product/service that they are going to use (this text has to be easy to understand for vulnerable users, as well). Some of the experts are saying that while users of smart home devices usually receive some sort of information about the product, these disclosures are often not detailed enough or are difficult for laymen to understand. Receiving easily understandable information about purchased smart devices is crucial in order to understand their risks and what to do to reduce these. In my questionnaire, more than half of the people agreed that they feel like they are being listened to when they are surrounded by items equipped with microphones such as laptops or speakers with virtual assistants, and the people who have smart home items have actually on average strongly disagreed with the fact that they have a voice assistant who is always on standby. This makes sense because in order for voice assistants to work, they need to be listening to you all of the time, and even if they do not necessarily have a recording of you, they can store data that they learn about you even if they are not being talked to. (Clauser 2019) The data that these devices store can help you see ads and content that are more relevant to you and many more things that can make your online presence more personalized, but it is a personal choice between privacy and comfortability one has to make. 66,7% of the smart-home owner participants also stated that they feel more secure in their home than they previously did in their regular homes. This can be due to many factors but this illustrates perfectly how people can feel secure in smart homes if they are setting boundaries for these devices and making the devices suit their needs instead of just going with the basic setting of these devices. The environment in which one feels secure differs for people. That is exactly the reason why it is

highly important for users to understand how their devices work and be able to configure them, in order to fit their personal needs.

VI. Cost and Energy Saving

When asked in the questionnaire, 16 out of the 31 people replied that they do not have smart homes because they are too expensive to install, although on average the participants do not believe that the upkeep of smart homes is more than the upkeep of regular homes. Many things factor into the price of home automation, including the size of the home, the quality of the devices, the types of devices one desires and whether or not they can install it themselves or they need to call someone to install it for them. (Today's Homeowner 2023) These smart home devices are usually investments which will make life cheaper in the long term, but more expensive in the short term. According to a statistic done by the *New York Times,* LED light bulbs (which all smart lights are) use 75% less energy than lamps did 25 years ago, not to mention the dimming function which can help save an additional 40% of energy (Cericola 2019). However, not all smart home items are good for saving energy. A comparative research done on the contrast of energy efficient and regular smart home items showed an average of 10-15% decrease in the energy consumption of smart washing machines, smart ovens and smart refrigerators (Malysheva et al. 2024). Even the devices that are not specifically energy efficient have eco programmes, which help save energy as well as water in some cases.

Another factor that plays a role in the expenses of a smart home is the servicing of these items, about which there is some speculation, since people started noticing that newer electric devices of all sorts need replacing much sooner than they did decades ago, which can be a marketing technique companies use, but there is no solid evidence for that. In 2018 HOP/Stop Planned Obsolescence, a French organization (Local Futures 2020) filed a legal complaint against Apple claiming that Apple makes its devices slower on purpose, so that people will need to replace their devices sooner and buy new ones (The Guardian 2018). While these lawsuits do not prove that all tech companies make their devices less durable than they could be on purpose, the fact that these cases were so strong that they could be taken to court emphasises the importance of this question in today's world. If allegations such as these prove to be true, they bring up the question whether or not these energy saving functions of new smart home devices are worth it in terms of costs and being environmentally friendly, as well. In the interview with Imre Balogh he said that there is a group of people who install smart home appliances for energy saving purposes but comfort is the main reason there is a demand for smart home appliances, not

energy and money saving. In the light of all this it seems that while it is physically possible to make smart home devices more energy friendly than traditional home appliances, that might not be the current focus of the companies that manufacture them, which does not make them outstanding when it comes to being cost effective and environmentally friendly.

VII. Discussion

Overall smart home systems can be a great way to improve comfortability and personalization in the home, and there is definitely a future for them if people are open enough to learn more about them. Results from the questionnaire show that 72,5% of people already have smart items at home, and 38,7% current non-smart home owners are thinking about investing in a smart home. With the solutions smart homes offer to a variety of problems they are bound to become even more popular in the future. Right now, only 22,5% of participants think that smart homes are more reliable than regular homes, however, the reasons which they justified their concerns with are scientifically not valid concerns, or are concerns that could be solved easily. For smart homes to become more reliable users need to educate themselves about what that reliability would look like and what they can do in order to achieve it. The biggest risks of smart homes like being hacked and unable to control your devices stem from people not actually understanding what precautions they need to take to prevent these accidents from happening. Most of the previously mentioned issues can be solved by buying devices from trustable brands and doing some research about them. Needless to say, it is also crucial for these devices to be up to a standard that all smart devices need to be in order for them to be available on the market. Smart homes can offer great solutions for people living with disabilities or some sort of physical limitations. Fall detection cameras, and health monitoring devices can save lives if used correctly, however, for users to only rely on them instead of caregivers, improvements need to be made. I truly believe that with the investment of time and resources smart homes will be the future of living, and they will make life quality better for everyone. The biggest function that needs improvement is the energy saving possibilities smart items offer. Most of them have some eco-friendly setting but it would be more beneficial if they were programmed eco friendly from the start, instead of users having to configure the devices themselves. Based on the articles I read, the research I conducted and the interview I had done, I found that people are still quite confused about what smart homes are, and how they work, and therefore it is crucial to raise awareness, but hopefully if smart homes become more normalized and common in the future,

people will understand them more, the same way it was with many innovations throughout history.

VIII. Conclusion

This research's objective has been to represent current opinions of different people about smart homes, along with an interview with Imre Balogh, expert in smart home systems and founder and director of ProGuard. The article aims to prove that smart homes are actually more reliable than people believe them to be, but only if the users know about the possible risks and how to resolve them. The interview with Imre Balogh justifies this hypothesis, as well as some answers in the questionnaire, although not all the participants' answers demonstrate clear understanding of how smart homes work and what threats they pose.

In conclusion, smart home systems can offer solutions to many problems and with technology improving, they might be able to solve many more problems that they cannot today. The world is constantly changing, and we as a society have to change with it, in order to not get stuck in one place, and smart homes will be a part of the future, whether we like it or not, so why not work on improving them?

Bibliography

Alzahrani, A Alshamlani, M Hsu, W-C Harish, S & Daim, T 2020 'Personal Transformation: Evaluation of Smart Home Hubs'*World Scientific Series in R&D Management*, 213–243.

Beresford Research 2025 'Age Range by Generation'*Beresford Research*.

Birchley, G Huxtable, R Murtagh, M Meulen, R Flach, P & Gooberman-Hill, R 2017 'Smart homes, private homes? An empirical study of technology researchers' perceptions of ethical issues in developing smart-home health technologies'*BMC Medical Ethics*, 18(1).

Cericola, R 2019 'How to Save Money and Energy With Smart Home Devices*' Wirecutter: Reviews for the Real World*.

Cericola, R 2022 'The Best Smart Home Devices to Help Seniors Age in Place'*The New York Times.*

Clauser, G 2019 'Amazon's Alexa Never Stops Listening to You. Should You Worry?' *Wirecutter: Reviews for the Real World*.

Dzierzek, K 2013 'The Use of KNX/EIB to Control Devices in an Intelligent Home'*Acta Mechanica Slovaca*, 17(3), 50–54.

Gøthesen, S Haddara, M & Kumar, K N 2023 ’Empowering homes with intelligence: An investigation of smart home technology adoption and usage’*Internet of Things*, 24(1), 100944.
Guhr, N Werth, O Blacha, P P H & Breitner, M H 2020 ’Privacy concerns in the smart home context’*SN Applied Sciences*, 2(2).
Jose, A C & Malekian, R 2015 ’Smart Home Automation Security: A Literature Review’ *ResearchGate.*
Kabir, M H Hoque, M R & Yang, S-H 2015 ’Development of a Smart Home Context-aware Application: A Machine Learning based Approach’*International Journal of Smart Home,* 9(1), 217–226.
Kelly, S M 2024. ’That security camera and smart doorbell you’re using may have some major security flaws’*CNN.*
KNX Association (KNX) n.d. ’The legacy of KNX’*KNX Association*.
Local Futures 2020 ’Stop Planned Obsolescence’*Local Futures.*
National Science and Media Museum 2020’ A Short History of the Internet’ *National Science and Media Museum*
Malysheva, A A Rawat, B Singh, N Jena, P C & Kapil 2024 ’Energy Efficiency Assessment in Smart Homes: A Comparative Study of Energy Efficiency Tests’*BIO Web of Conferences*, 86, 01083.
The Guardian 2018 ’France investigates Apple over claims of planned obsolescence’*The Guardian.*
Piasecki, S 2023 ’Expert perspectives on GDPR compliance in the context of smart homes and vulnerable persons’*Information and Communications Technology Law*, 32(3), 385–417.
Riley, A 2020 ’How your smart home devices can be turned against you’*BBC News.*
Silverio-Fernández, M Renukappa, S & Suresh, S 2018 ’What is a smart device? - a conceptualisation within the paradigm of the internet of things’*Visualization in Engineering*, 6(1).
Smartest Home 2024 ’The history of the smart home from 1963-2025: Milestones’*Popular Mechanics Magazine*.
Statista Research Department n.d. ’Penetration rate of the smart homes market in the United States from 2019 to 2028.’*Statista*.
Today’s Homeowner 2023 ’How Expensive Are Smart Homes?’*Smart Home.*
Zielonka, A Wozniak, M Garg, S Kaddoum, G Piran, Md J & Muhammad, G 2021 ’Smart Homes: How Much Will They Support Us? A Research on Recent Trends and Advances’*IEEE Access*, 9, 26388–26419.

Appendix 1: Questionnaire

Questions

2.

What is your age?

Under 18

18-24

25-34

35-44

45-54

55-64

65+

Disclaimer

If you have your own apartment/house, please fill this questionnaire out according to that. If you do not have your own apartment/house yet, please imagine what you would do if you did have your own apartment/house, and make decisions based on that.

3. I am interested in the topic of smart homes (reading articles, watching videos about it).

1-10 (strongly disagree-strongly agree)

4. If you do have some knowledge about smart homes, how did you gain it? Do you think that source is reliable? (newspaper, TikTok, friends?)

5. I understand how smart homes work.

1-10 (strongly disagree-strongly agree)

6. I have heard negative things about the reliability of smart homes.

1-10 (strongly disagree-strongly agree)

7. I believe smart homes offer more comfort than regular homes, but have no significant advantages.

1-10 (strongly disagree-strongly agree)

8. I believe smart homes offer other things beyond comfortability and personalization.
1-10 (strongly disagree-strongly agree)

9. I believe smart homes are safer than regular homes.
1-10 (strongly disagree-strongly agree)

10. I believe smart homes are more reliable than regular homes.
1-10 (strongly disagree-strongly agree)

11. I believe the upkeep of smart homes requires more time than the upkeep of regular homes.
1-10 (strongly disagree-strongly agree)

12. I believe that in cca. 30 years the vast majority of the population will live in smart homes.
1-10 (strongly disagree-strongly agree)

13. I have items in my home which are "smart" (e.g. light bulbs, TV, kitchen appliances, vacuum cleaner, etc.)
1-10 (strongly disagree-strongly agree)

14. I have a full smart home ecosystem.
1-10 (strongly disagree-strongly agree)

15. Please list a few "smart" items you have at home:

16. In simpler terms: Do you have a smart home system?
Yes
(Continue at question 27.)
No
(Continue at question 17.)

Questions for people who DO NOT have a smart home system

17. I am considering investing in a smart home system.
1-10 (strongly disagree-strongly agree)

18. I would invest in a smart home system, but I do not have enough knowledge about them.
1-10 (strongly disagree-strongly agree)

19. I would invest in a smart home system, but it is too expensive.
1-10 (strongly disagree-strongly agree)

20. I think I would easily be able to learn how to use smart home devices.
1-10 (strongly disagree-strongly agree)

21. I would invest in a smart home system, but I do not think the technology is reliable enough yet.
1-10 (strongly disagree-strongly agree)

22. I feel like I am being listened to when I am surrounded by smart devices with microphones (e.g. smartphones, laptops, speakers with virtual assistants)
1-10 (strongly disagree-strongly agree)

23. I would invest in a smart home system, but I have concerns about privacy in a smart home.
1-10 (strongly disagree-strongly agree)

24. I fear that if I had smart appliances they would require more servicing than non-smart appliances.
1-10 (strongly disagree-strongly agree)

25. I usually have a hard time trusting new technologies.
1-10 (strongly disagree-strongly agree)

26. Other reason for not wanting a smart home system:

Questions for people who DO have a smart home system

27. How long have you had a smart home system?

less than 1 year
1-2 years
3-5 years
5-10 years
10+ years

28. The "smartness" of the home has caused me problems before.
1-10 (strongly disagree-strongly agree)

29. I have a smart home for accessibility reasons (touchscreens instead of buttons, healthcare devices)
1-10 (strongly disagree-strongly agree)

30. I always do thorough research before buying a new device, to make sure that I get the most reliable one.
1-10 (strongly disagree-strongly agree)

31. There are certain brands that I trust, and I only buy devices from them.
1-10 (strongly disagree-strongly agree)

32. I have a voice assistant who is always turned on/on standby (e.g. Alexa, Google Assistant)
1-10 (strongly disagree-strongly agree)

33. I feel more secure in my home than how I felt before installing the smart home system.
1-10 (strongly disagree-strongly agree)

34. My smart appliances require more servicing than the non-smart appliances I had before them.
1-10 (strongly disagree-strongly agree)

35. I trust the service of a smart home device more than I trust the service of another person.
1-10 (strongly disagree-strongly agree)

36. I trust smart appliances so much, that I have had some kind of employee's service replaced with a smart technology (either occasional or regular employee, such as gardener, cleaner, repairperson)
1-10 (strongly disagree-strongly agree)

Appendix 2: Interview Questions

- What is the biggest security concern related to smart home devices?
- How can smart home devices be more secure and reliable?
- What can smart home companies do to ensure this safety?
- Is there any kind of regulation which states that smart home devices need to be able to be controlled manually, even if only to turn them off?
- Why do most people buy smart home items, are they mainly used for comfort?
- In what way are smart homes different from regular homes when it comes to energy saving and cost?
- Do new buildings have smart home systems installed in them right away, have you seen that concept in Hungary? (apartment complexes and such)

M E L L É K L E T

A konferenciakötet gerincét a tudományos igényű, lektorált tanulmányok adják, amelyek a rendezvény tematikájához kapcsolódó kutatási eredményeket, elméleti kereteket és módszertani megfontolásokat rögzítik. A jelen melléklet ezzel párhuzamosan olyan, a konferencia szakmai világához szorosan illeszkedő írásokat tartalmaz, amelyek eltérő műfaji és kidolgozottsági logikával készültek. E szövegek egy része nem klasszikus tanulmányként, hanem inkább dokumentáló, reflexív vagy gyakorlati fókuszú közlésként értelmezhető: céljuk nem egy teljes körű kutatási apparátussal alátámasztott érvelés felépítése, hanem egy adott megközelítés, tapasztalat vagy szakmai álláspont rövid, hozzáférhető összefoglalása.

A szerkesztőség a melléklet közreadásával azt kívánja jelezni, hogy a konferencia tudományos értékét nem kizárólag a tanulmány-formátumú közlések adják, hanem az a szakmai párbeszéd is, amelyben a különböző műfajú megszólalások egymást kiegészítik. A mellékletben szereplő írások így a kötet egészéhez kapcsolódó kontextust bővítik: betekintést adnak a témához kötődő alkalmazott szemléletekbe, a gyakorlati következtetésekbe, illetve a konferencia diskurzusának egyes hangsúlyaira. Közlésük célja a szakmai teljesség megőrzése és az eltérő nézőpontok dokumentálása, miközben a kötet tudományos törzsanyagától elkülönülten, mellékletként jelennek meg.

A hitelesség és bizalom helyreállítása a mai társadalomban

Pintér Gábor Attila[27]

Nagyon sok egyetemi, tudományos hátterű erőadást fognak hallani, és örülünk, hogy ennek mi is részesei lehetünk, ezzel is ezeket a gondolatokat népszerűsíthetjük együtt. Amit mindenképp kiemelnék, hogy itt többször felmerült már az előttem szólókban is, hogy a bizalom a fő témánk ma, és én arra próbálok majd útmutatót adni, hogy hogyan lehet helyreállítani, és mit kell tennünk ahhoz, hogy képben maradjunk úgymond, és hogy megfeleljünk ennek az új kihívásnak. Ami már tulajdonképpen nem is annyira új, mert már az ötvenes évek óta velünk van, de mint forradalom, ez most ért el minket technológiában, és nyilván minden ilyennél bizonyos kérdések felmerülnek. Nagyon röviden, a küldetésünk a Mind Mate-ben a digitális transzformáció, a mesterséges intelligencia, a fenntarthatóság, a KKV fejlesztés és ezekkel kapcsolatos témáknak a kibeszélése, különböző fórumoknak a szervezése, rendezvények szervezése és egyeztetése. Minden ilyen hasznos lehet abban, hogy aztán valamilyen konklúzió tudjon keletkezni, és akár a véleményeknek az ütköztetésével, akár a különböző nézőpontoknak az összehangolásával, szinergiájával valami újat sikerüljön kihozni.

Reméljük, hogy ez a mai nap is egy ilyen nap lesz, hogy itt a sok előadó zanzásított tudása végül is, akár itt a helyszínen lévőket, akár az online térben követőket elégedettséggel tölti el, és tudunk ebből közösen is építkezni. Üzletileg nagyon röviden nekem 16 év IBM-es hátterem van, amit kiemelnék az a Solution Design Center of Excellence-nek a vezetése 15 országra vonatkozóan, ahol minden fajta komplex megoldásokat raktunk össze a csapatunkkal. Többek között már akkor is AI alapú megoldások is voltak benne, illetve most a mérnök és IT dimenziókat segítem összekapcsolni a BEKO Engineering-ben, mint üzletfejlesztő.

Nagyon sok információ keletkezik, nagyon sok téma és könyv, különböző fórum foglalkozik a mesterséges intelligenciával, és már-már ez annyira sok és annyira elharapódzó, hogy nagyon nehéz ebben igazán követni, hogy mire is figyeljünk, mik azok a területek, amik igazán fontosak, mik azok a változások, amiket igazán kezelnünk kell, és hol is fogja érinteni az életünket a mesterséges intelligencia. Akár technológiai, akár humán szempontokat fel lehet eleveníteni, akár az életünkre, akár a munkavilágára vonatkozó hatását, vagy a szociális-

[27] Mind Mate Inspiráció Egyesület elnöke és Thinkers360 Thought Leader. 25+ éve van az IT tanácsadás, stratégiatervezés, megoldástervezés, üzletfejlesztés világában, a hazai és nemzetközi tudás összekapcsolásában.

társadalmi szempontokat lehet nézni, tehát nagyon sokféle megközelítésből elemezzük ezt. Szerencsére ehhez vannak inputok, ami egyben ez komplexitást is jelent, és ez növeli azt a fajta bizalmatlanságot, bizonytalanságot, hogy most akkor mit is tekintsünk releváns információnak, és hogyan is lehet ebből valamilyen leszűrt területet, leszűrt konzekvenciát kiszűrni.

A média kommunikáció, a közösségeknek a hatása, a generációk közötti együttműködések azok, amiket mindenképp szeretném kiemelni, hiszen még mi úgy szocializálódtunk 20-30 évvel ezelőtt, hogy azért volt egy egymásra épülés, és tényleg a tanár-diák viszony még egy meghatározó viszony volt az életünkben, abból a szempontból, hogy a tudás, meg a külvilágnak az arculata a szülőkön, meg a tanárokon keresztül jutott el hozzánk. A mai generációnak már annyi inputja és annyi féle csatornája van, amiben tényleg nagyon nehéz ezt biztosítani, hogy hiteles arcokat, hiteles forrásokat kapjanak, és az érződik a társadalomban sok felmérés és visszajelzés szerint, hogy kiégettek a különböző munkavállalók, a fiatalok is egyre inkább a közösségi térben próbálnak kapcsolódni és újratervezni magukat. Tehát nagyon-nagyon sok probléma felmerül, amiatt, hogy felgyorsult a világunk.

De mit lehet tenni? Egyrészt van egy információ dömping. Vannak olyan kimutatások, hogy évente megduplázódik az információ, tehát minden évben annyi információk keletkezik, mint a korábbi években összesen, és ezt már az emberi elme is képtelen befogadni alapvetően, és ennyiféle csatornán, ennyiféle témában szűrés nélkül operálni. Ehhez nyilván segítséget jelenthetnek ilyen mesterséges intelligencia jellegű előszűrések, vagy előfeldolgozó kis egységek, de itt megint előjön a hitelességi kérdés, hogy akkor tudjuk-e, hogy miket alkalmazzunk. Ahogy Andrea is megfogalmazta az a legérdekesebb, hogy mire, milyen eszközt tudunk használni, ami tényleg segítségünkre van, és ami nem méginkább bonyolultabbá teszi az életünket. A másik, hogy most már mindenki influenszer lett, tehát mindenki a közösségi médiában osztja meg a gondolatait, mindenki elmondhatja a véleményét, ami egyrészt megint csak nagyon jó, mert a vélemény szabadságot erősíti. Bár a szólásszabadságnak én érzem a korlátait, mert az etikai határait be kellene tartanunk, illetve a másoknak a megbántását is el kell kerülnünk. Tehát inkább csak maradjunk a vélemény szabadságnál, mert a szólászabadság egy sokkal erősebb fogalom szerintem. De mindenki influenszer, és hogyha 8 milliárdan vagyunk a bolygón, abból biztos, hogy most már 1 milliárdan fönt vagyunk legalább a neten, ha nem többen, lehet, hogy már 4 milliárd közeli ez a szám. És a másik, hogy vannak különféle algoritmusok, amik segítenek építeni a világunkat, de ez egyben rá is szorít minket, hogy egyre inkább virtuális világgá és virtuális térbe tesszük át a kommunikációt. Itt szintén elhangzott egy előző előadásban, hogy 80 százalékát a valódi információnak az érzelmi töltet, meg a személyes tapasztalat tudja adni, és hogyha ezt kivesszük az egészből, akkor nagyon-nagyon elsilányul az

a tartalom, meg az a hozzáadott érték, amit egyébként egy személyes kommunikáció vagy egy bizalmi kommunikáció tud nyújtani. Azokról az együttműködésekről beszélek, ahol a generációk közötti szakadék is megnehezíti a kommunikációt és az iskolai szerepek is megváltoztak, hiszen most már egy olyan világra kell felkészíteni a fiatalokat, ami sokkal szerte ágazóbb, mint a mi korunkban. Nem csak egy szerepre, egy munkakörre, hanem egy folyamatos tanulásra kell tulajdonképpen felkészíteni, egy folyamatos tudásépítésre kell felépíteni a fiatalokat, és ez még nagyobb kihívás, mint korábban.

De hogyan lehet ezt megtenni? Ehhez foglaltam össze pár gondolatot. Az egyik, amit mindenképp javasolnék az eddigi tapasztalatok alapján, hogy bizonyos keretrendszereket meg kell tartanunk. Ez azért is kell, hogy ugyanazt a nyelvet tudjuk beszélni, illetve, hogy tömegesíteni tudjuk az adott témában fennálló, vagy abban a témában információval rendelkező területeket. Például manapság mindenki hallott az ESG területről, tehát az Environmental Social Governance szemléletekről, mint egy ilyen összefogó terület. De ugyanilyen a SDG célok, ugye a fenntarthatósági témákban, az a Sustainability Development Goal-ok, fenntartható fejlődési céljai. Illetve ilyen a CSR mint terület, azaz a Corporate Social Responsibility felelősségvállalás, aminek most már ISO szabványa van. Nagyon sok ilyen módszertani, kiválósági, vagy minőségbiztosítási elem az újfajta keretrendszerekben testesül meg. De a mesterséges intelligenciára is vannak most már új szabályozások, új etikai elvárások, törvények, tehát ezeket is mindenképp figyelemmel kell kísérnünk, és ezekben a keretrendszerekben kell megtalálnunk a megoldásokat, meg az egyéni sikerességünket, az egyéni életfilozófiánkat, legyen az akár a kreatív oldalunk, a kommunikációs oldalunk vagy a hétköznapi kikapcsolódási oldalunk. A másik, amit mindenképp szeretnék kiemelni, az az, hogy muszáj, hogy közösséget építsünk, mert egy-egy embert már meghalad a tudásterület, illetve a komplexitás, sőt, akár a mesterséges intelligencia maga is, ami 2030 és 2050 között már át fogja lépni azt a szintet, ami az a szingularitási pont, amit Imre említett, hogy már okosabb lesz egy átlagos embernél maga a mesterséges intelligencia. Tulajdonképpen ezt csak úgy tudjuk feldolgozni, hogyha összetartunk és közösségben gondolkozunk, ahol ez a Nash-egyensúly érvényesül. John Nash egy híres közgazdász, matematikus volt, akinek van egy olyan elmélete, hogy az egyéni és a közösségi érdekeket egyszerre kell érvényesíteni, ami a legsikeresebb stratégia. Én ugyanezt próbálom népszerűsíteni itt a mostani világunkkal kapcsolatban, hogy szerintem társadalmi szempontból is az a legcélravezetőbb, hogyha a közösségi érdekeket összekapcsoljuk. Most már nem véletlen, hogy mindenki ökoszisztémaépítésről beszél, ahol a különböző felek partnernek tekintik egymást, és már nem csak ügyfél meg partner viszonylatban, hanem a munkavállalók, a különböző érdekelt felek, amik nagyon sokrétűek

lehetnek, ezek mind-mind az ökoszisztémának a részei tudnak lenni. Egyben kell az értékláncot néznünk és ahhoz illeszkednünk, ami megint csak azt mondom, hogy persze nehezebb egy picit, viszont, hogyha ez meg tud történni, akkor viszont ott van egy összetartás és nyer-nyer helyzeteket tudunk kialakítani. Így a közös fejlődés, a folyamatos fejlődés, sokkal biztonságosabb és megbízhatóbb tud lenni. Írtam nemrégiben egy olyan cikket is, hogy (G)ép-testben, (G)ép-lélek, ahol a G zárójában van, asszociálva az ép-tesben, ép-lélek szólásra, mert ma még a mesterséges intelligencia nem rendelkezik érzelmi töltettel. Szimulálni most már egyébként tudja azt, tehát hogyha valaki beszélget mondjuk egy ilyen AI fórumon egy mesterséges intelligencia egységgel, akkor szimulálja az érzelmeket, és teljesen meggyőző tud lenni, de valójában ezek a mi saját érzelmeink, amiket visszatükröz nekünk, vagy amikre rákapcsolódik. Itt fontos lenne az, hogy mi megtartsuk a valós emberek közötti kapcsolódást is, mert elég természetes, hogy tudatosan megtanuljuk használni a technológiát, de a technológia viszont szükséges ahhoz, hogy ezt az ipari forradalmat megfelelően tudjuk kordába tartani.

A másik fontos dolog a folyamatos tanulásnak a biztosítása. Itt van egy olyan rendszer, ami a különböző képességeknek az újragondolását segíti, a veleszületett adottságok, a fejleszthető képességek, megszerezhető készségek, alkalmazható jártasságok, vagy a tudományos és nemzetközi fejlesztői, kutatói mesterség szintekre lehet gondolni. Ezeknek mind külön jellemzői vannak. Amikor a képességfejlesztésről beszélünk, akkor ilyen árnyalatokban érdemes végig gondolni, hogy mit is fejlesztünk, és mire készítjük fel magunkat. Nyilván, amikor egy jártasságot építünk, az egy újrahasznosítható és egy váratlan helyzetben is segítségünkre jövő képesség, de egy adottság vagy egy tehetség, az meg leginkább egy magunkkal hozott terület, és ezeket is lehet bizonyos szempontból lehet készséggé vagy képességgé fejleszteni, de ezek sok gyakorlattal vihetők tovább. Itt fontos, hogy ezt a tanulás-tanulási képességet didaktikának hívjuk és ez a didaktikai tudás összekapcsolt legyen az innovációs képességünkkel. Én egyébként erre még 2014-ben egy műnyilvántartásba vétellel egy ilyen keretrendszert meg tudtam fogalmazni, ahol ezek összekapcsolhatók egy egységes rendszerbe és így a folyamatos tanulás, illetve a folyamatos megújulás az kézzel foghatóvá tud válni. Úgyhogy reméljük, hogy ebben is tudunk együttműködni itt a hallgatósággal, illetve a későbbiekben másokkal. A másik fontos dolog, hogy hitelesnek kell lenni a kommunikációban. Én azt vallom, hogy a hitelesség az tulajdonképpen három oldalról látszik biztosítottnak. Az egyik, hogy **logosz** legyen, tehát a logikai összefüggések rendben legyenek és valid legyen az az információ, amiből építkezünk. A másik az **étosz**, mint érzelmi töltet, hogy elhivatottak legyünk abban, amit mondunk, és bízzunk is benne, illetve hittel szolgáljuk az információt. A **pátosz** az meg a történelmi, historikus voltára utal, hogy nyilván az se baj, hogyha a korábbi

információkra alapozva hozunk és alkotunk újat és egy megbízható alapra tesszük a gondolatunkat, illetve egy kicsit a küldetéstudat is benne van, hogy valamit előre akarjunk vinni és megváltoztatni a jó irányába.

Úgyhogy ma egyszerre hoztam nektek ezt a négyfajta képességet, ami a társadalmi képességeknek a két szintje a közösségépítés és a keretrendszerek építése, illetve a tanulási képességekből, a rendszerezés, illetve az innovációs képességeknek a fejlesztése. Mi ezekben tudunk segíteni és partnerek lenni alapvetően mentorainkkal. A jövőtudatos gondolkodás, a mi megfogalmazásunkban, még különböző aldefiníciókat is jelent, akár az egyéni képességekre, vagy a sejt szinten történő összekapcsolódásokra gondolva. Ez az az együttműködés a jövőtudatos társadalom iránti kezdeményezésünkre, amit 2023 augusztusában, Magvető napon indítottunk el és szeretnénk egyszerre a környezetünkre, az életközösségünkre, a társadalmi nevelésre, illetve a sikerességre és a társadalmat együtt építő közös eredményekre alapozni. Úgyhogy, a mi szempontunkból a hitelesség az ezekből építkezik, és önöknek is, számotokra is azt javaslom, hogy ezt keressük a közeljövő időszakában, mert akkor itt a mesterséges intelligencia okozta változásokra is meg tudunk felelni.

Köszönöm szépen a figyelmet!

Digitális állampolgárrá válni nem kell félnetek jó lesz

Alföldi István[28]

Először is - én is egy kézfeltartás kéréssel kezdeném: Kinek van digitális állampolgársága, tegye fel a kezét! Jó hír ez a sok feltartott kéz. Ez egy projekt a Digitális Magyarország Ügynökség (DMÜ) vezetésével és számos majd itt elhangzó erőforrás bevonásával, amelynek az előkészítése 2023-ban kezdődött. Mostanra körülbelül kétmillió digitális állampolgár van, de ezek közül valóban használója ennek a lehetőségnek néhány százezer. Az elképzelés úgy szól, hogy 2026 végére teljeskörűen kiszélesedik a felhasználási lehetőség, a kérdés csak az, hogy vajon mennyien és milyen módon fogják tudni használni, illetve, hogy el tudjuk-e dönteni akár itt, akár ennél sokkal szélesebb körben, hogy hova tegyük a vesszőt. Én tudom, hogy hova kell tenni a vesszőt, a mostani előadás is azt próbálja megerősíteni, hogy ebben közös álláspontunk legyen, és minél szélesebb körben terjedjen.

Jövő héten az Infoparlamentnek is kiemelt témája lesz a **digitális állampolgárság**. Regisztráltam, ott leszek, és az eredményt érdemes lesz majd figyelni. Miután itt a konferencia címe a Mesterséges Intelligenciával kapcsolatos, úgy gondoltam, hogy tényleg arról is kéne néhány szót mondani, de az előző előadásokban már annyi fontos gondolat elhangzott ezzel a témával kapcsolatban, hogy én csak egyet tudok hozzátenni. Megkértem a Chat GPT-t, hogy foglalja össze, hogy szerinte mire jó a Mesterséges Intelligencia, és kivételesen éppen 12 pontban foglalta össze. Megnyugtatásul mondom, nem fogom felolvasni a 12 pontot. Van néhány nagyon fontos pont, de azt gondoltam, hogy miután az is elhangzott, hogy meglesznek osztva a prezentációk, szerintem érdekes és értelmes végig nézni. Amit én mindenképpen kiemelendőnek tartok, hogy az MI az emberiség szolgálatába álljon-ez az első pont-, a hármas

[28] Alföldi István, a Mind Mate Inspiráció Egyesület alelnöke.

MMI Egyesület, villamosmérnök, mérnök-tanár, mérnök-közgazdász

Nagy, országos projektek és témák vezetője. Mottó: „MI-RE KÉSZÜLSZ?!”

pont azt mondja, hogy mindenki számára legyen hozzáférhető a Mesterséges Intelligencia, és kiemelem a 12. pontot, amely azt hangsúlyozza, hogy a Mesterséges Intelligencia nem cél, hanem eszköz. (A többit mindenki olvassa majd el, és szándékosan nem AI-t mondtam. Mert az AI, hát az az én monogramom. :-))

A címadásról: a Gertrudis királynővel kapcsolatban elhangzottak késztettek erre, (és ez nem az első eset, mert már egyszer tartottam előadást a Mesterséges Intelligenciát megismernetek, nem kell félnetek, jó lesz címmel, úgyhogy saját magamat másoltam). Azt gondolom, hogy a kérdést azért kell feltenni, és azért kell jó helyre tenni a vesszőt, mert ez egy olyan projekt, amivel kapcsolatban még el sem kezdődött, de már olyan mértékű a füstölgés körülötte a majd lehallgatnak, majd megismernek, vagy az összes adatomat megismerik, meg minden magánéletem kútba esik, …. Csak azt felejtették el ezek és akik még ezt most is mondják, hogy gyakorlatilag tíz évvel, húsz évvel, harminc évvel ezelőtt is mindent tudtak rólunk, tehát ilyen értelemben ennek a dolognak nincs sok jelentősége. Különösen akkor nem, hogyha világossá válik, hogy a DÁP-pal mi mindent lehet könnyebbé, egyszerűbbé, magától érthetőbbé tenni az életünkben. Tehát ugye itt elhangzott, hogy vannak okosotthonok, elhangzott, hogy mi minden előnye lehet a mesterséges intelligenciának. Hogyha belegondolok abba, hogy egy kormányablaknál mennyi ideig kellett sorba állni , és milyen bonyolult, speciális intelligenciát igénylő megoldásokra volt szükség ahhoz, hogy ezeket az ügyeket az ember el tudja intézni... Vagy, ha csak arra gondolok, hogy a napi sajtó mennyire felkapta az ügyfélkapu plusz bevezetését, ezzel szemben, ha digitális állampolgár vagyok, akkor egy QR-kód beolvasással pillanatok alatt mindent el tudok érni, pillanatok alatt bejutok az ügyfélkapumba. Na és akkor most valaki fölteszi a kérdést, jó, de hát hány embernek nincs ügyfélkapuja, mert semmi szüksége nincsen rá a kilenc és félmillió magyar emberből. De - még ha ritkán is - ügyeket nekik is kell intézniük.

Nagyon nagy érdeklődéssel hallgattam az oktatással kapcsolatos nagyon jó előadásokat. Itt nem hagyományos oktatásról, hanem evangelizációról, edukációról van szó, itt arról van szó, hogy milliókat kell meggyőzni és a téma barátjává, együttműködőjévé tenni olyan formában, hogy ez számukra is érthető és világos legyen. De ennyi embert nyilván nem iskolában - persze ott is lehet -nem Tilos, de nem iskolai képzéssel, és nem tanfolyamokkal kell a DÁP használatára felkészíteni, és a vessző megfelelő helyére tevésének elkötelezettjévé tenni. Ugye hát játék a betűkkel, de lehet látni, hogy a DÁP-ot még meg is fordíthattam: Pakolj Át Digitálisra! Szóval a DÁP elsősorban arról szól, hogy mi mindent tud, miért tudja, hogy tudja. Az előbb már említettem, tehát nem akarnám felsorolni azokat a remélt előnyöket, amelyeket 2026 végével mindenképpen tudnia kell, amelyeket a projektet vezető ügynökség - DMÜ -is meghirdetett, és

azt gondolom, hogy talán még az is egy fontos szempont, illetve nem talán, hanem biztos, hogy az Európai Unióban is működik a Digitális Tárca (Digital Wallet) projekt. A digitális állampolgárság tömegesítéséhez nagyon sokaknak a bevonására van szükség. Itt csak megemlíteném, hogy például a Neumann János Számítógép-tudományi Társaság, amelyet 23 éven keresztül vezettem indította el Magyarországon a digitális írástudás tömegesítését. Ez alatt az időszak alatt több, mint 600 ezer emberhez sikerült az ECDL-t - ma már ICDL - eljuttatni, ami elég jelentős dolog volt, és ma is tart. Tehát a lényeg az, hogy itt -nem erről van szó, de arról igen, hogy mindenhol el kell érni az embereket, és ebbe nagyon sokaknak együtt kell működni az írástudók közül, hogy úgy tudjuk elérni az embereket, hogy valóban kedvük legyen hozzá és örömet találjanak abban, hogy digitális állampolgárra válhatnak.

Reméljük, hogy 2026-ra ez tényleg megvalósul. És ha ismét a megfordítom a DÁP betűk sorrendjét, még egy fontos értelmezésre nyílik mód: Polgárok Átállási Döntése. Tehát a polgárok átállási döntését befolyásolni, a pozitív irányban befolyásolni, az írástudók dolga. Azt nem mondhatom, hogy kötelessége, mert ebben a formában ez túlzás lenne, de azt igen, hogy Pakolj Át Digitálisan. Azt persze el kell mondani, hogy túl régi telefonokkal nem lehet dápolni, de lehet olcsón cserélni telefont, és ez, számos egyéb előnye mellett a dáp-használatot is lehetővé teszi.

A vesszőnek tehát itt a helye: **Nem kell félnetek, jó lesz.** Erről kell mindenkit meggyőzni, ez az edukáció azonban nagyon nehéz, és egyáltalán nem magától értetődő, ezt világosan lehet látni abból, hogy gyakorlatilag a legkisebb falutól Budapestig mindenhol szükség van erre. A fenntartásokkal kapcsolatban - pl., hogy nagyon régi telefonokkal nem lehet dápolni - Eszterházy Miklós nádor az 1600-as években azt mondta, hogy „őrültség, semmit sem tennünk, ha mindent nem tehetünk is”. Ez ma is érvényes. Sőt, én egy kicsit meg is erősítem. Nagyon fontos sok mindent tennünk. Vannak jó válaszok. Rajtunk nem múlik. Ebben remélem az együttműködésünket! Köszönöm szépen a figyelmet