

Észrevesszük a mesterséges intelligencia jelenlétét a mindennapjainkban?

– Empirikus kutatás az MI által generált tartalom felismerésére

Összefoglalás: A mesterséges intelligencia (MI) térnyerése a mindennapokban számos területen megfigyelhető, beleértve az online tartalmak generálását, az ügyfélszolgálati interakciókat és a digitális médiafogyasztást. Az írás azt vizsgálja, hogy az emberek milyen mértékben képesek felismerni az MI által generált tartalmakat a hétköznapi élet különböző szinterein. Az empirikus kutatás során kérdőív segítségével kísérletet tettünk annak felderítésére, hogy a résztvevők észreveszik-e a mesterséges intelligencia által létrehozott szövegek jellemzőit. Az eredmények azt mutatják, hogy az emberek időnként nehezen tudják megkülönböztetni az MI-tartalmakat az emberi alkotásoktól, bár bizonyos mintázatok, nyelvi sajátosságok felismerése alapján néhanyan képesek többé-kevésbé pontos azonosításra. A kutatás rávilágít az MI jelenlétének észlelésével kapcsolatos kihívásokra és annak lehetséges társadalmi, etikai vonatkozásaira.

Kulcsszavak: MI-generált szöveg, generációk közötti különbség, szövegfelismerés.

Abstract: The rise of artificial intelligence (AI) in everyday life can be observed in many areas, including online content generation, customer service interactions and digital media consumption. This presentation will explore the extent to which people are able to recognise AI-generated content in different settings of everyday life. In the empirical research, a questionnaire was used to test whether participants perceive the characteristics of AI-generated texts. The results show that people sometimes have difficulty distinguishing AI content from human-generated content, although some are able to identify it more or less accurately by recognising certain patterns and linguistic features. The research highlights the challenges of detecting the presence of AI and its possible social and ethical implications.

Keywords: AI generated text, generational difference, text recognition

* *Dunaújvárosi Egyetem, Informatikai Intézet, Szoftverfejlesztési és Alkalmazási Tanszék, Eszterházy Károly Katolikus Egyetem Neveléstudományi Doktori Iskola*

E-mail: mkollar@uniduna.hu

[1] Sarzaeim, P.–Doshi, A. M.–Mahmoud, A. M. (2023): *A Framework for Detecting AI-Generated Text in Research Publications*. Istanbul: 11th International Conference on Advanced Technologies (ICAT'23), pp. 121–127.

Bevezetés

A mesterséges intelligencia (MI) az életünk minden területén egyre hangsúlyosabb szerepet tölt be. Az MI által generált tartalmak, legyenek azok szövegek, képek vagy egyéb média-tartalmak, ma már a közösségi média posztjaitól kezdve a híroldalakon megjelenő cikkekig számos helyen található meg. Az MI-technológiák fejlesztése lehetővé tette, hogy a tartalmak sokszor emberi kéz által alkotottnak tűnjenek, amely új kérdéseket vet fel a felismerhetőség és az etikai vonatkozások kapcsán. Ez a tanulmány azzal foglalkozik, hogy mennyire képesek az emberek felismerni az MI által készített tartalmakat és milyen tényezők segítik, illetve nehezítik ezt a felismerést.

A mesterséges intelligencia alkalmazása a tartalomkészítésben egy új korszakot nyitott meg, ahol a számítógépes algoritmusok segítségével létrehozott szövegek, képek és videók egyre kifinomultabbak és nehezebben megkülönböztethetők az emberi alkotásoktól. Az MI általi tartalomgenerálás történeti áttekintése mellett érdemes megvizsgálni azokat a kutatásokat, amelyek a tartalom felismerhetőségével foglalkoznak. Ide tartoznak például az olyan tanulmányok, amelyek a GPT-szerű nyelvi modellek, a deep-fake-technológia és egyéb MI-eszközök alkalmazásának hatását vizsgálják a tartalom eredetiségének érzékelésére.

Az ezzel foglalkozó kutatások például az alábbi kérdésekre keresik a válaszokat:

- Mennyire képesek az emberek felismerni az MI által generált tartalmakat különböző típusú médiumokban (szöveg, kép, videó)?
- Milyen tényezők befolyásolják a felismerés képességét? (pl. nyelvi sajátosságok, vizuális mintázatok)?
- Vannak-e olyan demográfiai vagy egyéb jellemzők, amelyek hatással vannak az MI-tartalmak felismerésére?
- Mennyire tartják az emberek fontosnak, hogy felismerjék az MI által generált tartalmakat, és milyen etikai kérdéseket vet fel a tartalmak MI-jellegének felismerhetősége?

Elméleti háttér, témához kapcsolódó információk

A mesterséges intelligencia (MI) által generált tartalmak felismerhetőségének problémája kiemelt figyelmet kapott az MI-technológiák rohamos fejlődésével. Az ilyen tartalmak, mint például szövegek, képek és videók, jelentős társadalmi és tudományos kihívásokat vetnek fel, különösen azok hitelességének és eredetiségének meghatározásában. [1]

Az MI által generált szövegek felismerésére különféle gépi tanulási modelleket használnak, amelyek a nyelvi és stilisztikai jellemzők elemzésére építenek. Ezen modellek, például a BERT- és CNN-alapú megközelítések, a szöveg szintaktikai és szemantikai mintázatait elemzik, hogy az MI-eredetet meghatározzák. [2]

Sajnos azonban az újabb detektálási technikák, például a vízjelzés és az outlier detection, gyakran alulmaradnak a szövegek újrafogalmazása vagy módosítása esetén. [3]

Az MI-tartalmak felismerhetőségével kapcsolatos kihívások nem csupán technológiai, hanem etikai kérdéseket is érintenek. A tartalmak torzítása, félretájékoztatás és a szerzői jogok megsértése mind olyan problémák, amelyek az MI által generált tartalmak széles körű használatából erednek. [4]

A különböző demográfiai csoportok eltérően reagálnak az MI által generált tartalmakra. Egyes kutatások azt sugallják, hogy az idősebb generációk kevésbé érzékenyek az MI-tartalmak stilisztikai és nyelvtani sajátosságaira, míg a fiatalabb generációk nagyobb magabizottsággal észlelik ezeket az eltéréseket valószínűleg a digitális technológiákkal való nagyobb tapasztalatuk miatt. Azonban a tapasztalati tudás és a kontextuális feldolgozás néhány esetben előnyt jelenthet az idősebb generációk számára, különösen, ha a szöveg specifikus ismereteket igényel. [5] [6] [7]

A kutatáshoz kapcsolódó célok és módszertan áttekintése

Célkitűzés:

Kutatásom célja annak vizsgálata, hogy a különböző generációk milyen mértékben ismerik fel az emberi és mesterséges intelligencia által készített tartalmakat. A kísérlettel az alábbi kérdésekre kerestem a választ:

- Milyen különbség mutatkozik az MI és az ember által létrehozott szöveges állomány felismerési képességében generációként?
- Mennyire magabiztosak a kísérlet során adott válaszokban a csoportok?
- Milyen okok miatt döntöttek a válaszadók a választásuk mellett?

[2] Vora et. al. (2023): „A Multimodal Approach for Detecting AI Generated Content using BERT and CNN”. *International Journal on Recent and Innovation Trends in Computing and Communication*, 1. kötet 11., (1.), pp. 691–701.

[3] Krishna et. al. (2023): „Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense,” 37th Conference on Neural Information Processing Systems (NeurIPS 2023), pp. 27469–27500.

[4] Kraus, S.–Kanbach, D. K.–Krysta, P. M.–Steinhoff, M. M.–Tomini, N. (2022): Facebook and the creation of the metaverse: radical business model innovation or incremental transformation? *International Journal of Entrepreneurial Behavior & Research*, 28., (9.), pp. 52–77.

[5] Chakraborty et al. (2024): „Position: On the Possibilities of AI-Generated Text Detection,” 41st International Conference on Machine Learning.

[6] Alamleh, H.–AlQahtani, A. A.–ElSaid, A. (2023): „Distinguishing Human-Written and ChatGPT-Generated Text Using Machine Learning,” *Systems and Information Engineering Design Symposium (SIEDS)*. Charlottesville, pp. 154–158.

[7] Alam et al. (2020): „AI-Fairness Towards Activity Recognition of Older Adults,” *MobiQuitous 2020 – 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*.

Alkalmazott kutatási módszer:

A kísérletet 2024. májusa és októbere között végeztem. 1 db ember által és 4 db ChatGPT segítségével létrehozott általános témájú hirdetésről kellett a válaszadóknak eldönteni, hogy ki készítette, majd megadni, mennyire biztosak a döntésükben, valamint megjelölni a döntés okát. A válaszokat Google űrlap segítségével, önkitöltős módszerrel gyűjtöttem be.

Bár a kérdőív kérdései ezen kívül még egyéb területet is érintettek, de kutatásom jelen fázisában csak a két hirdetéssel kapcsolatos eredmény elemzésére térek ki. Az első esetben egy MI-által létrehozott szöveget olvashattak, a második esetben pedig ember által írt szöveget láttak.

Eredmények értékelése, következtetések

A kiválasztott kérdések az alapstatisztikai és a mesterséges intelligenciával kapcsolatos kérdéscsoportból kerültek ki.

ALAPSTATISZTIKAI KÉRDÉSCSOPORT

A vizsgálatba összesen 171 főt vontam be, akiket 4 csoportra (generációra) osztottam fel:

- Baby boomerek (1946–1964)
- X-generáció (1965–1979)
- Y-generáció (1980–1994)
- Z-generáció (1995–2010)

Mind a 4 generációból a válaszadás szempontjából első 15–15 fő került kiválasztásra és vizsgálatra, hogy a reprezentativitás ne okozzon problémát (így kiszűrhetők a különböző csoportok digitális írástudásban megjelenő különbségei). Tehát az összesen 60 fő – ebből 15 fő férfi (25%) és 45 fő nő (75%) – válaszainak elemzése történt meg.

A kiválasztottak nagyjából fele (55%) egyetemi/főiskolai, 42% középfokú/technikus/felsőfokú OKJ-s, és csak 3% alacsonyfokú végzettségű.

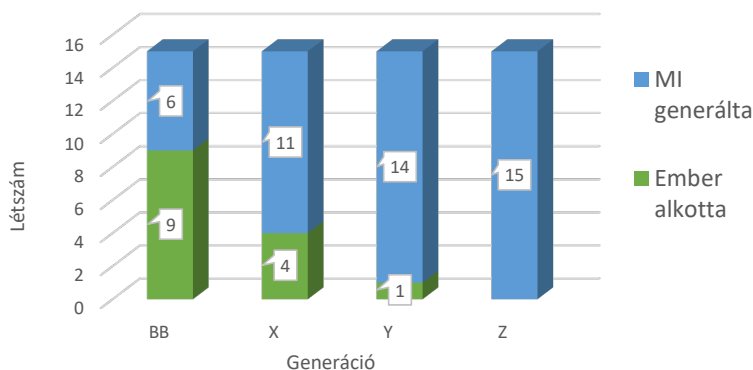
MESTERSÉGES INTELLIGENCIÁVAL KAPCSOLATOS KÉRDÉSCSOPORT

Ebben a kérdéscsoportban arra voltam kíváncsi, hogy a mintahirdetésekről a válaszadók felismerik-e, hogy ember, vagy az MI készítette-e, válaszaikban mennyire biztosak a válaszadók és mi volt a döntés oka.

Bár összességében 5 mintahirdetést kaptak, de jelenlegi elemzésemben csak az 1. és a 2. hirdetést vontam be (az elsőt a mesterséges intelligencia, a másodikat a szerző írta).

A válaszadók nem tudták, hogy a kapott mintafeladatok között mennyi MI-által generált került elhelyezésre, így ez sem befolyásolhatta a döntésüket.

1. ábra. Az első mintahirdetésre adott válaszok az ember vs. MI kérdésre

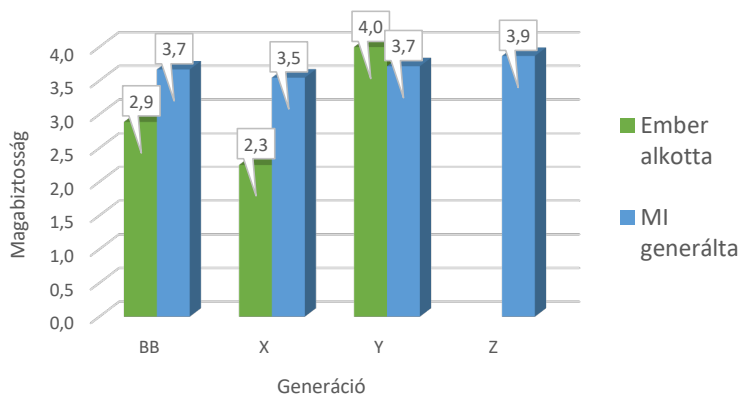


Forrás: Saját ábra.

Az 1. mintahirdetés esetében igen jól látható, ahogy haladunk visszafelé életkorban, úgy egyre inkább felismerték, hogy nem ember által fogalmazott szöveggel találkoztak (ez jól látható az 1. ábrán). Ebben a tanulmányban nem kerül ismertetésre a többi feladat, de azoknál nem teljes mértékben igaz ez a megállapítás, azonban általánosságban többnyire kijelenthető, hogy szignifikáns a kapcsolat az életkor és a MI által generált tartalom felismerése között.

A 2. ábrán látható, hogy az ötfokozatú Likert-skálán általánosságban kb. közepesnél egy kicsit jobban voltak magabiztosak a válaszukban a válaszadók. Azok, akik a helyes választ adták, inkább érezték biztosnak azt, mint akik helytelenül válaszoltak. Ennél a feladtnál szembe tűnő, hogy a hibásan választók közül az Y-generáció tagjai a többi korcsoporttól eltérően egy kicsit jobban bíztak a válaszuk helyességében (még a jól választóknál is nagyobb mértékben).

2. ábra. Az első mintahirdetésre milyen magabiztossággal válaszoltak a különböző generációk

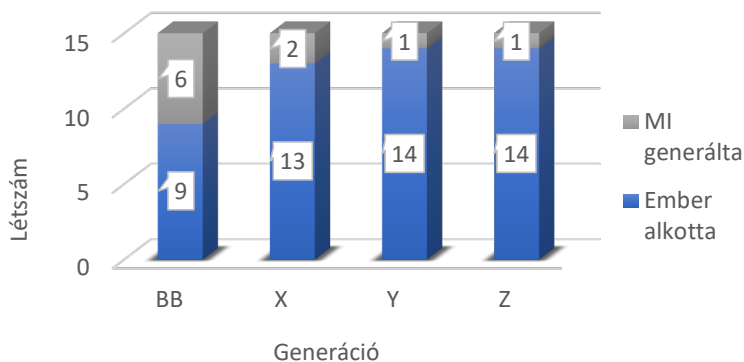


Forrás: Saját ábra.

Ahogy az első mintahirdetésre, úgy a másodikra is többségében eltalálták a választ a megkérdezettek. Ebben az esetben is a fiatalabb generációk inkább adtak helyes válaszokat, mint az idősebbek (ez a különbség jól látható a 3. ábrán is).

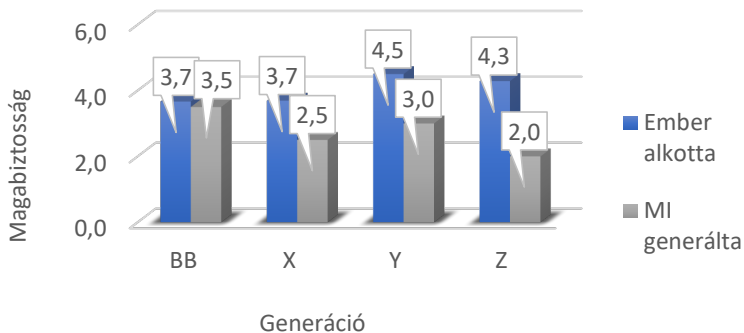
A 4. ábrán is látható, hogy a második mintahirdetés esetében is hasonló a helyzet az elsőhöz: akik a rossz választ adták, jóval kevésbé voltak magabiztosak, mint a másik válasz esetében. Itt is a legbiztosabbak az Y-generáció tagjai voltak, a 4,5-ös átlagos érték is ezt mutatja. Az 5 mintakérdés esetében ennél a feladatnál jött ki legmagasabbra a magabiztosság, ebben az esetben átlagosan 3,9-re értékelték a válaszadók a magabiztosságukat az ötfokozatú skálán, a többi hirdetésnél 3,4–3,6 közötti értékeket kaptunk.

3. ábra. A második mintahirdetésre adott válaszok az ember vs. MI kérdésre



Forrás: Saját ábra.

4. ábra. A második mintahirdetésre milyen magabiztossággal válaszoltak a különböző generációk



Forrás: Saját ábra.

Mindkét esetben a döntés indokaként a szokásos okok (nyelvezet, stílus, helyesírás, természetesség) mellett megadhattak egyéb, szabad szöveges válaszokat is.

Itt megmutatkozott, hogy mennyire különbözően gondolkodtak a válaszadók, milyen nyelvi elemeket kerestek a szövegekben. Például az alábbi válaszok érkeztek:

- „Ember fia nem tudja, mi az a pasztell.”
- „Generálhatta ember is, de a tagolása, a megfogalmazása mindenképpen gyanús.”
- „Pamutból farmer?”
- „Az árképzés is gyanús, emberek ritkán használják ezt a módszert sima hirdetésnél.”
- „Ez egy intelligensebb ember hirdetése. Ő tagol, de nem úgy, mint a fentiek. Az 500Ft-ra kerekítés is emberre vall.” (– ebben az esetben rossz volt a válasz és az ötfokozatú skálán négyes magabiztosságú volt a válaszadó).

Összefoglalás

A tanulmányom központi kérdése az volt, hogy az emberek képesek-e felismerni a mesterséges intelligencia (MI) által generált tartalmakat a mindennapokban. Az empirikus kutatás kérdőíves módszerrel vizsgálta, hogy a különböző generációk hogyan teljesítenek az emberi és MI által készített szövegek megkülönböztetésében.

A kutatásba bevont résztvevők 4 generációba kerültek besorolásra és öt mintahirdetés képezte a kísérlet alapját, amely ember, illetve mesterséges intelligencia által létrehozott szöveges állomány volt. A válaszok elemzése során figyelembe vettem a magabiztossági szinteket és a döntések mögötti indokokat is.

Az eredmények elemzése során az alábbi következtetésekre jutottam:

- Az idősebb generációk (Baby Boomerek és X-generáció) kevésbé pontosan ismerték fel az MI-által generált szövegeket, míg a fiatalabb generációk (Y és Z) jobb teljesítményt nyújtottak, vagyis az MI-által generált tartalmak felismerése generációs különbségeket mutatott.
- Az Y-generáció tagjai a legnagyobb magabiztossággal válaszoltak, még akkor is, ha a válaszuk helytelen volt.
- A válaszadók döntéseiben a nyelvi stílus, helyesírás és tagolás játszották a legnagyobb szerepet.

Kijelenthető, hogy az MI által generált szövegek egyre nehezebben megkülönböztethetők az emberi tartalmaktól, ami fontos társadalmi és etikai kérdéseket vet fel.