

<https://doi.org/10.17048/AM.2023.251>

Dr. T. Nagy László, Dr. Boda István Károly, Dr. Tóth Erzsébet: A mesterséges intelligencia multimédiás lehetőségei, médiakonverziók

Dr. T. Nagy László

Debreceni Református Hittudományi Egyetem (DRHE), Matematika és Informatika Tanszék

t.nagy.laszlo@drhe.hu

Dr. Boda István Károly

Debreceni Református Hittudományi Egyetem (DRHE), Matematika és Informatika Tanszék

boda.istvan@drhe.hu

Dr. Tóth Erzsébet

Debreceni Egyetem, Informatikai Kar, Adattudomány és Vizualizáció Tanszék

toth.erzsebet@inf.unideb.hu

Absztrakt: A jelenleg elérhető mesterséges intelligencia entitások számtalan formában vannak jelen az interneten, ezen megnyilvánulások kimenetén a multimédia minden elemét megfigyelhetjük. A multimédia és a különböző, egyre változatosabb formában megjelenő digitális média elemek tanulásra-oktatásra gyakorolt pozitív hatásait már évtizedes tapasztalatok és kutatások támasztják alá. Az MI és az általa generált médiatartalmak széles körű megjelenése és elérhetősége azonban olyan új lehetőségeket teremt vagy nyit meg számunkra, amelyek használhatósága és alkalmazása további vizsgálatokat igényel. Az MI által generált verbális és multimédiális tartalmak egészen más szerkesztési elvek és bemeneti adatok alapján jönnek létre, mint azok a tartalmak, amelyeket mi emberek alkotunk. Az MI által létrehozott kimenetek – mindezek ellenére – egyre kevésbé különböztethetők meg az ember alkotta tartalmaktól. Az előadás tárgyát adó kutatás olyan konkrét mesterséges intelligenciával támogatott médiageneráló eszközöket és azok kimeneteit vizsgálja empirikus módszerrel és az eredmények heurisztikus kiértékelésével, amelyek valamilyen adatkonverziót eredményeznek. Jelen tanulmányban célunk a leggyakoribb – főleg ingyenes – MI alkalmazások összegzése, valamint be és kimeneteik vizsgálata média/adat konverziós szempontból, néhány gyakorlati példán keresztül is szemléltetve.

Kulcsszavak: mesterséges intelligencia, multimédia, médiakonverzió, tanulástámogatás, digitális oktatás

THE MULTIMEDIA POSSIBILITIES OF ARTIFICIAL INTELLIGENCE AND MEDIA CONVERSION

Abstract: The currently available artificial intelligence entities are present on the internet in countless forms and the output of these manifestations can be observed in all elements of multimedia. The positive effects of multimedia and the increasingly diverse digital media elements on learning and teaching have been supported by decades of experience and research. However, the widespread experience and accessibility of AI and the media content it generates creates or opens up new possibilities for us, the usability and application of which require further investigation. The verbal and multimedia content generated by AI is based on very different editing principles and inputs from the contents we humans create. In spite of this, AI generated output is increasingly difficult to distinguish from human-generated content. The research that is the subject of this talk investigates specific AI-supported media generation tools and their outputs using empirical methods and heuristic evaluation of the results that conclude in some kind of data conversion. In this paper, we aim to summarize the most common, free-of-charge AI applications and to investigate their inputs and outputs from a media/data conversion perspective, illustrated by some practical examples.

Keywords: artificial intelligence, multimedia, media conversion, learning support, digital education

1. Bevezetés

A multimédia eszközeinek, mint lehetőségeknek a tanulás vagy oktatás támogatásban való alkalmazhatóságát régóta és sokan vizsgálták, többek között mi is foglalkoztunk a témával. Később, a közösségi hálózatok megjelenésével és széleskörű elterjedésével, számos új eszközt kaptunk, amelyek újdonsága és fő ereje a közösségben rejlő tudáspotenciál, ami természetesen szervesen a multimédia eszközeire és az internet hálózatára épül. (T. Nagy–Boda 2014) A közösségi médiák után, a mesterséges intelligencia színrelépésével, a fókusz ma a szintetizált tartalmakra irányul. Habár az egyik legismertebb alkalmazás a ChatGPT ingyenes verziója – mint nyelvi modell – alapvetően szöveges kommunikációra koncentrál, léteznek egyéb olyan MI-n alapuló megoldások, amelyek multimédiások vagy a karaktereken alapuló kommunikáció mellett képesek hang, kép, videó stb. alkotására, konverálására vagy ezek tartalmának megváltoztatására.

A mesterséges intelligencia – kutatás eredményei az elmúlt évtizedekben, de különösen az elmúlt 5-10 évben szinte észrevétlenül jelentek meg a mindennapjainkban. Az is megfigyelhető, hogy a legtöbb ember sok esetben úgy kezdett mesterséges intelligencián (MI) alapuló terméket vagy szolgáltatást használni, hogy igazából nem is sejtette vagy el sem gondolkodott rajta, hogy a háttérben az eredményeket részben vagy egészben mesterséges értelem alkotja. A ChatGPT 2022 őszi megjelenésével azonban sok minden megváltozott e téren. A nyelvi modell ingyenes elérhetővé tétele, majd az azt követő erős médiavisszhang, olyan rétegek látókörébe is eljuttatta a mesterséges intelligenciát és annak vívmányait, akik eddig kevés tudomást szereztek róla vagy egyáltalán nem érdeklődtek iránta.

2. A kutatás

A mesterséges intelligencia használatának bizonyos célból, olyan esetekben van igazán gyakorlati értelme, amikor nem tudunk (vagy nem is akarunk) az összes elképzelhető releváns be és kimenetre felkészülni (pl. azok várható nagy száma miatt), vagy olyan feladatot akarunk rábízni az algoritmusra, amelyek hagyományos programozói megoldása is igen nagy emberi/gépi/stb. erőfeszítéseket igényelne. A természetes nyelvek feldolgozása (pl. a szövegalkotás) vagy a képfeldolgozás ilyen feladat. A közösségi hálózatokban megjelent (közösségi) tudást ez esetben a mesterséges értelem által generált kimenet helyettesíti. Szándékosan a kimenet szót használva az előbb a tudás szó helyett, hiszen a MI – legalábbis mai szintjén – nem érti a kérdést, ahogy a választ sem, legalábbis abban az értelemben biztosan nem, ahogy mi emberek tesszük.

Kutatásunkban az tűztük ki célunk, hogy empirikus módszerrel, a válaszok heurisztikus kiértékelésével, megvizsgáljuk az interneten elérhető (főleg ingyenes) mesterséges intelligenciát használó eszközök és megoldások milyen multimédiás és médiakonverziós képességekkel rendelkeznek. Mindemelllett áttekintjük az előbbi képességeknek a gyakorlati alkalmazásait – amelyek külön vizsgálat tárgyát képezve –, várhatóan eredményesen tudnak támogatni bizonyos oktatási, tanulási folyamatokat vagy részfeladatokat. A vizsgálat módszerét az is indokolja, hogy a jelenleg hozzáférhető MI alkalmazások kutatáselméleti háttér szempontjából olyan fekete dobozoknak tekinthetők, amelyek működéséről legjobb esetben is csak korlátozott információk érhetők el, és amelyekre alkalmazható a Turing teszt módszertana. (Turing teszt 2023a, 2023b)

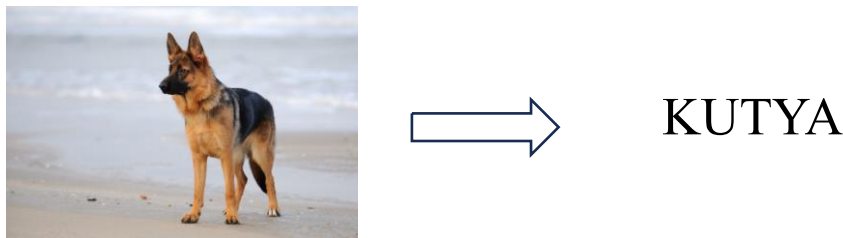
3. Médiakonverziók

A mesterséges intelligencia mint entitás, működési alapelvét tekintve deep learning-et alkalmazó neurális hálózat, ami bemeneti feladat alapján szintetizálja a várt (például szöveges) választ a kimeneten. Fontos megjegyezni, hogy – a jelenlegi megoldásoknál – lényegében nem a feladat (bemenet) értelmezése történik, hanem a rendszer a kérdés szerkezetére alapján próbálja megjósolni a lehetséges kimeneti választ (predikció) az adatbázisban fellelhető (kapcsolódó) releváns információk alapján. Éppen ezért, a működés jellegéből adódóan előfordulhat, hogy a MI (pl. ChatGPT) olyan választ ad a kérdésre „nagy magabiztossággal”, ami nyilvánvalóan nem helyes. Ilyen esetben nem szándékos a félrevezető eredmény, hanem szakszóval élve az MI „hallucinál”. Mivel nem tudja értelmezni a feltett feladatot/kérdést – tehát valójában nem érti mit kérdeztünk –, nem tudja elbírálni a válasz jóságát sem. A felmerülő lehetséges adatforrások felhasználása és a válaszok (kimenetek) kiválasztása vagy összeállítása, valószínűleg valamiféle rangsorolás alapján történik pl. hivatkozások gyakorisága, találatok népszerűsége vagy értékelése, egyéb relevancia stb. Egy a ChatGPT-hez hasonló nyelvi modell esetében a kimenet ugyanis a kérdésre: az adatbázisban talált információk alapján összeállított lehetséges (valószínű) szórend/mondathalmaz csupán. Mindezeket figyelembe véve (vagy mondhatjuk, hogy mindezek ellenére) az MI kimenetei a legtöbb esetben olyan

válaszok, amelyek azt az érzetet keltik bennünk, mintha a feladatot értelmező értelem (pl. ember) oldotta volna meg. (Laky 2023)

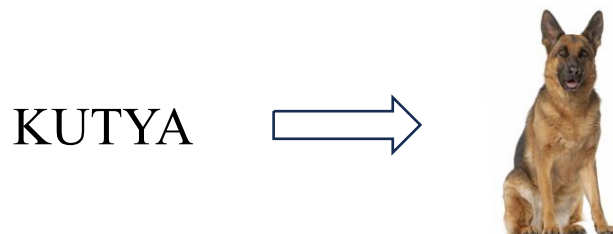
A deep learning neurális hálók fontos képessége a hatékony mintafelismerési vagy mondhatjuk úgy is, összefüggésfelismerési tulajdonsága, ezzel bizonyos területeken (pl. a kép és a beszéd/szövegfeldolgozás) sokkal jobb eredményeket érnek el, mint a korábbi gépi tanulási megoldások. (Tóth 2020) Ebből a mintafelismerési tulajdonságból adódóan pedig képesek az adattranszformációra és az irányított tartalomgenerálásra (adatszintetizációra).

Meg kell azonban jegyezni, hogy igazán élesen nem minden esetben választható szét e két fogalom, inkább azt mondhatjuk valamelyik jobban jellemző. Például tetszőleges képet adva az MI bemenetére egy kutyáról, egy kondicionált, mély neurális háló képes azt a szöveges kimenetet adni: „kutya”. Ez esetben a képpontokból összeálló adathalmazt, mint képi információt („értelmezi”) transzformálja a nyelv „kutya” szavára, amelyet karaktorsor formájában prezentál kimeneti eredményként. Ez esetben inkább adattranszformációnak értelmezzük a műveletet (a kutyát ábrázoló képből (pixelhalmazból), a kutya – betűkből álló – szóképe lett).



1. ábra Transzformációs adatkonverzió

Ha viszont fordítva: azt kérem szöveges formában az MI-től: „Rajzolj kutyát!” inkább tartalomgenerálásnak definiáljuk a konverziós műveletet.



2. ábra adatkonverzió tartalomgenerálással

Azaz, ha csökken a jellemzőtér transzformációnak, ha pedig nő, generálásnak értékeljük a kimenetet. Azt is meg kell jegyezni, hogy konverzióknál általában médiaváltás (pl. képből>szöveg: jpg>txt) is bekövetkezik, de nem minden esetben! A következő fejezetekben azokat a MI-át használó konverziós irányokat ismertetjük, amelyek az interneten is elérhetőek.

Megállapítható, hogy a mesterséges intelligencia lehetséges és alkalmazott felhasználási területeit az adatelemzés/adatfeldolgozás műveletein kívül, a kimeneti oldalon alapvetően két konverziós, működési formára tudjuk bontani, ezek a transzformáció és a tartalomgenerálás.

3.1 Transzformációk

<u>bemenet</u>		<u>kimenet</u>
• karakter	>	karakter (fordítás)
• képből	>	karakter (karakterfelismerés)
• képből	>	szöveg (objektumfelismerés)
• beszédből	>	írott szöveg (beszéd felismerés)
• hangból	>	szöveg, (hangfelismerés)
• zenéből	>	szöveg (zenefelismerés)
• szövegből	>	beszéd (gépi felolvasás)

karakter > karakter (fordítás)

A (természetes) nyelvek közötti fordítás az egyik legismertebb és leggyakrabban használt MI alkalmazási forma. Ez esetben a karakterekből álló forrásnyelvi szöveget kívánjuk átalakítani egy tartalom szempontjából az eredetivel megegyező célnyelvi szöveggé. Ennél az adatkonverzióval médiakonverzió nem történik. Habár fordítóprogramok már régóta léteznek, a nyelvek közötti „jó minőségű” fordítás igen sokáig váratott magára. Mindazok ellenére, hogy az elmúlt években jelentős és ugrásszerű fejlődést tapasztalhattunk, ennek a feladatnak a gépiesítése még mindig kihívásokkal küszködik sok esetben, tartalomtól, nyelvtől, témától stb. függően. A legismertebb alkalmazás a Google fordító, ami online elérhető több mint 120 nyelven. A 31 nyelvet ismerő Deepl a fordítás pontosságában, az árnyalt jelentések értelmezésében igen erős feltörekvő alkalmazás, amely magyarul is ért. (Deepl 2023)

képből > karakter (karakterfelismerés)

A karakterfelismerés funkció több dokumentumkészítő alkalmazásban és a Google Lens azonnali fordítás opciójában is megtalálható. Jelen esetben gondoljunk az optikai karakterfelismerésre (OCR) ahol a releváns (szöveges) információ változatos grafikus (képi) formában jelenik meg a bemeneten, például a képen egy szó vagy szöveg részeként, a kimeneten pedig a betűk felismerésének megfelelő karaktert várjuk. (Google Lens 2023) E médiakonverzió esetében a lefotózott szöveg tartalmú képek – karakteres transzformáció után – akár szerkeszthetők, másolhatók lesznek, vagy másik szöveges alkalmazás bemenetére küldhetők.

Ez a tanulás támogatásához is számos módon segítséget nyújtó médiakonverzió, az MI alkalmazásának egy igen fontos területe. A képen megjelenő szöveg helyes felismerése igen sok dologtól függ: pl. kontraszt, megvilágítás, betekintési szög, méret, karakterkészlet, stb. Az MI deep learning mintafelismerési képessége által ugrásszerűen megnő a pontos felismerési arány akár kézírások esetében is. Tehát nem szükséges (és nem is lehetséges) minden várható szöveg képi megjelenésére felkészülni, a MI megfelelő kondicionálás után kitalálja, azaz „értelmezi” a szöveget (karaktereket).

Azon túl, hogy szerkeszthető szöveget kapunk a képi szöveg helyett, az OCR alkalmazása után, (ami már önmagában is előnyös) további adatkonverzióval újabb funkciókat nyerhetünk. A Google Lens alkalmazása az azonnali fordítás használatához a virtuális valóságot is segítségül hívja.

A Google Lens esetében a telefon kamerája „látja” a megmutatott idegen nyelvű szöveget képi (ez esetben élő videó) formájában. Az MI-vel támogatott OCR karakteres formába konvertálja a szöveget (képből>szöveg) ezután a Google fordítóprogram lefordítja a kívánt nyelvre. Ezen adatkonverzióval médiakonverzió nem történik: karakter>karakter. Majd a kívánt nyelvű szöveget képi (élő videó) formában megjeleníti az eredeti szöveg helyén. Tehát ez esetben: (eredeti)képből>(eredeti nyelvű)karakteres szöveg>(célnyelvi) szöveg>(virtuális)kép, amely kép a lefordított szöveget jeleníti meg, ha az adatáramlás irányát és konverzióit tekintjük.

képből > szöveg (objektumfelismerés)

A nem szöveget tartalmazó vizuális információk (pl. ember, közlekedési táblák stb.) felismerése az önzetű járművek egyik alapvető tulajdonsága. A (médiá)konverzió iránya ez esetben: a képi pixelhalmazból szöveg (karakterhalmaz) lesz a képi tartalomnak megfelelően a kimeneten. A hatékony és pontos objektumfelismerés az MI kutatás egyik fő iránya, ahol katonai, navigációs, közlekedési és biztonsági célú alkalmazásai a legismertebbek.

ESZTERHÁZY KÁROLY KATOLIKUS EGYETEM
INFORMATIKA KAR • DIGITÁLIS TECHNOLÓGIA INTÉZET
AGRIA MÉDIA KONFERENCIA 2023

Ezen túl az eszközök, tárgyak, egyéb dolgok neveinek beazonosításában is segíthet akár rögtön idegennyelvű válaszokkal. Széleskörűen alkalmazható tanulás támogató eszköz, akár kezdő vagy kisgyermek esetében is (pl. tárgykép-szóképfelismerés összekötés, szótanulás stb.). A Google Lens-ben szintén integrált megoldás az objektumfelismerés. (Google Lens 2023)

beszédből > írott szöveg (beszédfelismerés)

Ez esetben az élő beszéd>karakteres (szöveges) verziójának előállítását várjuk az MI-től. Technikailag azért nehéz feladat ez a fajta konverzió és nem igazán standardizálható, mert a hangzó beszéd – egy nyelvénél – beszélői többféle akcentussal, kiejtéssel, tempóval, hangerővel, beszédhibákkal, háttérzajokkal stb. terhelten beszélnek. Hasonlóan a karakter és a képfelismeréshez itt sem lehet programozottan az összes várható lehetőségre felkészülni, azaz ez esetben is az MI mintafelismerési képességét hívjuk segítségül. Eleinte az iparban hangvezérlésre használták ezt a funkciót.

Ma már akár egybefüggő szövegeknél diktálás funkciók is kiválthatóak vele klasszikus értelemben, (pl. jegyzet, vázlat, feljegyzéskészítés mint tanulástámogató dokumentumok) vagy meglévő hangfájl szöveges leírata, továbbá videó adatfolyam feliratozása is elkészíthető vele. Újabban néhány nyelvoktató alkalmazás a kiejtés gyakoroltatására is alkalmazza a MI beszédfelismerő képességét (pl. Duolingo, Xeropan), azaz a nyelvtanuló által kiejtett szöveget összehasonlítják (mintaillesztik) egyfajta szabványkiejtéshez. Ha bizonyos határértékek között hasonlít a referenciához a tanuló által kimondott szöveg, a rendszer elfogadja azt helyesnek. (Duolingo 2023), (Xeropan 2023)

hangból > szöveg, (hangfelismerés)

Ennél a médiakonverziónál az MI hangfelismerés képességét hívjuk segítségül, ahol a nem beszéd jellegű hangoknak, zajoknak várjuk a szöveges leírását a kimeneten: pl. taps, motorzaj, égdörgés, autókürt stb. Hallássérülteknek szöveges leírás (felirat) készíthető pl. videófájlra, vagy általános vezérlő parancsoknak is felhasználható a szöveges kimenet.

zenéből > szöveg (zenefelismerés)

Az úgynevezett zenefelismerés funkcionál, a lejátszott zeneszám esetében a konvertált kimenet lehet metaadat; pl. szerző/előadó, cím: Queen – Bohemian Rhapsody, vagy más esetben az énekhang szöveges leírását kaphatjuk eredményként, hasonlóan ahogy a beszédből>szöveg bekezdésnél már említettük. E médiakonverziós esetben énekek, dalok szövegének leiratát kapjuk, amelyet használhatunk a szövegértés támogatásához vagy pl. dalszöveg tanulásához is.

szövegből > beszéd (gépi felolvasás)

Az emberi beszédet szimulálva, az írott szöveget olvassa fel a MI. Betűk, számok vagy szavak kiejtése egyszerű, – akár analóg – hozzárendeléssel is megoldható (pl. az egyes hangokhoz vagy számokhoz előre rögzítettek egy-egy emberi hangot, szótagot vagy szót, amit később összevágtak a célnak megfelelően). Később jelent meg a szöveg „felolvasása” beszédszintetizációval, ahol minden betűnek vagy szótagnak külön képezi a kiejtését a gép (pl. digitális) hangszintetizátor segítségével. Ezek a „felolvasási” formák azonban ma már elavultnak és igen gépinek hatnak.

Az előző módszereknél jóval fejlettebb megoldás a MI deep learning képességeinek alkalmazása a szövegek felolvasásánál. Ez technikailag úgy valósul meg, hogy élő személy(ek) tetszőleges (pl. 10-100 oldalas) felolvasásának hangfájlját használva (mint tanítási adatbázis), az MI bármilyen szöveget „fel tud olvasni” a tanító adatbázisban szereplő ember(ek) természetes hangján, amelyet az valójában soha nem olvasott fel. Ez esetben is a deep learning hálózat fejlett mintafelismerési és predikciós képességét használjuk fel a kívánt szintetizált kimenet elérése érdekében.

Ez a fajta adatkonverzió is hasznos, hiszen akár utazás közben is meghallgathatjuk, (megtanulhatjuk, átisméltelhetjük) a leírt szöveges tartalmakat, mindemellett látássérülteknek is igen fontos lehetőség a gépi felolvasás. Például Natural Readers, Voicemaker (Natural Readers 2023), (Voicemaker 2023)

3.2 Tartalomgenerálás

- szövegből > kép (képgenerálás)
- képből > szöveg (képleírás)
- szövegből > videó (videógenerálás)
- képből > kép, képből > videó (kép és videóátalakítás)
- szövegből > hang (zeneszerzés)
- kérdésre > válasz (chatbot) (szöveg > szöveg)

szövegből > kép (képgenerálás)

A képgenerálás funkcionál szöveges bemenet utasításai alapján alkot az MI, azaz képet, ábrát generál szövegből. Pl. „Rajzolj vágatató lovat vízparton naplementében, romantikus stílusban, fotorealistikusan.” A virtuális művész, vagyis az MI, tehát szöveg alapján állít össze alkalmazástól függően egy vagy több képet az adatbázisából az utasításoknak leginkább megfelelően. Az első próbálkozásokat aztán lehet finomítani az igényeinknek megfelelően. Az utóbbi időben több olyan publikusan elérhető alkalmazás jelent meg, amely igen valóság-hű képi világot képes előállítani leírás alapján akár fotorealistikusan is. (pl. MidJourney, DALL-E, Stable Diffusion, NightCaffe, replicate, stb.) (MidJourney 2023), (DALL-E 2023), (Stable Diffusion 2023), (NightCaffe 2023), (replicate 2023)

Az alkalmazásokkal készíthetünk ábrákat, rajzokat, változatos formában vagy fotószerű képeket is. Ezeket felhasználhatjuk illusztrációként, példaként, akár szóképként is. Igen szórakoztató az utasításokkal (ez esetben zömmel angol nyelvű utasításokkal) való rajzolás, többek közt a szótanulás esetében, ahol nemcsak a szavak képe, hanem például a képelemek egymáshoz való viszonya és helyzete is irányítható.

képből > szöveg (képleírás)

Az előzőhöz képest másik irányú adatkonverzió, amikor a MI a feltöltött képről szöveges leírást, azaz képleírást készít. Ez a lehetőség az objektumfelismeréshez hasonlatos, de annyival több nála, hogy a képi tartalmak viszonyát, kapcsolatát, helyzetét, arányát, jelentését stb. akár szimbolikusan is értelmezi. Mindkét konverziós esetben médiaváltás történik (szövegből>kép, képből>szöveg).

szövegből > videó (videógenerálás)

A szöveges bemenet utasításai alapján a MI videó médiát alkot. Ez a médiakonverziós forma – a bonyolultsága miatt – ma még kevésbé elterjedt, de számos megoldása létezik. Pl. replicate.com (replicate 2023)

képből > kép, képből > videó (kép és videóátalakítás)

Az MI képből egy másik képet vagy videót alkot (vagy átalakít), az utasításoknak (pl. szöveges) megfelelően. Például hozzáad, elvesz, módosít képet vagy képi objektumokat. Kiszínezhetsz fekete fehér képet, feljavíthatja a felbontást, kijavíthat sérült vagy hiányzó részeket, mozgóvá alakíthat állóképeket stb.) A szövegből kép bekezdésben említett, jelenleg elérhető megoldások közül (lásd fentebb), több is képes bemeneti képek felhasználására, átalakítására a feladatnak megfelelően.

szövegből > hang (zeneszerzés)

A szöveges bemenet utasításai alapján az MI hangot(kat) vagy akár zenét alkot. A szövegből „zeneszerzés” mint médiakonverzió igen érdekes kísérleti területe az MI kutatásának. Több irányban folynak kísérletek: az MI utasítások alapján ír zenét pl. egy adott stílusban, vagy akár egy adott előadót utánozva alkot.

kérdésre > válasz (chatbot) (szöveg > szöveg)

A csevegő alkalmazásokban használt szöveges kommunikáció esetében nincs médiaváltás, azaz egy szövegesen megfogalmazott kérdésre szöveges választ kapunk. (pl. ChatGPT 3.5 nyelvi modell) Az elvárt kimenet itt az, hogy a kérdésünkre „emberi nyelven” értelmes választ kapjunk. Mint már az előzőekben is említettük az MI a kérdés alapján szintetizálja (generálja) a várt szöveges választ, azaz a ma elérhető mesterséges intelligenciák nem „értik” valójában a kérdést, hanem a feltett kérdés alapján próbálják megjósolni a lehetséges kimeneti választ. (Laky 2023) Egy a ChatGPT-hez hasonló nyelvi modell esetében a válasz, egy lehetséges (valószínű) szósortrend/mondathalmaz csupán.

ESZTERHÁZY KÁROLY KATOLIKUS EGYETEM
INFORMATIKA KAR • DIGITÁLIS TECHNOLÓGIA INTÉZET
AGRIA MÉDIA KONFERENCIA 2023

Mindezek „ellenére” a ChatGPT megjelenésével elérhetővé tett komplex predikciós nyelvi modell általános, összetett kérdésekre is képes válaszolni és azt is fontos megjegyezni, hogy a modellt fejlesztik, bővítik az adatbázisát, valamint a felhasználói interakciók alapján is folyamatosan tanul és fejlődik. A jelenlegi 3.5 verzió már olyan kérdésekre ad jó válaszokat, amelyekre a megjelenése előtt lévő korábbi megoldások válasza (vagy eredményei) nem igazán voltak relevánsak vagy komolyan vehetőek. A ChatGPT legutóbbi verziói (3.5 vagy 4) olyan eszközt adtak a kezünkbe, amelyeket nem lehet figyelmen kívül hagyni sem az oktatás, sem általánosságban a tanulás támogatás esetében. A nyelvi modell részletesebb elemzése messze meghaladja a tanulmány kereteit, azonban az kijelenthető, hogy a ChatGPT igen széles körben alkalmazható általános tanulástámogató eszközként is, amely képes – a legtöbb tudományterületen – konkrét kérdések megválaszolására, elemzésekre, dolgozatok, esszék elkészítésére, vagy akár a nyelvtanulás támogatására is. A 4-es verziótól a modell „lát”, azaz képes értelmezni a bemenetere adott képeket is, ezzel mind szélesebb körben és komplexebb válaszok lehetőségét biztosítva. 2023 szeptemberétől a ChatGPT Plus már tud „beszélni” és értelmezi a beszédet, azaz „hall”. Ez például azt is jelentheti, hogy a megoldandó probléma egy képen szerepel, amelyre a megoldást írott vagy akár elmondott válaszban fogjuk megkapni. Tehát ha tanulástámogatási/oktatási irányból közelítjük meg a dolgot, tekinthetjük metaforikusan egy "magántanárnak" vagy "konzulensnek" az MI-t, akivel "meg lehet beszélni" sok mindent, akár idegen nyelven is! (ChatGPT 2023), (ChatGPT4 2023)

Összegzés

Az előzőekben áttekintettük a jelenleg elérhető mesterséges intelligencia alkalmazások médiakonverziós képességeit és lehetőségeit. Azt tapasztaltuk, hogy alapvetően két részre bonthatjuk a konverziókat. Az egyik megközelítés szerint a bemeneti adathoz képest a kimenetet transzformációnak tekinthetjük, a másik esetben tartalomgenerálásnak (más néven szintetizációnak). Ha csökken (vagy nem változik számottevően) az úgynevezett jellemzőtér, akkor (adat)transzformációról (pl. képből> szöveg), ha pedig növekszik – azaz csökken az absztrakció – akkor tartalomgenerálásról beszélhetünk (pl. szövegből>kép). Azt is megállapítottuk, hogy nem minden konverziónál történik médiaváltás. (ld. A fordítóprogramok vagy a szöveges chat alkalmazások be és kimenetei esetében nem minden esetben van médiaváltás, azaz: pl. szöveg>szöveg.) Az esetek többségében azonban az adatkonverzióval médiaváltás is létrejön. (pl. szövegből>kép, képből>videó, beszédből> szöveg, stb.) A vizsgálat alapján azt is megállapíthatjuk, hogy több alkalmazás rendelkezik multimédiás képességekkel, azaz képes egyszerre többféle médiát (pl. szöveg, kép, hang, videó) kezelni és akár átalakítani. Az is tendenciának tűnik, hogy egyre több alkalmazás válik multimédiássá. Gondolunk itt pl. a ChatGPT evolúciójára, ahol az első verziók (szöveges kérdés> szöveges válasz) be és kimenetéhez képest a fizetős verzió ma (2023 ősz) már képes „látni”, „hallani” és „beszélni” vagy a Google Lens alkalmazásra, amely többszörös médiakonverzióval, a virtuális valóságot segítségül hívva, fordítja le a „látott” szöveget a kívánt nyelvre. Szintén tendencia, az egyes entitások kimeneteinek drasztikus fejlődése mind szubjektív mind objektív szempontból is. Tehát egyre jobb és emberszerűbb szöveges válaszokat kapunk a csevegőalkalmazásokban, egyre szebb képeket, rajzokat alkot az MI, valamint a hangok, zenék, videók terén is látványos fejlődésnek lehetünk szem és fültanúi. Mint tudjuk, az MI által generált verbális és multimediális tartalmak egészen más szerkesztési elvek és bemeneti adatok alapján jönnek létre, mint azok a tartalmak, amelyeket mi emberek alkotunk. Mégis, az MI által létrehozott kimenetek – mondhatjuk mindezek ellenére – egyre kevésbé különböztethetők meg az ember alkotott tartalmaktól. Sok esetben éreztük úgy, hogy az eredmény „átmenne” a Turing teszten, azaz a kérdező nem tudná eldönteni, hogy ember vagy gép adta a választ.

A vizsgálat során azt is megállapítottuk, hogy az MI intelligencia sok esetben úgy jelenik (vagy jelent) meg a mindennapi életünkben bizonyos alkalmazások által, hogy „észre sem vettük” vagy el sem gondolkoztunk rajta, hogy bizonyos rész vagy célfeladatokat az MI old meg. (Pl. Feladatok azonnali kiértékelése, vezérlés hangutasításokkal, képminőség javítás stb.) Jelen alaputatásnak nem volt konkrét célja az egyes lehetőségek oktatási/tanulási oldalról való részletes vizsgálata, de úgy érezzük, hogy általánosságban a legtöbb alkalmazás valamilyen módon, kreatívan beilleszthető és széleskörűen alkalmazható megoldás lehet a tanulástámogatásban vagy bizonyos esetekben az oktatásban is.

Irodalomjegyzék

- Laky, Z. (2023). „Az embereket nem nehéz átverni” – a Google magyar tudósa a mesterséges intelligencia öntudatáról és veszélyeiről, Válasz Online, 2023.03.10 interjú Szegedy Krisztiánnal, Elérés (2023. 10. 12.) <https://www.valaszonline.hu/2023/03/10/szegedy-krisztian-matematikus-google-mesterseges-intelligencia-interju/>
- T. Nagy, L., Boda, I. (2014). Nyelvtanulás és nyelvtanulási lehetőségek a web 2.0 közegében In: Bányi, Sz., Navracsics, J., Vigh-Szabó, M. (szerk.) (2014). Nyelvelsajátítási-, nyelvtanulási- és beszédkutatások = Papers in language acquisition, language learning and speech research Budapest, Magyarország, Veszprém, Magyarország: Gondolat Kiadó, Pannon Egyetem Modern Filológiai és Társadalomtudományi Kar 199-211.
- Tóth, L. (2020). Mesterséges neuronhálók és alkalmazásai, Elektronikus tananyag, Szegedi Tudományegyetem, Elérés (2023. 10. 10.) <https://www.inf.u-szeged.hu/~tothl/bevmely/01.%20Gepi%20tanulas.%20neuron.%20neuronhalo.pptx>
- Turing-teszt (2023a). Turing teszt ismertető leírás, Elérés (2023. 08. 18.) <http://www.mestersegesintelligencia.hu/doc/Turing%20teszt.php>
- Turing-teszt (2023b). Wikipédia szócikk, Elérés (2023. 10. 18.) <https://hu.wikipedia.org/wiki/Turing-teszt>
- ChatGPT (2023). Elérés (2023. 09. 28.) <https://openai.com/blog/chatgpt>
- ChatGPT4 (2023). Elérés (2023. 09. 28.) <https://openai.com/research/gpt-4>
- DALL-E (2023). Elérés (2023. 09. 28.) <https://openai.com/dall-e-2>
- Duolingo (2023). Elérés (2023. 10. 01.) <https://hu.duolingo.com/>
- Google Lens (2023). Elérés (2023. 09. 28.) <https://lens.google/intl/hu/#translate>
- MidJourney (2023). Elérés (2023. 09. 28.) <https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F>
- Natural Reader (2023). Elérés (2023. 10. 18.) <https://www.naturalreaders.com/>
- Nightcaffe (2023). Elérés (2023. 10. 18.) <https://creator.nightcafe.studio/text-to-image-art>
- Replicate (2023). Elérés (2023. 10. 18.) <https://replicate.com/>
- Stable Diffusion (2023). Elérés (2023. 10. 18.) <https://stablediffusionweb.com/>
- Voicemaker (2023). Elérés (2023. 10. 18.) <https://voicemaker.in/>
- Xeropan (2023). Elérés (2023. 10. 01.) <https://hu.xeropan.com/>