



Hangoskönyvek az információs szolgáltatásban

Témavezető

Lengyelné Dr. Molnár Tünde

Egyetemi docens

Készítette

Ujhelyi Gábor

Könyvtár- és Információtudomány MA

EGER, 2023

Lezárás dátuma: 2023. december 4.

Tartalomjegyzék

Tartalomjegyzék	2
Ábrajegyzék	3
Bevezetés	5
Papír könyv és hangoskönyv kapcsolata.....	5
Megvalósítás, előállítás, technikai feltételek	8
Könyvtár és oktatás	14
*-learning	19
Mesterséges intelligencia	20
Interakció	28
Jogi vonatkozás.....	29
Hangoskönyvek és ember-gép kommunikáció jövője.....	30
Kutatás bemutatása	33
Kutatási eredmény.....	34
Összegzés.....	38
Nyilatkozat	39
Irodalomjegyzék	39

Ábrajegyzék

1. ábra – Beszélő gép	12
2. ábra – Tanulótípusok.....	16
3. ábra – Bloom taxonómiája	17
4. ábra – SAMR modell	18
5. ábra – R2D2 modell	18
6. ábra – Digitális tanulás formái	20
7. ábra – Nyelvi modell tanítás számítási kapacitás igénye.....	22
8. ábra – Nagy nyelvi modellek.....	23
9. ábra – Neurális hálózat.....	24
10. ábra – Reinforcement Learning by Human Feedback	25
11. ábra – LangChain.....	26
12. ábra – Uncanny valley effektus	31
13. ábra – Adatforrásként felhasznált könyvek.....	33
14. ábra – Proof-of-concept alkalmazás.....	34
15. ábra – Teszteredmények (összes)	35
16. ábra - Szóbeli összes teszt megoszlása	36
17. ábra - Írásbeli összes teszt megoszlása	36
18. ábra – Teszteredmények (csak magyar nyelvű tesztek).....	36
19. ábra - Szóbeli magyar teszt megoszlása.....	37
20. ábra - Írásbeli magyar teszt megoszlása.....	37

Absztrakt: A hangoskönyvek számára az utóbbi évek technológiai fejlesztései számos új alkalmazási, bővítési lehetőséget nyitottak meg, továbbá a társadalmi trendek is támogatják terjedését, azonban az információszoftvarak szolgáltatási eszköztárának nem képezi részét. Jelen kutatás ezen helyzetre kínál megoldást, egy egyedi fejlesztésű szoftver adaptálásával, ami megvalósítja komplett könyvek feldolgozását és azok tartalmukra vonatkozóan párbeszéd lehetőségét. A jelenleg széles körben elérhető és egyre fejlettebb nagy nyelvi modellek, valamint felhő szolgáltatásokon keresztül elérhető text-to-speech (gépi felolvasó) és speech-to-text (leírató) megoldások segítségével megvalósítható az információszoftvarak érdekében a hangoskönyvek jelenlegi koncepciójának kiterjesztése. Egy digitálisan rendelkezésre álló írott könyv automatikusan előállhat hangoskönyvként úgy, hogy ne csak felolvastatható és meghallgatható, de gépileg interaktívvá tehető, összefoglalható, magyarázható és kérdezhető legyen, ami nagy mértékben hozzájárulhat ahhoz, hogy a könyvek az auditív tanulótípusokhoz is közelebb kerüljenek és segítsék az információközlést, a befogadást, valamint segítséget nyújthat az olvasási nehézségekkel küzdők számára, és megvalósul az igazi „beszélő könyv”. A hangoskönyvek eredetének, történetének, átalakulásának elemzésén túl számba kell venni a kapcsolódó tudományterületeket, a technikai problémákra nehézségeket, megoldott és megoldandó feladatokat, a hangoskönyvek jelenlegi helyzetét, használatát. Fontos számba venni a jelenleg elérhető fejlett felhő alapú és helyben futtatható nagy nyelvi modellek tulajdonságait, finomhangolhatóságát, különböző méretű dokumentumok feldolgozhatóságát és azokról egy minta alkalmazásban írott vagy szóbeli formában természetes nyelven feltett kérdésekre általuk adott válaszok tartalmi validitását. Ezt követően határozható meg a hangoskönyvek meglévő és lehetséges felhasználási lehetőségei a közvetlen civil felhasználás mellett a könyvtárak és az oktatás területén. A saját fejlesztésű szoftver adaptálása lehetőséget biztosított tesztelésre, és a kapott adatok elemzése validálja a hangos könyvek oktatásban történő alkalmazásának indokoltságát, alkalmas a szakmai tartalmak élőszóban interaktívvá tételére és ezáltal segítséget nyújthat a tanároknak a tananyag közvetítésére és a diákoknak a tudásanyag jobb megértésére és könnyebb befogadására.

Kulcsszavak: hangoskönyv, könyvtár, interaktivitás, mesterséges intelligencia, nagy nyelvi modellek, auditív tanulótípus, olvasási nehézségek

Bevezetés

Annak ellenére, hogy a hangoskönyvek létrejötte nem újkeletű, kevés fejlődésen mentek keresztül, azonban az utóbbi évek technológiai fejlesztései számos új alkalmazási, bővítési lehetőséget nyitnak meg. Az írott ismeret anyagok hangon keresztül befogadhatóságához több út vezet. Az információszolgáltatás felőli megközelítésből a figyelem középpontjába elsősorban a tudományos, ismeretterjesztő, szak- illetve tankönyvek kerülnek. Ebben az esetben arányaiban kisebb jelentősége van az audio előadás módjának, stílusának, hangszínének, dominánsabb az információ jól érthető átadásának szándéka, aminek köszönhetően a maga hiányosságaival együtt is van létjogosultsága a gépi felolvasásnak, beszédszintézisnek.

Kutatásom részét képezi egy egyedi fejlesztésű szoftver adaptálása komplett könyvek feldolgozására és azok tartalmára vonatkozóan párbeszéd lehetőségének megteremtésére, melynek eredményeit szeretném bemutatni. Kutatásom célja annak igazolása, hogy a jelenleg széles körben elérhető és egyre fejlettebb nagy nyelvi modellek, valamint felhő szolgáltatásokon keresztül elérhető text-to-speech (gépi felolvasó) és speech-to-text (leírató) megoldások segítségével megvalósítható az információszolgáltatás érdekében a hangoskönyvek jelenlegi koncepciójának kiterjesztése. Ennek keretein belül egy digitálisan rendelkezésre álló írott könyv automatikusan előállhat hangoskönyvként úgy, hogy ne csak felolvastatható és meghallgatható, de gépileg interaktívvá tehető, összefoglaltatható, magyaráztatható és kérdezhető legyen, ami nagy mértékben hozzájárulhat ahhoz, hogy a könyvek az auditív tanulótípusokhoz is közelebb kerüljenek és segítsék az információközlést, a befogadást, valamint segítséget nyújthat az olvasási nehézségekkel küzdők számára. A tágran értelmezett hangoskönyv meghatározással közelebb hozom az angol irodalomban elterjedtebb „talking books” tükörfordítását igazi „beszélő könyv” megvalósításával.

A kutatás megvalósítása során a jelenleg elérhető fejlett felhő alapú és helyben futtatható nagy nyelvi modellek tulajdonságait, finomhangolhatóságát, különböző méretű dokumentumok feldolgozhatóságát és azokról egy minta alkalmazásban írott vagy szóbeli formában természetes nyelven feltett kérdésekre általuk adott válaszok tartalmi validitását vizsgálom.

Tanulmányomban bemutatom a ma ismert hangoskönyvek eredetét, történetét, átalakulását. Kitérek a kapcsolódó tudományterületekre, a technikai problémákra, nehézségekre, megoldott és megoldandó feladatokra, a hangoskönyvek jelenlegi helyzetére,

használatára. Vizsgálom a hangoskönyvek meglévő és lehetséges felhasználási lehetőségeit a közvetlen civil felhasználás mellett a könyvtárak és az oktatás területén. Ezt követően bemutatom a mintaalkalmazással elért eredményeimet, következtetéseimet, foglalkozom a továbbfejlesztési lehetőségekkel, kutatási irányokkal.

Papír könyv és hangoskönyv kapcsolata

A hangoskönyvekre sokan az írott könyvek egyszerűen felolvasott, újképletű változataként tekintenek, mi több, hajlamosak egyfajta alacsonyabb rendű megjelenési formát, „a lusta emberek olvasási módját” látni benne, ami erodálja az olvasási szokásokat és hátrányosan hat a kultúrára. Azon túl, hogy látási problémákkal, olvasási nehézségekkel küzdő embertársaink számára milyen mértékben nyithatják ki a világot, vagy miért lehet létjogosultsága a különböző tanulótípusokhoz való adaptációban, az embernek első közelítésben elképzelése sincs, milyen lehetőségek rejlenek a hangoskönyvek jelenlegi és jövőbeni felhasználási lehetőségeiben akár az oktatás területén, akár azon kívül, valamint a technológia, a mesterséges intelligencia fejlődése milyen további kapukat nyit ki. Természetesen sok megoldandó feladat, probléma van még ezen az úton, ugyanakkor az elért eredmények nem csak széleskörű azonnali alkalmazási területet biztosítanak, hanem újabb továbbfejlesztési, kutatási lehetőségek alapjait teremtik meg.

Az első teljes hosszúságú hangoskönyvet 1930-ban vették fel, majd a második világháborút követően kezdték támogatással terjeszteni, elsősorban a látássérült háborús veteránok megsegítése érdekében, de segítséget nyújtva a más okból kifolyólag olvasási nehézségekkel küzdők számára is (Smith, 2022). A hangoskönyvek lemezen terjesztése meglepő módon a zeneipart is megelőzte. A hangoskönyvek hamarosan nagy népszerűségnek kezdtek örvendeni, de igazán akkor kezdett csúcsra érni a terjedésük, amikor megjelentek előbb a szalagok és a kompaktabb kazetták, CD lemezek és az ezeket használó, már hordozható lejátszó egységek, a walkmanek, discmanek, majd az mp3 fájlok megjelenésével a digitális lejátszók, és ugrásszerű növekedésnek indult a digitális formátumok online terjesztésével az internet elterjedésének köszönhetően.

A hangoskönyvek elsősorban az írott könyvek felolvasott változataként alakultak ki eleinte teljes terjedelemben, de rövidesen megjelentek a tömörített, rövidített változatok és kivonatok. Bár túlnyomó részben a hangoskönyvet mint valamely írott könyv hangalapú változatát ismerjük, ha nem is tömegével, de előfordulnak olyan művek is, amelyek csak audio formátumban, hangoskönyvként elérhetőek, vagy legalábbis előbb léteztek hangban, mint írásban. Ennek oka elsősorban a formátum növekvő népszerűségében, az újdonság

varázsában, a ritkaságban, a kiemelkedési lehetőségben, valamint a célcsoport jellemzőiben és olvasási szokásaiban keresendő, de természetesen vannak a hangoskönyveknek kifejezetten olyan tulajdonságai is, amelyekkel az írott alkotások formátumukból adódóan nem tudnak rendelkezni. Elég csak olyan jellegű művekre gondolni, mint azok az ismeretterjesztő kiadványok, ahol a hagyományos képi illusztráció helyett nagyobb jelentősége van a hangeffektusoknak, ilyenek például a nyelvjárásokat, dialektusokat bemutató dokumentumok. Ezekben az esetekben nem az a kérdés merül fel, hogy a hangoskönyv egyenértékű-e az írott könyvvel, hanem hogy mi módon tudja visszaadni egy írott változat a hangban elérhető többlet értéket. Ez mindenképp megerősíti, hogy a hangoskönyvekre ne csak az írott könyvek felolvasott változataként tekintsünk, hanem olyan művekként, amelyek önállóan is megállják a helyüket. Kicsit tovább vizsgálva lényegesen több előnyét is felfedezhetjük, mint elsőre gondolnánk. A már említett többlet tartalom nyújtásának lehetősége mellett a szöveg változatlan tartalmú hang reprezentációja magától értetődő módon nyújt segítséget a látási problémákkal és olvasási nehézséggel küzdő emberek részére, akik számára ezáltal jobban kinyílhat a világ, a hasonló problémákkal nem szembesülő embertársaik által elérhető megközelítő lehetőséget szerezve az információhoz és az élményhez jutásban. Ennek köszönhetően a “beszélő könyv”-nek is nevezett kiadványok az esélyegyenlőség, az akadálymentesség jelentős eszközei lehetnek. Az olvasásban korlátozottság azonban nem feltétlenül statikus korlát. A mai, kifejezetten felgyorsult világban egészséges, olvasni szerető ember is kerülhet olyan szituációba, amikor akadályokba ütközik az olvasás lehetősége, azonban mégis szeretne hódolni egy szöveges műalkotás élvezetének, az ismeretszerzésnek, vagy konkrétan tanulásnak. Ilyen helyzetek lehetnek átmeneti testi akadályok, egy szemműtét, fáradtság, rossz fényviszonyok, de a figyelem megosztásával olyan esetekben is felmerülhet az igény, amikor egyéb, nem teljes figyelmet lekötő tevékenység közben nyílik lehetőség a hanggal reprezentált tartalom befogadására, például autóvezetés, főzés, takarítás közben. Monoton cselekvések közben ez a többlet figyelem igény nem is feltétlenül okoz hátrányt, sőt, az autópályákon akár balesetveszély elkerülésében is segítség lehet az ingerszegény környezet és fáradtság kombinációja által okozott elalvás megelőzésével. A műalkotások hallás általi élvezete semmiképp sem tekinthető ördögtől való dolognak, a gyermekek számára olvasott mesék pozitív fejlesztő hatásával kapcsolatban teljes az egyetértés a kutatók között (VisikoKnox-Johnson, 2016). Az emberi szokások változása, a technológia fejlődése, a digitalizáció mindennapi életbe beépülése egyre közelebb hozza ezeket a lehetőségeket, amik észrevétlenül épülnek be a hétköznapijainkba.

Megvalósítás, előállítás, technikai feltételek

A hangoskönyvek létrejöttekor a felvett hangok lejátszása még közel sem volt olyan magától értetődő, mint az olvasás általi befogadás. Míg egy könyv bármikor elérhető volt, elfért egy kisebb táskában, bárhová magával vihetette az ember, addig a lemezek lejátszása nagyméretű eszközt igényelt, ami közel sem állt mindenhol rendelkezésre, mozgatása pedig annál nagyobb nehézséget okozott, fel sem merülhetett például egy utazásra elvinni. Amilyen szabadságot adott a szemnek és a kéznek, ugyanolyan korlátot jelentett egyszerre helyben és időben az eszközigény miatt. Természetesen a későbbi hanghordozók és lejátszó eszközök megjelenésével ez a különbség rohamosan csökkenni kezdett. Az elterjedésük legnagyobb ugrását épp ezért a hordozható lejátszók megjelenése jelentette, amit a későbbi digitális formátumok és a kommunikációs hálózat elterjedése tovább tudott erősíteni. Érdekes megvizsgálni, hogy a könyvek tárolása a különböző formátumokban mennyire eltérő helyszükségletet igényel. A papír alapú könyvek mérete mindenki által könnyen becsülhető határok között mozog. Természetesen vannak kötetenként eltérő terjedelmű, illetve különböző kötetszámú kiadványok, azonban a forgalomban lévő könyvek fizikai dimenziói egységes nagyságrendben mozognak. Hangoskönyvek analóg fizikai mágneses adathordozón vagy kazettán, szalagon, lemezen való tárolása már kisebb helyet is igényelhet, mint a papír alapú könyv, de nagy számosságú kiadvány azért így is tetemes fizikai helyett tud elfoglalni. A nagyságrendi ugrás a helyfoglalásban a digitális formátumokra váltásnál történik. Mind az írott könyvek, mind a hangoskönyvek digitális reprezentációja a mai tároló eszközökön elhanyagolható méretű fizikai szükségletnek feleltethető meg. Hozzá kell tenni, hogy a formátumok, illetve a feldolgozás módja nagyságrendi eltéréseket okozhat. Egy papír alapúból nagy felbontásban beszkenelt könyv százszor, ezerszer akkora tárhelyet is foglalhat, mint egy karakteresen ábrázolt írott anyag. Ha nem is ekkora, de jelentős különbségek vannak a hangfelvételek méretében is a különböző mintavételezési és pontossági adatok, valamint a veszteséges és veszteségmentes tömörítési módok és azok paramétereinek függvényében. A változó paramétereket tovább bővíti, hogy a szélessávú internet elterjedésével újabb szempont jelent meg, miszerint nem szükséges a hanganyagoknak a lejátszás során egyfolytában és teljes terjedelemben jelen lenniük, hanem elegendő, ha egy szolgáltató tárolja a nagyméretű állományokat, és éppen csak a pillanatnyilag lejátszott adatsomagokat küldi át a hálózaton keresztül a lejátszó eszköz számára. Visszatérő döntési helyzetet teremt és sokszor kompromisszumot igényel, hogy az elkészülő hanganyag tárolásához szükséges nagyobb tárhely, a vele összefüggésben

álló átvitelhez szükséges sávszélesség és az ezekhez szükséges eszközök teljesítménye és költsége, valamint a tökéletesre törekvő hangminőség között hol van az a határ, amelyet célszerű elérni, és azt milyen mintavételi frekvencia, bitmélység, illetve filterezési és tömörítési mód és bitráta választásával lehet megfelelően megvalósítani. Mindenesetre a mai modern eszközök korában a korszerű technika már lehetőséget biztosít arra, hogy az otthoni és mobil zenelejátszókon, az asztali és hordozható számítógépeken, valamint akár a hazánkban kevésbé felkapott, de a tengerentúlon nagy népszerűségnek örvendő okoshangszórókon túl a szinte minden pillanatban és mindenhol velünk levő mobiltelefonon keresztül számos élethelyzetben helytől függetlenül élvezhető minőségben tudjunk rengeteg hangoskönyvet magunkkal vinni és hallgatni akár a telefonokban elérhető tárolókapacitás, akár a szinte bárhol jelenlévő nagysebességű vezeték nélküli hálózaton keresztül.

A digitálisan tárolt hangoskönyvek magukban hordozzák az elektronikusan nyilvántartott adatok által hozott előnyöket, mint a kereshetőség, könnyebb és gyorsabb hozzáférhetőség, a külön is kezelhető, de magával is hordozható metaadatok, ugyanakkor vannak hátrányai is. Megszűnik a könnyen és vitathatatlanul azonosítható példány, a másolatok a korábbi példányokkal teljesen egyenértékű, identikus, akár megkülönböztethetetlen példányokként élhetik tovább az életüket. Ez felvet problémákat, megoldandó feladatokat a hangoskönyvek értékesítése, valamint a szerzői jogok biztosítása során. Jogosan merül fel az igény, hogy egy szerző szellemi alkotása ne kerülhessen kontrollálhatatlanul terjesztésre a kifejezett hozzájárulása nélkül, ez azonban a digitális adatok körében nem csak hanganyag, hanem képek és írott dokumentumok, valamint videók esetén sem triviális feladat. A másolás elleni védelem egyik lehetséges iránya a hozzáférés korlátozása. Ez akkor könnyebben megvalósítható, ha a jogosult használó nem töltheti le saját eszközére az anyagot, csupán online elérhetőséget kap, amely lehetőséget teremt nem csak egyszeri megvásárlás keretein belül értékesítésre, hanem használat, vagy előfizetés alapú igénybevételre is. Itt is felmerül az autentikációs adatok illetéktelen megosztásának kérdése, de IP cím, geolokáció alapú korlátozás és konkurens hozzáférés limitálás segítségével ezt meg lehet nehezíteni. Másik irány a dokumentum példány másolhatóságának védelme. Ezt azonban maga az adat önmagában nem tudja megvalósítani, a blokkolás az adatot olvasó alkalmazás feladata, és attól kezdve, hogy az adattartalom a felhasználó gépén megtalálható (akár jelszóval védve is), csak eszköz kérdése, hogy az kinyerhető és más formában továbbadható legyen. Tágabb értelemben itt össze is ér a kétféle védelem iránya, ha az adat megjelenítésre, lejátszásra (akár online közvetített, akár helyben tárolt formában) a felhasználó gépén bármilyen formában előáll, akkor az, vagy annak

legalábbis veszteséggel kinyert adattartalma megszerezhető, másolható (szélsőséges esetben akár analóg eszközök közbeiktatásával). Speciális eset az adatállományok vízjellel való védelme, itt azonban nem a képeknél megszokott látható vízjel játssza a szerepet, hanem bármilyen olyan transzformáció vagy többletadat, ami nem az adatállomány metaadatai között helyezkedik el, hanem a tényleges adattartalomra van alkalmazva, akár érzékelhető módon, akár teljesen észrevétlenül (pl. képek vagy képként közvetített írott dokumentumok esetén az emberi szem által nem észrevehető árnyalat módosításokba, vagy hangok esetén az emberi fül által nem hallható hangmanipulációkba kódolva), amelyek a tartalomfogyasztás élményét nem rontják, de a megjelenített, lejátszott média másolásakor, konverziójakor is továbbkerülnek. Ez a másolást ugyan meg nem akadályozza, azonban az ellenőrizhetőségre, számonkérhetőségre biztosít lehetőséget, egy illegális másolat esetén azonosíthatóvá teszi az eredetileg legálisan megszerzett dokumentum forrást, ami a megfelelő jogi szabályozás és retorzió lehetősége mellett kellő visszatartó erő lehet. Az egyedi digitális példányok hiteles azonosítására hozták létre a blokklánc technológiára építve az NFT-ket (non-fundible tokens). Az elosztott technológia által biztosított hitelesítés garantálja a vevő birtokában lévő digitális műalkotás egyediségét és eredetiségét. Ez a technológia alkalmas bármilyen digitális műalkotás kezelésére, legyen az kép, hang vagy videó, biztonságos és hiteles platformot teremtve a műalkotások digitális kereskedésének (Fernandes & Morais, 2022). Megjelenésekor hatalmas potenciált láttak benne, de ez a lelkesedés hamar kifulladásra látszott. A lehetőségek ellenére hangoskönyvek ezen reprezentációja nem terjedt igazán el.

Hangoskönyvek előállításának legkézenfekvőbb módja az emberi felolvasás, néhány évtizeddel ezelőttig nem is nagyon volt más elérhető lehetőség rá, ugyanakkor egyben ez a legköltségesebb módja is. Szükség van hozzá egy jól képzett, megfelelő orgánummal, hangképzési technikával, rutinnal rendelkező színészre, előadóművészre, aki a hangoskönyv nyelvét anyanyelvi szinten műveli, valamint professzionális stúdió felszerelésre, ahol a lehető legkisebb mértékű zaj és zavaró tényező mellett, legjobb minőségben lehet rögzíteni hangfelvételt. Az elkészült hangfelvételen nem elhanyagolható utómunka végzése is szükséges, úgy mint hangmérnöki tevékenység, vágás. Kétségtelen, hogy jelenleg ezen a módon érhető el a legjobb minőség, a legmagasabb szinten biztosítható élmény, azonban arra számíthatunk, hogy a hangoskönyvek előállítása forradalom előtt áll. Az emberi hang mesterséges előállítására régóta történnek próbálkozások. Kezdetben mechanikus és akusztikus eszközökkel próbálták az emberi nyelv hangjait utánózni (Kempelen Farkas beszélő gépe, ld. 1. ábra), később a hangfelvételi eljárások fejlődésével rögzített

hangdarabok összeillesztésével valósult meg a gépi beszédelőállítás. Ezek bonyolult szabályalapú, az adott nyelv nyelvtanát részleteiben ismerő, a kivételekre is felkészített leprogramozott algoritmusok segítségével tudtak megvalósulni. A technika jelenlegi állása, a mesterséges intelligencia, a nagy nyelvi modellek utóbbi években lezajló rohamos fejlődése azonban lehetővé tette, hogy gépi beszéd szintézist számos különböző nyelven jó minőségben valósítsanak meg nagy mennyiségű adatmintán betanított neurális hálózatok segítségével. Bár ezek az eszközök még nem tökéletesek, így is nagyon sok előnyt biztosítanak az emberi felolvasásokkal szemben. Lerövidül az előállítási idő, csökkennek az előállítási költségek, minimálisra redukálható az utómunka szükségessége, az előállított beszéd jellemzői anélkül változtathatók, hogy arra külön felolvasó embereket kellene keresni és foglalkoztatni. Kiválasztható a hangszín, az orgánus, a sebesség és természetesen a nyelv. A gépi felolvasó szoftverek nagy szórást mutatnak az általuk előállított hanganyag minőségében és élvezhetőségében. Természetesen lényegesen kifinomultabb megoldások léteznek a nagyobb világnyelvekre (különös tekintettel az angolra), mint a kisebb gyakorisággal előforduló nyelvek esetén. Ennek magyarázata kereshető egyfelől az azokon a nyelveken felmerülő igényekben (az azt beszélők számában), ezáltal az azokon készülő megoldásokra fordítható fejlesztési költségekben, másfelől a neurális hálózatok számára ezeken a nyelveken nagyságrendekkel nagyobb mennyiségben elérhető betanító minták jelenlétében. Míg magyar nyelven inkább csak botladozó megoldásokat találni, addig angol nyelven léteznek az emberi beszédet a megtévesztésig utánozni képes szoftverek, megoldások, melyek olyan finomságokra is képesek, mint a beszédsebesség menet közbeni dinamikus változtatása, vagy az emberre jellemző nüansznyi szünetek beiktatása a folyó szövegbe. Természetesen ezek költsége sem nulla, a rendszereket kiszolgáló infrastruktúra komoly számítási teljesítményt igényel, azonban az ezek használata során felmerülő költségek szinte elhanyagolhatóak az emberi munkával előállított hangoskönyvek költségéhez képest. A költségeket nagyban meghatározza a szoftverek igénybevétele módja. Két alapvetően eltérő modell a felhő szolgáltatásként (SaaS - software-as-a-service) igénybe vett és a saját infrastruktúráján helyi példányban (on-premise) üzemeltetett architektúra (Sun, Zhang, Chen, Zhang, & Liang, 2007). Míg előbbi esetben nagyobb szolgáltató gondoskodik a megfelelő elosztott erőforrások biztosításáról, rendelkezésre állásukról, rendszeres frissítésükről és skálázhatóságukról a szolgáltatást azonnal igénybe vehetően elérhetővé téve, amelyért használat alapú vagy átalány díjat számol fel, addig a saját hardver és szoftver környezet kialakítása magasabb kezdeti költséget és az üzemeltetés állandó terheit vonja magával és jóval komolyabb szakértelmet

igényel. Egyre több szolgáltató biztosít felhő megoldásokat, melyek paraméterei (pl. elérhető nyelvek, hangszínek), minősége, sebessége, költsége nagy szórást mutat, valamint gombamód szaporodnak a megvásárolható vagy esetenként akár kereskedelmi célra is ingyenesen használható nyílt forráskódú megoldások. A felhasználás célja, komplexitása, gyakorisága, a rendelkezésre álló erőforrások és kompetencia, valamint az elvárt minőség határozza meg, hogy milyen esetben melyik megoldás a gazdaságosabb és biztosítja a megfelelő eredményt.



1. ábra – Beszélő gép

forrás: <http://www.kempelen.hu/beszelogep.html>

A gépi beszédfelolvasó szoftverek jelenleg számos kihívással küzdenek, melyek egy része várhatóan rövid időn belül megoldódik, más részük feltehetően még hosszabb időt fog igénybe venni. Tipikus hibaforrások a különleges kiejtést igénylő karaktersorozatok, mint számok, római számok, dátumok, időpontok, képletek, címek, telefonszámok, rövidítések, mozaikszavak, ritkább tulajdonnevek. Ezekre megoldást jelenthet speciális szótárak alkalmazása, modellek finomhangolása, az emberi visszacsatolás általi megerősített tanítása (RLHF – reinforcement learning by human feedback), illetve a készítés közbeni emberi közreműködéssel orvosolni lehet a problémák egy részét szinonimákkal való helyettesítéssel és kézi fonetikus átirat segítségével, amellyel könnyebbé tehető a beszédszintetizátor feladata (Casper, és mtsai., 2023). Ennél kifinomultabb, de szintén kézi, vagy a szóalak és a kontextus alapján csak közel megfelelően automatizálható köztes feldolgozási lépést tesz lehetővé egy erre kialakított leíró nyelv (SSML – Speech Synthesis Markup Language), melyben a szintaxis lehetőséget biztosít, hogy a felolvasó algoritmus számára speciálisan kiejtendő kifejezések előre definiált kategóriákkal felcímkézhetőek legyenek (Jin, Lee, & Park, 2004). A felolvasó szoftverek számára komolyabb kihívást jelent az élő emberi beszéd sajátosságainak utánzása, a hangsúlyozás, hanglejtés, a dinamikusan változó beszédtempó,

nem is beszélve az átélés és korhűség (nyelvezet, dialektus) kevéssé kézzelfogható jellemzőiről. Azonban az ebben való fejlődés rohamtempóban halad előre és szinte napi gyakorisággal jelennek meg jobbnál jobb teljesítményű megoldások. A technológia nem csak azt teszi lehetővé, hogy néhány előre rögzített mintából válasszuk ki a felolvasó hangot, illetve annak magasságát, hangszínét. sebességét módosítsuk, de arra is lehetőség van aránylag rövid minta megadásával, hogy tetszőleges ember hangjára betanítsuk a modellt és a felolvasás az általunk kiválasztott személy hangján szólaljon meg. Ez természetesen felvet jogi kérdéseket is, egyrészt a tulajdonosának hozzájárulási szükségességéről, személyiségi jogairól, másrészt a rosszhiszemű felhasználás előtt is megnyitja a lehetőséget. Ismert emberek, közéleti szereplők hangjának, beszédstílusának utánzásával a deep fake technológiát előszeretettel alkalmazzák szándékos félrevezetés céljával.

A felolvasás jellemzőinek megválasztása nem kizárólag az élvezeti érték figyelembevételével történik, a paraméterek tudatos megválasztásának nagy jelentősége van a felhasználási cél szempontjából is. Más hangszín érthető jól különböző lejátszó eszközökön eltérő körülmények között, hasonlóan más beszédsebességet igényel a felolvasott információk érthetősége egy tudományos szöveg és egy szépirodalmi alkotás esetén. Az egyes eszközök eltérő frekvencia átvitele miatt más jellemzőkkel rendelkező hanganyag marad az emberi fül számára jól érthető egy analóg közvetítésű telefonvonalon, egy alacsonyabb sáv szélességet veszteséges tömörítéssel kompenzáló digitális kapcsolaton keresztül továbbított jelfolyam esetén, vagy egy teljes terjedelmű hangállomány stúdió minőségű hang átalakító és hangszórón vagy fejhallgatón keresztüli hallgatása esetén. A jól érthetőség nyelvenként is eltérő beállításokat igényelhet a nyelvenként változó jellegű és gyakoriságú fonémákból kifolyólag. Az orgánumnak, hangszínnek a befogadás szempontjából is van jelentősége, fontos a témához, adott esetben (szak)területhez, vagy épp a szórakoztatáshoz illeszkedő hangszínt választani, a pszichológiai kutatások szerint az ember a kapott információt annak értékelésekor nem képes teljesen különválasztani az információforrástól, amivel kapcsolatban a tudat alatt kialakult benyomásra a hallott hang jellemzői is hatást gyakorolnak (Cialdini, 2009). A sebesség tökéletes megválasztásának ma már olyan szempontból kisebb a jelentősége, hogy a számítási kapacitások növekedésével a kész audio anyag lejátszási sebessége tág határok között változtatható anélkül, hogy a mintavételezett hang jeleit eltérő sebességgel adnánk ki, ezáltal a lejátszott hang magasságának változását is okozva. Fourier transzformációk segítségével a felvételek úgy gyorsíthatók és lassíthatók, hogy a hang szeletek alap- és felharmonikusainak frekvenciája változatlan maradjon, és csak az egymást követő apró részek időtartama változzon (Veldhuis

& He, 1998), ezáltal az eredeti hangszín megőrzése mellett változtatható a lejátszási sebesség (zene esetében az interferenciák miatt nagyobb mértékű sebességváltoztatás esetén ez okozhat azért problémát, de a beszéd tag határok között jól érthető marad). Anyanyelven egy átlagos tempójú, könnyed tartalmú beszéd 1,7-1,8x-os gyorsítás mellett is könnyen befogadható, lassabb eredeti beszédtempó esetén akár 2x-es sebesség felé is növelhető. Nagyobb koncentrációt igénylő mélyebb tartalom, esetleg idegen nyelv esetén ez a tematikus kompetencia és a nyelvtudás függvényében jelentősen kevesebb lehet az információ- és a szöveg követhetőségének elvesztése nélkül. A gyorsított lejátszás azért is kap fontos szerepet, mert az írott anyagokkal szemben a hangnak egyidejűleg csak egy pillanatnyi szeletét halljuk, elveszítjük az áttekintés, pásztázva, “szkennelve” olvasás lehetőségét, a leginkább az online tartalmak olvasásánál kialakult ún. F-minta szerinti feldolgozást (Shrestha, Lenz, & Owens, 2007).

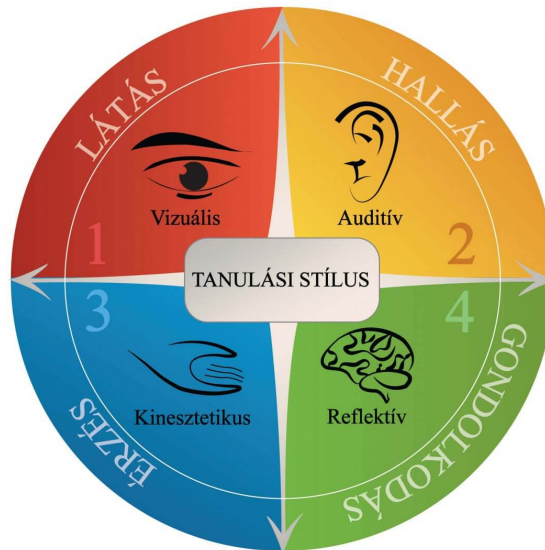
Könyvtár és oktatás

A könyvtárak idejekorán felismerték a hangoskönyvek jelentőségét, gyűjteményükben régóta megtalálhatóak hangoskönyv kiadványok. Ahogyan az ebookok, a hangoskönyvek is felfoghatók az analóg könyvek egyfajta digitális reprezentációjaként, természetesen eltérő célközönséget kiszolgálva. A könyvtárak számára nagy kihívást jelent az olvasótáboruk megtartása és bővítése az átalakuló felhasználói szokások mellett. Nagyon fontos lépést tartani a technológia fejlődésével, újabb és újabb szolgáltatásokat nyújtani és a meglévő gyűjteményt egyre szélesebb módon elérhetővé tenni. Hangoskönyvek segítségével egyrésztől adaptálódni lehet a technológiai kihívásokhoz, másrésztől változatos, a fiatalabb generációk számára vonzóbb új csatornákat lehet megnyitni a könyvtárak felé. Ez lehetőséget teremt számukra az “olvasó” közönség szélesítésére, olyan tábor elérésére és könyvtári kiadványokkal való kapcsolatba kerülésre, akiket a hagyományos módon való olvasáson keresztül nem sikerül. A hangoskönyvek számossága eltörpül az írott könyvek mellett, példaként említem az innováció területén élen járó Csorba Győző Könyvtárat, mely 600 darabos, színészek által felolvasott hangoskönyv gyűjteménnyel rendelkezik. Amíg élvezhető eredményt csak képzett színészekkel stúdió körülmények között felvéve lehet elérni, a legtöbb könyvtártól nem várható el, hogy saját maga állítson elő hangoskönyveket így a gyűjteményünk bővítése érdekében, azonban a beszédsszintézis fejlődésével várhatóan rövid időn belül jó minőségű hangoskönyvek állíthatók majd elő alacsony eszközígeny és költségteher mellett a meglévő írott digitális anyagokból. Természetesen a szolgáltatások bővítése nem merül ki a hangzó anyagok beszerzésében és tárolásában, a versenyképesség

érdekében ki kell alakítani olyan online szolgáltatásokat is, amiken keresztül helyszíni jelenlét nélkül online is elérhetővé válnak a digitális tartalmak. Ezek megvalósításának előfeltétele olyan szoftver és hardver infrastruktúra kialakítása, amely képes a konkurens igényeket nagy számban is kiszolgálni, tárolni és gyorsan elérhetővé tenni a digitális állományokat és a hang alapú tartalmakat a hálózaton keresztül valós időben egyidejűleg sok példányban streamelni a felhasználók számára. Különösen nehéz helyzet elé állítja ez az iskolai könyvtárakat, ahol lényegesen kevesebb az elérhető anyagi forrás és informatikai infrastruktúra.

Az oktatás számára azonban kiemelkedő lehetőséget rejt a hangoskönyv alkalmazása. Egyrészt az e-learning rendszerek a legkönnyebben elérhető platformok a hangoskönyv alkalmazásához az önálló otthoni felkészülés eszközeként, melyek akár kellően felkészített digitális könyvtári szolgáltatásokkal is integrálhatóak a teljesen nyíltan elérhetővé tétel nélkül is (pl. eduID segítségével). Másrészt fontos felhasználási területe a hangoskönyveknek az írott könyvekkel való egyidejű alkalmazása tantermi oktatás során is. A diszlexia sok tanuló számára nehezíti meg az olvasás tanulást, amelynek kompenzálásában segítség lehet a képi és hang információ parallel befogadása, de emellett általános iskolások között végzett felmérések kimutatták, hogy a hangoskönyv és írott könyv egyidejű alkalmazása nagyban megkönnyíti a folyékony és értő olvasás elsajátítását nem csak az olvasási nehézséggel küzdő diákok számára (Nash, 2023).

Hangoskönyvek alkalmazással nagy segítség tud lenni az egyes tanuló típusokhoz való alkalmazkodó tanítás során is, az auditív tanulótípusok (2. ábra) sokkal hatékonyabban tudják elsajátítani a hangban kapott ismeretanyagot, mint az írásos anyagok olvasásán keresztül. A különböző modalitások kombinációjával lefedhető a vizuális és auditív tanulótípusok preferenciái is, ezáltal kiegyensúlyozottabb eredmény érhető el az eltérő készségekkel rendelkező tanulók között. A színvonalasan elkészített hangoskönyvek szerepet játszhatnak a szépirodalom megismertetésében és megszerettetésében, közvetve az alfa generáció olvasáshoz való alacsony affinitásának növelésében is, és bár nem célszerű a hangoskönyvekre az emberi élő szó helyettesítőjeként tekinteni, a gyermekek képzelőerejének fejlődésében a meséléshez hasonlóan jótékony hatással rendelkezhetnek.



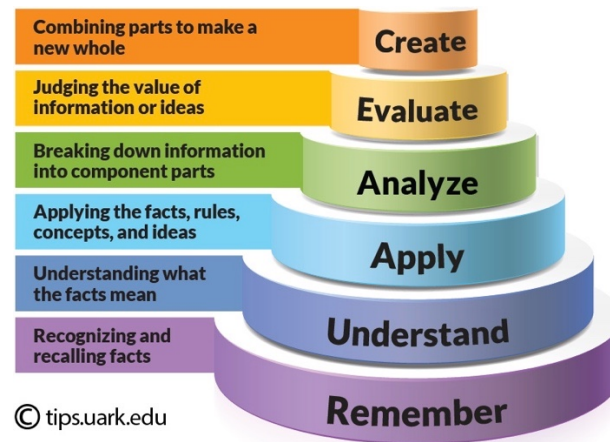
2. ábra – Tanulótípusok

forrás: <https://egyensulyvert.hu/szolgaltatasok/tanulasi-stilus-felmeres>

A hangoskönyvek oktatásban jelenleg elérhető szintén kiemelkedő területe az idegennyelv oktatás. Az angol nyelv mint idegen nyelv oktatása során végzett kutatások igazolták, hogy a nyelv elsajátítása (különös tekintettel a kiejtés tanulására) sokkal hatékonyabb hangoskönyvek, mint kiegészítő eszközök alkalmazása segítségével (Singh & Alexander, 2022). Az órába beépítve egy képzett angol anyanyelvű művész által felolvasott tartalom magától érthető módon sokkal autentikusabb kiejtést tud bemutatni a diákoknak, mint a legjobban felkészült tanár, aki a nyelvet csak tanult nyelvként ismeri, így mindenképp ajánlható a repertoár ezirányú bővítése mind tantermi, mind különórás foglalkozások során. Bár a statikus hangoskönyvek nem biztosítják az élő szituációs gyakorlatokkal megegyező hasznosságot, ugyanakkor – főleg a tanulás kezdeti fázisában – kiegészítő otthoni tanulás és gyakorlás során kiváló segítséget adhatnak az introvertált tanulóknak, akiknek nagyobb nehézséget okoz az előszóbeli kapcsolatteremtés és reagálás idegen nyelven elhangzott mondatokra.

Az oktatás és ismeretterjesztés gyorsabb térhódítást tesz lehetővé a számítógépes beszéd szintézissel előállított hangoskönyvek számára, ugyanis a szépirodalmi műalkotásokkal szemben a szakmai anyagok illetve tankönyvek, vagy akár cikkek, folyóiratok esetén kisebb jelentősége van a felolvasás élményszerűségének, lényegesen nagyobb tolerancia várható el az információközlés, tudásátadás célú felhasználása során. Mivel a hangoskönyvek ilyen módon történő előállítása nagyságrendekkel gyorsabb és költséghatékonyabb, mint emberi felolvasással, sokkal könnyebben és gyorsabban adaptálható az oktatás területén.

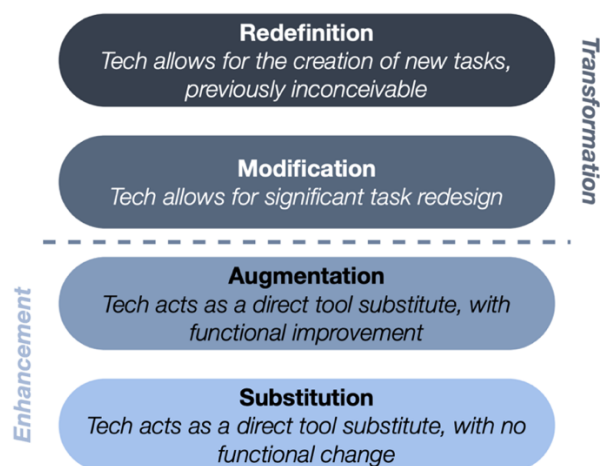
Az hangoskönyvek interaktívvá tétele az oktatásban betöltött szerepét átranzformálja, a módszerekben, modellekben elfoglalt helyét áthelyezi. Bloom taxonómiájának (3. ábra) megfelelően míg az írott könyv olvasás, a hangoskönyv hallgatás leginkább a megértés, addig az tartalommal való interaktív kommunikáció, kérdezés és utasítás megfogalmazás az alkalmazás, analízis szintjeihez mindenképp kapcsolhatók, de értelmezéstől és a használati aktivitás mértékétől függően megvalósulhat a kiértékelés, és szerepet játszhat a létrehozásban is (Huitt, 2011).



3. ábra – Bloom taxonómiája

forrás: <https://tips.uark.edu/using-blooms-taxonomy/>

A nagy nyelvi modellek ezúton bevonása, az oktatás során történő alkalmazása a technológiának a tanulási és tanítási folyamatára gyakorolt hatásában a SAMR-létra befolyásolási szintjei közül az átalakulás (transformation) szakaszon belüli újraértelmezés (redefinition) ér fel. Ez a SAMR (4. ábra) szerinti átalakulás legmagasabb szintje, az újradefiniálás a technológia segítségével teljesen új tanulási lehetőségeket tesz lehetővé, emellett felülről lefelé haladva magában foglalja az alacsonyabb szinteket is. Erős új technológiai soft skilleket ad a tanulóknak, mint a digitális együttműködés, a kommunikáció, a technológiai műveltség, valamint az új rendszerekhez és folyamatokhoz való alkalmazkodás képessége (Puentedura, 2012).



4. ábra – SAMR modell

forrás: <https://blog.zhaw.ch/papierlosesstudium/samr-modell-und-weiterbildung-fuer-lehrpersonen/>

A klasszikus 4 tanulótípus (vizuális, auditív, reflektív, kinezetikus) az írásos válaszok mentén a vizuális, szóbeli válaszok során az auditív tanulókat támogatja kifejezetten, de az interakció során a jól megfogalmazott kérdés feltevés szükségessége kapcsolódik a reflektív típushoz is. A kicsit másképp strukturált R2D2 modell (5. ábra) tanulótípusai (reading, reflecting, displaying, doing) közül a verbális és auditív (reading) tanulótípusoknak kedvez az írásbeliség, az olvasható válaszok, a természetes nyelvi környezet. Ezzel egyidejűleg a gyakorlatias (doing) és reflektív tanulótípusok ott kerülnek előnybe, hogy az információk nem készen, strukturálva kerülnek eléjük, hanem azt ki kell nyerni, a kérdés feltevéshez megfelelő megfogalmazást kell találni, párbeszédbe kell lépni a tartalommal, a nyelvi modellel (Muktha, Mumma, & Willis, 1999).



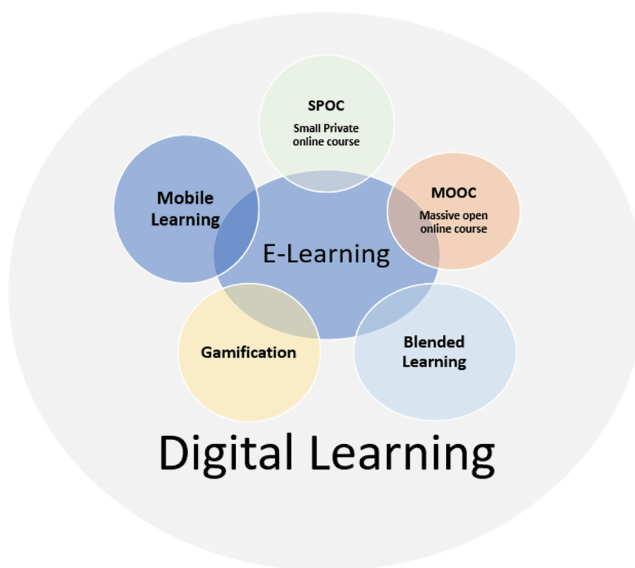
5. ábra – R2D2 modell

forrás: <https://edtech-insights.blogspot.com/2012/05/r2d2-good-bonk-re-think-before.html>

*-learning

Az elektronikus tanulási környezeteknek több olyan elnevezése van, amelynek egymáshoz való kapcsolatáról megoszlik a szakirodalom. Már az elnevezések is megosztóak, míg a *-learning kifejezések alapvetően tanulást jelentenek, a használatuk sokkal inkább tanulási környezetre, rendszerre, illetve platformra irányul. Az egyes fogalmakat szokták egymásba ágyazott halmazokként és egymás mellett jelenlévő, egymással metszeteket képező kategóriákként is ábrázolni (6. ábra). E-learning alatt tágan értelmezve olyan tanulási környezeteket értünk, amelyek valamilyen elektronikai eszközt alkalmaznak (elektronikus eszközzel támogatott tanulás). A napi szóhasználatban azonban általában olyan online rendszereket értünk alatta, amely valamilyen internetre kötött számítástechnikai berendezés (számítógép, tablet, okostelefon) használatával igénybe vehetően biztosít digitális tananyagot, illetve komplexebb esetben komplett kurzus hierarchiákat számonkérési lehetőségekkel együtt. A d-learning a digitális tanulási környezetre utal, mai viszonylatban, amikor szinte minden eszközünk digitális, az e-learning alatt is d-learninget értünk, azonban a szó szoros értelmében tanulást támogató elektronikus eszközök lehetnének analóg eszközök, mint pl. a hagyományos lemezjátszók (amik még nem tartalmaztak digitális elektronikát). M-learning (mobile-learning) alatt a mobil eszközzel támogatott (egyes értelmezések szerint kizárólag mobil eszközön végzett) tanulást értjük. Ez alapvetően az e-learning és a d-learning részhalmazaként értendő, mind tágabb értelemben, a digitális (elektronikus) eszközök közül a hordozható készülékeket értve alatta, mind pedig az e-learning hétköznapiabb jelentését tekintve, amikor az online e-learning rendszerek által nyújtott szolgáltatásokat mobil eszközről (jellemzően okostelefonról) veszik igénybe (Basak, Wotto, & Bélanger, 2018). Az u-learning (ubiquitous-learning) egy absztraktabb fogalom, az embert körülvevő világ kiterjesztése komplex tanulási környezetté, ami a szó konkrét jelentése alapján nem kellene, hogy feltétlenül digitális tanulásra vonatkozzon, beleérthető az “unplugged” világ minden eleme, azonban a kifejezést mégis inkább az embereket minden területen körülvevő digitális elemek tanulásba bevonására használják (Zhang, 2008). A szintén gyakran használt blended learning pedig ismét egy más megközelítésű csoportosítás, alapvetően nem a használt eszköz határozza meg technológiája alapján, hanem (itt is szemben a learning szó szerinti tanulás jelentésével) olyan hibrid oktatási módszert jelöl, amelyben a hagyományos tantermi oktatást kombinálják a digitális eszközök és online elérhető szolgáltatások alkalmazásával (Sharma, 2010). Ezen módszerek és eszközök mindegyikében közös, hogy a hangoskönyvek valamilyen formájának

szerepeltetése kiválóan illeszkedik az eszköztárunkba függően az adatforrástól, médiától, lejátszó eszköztől és a beépítés módjától.



6. ábra – Digitális tanulás formái

forrás: <https://f1000research.com/articles/10-1106>

Mesterséges intelligencia

A mesterséges intelligencia fogalmának értelmezése az idők során jelentős változáson ment keresztül. A legkorábbi és legtágabb meghatár szerint minden számítógéppel megvalósított megoldás mesterséges intelligenciának tekinthető, ami valamilyen tevékenységet végez el, amit emberi intelligenciával lehet megoldani, legyen az bármilyen egyszerű feladat is statikusan előre programozott algoritmus által lefuttatva. Az újabb értelmezések szerint már elvárás valamilyen mértékű komplex megoldás, automatikus adaptálódás, öntanulás megvalósítása. Szűk mesterséges intelligenciának tekintjük azokat az algoritmusokat, amelyek bár az ember által nyújtható teljesítményt elérik, vagy meghaladják, de csak egy jól körülhatárolható területen, feladatkörön, kontextuson belül képesek megvalósítani. Általános mesterséges intelligenciának (AGI - artificial general intelligence) nevezzük azt a teljesen univerzális eszközt, amely az ember által ellátható minden feladatot ki tud váltani (Goertzel, 2014). Ezt jelenleg még nem sikerült elérni, és a kutatók véleménye megoszlik abban, hogy mennyi időn belül válhat elérhetővé, vagy egyáltalán megvalósítható lesz-e valaha. Az intelligencia definíciójában sincs egyértelmű konszenzus, az előre definiált feladat végrehajtásától az önálló kreativitás szükségességéig számos álláspont található. Az intelligencia hányados kapcsán használt definíció szerint az az intelligencia, ami intelligencia teszten mérhető. Az egyik legrugalmasabb megközelítés alapján ami úgy működik, mint ha intelligens lenne, azt intelligensnek kell tekinteni. A

mesterséges intelligencia történetében legkorábbi és legvitatottabb megjelenés Kempelen Farkas 1769-ben bemutatott, "Török" névvel illetett sakkozó automatája, amellyel szemben erős gyanú áll fenn, hogy valójában egy elrejtett ember működtette és nem tartalmazott gépi intelligenciát. A gépi intelligencia mérésében egyik legjelentősebb mérföldkő az Alan Turing által 1950-ben lefektetett, róla elnevezett Turing-teszt, mely során teszt alanyok konzolon keresztül kommunikálnak előlük elrejtett emberi és gépi partnerrel, és az az eszköz teljesíti sikeresen a tesztet, amelyik esetén az alanyok nem ismerik fel a túloldalon válaszokat adó gépet (Turing, 1950). Történetileg kiemelendő a Deep Blue nevű sakkozó program, mely 1997-ben legyőzte Garri Kaszparov világbajnokot, vagy az IBM Watson programja, mely az amerikai Jeopardy vetélkedőt minden humán vetélytársát legyőzve megnyerte 2011-ben.

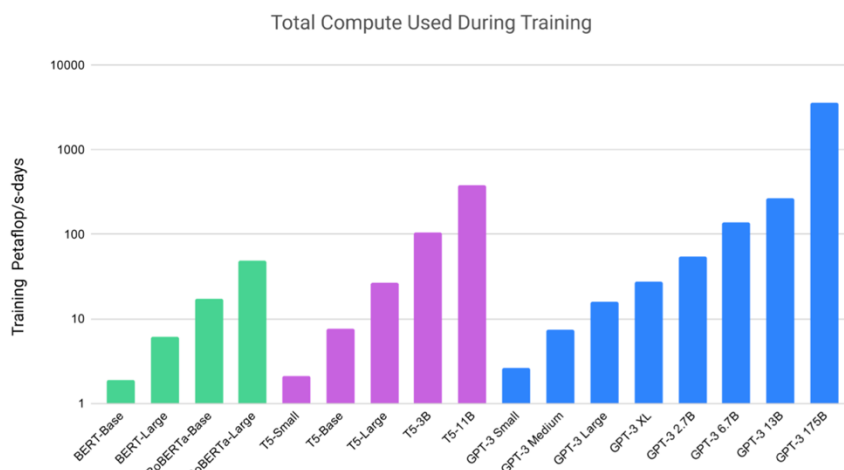
A 2000-es évektől kezdve az MI továbbra is gyors fejlődésen ment keresztül, és olyan területeken is elterjedt, mint a beszédfelismerés, a gépi látás és a gépi tanulás. Az MI ma már számos területen használatos, például az egészségügyben, az autópárházban és a tudományos kutatásokban. A jövőben az MI további fejlődése várható, és új lehetőségeket fog nyitni az emberi tudás és a technológiai fejlődés terén.

A legtöbben nincsenek tudatában annak, mennyire jelen van a mesterséges intelligencia az életünkben. Tekintve csak az olyan egyszerű használati eseteket, mint okos hangszóróval zenehallgatás, a modern okostelefonok kamerájával fényképezés során az automatikus beállítások és utólagos képjavítások, a video- és zenelejátszó streaming szolgáltatók ajánló algoritmusai, a navigáláshoz használt útvonaltervezők, vagy a kéretlen leveleket automatikusan kiszűrő rendszerek. A mesterséges intelligencia mindenütt jelen van. A civil élet mellett azonban ezidáig nagyobb jelentősége volt a mesterséges intelligencia gazdasági felhasználásának. Például a bankok és biztosítók az egyik legnagyobb felhasználói az ügyfél profilok kialakításában, minősítésében, szokások kiismerésében, folyamatok optimalizálásában, de nem utolsósorban számukra is az ügyfeleknek biztosított tájékoztató anyagok közvetítésében. A széles nyilvánosság közvetlen felhasználása számára az openAI 2022-ben közreadott GPT-3 nagy nyelvi modellje hozta át az áttörést, és robbantott ki igazi LLM (large language model) forradalmat a mindennapi életben.

A gépi felolvasással előállított hangoskönyvek és tágabban a beszéd szintézis oktatásban való felhasználásához a mesterséges intelligencia, a nagy nyelvi modellek fejlődése által elérhetővé váló lehetőségek kiaknázása ad igazán kimagasló potenciált. A mesterséges intelligencián belül a mélytanuló algoritmusok és a nagy nyelvi modellek használata már

több évtizedes múltra tekint vissza, a szakértők jelentős eredményeket értek el eddig is nemcsak klasszifikáció és szövegelemzés, de a szöveg generálás területén is, azonban a széles nyilvánosság tudatába nagyjából egy évvel ezelőtt, az OpenAI amerikai vállalat chatGPT szolgáltatásának nyilvános megnyitását követően került be, amittől kezdve népszerűsége és széleskörű használata ugrásszerű növekedésbe kezdett, minden eddigi rekordot megdöntve már két hónap alatt 100 millió felhasználót tudott magáénak (Hu, 2023).

Az openAI által aktuálisan használt legújabb modell a GPT-4 (a GPT-3 és GPT-3.5 életben tartásával egyidejűleg). Ez a modell a feladatok széles skáláját képes megoldani, például beszélgetést, szövegkiegészítést és még a kódolást is megdöbbentően jó teljesítménnyel (Kublik & Saboo, 2022). A modell elődje (GPT-2) nem csak nyílt forráskódú, de a betanított modell is ingyenesen használható. Fontos megjegyezni, hogy a neurális hálózatra alapuló (nem csak nyelvi) modellek esetében a legnagyobb érték nem a programkód, az algoritmus, hanem a hatalmas befektetett erőforrás (7. ábra) a modell betanítására (az öntanuló algoritmus súlyainak, paramétereinek kialakítása, hangolása).



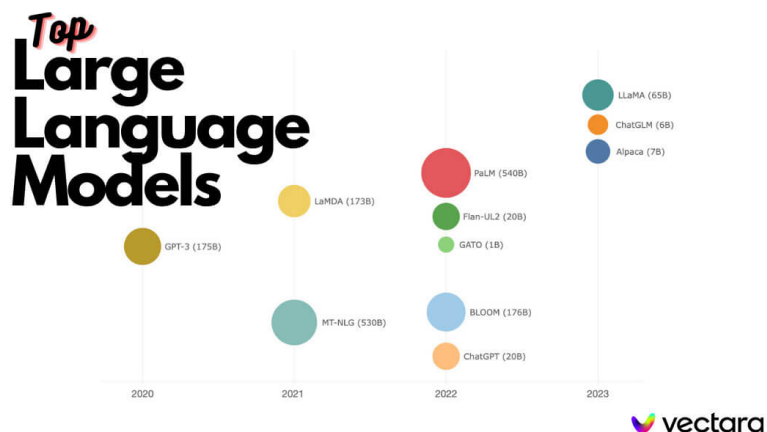
7. ábra – Nyelvi modell tanítás számítási kapacitás igénye

forrás: <https://sh-tsang.medium.com/review-gpt-3-language-models-are-few-shot-learners-ff3e63da944d>

A betanított GPT modellt azt openAI jelenleg korlátozott ingyenes felhasználáson túl token (a modell által felbontott szöveges egységek, általában a szónál kisebb darabok) alapú, valamint a chatGPT formájában átalány jellegű díjazás mellett teszi elérhetővé.

A chatgpt mellett számos más nagy cég hasonló szolgáltatása is elérhetővé vált azóta a piacon, valamint rengeteg kisebb szervezet, csoportosulás tett letölthetővé nyílt forráskódú változatokat is (8. ábra). Napi-heti szinten hozzák ki a nagy gyártók és a kisebb kutatócsoportok az újabb és újabb megoldásaikat, modelljeiket. Számos elérhető modell közül vannak felhő szolgáltatásban elérhetőek és letölthető modellekkel és súlyokkal on-premise telepíthető és üzemeltethető (Sun, Zhang, Chen, Zhang, & Liang, 2007), nyílt

forráskódú változatok, melyek paraméterszáma és tanítókörpuszuk mérete 2-3 nagyságrenddel is eltérhet egymástól. Jelenleg a legnagyobb figyelmet az OpenAI ChatGPT szolgáltatása kapja, mely megjelenése után már két hónappal minden rekordot megdöntve 100 millió felhasználóval rendelkezett (Hu, 2023).



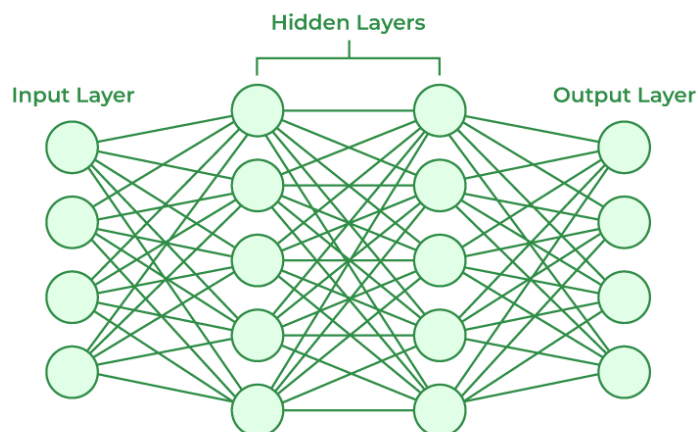
8. ábra – Nagy nyelvi modellek

forrás: <https://vectara.com/top-large-language-models-llms-gpt-4-llama-gato-bloom-and-when-to-choose-one-over-the-other/>

Ezek az előtanított generatív transzformer (GPT - generative pretrained transformer) modellek milliárdos nagyságrendű paraméterrel rendelkeznek és akár dollármilliárdos költségek árán óriási méretű tanító halmazokon lettek trainelve, aminek eredményeként a neurális hálózatok tanítás által behangolt súlyaik alapján alapvetően valószínűségi számítási módszerekkel külön nyelvtani szabályrendszer implementálása nélkül elsajátították a természetes emberi nyelv mondataalkotásának képességét. A modellek a szöveget token egységekben kezelik, melyek megfeleltetése nyelvenként eltérő, angol esetén közelít a szavakkal való megfeleltetéshez, magyar nyelv esetén egy szót általában több token reprezentál.

A neurális hálózatok, vagy mesterséges neuronhálók alapvetően az emberi agy működését hivatottak utánozni, ahol az agyban lévő neuronokat kis funkcionális programegységek reprezentálják, amik között a szinapszisokat súlyozott összeköttetéseknek feleltetik meg. Ezek a neuronok rétegekbe vannak csoportosítva, és a gyakorlati megvalósításban a bemeneti és kimeneti rétegek között nagy számosságú köztes rejtett réteg van (9. ábra), a szomszédos rétegekben elhelyezkedő neuronok közötti összeköttetések súlyai pedig a modell paraméterei. A hálózatokat tanító nagy mennyiségű, tanító és ellenőrző adatsoporra osztott adathalmazokkal tanítják, ami leegyszerűsítve annyit jelent, hogy iterálva újra és újra kiértékelik a modellt az aktuális paraméterekkel a kiértékelések között

módosítva a paramétereket, amíg a modell által adott eredmény pontossága el nem ér egy előre definiált mértéket.



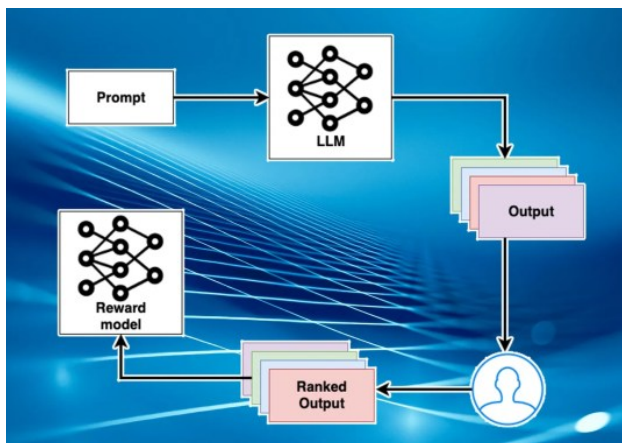
9. ábra – Neurális hálózat

forrás: <https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/>

A mesterséges intelligencia természetesen nem érti a szó szoros értelmében az általa feldolgozott és előállított szöveg értelmét, jelentését, azonban a nagyszámú mintából tanulva nagyon jó hatékonysággal utánozza le az emberek által előállított kifejezés sorozatokat és a figyelem (attention) funkcionalitásnak köszönhetően kontextus specifikus összefüggő szöveget tud alkotni utasításokra, kérdésekre adott válaszok formájában is (Xu, Liang, Huang, & Xiang, 2021). Megfelelő prompting megoldások alkalmazásával tág határok között befolyásolható a generált válasz iránya, stílusa, terjedelme, részletessége. Lehetővé teszi rövid instrukciókkal megfogalmazott téma vagy tématerület köré épített szöveg előállítását, meglévő szöveg tömörítését, összefoglalását, kulcsszavak kiemelését, entitások azonosítását és konkrét kérdésekre egzakt válaszok megfogalmazását.

Elsősorban a fejlesztők számára elérhető (alapvetően az internet aratásával begyűjthető és a nyilvános digitális könyvtári állományok körében fellelhető) betanító adatok összetételében a nagy nyelvek, főleg az angol nyelv lényegesen felülreprezentált, ezért a modellek angol nyelven lényegesen jobb teljesítményt mutatnak, mint a kisebb, ritkább nyelvek, például magyar esetén. Ennek ellenére a legtöbb élő (de esetenként holt) nyelv között nagyon jó eredménnyel képesek automatikusan fordításokat is végezni, valamit adott szöveg nyelvét felismerni. A generált vagy fordított tartalmak esetenként még lektorálást igényelnek, de előállnak teljesen készen felhasználható, átlagember beszédképességét hibátlanul reprodukáló (egyes implementációkban akár konkrét személy kommunikációs stílusát felvevő) kimenetek is, továbbá meglévő szöveg egyszerű hibáinak javítására is alkalmasak tudnak lenni. A modellek kiemelkedően jó teljesítményét a mennyiségi

tanuláson felül a kiértékelések során a modellekbe épített jutalom (reward) függvényeknek emberi visszajelzések (RLHF) alapján történő finomhangolása biztosítja (10. ábra).

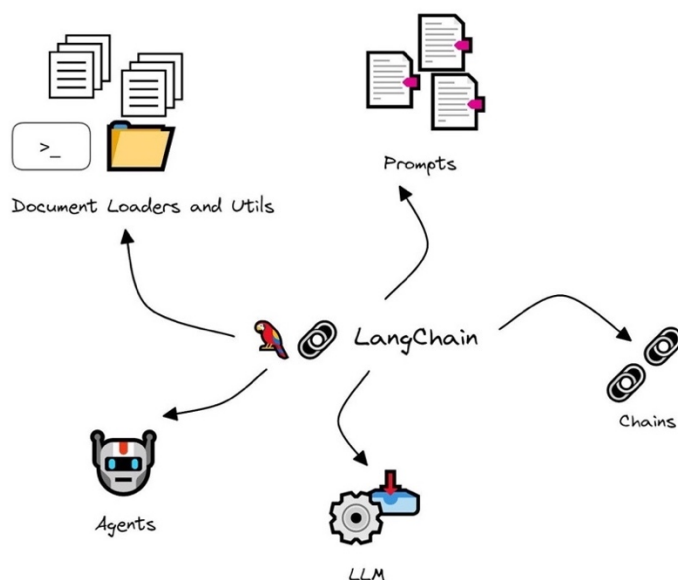


10. ábra – Reinforcement Learning by Human Feedback

forrás: <https://bdtechtalks.com/2023/01/16/what-is-rlhf/>

A GPT modellek három legnagyobb és sokat kritizált korlátja az időben korlátozott lexikális tudás, a specifikus tudásbázis kezelésének hiánya, valamint a hallucináció. A korlátozott tudás oka, hogy (a modell kontrollálhatatlan “elfajulásához” vezető korábbi negatív kimenetek tapasztalataiból következően (Davis, 2016)) az előtanításhoz használt adathalmaz előre összeállított, statikus, nem bővülő információtömeg, valamint a használat közbeni interakciók nincsenek automatikus visszacsatolásra és öntanulásra felhasználva. A specifikus tudásbázis kezelésének hiánya azt jelenti, hogy a nyelvi modell alapképessége szerint az előtanított adatforrásokból dolgozik, valamint a nagy számítási kapacitás igény miatt az interakció alapját adó, információt befogadó képességét biztosító (a “tudását” visszaható tanítás nélkül befolyásoló) kontextus-ablaka erősen korlátozott méretű. A modellek hallucinációja alatt azt a jelenséget értjük, amikor a modell nem ismeri fel a kontextusban szereplő témához tartozó ismeret hiányát, és a nyelvi forma betartásával látszólag értelmes válaszokban valótlan információ tartalmat állít elő illetve kombinál. Ezeket a problémákat a modellek kiterjesztésével folyamatosan korrigálják, különböző kapcsolódási pontokat, interfészeket alakítanak ki, amelyek segítségével külső információ forrásokhoz, adott esetben a publikus internet tartalmához is hozzáfér a modell, illetve a gyakorlati célzott alkalmazások során olyan hibrid rendszerekkel építik körbe, melyek több lépésben és nagyobb egységek kisebbekre bontásával lehetővé teszik a külső tudásbázisok feldolgozását, valamint a generált eredmények tartalmi validációját. Többségében olyan komplex architektúrák építését jelenti, ami interfész hívásokkal egészíti ki a modellt és többlépéses feldolgozást biztosít. Így ma már lehetőség van a modellek élő online tartalommal való összekapcsolására, dokumentum elemzésre, egyedi komplex rendszerek

építésére. A kész szolgáltatások által még nem megoldott problémák áthidalására több lehetőség kínálkozik. A finomhangolás során prompt-completion párokat “tanítanak rá” a modellre, melyek hatására a mély neurális hálózat felső rétegei között szereplő paraméterek módosulnak, ezáltal hatást gyakorolva a generált válaszok adattartalmára és jellegére. A hibrid megoldásokkal kombinálhatók felhő szolgáltatásban elérhető nagyobb modellek kisebb specifikus on-premise modellekkel, valamint több lépésben elő- és utófeldolgozás, ezek segítségével prompt újrafogalmazás, többszörösen előállított és abból kiválasztott válaszképzés, válasz validálás valószínűsíthető meg, nagyobb adatmennyiségek darabolást követő iterációkkal vagy hierarchiába szervezett vektorizálással, kontextus tömörítéssel, részenkénti hasonlóságkereséssel kezelhetők. Ilyen hibrid megoldások körében elterjedt a LangChain (11. ábra) architektúrája (Topsakal & Akinci, 2023). Emellett folyamatosan bővítik a modellek mögöttes tudása és a felhasználók közé olyan mesterséges korlátok beépítését, melyek megakadályozzák a rossz szándékú felhasználást, diszkriminatív tartalom előállítását, vagy adott esetben bűncselekményben való segítségnyújtást. A legnagyobb nyelvi modellek működtetése nagyon költséges feladat, és bár általánosságban a modell jósága arányos a paramétereinek számával és a tanító adathalmaz méretével, sok esetben adott célra hatékonyabb megoldások érhetők el specifikus kisebb modellek és azok jól tervezett kombinációinak több lépéses alkalmazásával.



11. ábra – LangChain

forrás: <https://coinsbench.com/chat-with-your-databases-using-langchain-bb7d31ed2e76>

A generatív modellek így a hangoskönyvek felhasználása területén is lehetővé teszik, hogy hangként ne csak az eredeti mű változatlan formája képződjön le, hanem a beszéd szintézis alapját képező írott forrás anyagból más szempontok szerint előállított tartalom

keletkezzen. Ennek segítségével megvalósítható az írott forma mellett felolvasva a tartalom automatikus kivonatolása, összegzése, átfogalmazása, egyes részleteinek magyaráztatása, amely adaptálódni tud a hallgató kompetenciájához, képességeihez is. Természetes nyelv feldolgozás és értés (NLP – natural language processing, NLU – natural language understanding) módszerei és a generatív modellek segítségével forrás anyagból előállított származtatott információk felhasználása számos lehetőségeket nyit meg az oktatási célú felhasználás előtt. Ezek a tanulás, tanítás, vagy tágabban az oktatás támogatás különböző területén és fázisában tudnak segítséget nyújtani akár a tervezés, felkészülés, a személyes óra tartás, az otthoni tanulás, vagy a számonkérés területén. Példaként említhető a vázlat generálás, a multimodális automatikus illusztrálás, tananyag gyártás, a feladatsor előállítás, a teszt vagy beszámoló kiértékelés lehetősége.

A mesterséges intelligencia hangoskönyvek esetén alkalmazható másik irányú felhasználási területét képviseli az az irány, amikor a feldolgozás nem az írott szöveg irányából halad a beszédhang előállítása felé (TTS – text-to-speech), hanem a verbális kommunikációt transzformálják írott formára (STT – speech-to-text). Ebben az esetben a hangon alkalmazott transzkripcióval az összefüggő szövegből vagy egy-egy kisebb szövegelemből leirat készül, és a további feldolgozás az írott szöveges formán történik. Sok esetben a leiratozáshoz nem szükséges a szöveg nyelvét sem megadni, néhány szavas mondatból nagy pontossággal sikerül automatikusan megállapítani a nyelvet. Az automatikus transzkriptálás nehézségei nagyban hasonlítanak a beszéd-szintézis problémáihoz, a hibák egy része a ritkább nyelvekből, a speciális kifejezésekből, azonosítókból, rövidítésekből, tulajdonnevekből ered, melyeket a beszéd-előállításhoz hasonlóan finomhangolással, egyedi szótárakkal lehet csökkenteni. A hibák másik forrása az alacsony hangminőség, a magas alapzaj, illetve a több beszélő egyidejű jelenléte. A háttérzaj kiszűrésére már nagyon jó hardveres és szoftveres megoldások léteznek, a több beszélő megkülönböztetését (és akár beszélgetések szereplőnkénti leiratozását) diarizációs megoldásokkal lehet biztosítani. A hang-szöveg konverzió közel valós időben is elérhető, pontosságukhoz, a “félrehallás” korrekciójához nagyban hozzájárulnak a nyelvi modellek képességei, de további feldolgozás esetén (pl. szóban feltett kérdésre adott válasz) a nyelvi modellek jól tolerálják a kisebb bemeneti hibákat (lemaradó ragok, szavak, szótévesztések), így ha a leirat nem közvetlen végtermék (pl. diktálás), akkor a processzási szekvencia jól kompenzálja a hiányosságokat. Ilyen felhasználási esetek a hangoskönyv vagy videó lejátszás közbeni élő feliratozása másik nyelvre fordítva, beszéd-szintézissel kombinálva kvázi szinkrontolmácsolása is. Ezek a megoldások lehetőséget biztosítanak akár az idegen

nyelv elsajátítás támogatására, akár anyanyelven nem elérhető ismeretanyagok, oktatási adattartalmak megelőző hosszas emberi munka nélküli feldolgozására az oktatásban, vagy azon kívül is.

Interakció

Egy további lépés a felhasználási lehetőségek terén a tartalom interaktívvá tétele. A nagy nyelvi modellek beépítésével kialakított megoldásokban lehetőség nyílik a betöltött forrásanyag, könyv tartalom részeinek felolvasásán vagy összefoglalásán túl konkrét kérdések megválaszoltatására is, ami azon túl, hogy a könnyen kereshetőség hiányát is pótolja, olyan kapcsolatba lépést tesz lehetővé egy művel, mint ha azzal, vagy azon keresztül szinte a szerzőjével folytatnák írásban vagy szóban párbeszédet. Kérhetünk újrafogalmazást, magyaráztatást, forrás hivatkozást egy állításról. Fokozva ezt lehetővé válik a számonkérés lebonyolítása is a mesterséges intelligencia segítségével, akár gépi kérdés összeállítással emberi kiértékelés mellett, akár ember által előkészített feladatokra adott válaszok gépi kiértékelésével, akár a kettő kombinációjával. Az interaktív szóbeli kommunikáció egyik gyenge pontja jelenleg a válaszdíő kérdése, különösen komplexebb folyamatokba ágyazva nehezen megvalósítható az emberi kommunikációnál jelenlévő valósidejűség. Ennek oka az egyenként is számításigényes lépésenkénti átfutási idők összeadódása, amin a komplexitás és lépésszám csökkentése, a helyben biztosított vagy hálózaton keresztül igénybe vett hardver infrastruktúra felskálázása és az interakciók elemeinek terjedelmi korlátozása mellett az jelenthet segítséget, ha a folyamatlépéseket egymással átfedésbe tudjuk hozni, a transzkriptálást és kérdésértelmezést, valamint a válasz generálást, hang előállítását és lejátszást folyamatos streameléssel azelőtt elkezdjük végrehajtani a részeredmények alapján, hogy az előző lépés teljeskörűen befejeződött volna. A gyakorlati hasznosságon túl ez szerepet játszhat a befogadó fél, konkrétan a tanuló és a forrás, tananyag között kialakuló kapcsolat minőségében. Az ezekre a technológiákra épülő megoldás akár önálló alkalmazásként, akár egy integrált e-learning rendszer részeként a tanítási folyamatban asszisztensként segítheti a tanár munkáját, vagy blended learningként bevonható külső szereplőként az óramenetbe, de az otthoni felkészülés vagy gyakorlás eszköze is lehet. A kapcsolatot biztosító digitális közvetítő lehet (akár kivetítővel vagy digitális táblával kiegészített) asztali vagy hordozható számítógép, tablet, vagy okos telefon, ami abban is segítség lehet, hogy a mai éra mobilfüggő fiatal generáció számára szerethetőbbé tegye és ezen keresztül közelebb hozza a tanulást. Felmerül a lehetősége az okos hangszórók (pl. Amazon Alexa, Google Nest smart speakerek) integrációjának, azonban a korlátozott nyelvi

képességek (magyar nyelv támogatásának hiánya) és a csekély szabadságot biztosító platform (saját rendszeren belül megvalósított beszéd- és szándékfelismerés, merev folyamatkezelés, limitált beavatkozási lehetőség) nem teszi ideálissá a smart speakereket az ehhez hasonló megoldások kialakítására, de ez az Egyesült Államokban tapasztalható nagy népszerűségükkel szembeni alacsony hazai penetrációjuk miatt nem okoz érdemi hiányt.

Jogi vonatkozás

A mesterséges intelligencia egyre szélesebb körű elterjedése magával hozza a jogi szabályozás szükségességét, számos eddig ismeretlen kérdést vet fel például a felelősség és az etikai területén. Fontos szempontok a szerzői és személyiségi jogok, amellyel kapcsolatban különösen a nagy nyelvi modellek tanítására felhasznált, valamint a nagy nyelvi modellek által előállított adatok vonatkozásában számos tisztázandó kérdés van. Jelenleg a terület még nagyon gyerekcipőben jár, a szabályok lefektetése az elkövetkező időszak feladata.

Egyfelől – mint minden más technológiai fejlődés – csak úgy lehet integrálható, ha a társadalom javát szolgálja, ha az ember képességeit javítja. Másfelől komoly kockázatokat is rejt magában. A mesterséges intelligencia mint technológia szabályozása – más technológiák szabályozásához hasonlóan – azzal a kötelezettséggel jár, hogy jólétnövelő hatások a kockázatok minimálisra szorításával álljanak elő. A korszakalkotó technológia társadalom javát szolgáló elterjedése nemcsak szabályozási eszközöket igényel (pl. befektetés és kutatás-támogatás, szakemberképzés), de a kockázatok kiküszöbölésénél, a megbízhatóság erősítése kapcsán megkerülhetetlen (Tóth A. , 2019).

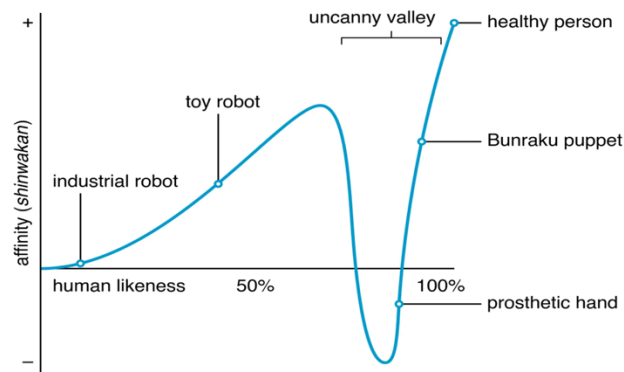
Az Európai Bizottság 2022-ben – most első alkalommal – javasolja mesterséges intelligenciára vonatkozó nemzeti felelősségi szabályok célzott összehangolását. Az egységes szabályozás lehetővé tenné, hogy a mesterséges intelligenciát használó termékek vagy szolgáltatások miatt bekövetkező károk károsultjai kártérítést kapjanak (Európai Bizottság, 2022).

Az Európai Tanács következtetése, hogy az Uniónak sürgősen foglalkoznia kell a felmerülő trendekkel, ideértve az olyan kérdéseket, mint a mesterséges intelligencia (MI) és az elosztott nyilvántartási technológiák (pl. blokklánc), biztosítva ugyanakkor az adatvédelem magas szintjét – az (EU) 2016/679 rendelet maradéktalan tiszteletben tartásával –, továbbá a digitális jogok, az alapvető jogok és az etikai normák magas szintjét (EUR-Lex, 2021).

Hangoskönyvek és ember-gép kommunikáció jövője

A távolabbi jövőben elképzelhető hogy minden minket körülvevő eszköz rendelkezik egy olyan ember-gép interfésszel, aminek segítségével beszélgethetünk vele. A nyelvi kommunikáció és az ember-gép kapcsolat továbbfejlesztésének egyik iránya, hogy ezek az eszközök ne csak neutrálisan tudjanak beszélni az emberrel annak saját nyelvén (mi több, anyanyelvén), hanem egyre emberibb jellemzőkkel rendelkeznek. Az egyik ilyen törekvés az antropomorf mesterséges intelligencia megvalósítása, azaz olyan digitális személyiségek megvalósítása, melyek emberi, vagy legalábbis emberszerű személyiségjellemzőkkel rendelkeznek, van céljuk, van hangulatuk, van stílusuk. Ezáltal lehetővé válik, hogy akár személyre szabottan más és más módon kommunikáljon a gép az emberrel, például a tanításban képes legyen a mesterséges intelligencia egy feladatot egy harmadikos diák szintjén érthetően elmagyarázni, ugyanakkor egy másik kontextusban akár egyetemi szintű szakmai diskurzust folytasson egy szakértővel egy adott témában. Másik irányból megközelítve egy ilyen digitális személyiség lehet egy szakértő, lehet egy értékesítő, akinek kifejezett célja a fogyasztót vásárlásra ösztönözni, de lehet akár egy beszélgetőtárs, egy kedves barát is. Rosszabb esetben természetesen itt is elképzelhető negatív felhasználási mód. ha ez a személyiség agresszív, befolyásoló és olyasmire próbálja manipulálni a felhasználót amit ő nem szeretné, különös tekintettel az illegális tevékenységekre.

Érdekes megemlíteni azt az ellentmondást, hogy míg fejlesztők törekednek a leginkább emberszerű mesterséges intelligencia, az AGI (általános mesterséges intelligencia) megvalósítására, addig az ember-gép kapcsolatban felfedezhető ezzel kapcsolatban egy ellentmondás, miszerint az emberek sokkal elfogadóbbak egy "gépszerű" géppel, egy csetlő-botló, esendő, hibákkal küzdő eszközzel, szívesebben alkalmazkodnak a hibáihoz, tolerálják a hiányosságait, azonban ha egy szintnél jobban megközelíti az emberi jellemvonásokat, akkor ez már ellenérzést vált ki belőlük és kontraproduktívvá válik (uncanny valley, ld. 12. ábra).



12. ábra – Uncanny valley effektus

forrás: Mori, M., MacDorman, K., & Kageki, N. (2012). *The Uncanny Valley [From the Field]*. *IEEE Robotics & Automation Magazine*, 98-100

A gépi beszéd előállítás valós idejű változata lehetővé teszi asszisztensek létrehozását, azonban általában csak olyan megoldás tud teljesértékű lenni ahol a digitális asszisztens képes felismerni a saját határait és ebben az esetben a párbeszéd vagy ügyintézés folytatható humán kapcsolattartóval is. Érdekes jelenség, hogy míg egyrésről nagy segítség, költségcsökkentés és folyamatos elérés biztosítása válik lehetővé gépi ügyfélszolgálatokkal, úgy egyre inkább értékelődik fel a humán ügyfélszolgálat jelentősége. Az ügyfelek szeretnék magukat fontosnak érezni, melynek részeként elvárják, hogy ne egy gép beszéljen velük, hanem egy hús-vér ember. Ez azonban lehetőséget teremt más hibrid megoldásokra is, lehetséges úgynevezett szellemi superhuman létrehozása kiborgok előállítása nélkül is. Kanadai (Agents Republic Inc.) és magyar (Mynds.ai Kft.) kooperációval épp egy olyan megoldás prototípusán dolgoznak a fejlesztők, amelynek segítségével egy humán ügyfélszolgálatos munkatárs megőrzi az emberi kapcsolattartást az ügyféllel, képes az emberi előnyöket megtartani, például megnyugtatni az ügyfelet, de a valós időben működő mesterséges intelligencia közvetlenül a keze alá dolgozik és az ügyféllel folytatott telefonbeszélgetést menet közben figyelve az ügyfél kérdéseire segít előkeresni a választ és előkészíteni az ügyfélszolgálatos munkatárs számára. Ez lehetőséget teremt az ügyfél élmény növelésére a kommunikáció emberi tökéletességével és az információkeresés gyorsaságával és pontosságával, teljeskörűségével kombinálva, valamint a mesterséges intelligencia esetleges botladozásait korrigálva az emberi készségekkel. Hasonló hibrid megoldás nagy segítség lehet a tanároknak első sorban online, de megfelelő előkészületek esetén akár tantermi oktatás során is.

A válaszok, tartalmak generálása azonban ismét újabb problémákat vet fel. A nagy nyelvi modellek teljesítményének javulása rohamtempóban nyitotta meg a teret az emberi erő helyett géppel generált szöveges tartalmak előállítására. Ezek szinte minden területen

megjelentek, beleértve a tömegkommunikációt, az újságírást, illetve az oktatást is különböző dolgozatok előállítása során. Ahogyan fejlődnek a nyelvi modellek, úgy válik egyre nehezebbé felismerni egy szövegről, hogy azt számítógép írta, vagy ember. Vannak olyan szövegek, amiről emberi szemmel nem, vagy legalábbis nem könnyen felismerhető a gépi forrás. Készülnek olyan alkalmazások, aminek célja ennek felismerése. Jelenleg az ad erre lehetőséget, hogy a nagy nyelvi modellek statisztikai alapon működnek, a szöveg generálás során minden lépésben a betanító adathalmazon való tanulás által kialakított súlyok szerint a hálózat kiértékelésekor a mindenkor legvalószínűbb következő szót választja. Ezzel szemben az emberi alkotás során nem ennyire determinisztikus a működés. A felismerő algoritmusok azt próbálják meg azonosítani, hogy az adott szövegre igaz-e, hogy a neurális hálózat szerinti legvalószínűbb szavak követik egymást. Azonban ahogy a hackerek és az infrastruktúra biztonsági szakemberek között is örök verseny van és a hackerek mindig egy lépéssel előrébb járnak, itt is elmondható, hogy a nyelvi modellek szöveg generálása mindig az azt detektáló algoritmusok előtt jár egy lépéssel és a verseny nem áll meg.

Magyarország jelenleg elég erős lemaradásban van a nemzetközi mesterséges intelligencia kutatásban, ugyanakkor komoly törekvések tapasztalhatók a felzárkózásra, például a PULI nevű magyar nagy nyelvi modell fejlesztése (Tóth, Oszkó, & Zijian, 2023). Ennek gazdasági és politikai motivációi egyaránt láthatók, és az oktatás is egyre nagyobb hangsúlyt fektet a területre. Ha számosságban messze is vagyunk a tudományterületen belüli kiemelkedéstől, a hazai fejlesztők között is szép számmal találhatók olyan szereplők, kisebb-nagyobb csapatok, akik időt és energiát nem kímélve dolgoznak a kutatási eredmények alkalmazásán és a technológia továbbfejlesztésén.

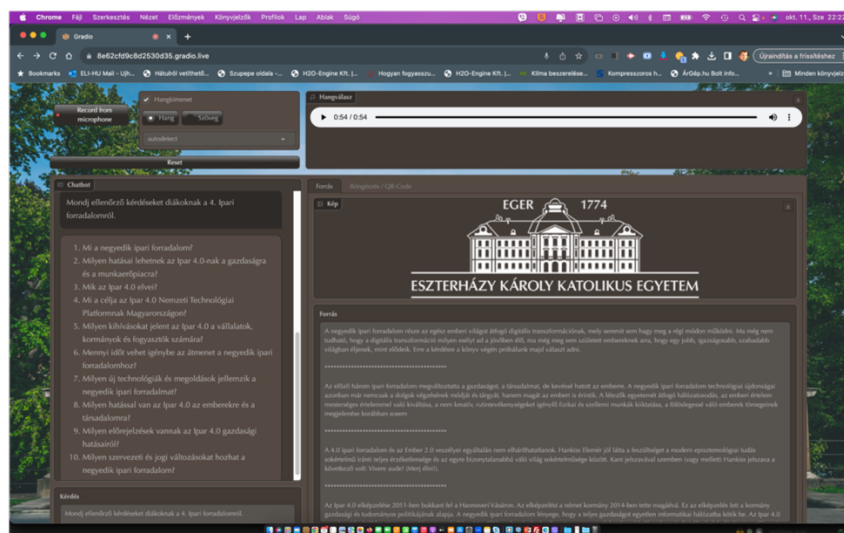
Kutatás bemutatása

A kutatásom során megismert információk felhasználásával meg terveztem mutatni, hogy a tágabban értelmezett hangoskönyveknek a mesterséges intelligencia felhasználásával olyan új lehetőséget nyitnak meg az oktatásban, melyek egyaránt segítséget nyújtanak a tanároknak és a diákoknak a könyvekben szereplő információ befogadása és értelmezése során. A kutatás során adaptáltam a Mynds.ai Kft. kutató-fejlesztő kollégáival készített proof-of-concept alkalmazást, mely lehetővé tette a komplett könyvtartalmakra vonatkozó szóbeli és írásbeli kommunikációt. Az alkalmazásban két jelentősen eltérő tematikájú és stílusú könyv, Csepeli György: Ember 2.0 (Csepeli, 2020) és Simon Harris, James Ross: Kezdőkönyv az algoritmusokról (Harris & Ross, 2006) című művének tartalma került feldolgozásra (13. ábra).



13. ábra – Adatforrásként felhasznált könyvek

A hálózaton keresztül böngészőből futtatható vékonyklienses megoldás felhasználói felületén a számítógép vagy mobiltelefon mikrofonján keresztül diktált vagy leírt kérdéseket tehetünk fel, melyre a könyvek tartalma alapján kiírt és felolvasott választ kapunk. Az alkalmazás angolul, magyarul és németül automatikusan felismeri a kérdező nyelvét és azon válaszol, miközben felületén az információ forrását is feltünteti. Az volt az elvárás, hogy a program megtalálja a kért információt, helyes választ adjon, és ismerje fel, ha olyan információt kérdezzünk tőle, ami nem szerepel a könyvekben. A működést biztosító hibrid szoftver architektúra gradio prototipizáló keretrendszerben készült python nyelvben, Azure felhő szolgáltatáson keresztül TTS és STT szolgáltatásokat, valamint az OpenAI chatGPT interfészét veszi igénybe kiegészítve linux környezetben futtatott LangChain, Weaviate komponensekkel és chromadb adatbázissal (14. ábra). A működéséről egy vágatlan videofelvétel megtekinthető a következő url-en: <https://youtu.be/Y72JsRD4cQA>.



14. ábra – Proof-of-concept alkalmazás

Kutatási eredmény

A törekvés sikeres volt, az elkészítés során nagy számban próbáltakon kívül százas nagyságrendben elvégzett angol és magyar nyelvű tesztek során az esetek többségében az elvárt módon helyes választ adott, megfelelően találta meg a rendelkezésre álló információkat, közölte a hiányzó adatok forrásokban meg nem találhatóságát, és releváns ellenőrző kérdéseket adott a kívánt témában. A hanglejtés magyar nyelv esetén kissé természetellenes volt, valamint a válaszdíők hosszabbak voltak, mint egy élő szereplős párbeszéd esetén megszokott. A néhány sikertelen tesztet többnyire az okozta, hogy az artikuláció hiányossága vagy a beszéd-szünet váltások pontatlan érzékelése miatt a transzkripció során elhagyott egy szót vagy szórészt, vagy tévesen ismerte fel a kérdés nyelvét. A chatGPT nemdeterminisztikus működése nehezíti a módszeres és következetes kiértékelést, de az arányok és a szubjektív tapasztalatok alapján ki merem jelenteni a gyakorlati alkalmazásra való alkalmasságot. A magyar nyelvű leiratozás sokszor nélkülözte a mondatvégi írásjeleket, valamint számos alkalommal hibázott szóvégi toldalékok esetében, azonban a nyelvi modell a kérdések értelmezésekor jól kompenzálta ezeket a hiányosságokat és tolerálta az angolul feltett kérdésekben időnként elkövetett nyelvtani hibákat. Következtetésem szerint a nyelvfelismerés nagyon érzékeny a kiejtésre és a háttérzajra, ez kompenzálható manuális nyelv kiválasztási lehetőséggel. A szünet érzékelési és transzkripciósi hibák jelentősen megváltoztathatják a kérdés eredeti szándékát, ez szükség esetén teljes mértékben kizárható írásos kérdés bevitellel.

A kutatás eredményének mérése során a rendszer szóban és írásban feltett kérdésekre adott válaszait vizsgáltam. Legmagasabb prioritás volt, hogy a forrás könyvekben

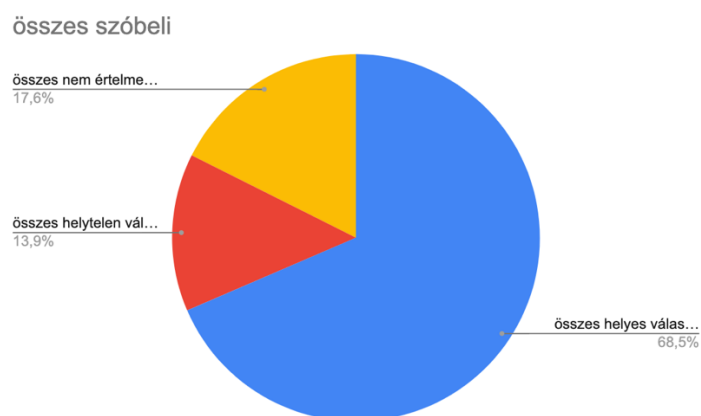
megtalálható információt megtalálja, arra kérdezve valóban olyan választ adjon, ami a könyv tartalmán alapul. Második prioritásom, hogy a könyvek határain belülre korlátozzam a modelt, azaz a könyvekben nem szereplő ismeretre kérdezve ismerje fel és jelezze, hogy az adott témáról nem rendelkezik forrással. Harmadik prioritás volt a kiválasztott témára olyan nyitott vagy eldöntendő ellenőrző, gyakorló kérdéseket tudjon feltenni, amelyekre a választ a könyv tartalmazza. Negyedik prioritás volt a magyar nyelvű könyvekről idegen nyelven feltett kérdések megértése és azokra a kérdésfeltevés nyelvén válaszok adása, ezeket néhány angol illetve német kérdéssel teszteltem. A tesztelés menete során járulékosan ellenőriztem, hogy a kontextus- és nyelv váltások nem gyakorolnak-e kedvezőtlen hatást a válaszok megfelelőségére.

A saját idegennyelv tudásom és kiejtésem korlátai miatt a dokumentált 108 tesztesetet két körben értékeltem ki, a tesztesetek eredménye az I. mellékletben található. Első körben a teljes teszteredmény halmaz eredményeit összesítetten számszerűsítettem (15. ábra). A válaszok helyességét három értékkel reprezentáltam: helyes, helytelen és nem értelmezhető. Utóbbit az ebbe szükségessé, hogy egyrészt írásbeli reprezentációja minden kérdésnek volt, szóbeli azonban nem, másrészt a nyelv felismerési hibák esetén a tévesen transzkriptált, sokszor értelmetlen kérdésre adott válasz helytelennek minősítése torzította volna a nyelvi modell válaszadási képességeinek mérését. Az így kapott helyes, helytelen és nem értelmezhető eredményeket két csoportban összesítettem, külön szóban feltett kérdésekre, és külön a leírozott, vagy eleve írásban feltett kérdésre adott válaszok számát. A kapott eredményeket feltüntettem abszolút értékben és a kiértékelési körben figyelembe vett összes eset számára vetítve százalékosan.

megnevezés	esetszám	arány
összes teszteset	108	100,00%
összes helyes válasz szóbeli alapján	74	68,52%
összes helytelen válasz szóbeli alapján	15	13,89%
összes nem értelmezhető helyességű válasz szóbeli alapján	19	17,59%
összes helyes válasz transzkriptált/írott alapján	99	91,67%
összes helytelen válasz transzkriptált/írott alapján	8	7,41%
összes nem értelmezhető helyességű válasz transzkriptált/írott alapján	1	0,93%

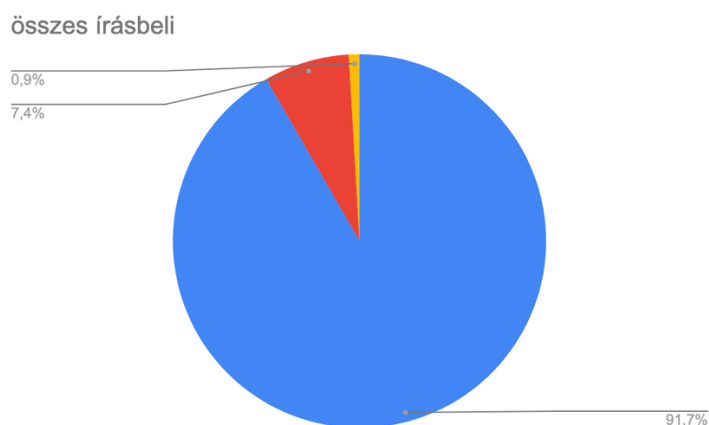
15. ábra – Teszteredmények (összes)

Az eredményekből az látható, hogy 68% feletti helyességet ért el az összes szóbeli kérdés esetén és közel 14% volt a hibás, és 17% feletti a nem értelmezhető válaszok száma (16. ábra). Tekintve, hogy ebbe az esetben az írásban feltett kérdések is az utóbbi kategóriában szerepelnek, a nem hibás válaszok száma 82%.



16. ábra - Szóbeli összes teszt megoszlása

Az írásban feltett vagy leíratott kérdésekre adott válaszok a vártak megfelelően magasabb pontosságot mutattak, itt közel 92% volt a helyes válaszok aránya, és 7% körüli a hibásaké (17. ábra).



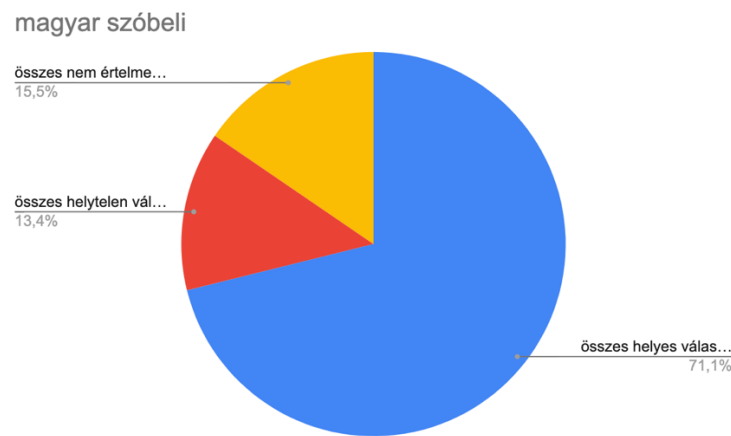
17. ábra - Írásbeli összes teszt megoszlása

Második körben a teljes tesztsorozat részhalmazán, a magyar nyelvű kérdések körében összesítetten az eredményeket (18. ábra). A válaszok helyességének értékkészlete, csoportosítása és aggregálása megegyezett az első kör módszerével.

megnevezés	esetszám	arány
összes teszteset	97	100,00%
összes helyes válasz szóbeli alapján	69	71,13%
összes helytelen válasz szóbeli alapján	13	13,40%
összes nem értelmezhető helyességű válasz szóbeli alapján	15	15,46%
összes helyes válasz transzkriptált/írott alapján	89	91,75%
összes helytelen válasz transzkriptált/írott alapján	8	8,25%
összes nem értelmezhető helyességű válasz transzkriptált/írott alapján	0	0,00%

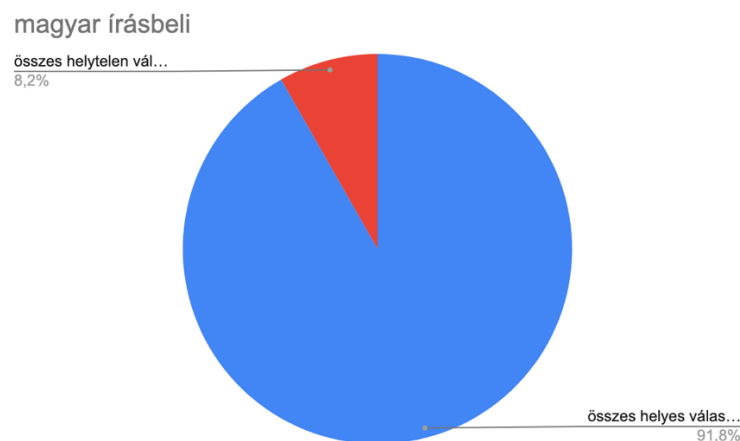
18. ábra – Teszteredmények (csak magyar nyelvű tesztek)

Az eredmények nem térnek el nagyságrendileg a bővebb halmaz eredményeitől, a különbségek pedig minden esetben javulást mutatnak. A szóbeli kérdések esetén 71% feletti a helyes, 13% a helytelen válaszok aránya (19. ábra).



19. ábra - Szóbeli magyar teszt megoszlása

Az írásbeli és transzkriptált kérdések esetén a teljes halmazával majdnem pontosan megegyező, azt 0,08 százalékponttal meghaladó közel 92%-os helyes válasz arányt ért el (20. ábra). Feltételezhető, hogy a különbség a kiejtésbeli hiányosság következményei.



20. ábra - Írásbeli magyar teszt megoszlása

Bár a hangminőséget nagyban befolyásolja a használt hardver környezet (beépített vagy külső mikrofon, headset) és az operációs rendszer és böngésző szintű mikrofon kezelés, továbbfejlesztési lehetőségekként azonosítottam fejlett zajszűrés alkalmazását, valamint diarizáció beépítését háttérzajból származó másodlagos beszélő leválasztására. Innovatív további kutatási terület lehet szájról olvasással való kiegészítés lehetősége mind a háttérzaj kompenzálására, mind a használati környezet kiterjesztésére olyan esetek és helyek irányában, ahol a beszédhang nem elfogadható (Fernandez-Lopez & Sukno, 2018). A válaszügy csökkentésre folyamatos streaming beépítése lehet célszerű a folyamatba (TTS megkezdése és hanglejátszás elindítása a teljes válasz elkészülése előtt), tovább

optimalizálható a promptképzés és válaszgenerálás során a tokenszám, amely az idő mellett a költségekre is jótékony hatást gyakorolhat. A nagy nyelvi modellek multimodalitásának kiaknázása tovább tágíthatja a lehetőségeket a magyarázatok képi illusztrálási lehetőségével. A kísérleti, proof-of-concept megvalósításon túllépve produktív környezetben való alkalmazása elképzelhető önálló szigetüzemű működésként, de megfontolás tárgyát képezheti komplex verbális e-learning keretrendszer köréépítése.

Összegzés

Az eddigiekben taglalt, külön-külön, vagy egy-két elemében kombináltan már elérhető megoldások átfogóbb integrálásával a tanítást széleskörűen támogató komplex megoldások hozhatók létre, amelyek már most is folyamatban vannak és a közeljövőben nagy fejlődés előtt állnak. A tananyagok várhatóan gazdagabbak, szórakoztatóbbak, alkalmazkodóbbak és testreszabhatóbbak, könnyebben befogadhatóak lesznek egyre jobban illeszkedve a tanulótípusokhoz, a készségekhez és a szociális szokásokhoz, de nem utolsó sorban egyre jobban kompenzálhatóak lesznek az emberi képességbeli korlátozottságok és a környezet okozta akadályok, például hang mellett akár szájról olvasással (Fernandez-Lopez & Sukno, 2018). A technológia az idő előrehaladtával az egyes komponensek szintjén egyre pontosabb, egyre gyorsabb és feltehetően egyre olcsóbb lesz. A megoldások feltehetőleg egyre nagyobb arányban épülnek be a mindennapokba, hatékonyabbá téve az oktatást és könnyebbé téve a pedagógusok munkáját.

A kutatás során nagy nyelvi modell alkalmazásával sikerült a hangoskönyvek alkalmazási területének egy olyan kiterjesztési lehetőségét azonosítanom és megvalósíthatóságát validálnom, ami az oktatásban alkalmazható szakmai tartalmak előszóban interaktívra tételére és ezáltal segítséget nyújthat a tanároknak a tananyag közvetítésére és a diákoknak a tudásanyag jobb megértésére és könnyebb befogadására. A mesterséges intelligenciára épülő technológiai megoldások jelenlegi fejlődési üteme alapján várhatóan hónapokon belül sokkal pontosabbá és élményszerűbbé tehetők az ehhez hasonló alkalmazások.

A fent bemutatottakat áttekintve elmondható, hogy napjaink fejlesztéseinek köszönhetően hangoskönyvek közvetlen és közvetett alkalmazhatósága rengeteg hasznos új lehetőséget teremt meg. A nagy nyelvi modellek utóbbi években történt és jelenleg is folyamatos várakozásokat felülmúló ütemű fejlődése már most nagyon sok területen nyújthat segítséget és várhatóan sok meglepetést fog még okozni. Érdemes belátni, hogy a sok aggasztó hangulatkeltés ellenére a technológiai fejlődés ellen nem küzdeni célszerű,

hanem megtalálni azokat a lehetőségeket, ahol azok minket szolgálhatnak, az életünket könnyíthetik és a munkánkat tehetik eredményesebbé. A nyilvánosságban felismerhető népszerűséget előnyünkre kell fordítani, és a hatékonyságon túl kapcsolódási pontként és motivációs eszközként is felhasználni. Zárszóként fontosnak tartom megerősíteni Csepeli György véleményét, miszerint a mesterséges intelligencia térhódításával nem kell a tanároknak félniük a munkájuk elvesztésétől, a tanári szerep ugyan átalakuláson megy keresztül, de a gyerekek tudáshoz való hozzájárásában továbbra is nélkülözhetetlen (Csepeli, 2020).

Nyilatkozat

A dolgozat szövegezéséhez mesterséges intelligencia használatára nem került sor.

A mesterséges intelligencia a kutatás fókuszában volt, a proof-of-concept applikáció API-n keresztül az openAI GPT3.5 modelljét hívta.

Irodalomjegyzék

- Basak, S., Wotto, M., & Bélanger, P. (2018). E-learning, M-learning and D-learning: Conceptual definition and comparative analysis. *Sage Journals*.
- Casper, S., Davies, X., Shi, C., Krendl Gilbert, T., Scheurer, J., & Rando, J. (2023). Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. *arXiv*.
- Cialdini, R. (2009). *Hatás*. Budapest: HVG Könyvek.
- Csepeli, G. (2020). *Ember 2.0: A mesterséges intelligencia gazdasági és társadalmi hatásai*. Budapest: Kossuth.
- Davis, E. (2016). AI amusements: the tragic tale of Tay the chatbot. *AI Matters*.
- EUR-Lex. (2021. 05 11). *EUR-Lex*. Forrás: Az Európai Parlament és a Tanács (EU) 2021/694 rendelete (2021. április 29.) a Digitális Európa program létrehozásáról és az (EU) 2015/2240 határozat hatályon kívül helyezéséről (EGT-vonatkozású szöveg): <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX%3A32021R0694&qid=1670782812870>
- Európai Bizottság. (2022. 09 28). *Európai Unió*. Forrás: A termékekre és a mesterséges intelligenciára vonatkozó új felelősségi szabályok biztosítják a fogyasztók védelmét és a folyamatos innovációt: https://hungary.representation.ec.europa.eu/termekekre-es-mesterseges-intelligenciara-vonatkozo-uj-felelossegi-szabalyok-biztositjak-fogyasztok-2022-09-28_hu
- Fernandes, C., & Morais, R. (2022). Do NFTs Sound Good? An Exploratory Study on Audio NFTs and Possible Avenues. *Informatics*.
- Fernandez-Lopez, A., & Sukno, F. (2018). Survey on automatic lip-reading in the era of deep learning. *Image and Vision Computing*.
- Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence*.
- Harris, S., & Ross, J. (2006). *Kezdőkenyv az algoritmusokról*. Bicske: Szak.

- Hu, K. (2023. 02 02). *ChatGPT sets record for fastest-growing user base - analyst note*. Forrás: Reuters: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
- Huitt, W. (2011). *Bloom et al.'s Taxonomy of the Cognitive Domain*.
- Jin, L., Lee, H.-J., & Park, J. (2004). Constructing SSML Documents with Automatically Generated Intonation Information in a Combinatory Categorical Grammar Framework. *International Journal of Computer Processing of Languages*.
- Kublik, S., & Saboo, S. (2022). *Gpt-3: Building Innovative NLP Products Using Large Language Models*. O'Reilly Media, Incorporated.
- Muktha, B., Mumma, P., & Willis, J. (1999). *R2D2: A Constructivist/Interpretivist Instructional Design Model*. Iowa: Iowa State University.
- Nash, B. (2023). Attending to the Sounds of Stories: The Affordances of Audiobooks in the English Classroom. *Changing English: Studies in Culture & Education*(30), 99-106.
- Puentedura, R. (2012). *The SAMR Model: Six Exemplars*.
- Sharma, P. (2010). Blended learning. *ELT Journal*, 456–458.
- Shrestha, S., Lenz, K., & Owens, J. (2007). “F” Pattern Scanning of Text and Images in Web Pages. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.
- Singh, A., & Alexander, P. (2022). Audiobooks, Print, and Comprehension: What We Know and What We Need to Know. *Educational Psychology Review*.
- Smith, K. (2022). All Ears: An Examination of the Documentality of Audiobooks, Podcasts, and Oral Histories with Extended Research into the London History Workshop Centre Oral History Collection in Collaboration with the Museum of London. London.
- Sun, W., Zhang, K., Chen, S.-K., Zhang, X., & Liang, H. (2007). Software as a Service: An Integration Perspective. *Service-Oriented Computing*.
- Topsakal, O., & Akinci, T. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. *All Sciences Proceedings*.
- Tóth, A. (2019. 2). *A mesterséges intelligencia szabályozásának paradoxonja és egyes jogi vonatkozásának alapvető kérdései*. Forrás: Infokommunikáció és jog: <https://infojog.hu/toth-andras-a-mesterseges-intelligencia-szabalyozasanak-paradoxonja-es-egyres-jogi-vonatkozasainak-alapveto-kerdesei-2019-2-73-3-9-o/>
- Tóth, N., Oszkó, B., & Zijian, G. (2023). Hungarian Sentence Analysis Learning Application with Transformer Models. *12th International Conference on Applied Informatics*. Eger.
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*.
- Veldhuis, R., & He, H. (1998). Time-scale and pitch modifications of speech signals and resynthesis from the discrete short-time Fourier transform. *Speech Communication*.
- VisikoKnox-Johnson, L. (2016). *The Positive Impacts of Fairy Tales for Children*. University of Hawai‘i at Hilo.
- Xu, P., Liang, D., Huang, Z., & Xiang, B. (2021). Attention-guided Generative Models for Extractive Question Answering. *arXiv*.
- Zhang, J.-P. (2008). Hybrid Learning and Ubiquitous Learning. *Lecture Notes in Computer Science*.