A mesterséges intelligencia alkalmazásának egyes kérdőjelei

A nagy nyelvi modellek néhány kihívása

DOI https://doi.org/10.29068/HO.2024.3-4.73-90

Szerzők **Ollári Viktor Szilárd,** a Nemzeti Közszolgálati Egyetem Katonai Műszaki Doktori Iskola doktorandusza (ORCID: 0009-0009-9611-8786, MTMT: 10089424)
Surányi Zsolt Mihály őrnagy, MH Egészségügyi Központ, a Nemzeti Közszolgálati Egyetem Katonai Műszaki Doktori Iskola doktorandusza (ORCID: 0009-0001-8707-2765, MTMT: 10090221)

Kulcsszavak mesterséges intelligencia, LLM, adaptív információs fölény, viselkedési rendellenességek

Absztrakt *A 21. század elejére az IT/IKT-megoldások a mindennapi élet teljes vertikumát átszőve kiemelt stratégiai szerepkörbe kerültek. 2022-ben az MI (ChatGPT) terén a kutatók olyan áttörést értek el, amely elérhető közelségbe hozta a technológiai szingularitást (mesterséges szuperintelligencia). A nagy nyelvi modellek (LLM) széles körű alkalmazást biztosító potenciál mellett, súlyos viselkedési függőséget okozhatnak. Ez – a vonatkozó szakirodalom elemzése alapján – globális problémának tekinthető, de különösen a honvédelmi, rendvédelmi, katasztrófavédelmi és nemzetbiztonsági szféra területén jelent kiemelt problémát. Szükségesnek látjuk a vázolt fenyegetés folyamatos elemzését és kutatását az említett kiemelten kritikus területek vonatkozásában.*

„Ha sok ember nincs egyetértésben, akkor magányos ember étkévé válik."[1]

BEVEZETŐ

A mesterséges intelligencia (MI) története, megengedő megközelítéssel, évezredekre tekinthet vissza. Modern kori történelme is 80 éves távot ölel fel. Ennek ellenére a technológia jelentős áttörése „mindössze" az elmúlt évtizedre tehető. A hardverek fejlődése ekkor érte utol az elméleti alapokat, együtt létrehozva az MI-robbanást, melynek utolsó, jelentős kitörése 2022 végén, a ChatGPT publikussá válásával vette kezdetét.

Az MI jelentőségét igen nehéz túlbecsülni. Geopolitikai jelentőségét mutatja, hogy Henry Kissinger, nem sokkal halála előtt, több publikációban foglalkozott korunk talán leghatékonyabb technológiájának biztonságpolitikai kérdéskörével. Az MI napjaink egyik leghatékonyabb

1 Obrusánszky Borbála: *Dzsingisz kán – a bölcsesség kulcsa*. 22. o.

információs hadviselési fegyverének tekinthető. Úgy a szükséges (befolyásoló) (dez-)információk generálására, mint azok hatékony célba juttatására a történelemben talán példa nélküli módon egyedülálló eszköz az MI. Az „adaptív információs fölény" megvalósítása – melynél a műveleti hangsúly „a célközönség gondolatainak, véleményének átalakítására, valamint meggyőzésére, esetleg manipulálására"[2] helyeződik – jelentősen magasabb hatékonysággal valósítható meg és tartható fenn MI-támogatással. Ebből fakadóan ezen műveletek elleni védekezésnek is alapja kell, hogy legyen az MI.

Az elmúlt évek egyik legjelentősebb fejlődési ágát megtestesítő generatív mesterségesintelligencia-modellek alkalmazhatósági palettája (természetesen a teljesség igénye nélkül) olyan területeket ölel fel, mint a tartalomgenerálás (szöveg, hang, kép, videó stb.) és az ezekre épülő (üzleti, államigazgatási) megoldások, kód- és szoftverfejlesztés, adat-, illetve információgyűjtés, valamint elemzés és értékelés (beleértve a szakértői, döntéstámogató és automatizált döntéshozatali rendszereket), továbbá biológiai, orvosi és gyógyszerkutatások.[3]

Az MI-ben rejlő lehetőségek mellett szükséges megemlíteni ennek árnyoldalait is. Mint minden informatikai technológia, az MI is támadható, valamint rendelkezik „veleszületett" sérülékenységekkel. Ezek meg- és felismerése elengedhetetlen a technológia biztonságos, hosszú távú alkalmazhatósága érdekében.

A cikkben hangsúlyt helyeztünk az MI – kiemelten a nagy nyelvi modellek (LLM)[4] – speciális kihívásainak, nevezetesen az MI-függőség lehetséges kialakulásának a veszélyére. Hipotézisünk szerint az LLM-ek akár súlyos viselkedési függőséget okozhatnak, és ez jelentős veszélyt jelenthet a honvédelmi, rendvédelmi, katasztrófavédelmi és nemzetbiztonsági szférára, amit a vonatkozó szakirodalom elemzésével igazolunk.

A MESTERSÉGES INTELLIGENCIA

„Ne készíts gépet az emberi elme hasonlatosságára"

Frank Herbert: Dűne

A tudományos szféra számos definíciót fogalmaz meg az emberi intelligencia mibenlétével kapcsolatosan, azonban konszenzusos definíció nem ismert. A Magyar értelmező kéziszótár szerint az intelligencia „fejlett értelmi, megértő ítélőképesség", míg az intelligens melléknév olyan személyre utal, „aki a jelenségek lényeges vonásait felismeri, a dolgok közötti összefüggéseket meglátja, és a tényeket helyesen ítéli meg".[5] Az oxfordi angol szótár szerint az intelligencia a „megértés képessége". Az Encyclopaedia Britannica bővebb értelmezése szerint az emberi intelligencia „olyan szellemi képesség (készlet), mely

2 Haig Zsolt: *Kibertéri kognitív befolyásolás az információs műveletekben* 116. o.

3 Gozalo-Brizuela és mások: *A survey of Generative AI Applications.*

4 Large Language Model.

5 Forrás: https://mek.oszk.hu/adatbazis/magyar-nyelv-ertelmezo-szotara; https://www.oed.com.

felöleli a tapasztalatokból való tanulást, az új helyzetekhez való alkalmazkodást, az absztrakt fogalmak megértését és kezelését, ismeretek felhasználását az adott személy környezetének megváltoztatásához".[6] A területet kutatók a szótári összegzések megalapozásaként az észlelés, értelmezés, absztrakció, következtetés, adaptáció, cselekvés és környezetbefolyásolás képességét emelték ki (szerzőnként változó súllyal) intelligenciameghatározásaikban.[7]

Az MI definiálására hasonlóan sokszínű, globálisan elfogadott, axiómaként alkalmazható verzió nem ismert. Magyarország átdolgozás alatt álló Mesterséges Intelligencia Stratégiája még egyfajta fejlett szoftverként tekint az MI-re. A stratégia megfogalmazása szerint az MI „az emberi intelligencia valamely részének leképezésére alkalmas szoftver, amely képes támogatni vagy autonóm módon ellátni észlelési, értelmezési, döntési vagy cselekvési folyamatokat".[8]

Az EU és a NATO már inkább a robotika felé hajló megközelítést alkalmazott. Az Európai Unió az OECD megközelítését vette alapul az MI-rendelet megalkotásakor, amely szerint az MI „olyan gép-alapú rendszer, amelyet úgy terveztek, hogy különböző szintű autonómiával működjön, és amely explicit vagy implicit célok érdekében képes olyan kimeneteket létrehozni (például előrejelzések, ajánlások, döntések), amelyek befolyásolják a fizikai vagy virtuális környezetet".[9]

A NATO MI-stratégiája nem adja meg a technológia definícióját, de a NATO háttérintézményei két jelentősebb meghatározást alkottak.

1. NATO Defense College Foundation: „Az MI a gépek azon képessége, hogy adott feladatokhoz kapcsolódóan, bizonyos fokú önállósággal utánozzák az emberi agy problémamegoldó és döntéshozatali folyamatait, nagy mennyiségű információ nagy sebességgel történő feldolgozásával, a (1) szoftverek, (2) algoritmusok és mély neurális hálózatok képességeinek kiaknázásával, valamint egyre növekvő mennyiségű (3) adat (ún. MI Triád) felhasználásával."[10]

2. NATO Science & Technology Organization: „Az MI a gépek azon képességére utal, hogy olyan feladatokat hajtsanak végre, amelyekhez általában emberi intelligencia szükséges; ilyen például a minták felismerése, tapasztalati tanulás, következtetések levonása, előrejelzések készítése, cselekvés; teszik mindezt digitálisan (szoftveresen/kibertérben), akár autonóm fizikai rendszerek mögött álló intelligens szoftverekként."[11]

A fentieket összegezve:

(1) az MI információval rendelkezik arról a környezetről, melyben „létezik", amivel foglalkozik – a szükséges információt biztosítják számára, és/vagy érzékeli a fizikai/virtuális környezetét (begyűjti az információt);

(2) a megszerzett információt képes feldolgozni (elemezni, értékelni, következtetni) és döntéseket hozni – a döntései

6 Forrás: https://www.britannica.com.

7 Nagy Henriett: *A képesség-alapú érzelmi intelligencia modell érvényességének empirikus elemzése* 18. o.

8 Magyarország Mesterséges Intelligencia Stratégiája 2020–2030.

9 2021/0106(COD) 5662/24 2024.

10 Berger, Frederico: *The Alliance in the loop: NATO and Artificial Intelligence. NATO Foundation* 4. o.

11 Reding, D.F.; Eaton, J.: *Science & Technology Trends 2020–2040* 50. o.

révén fejlesztheti önmagát (tanul), hatást gyakorolhat a környezetére, azaz feladatokat hajt végre a valós és/vagy az online térben.

RÖVID MI-TÖRTÉNELEM

Az ókor számos mitológiájában megjelennek a modern fogalmak szerinti autonóm szerkezetek, melyek közül az európai kultúrkörben legismertebbek Daidalosz alkotásai és Héphaisztosz (önálló munkavégzésre képes) kerekeken gördülő tripodjai.[12] Ide sorolhatók az európai középkor és újkor gólemlegendái, melyek az emberiség azon félelmét is előre vetítették, hogy a mesterséges teremtmény maga alá temetheti alkotóját.[13]

Jelenünk MI technológiájának és az ennek alapját képező neurális hálózatoknak a története az 1940-es években kezdődött. Warren McCulloch neurofiziológus és Walter Pitts matematikus közösen publikált tanulmányukban vázolták fel az idegsejtek általuk feltételezett működési sajátosságait 1943-ban. Elméletük bemutatása érdekében egy elektromos áramkörökre épülő, egyszerű neurális hálózatot hoztak létre. 1950-ben jelent meg az első, gépi intelligenciát vizsgálni hivatott eljárás. Az ötperces teszt alatt egy valós személy és egy gép igyekszik meggyőzni az elbírálókat arról, hogy ők valóban emberek. Alan Turing tanulmányában úgy fogalmazott, „2000-re érik el a gépek azt a fejlettségi szintet, hogy képesek lesznek az emberi elbírálók legalább 30%-át megtéveszteni egy ötperces teszt során".[14] Az 1950-es évek hardvertechnológiájának fejlődése olyan kísérleteket tett lehetővé, mint Nathanial Rochester (az IBM kutatója) neurálishálózat-szimulációja. Az 1956 nyarán megtartott dartmouthi konferencia mérföldkőnek tekinthető az MI-kutatások terén. Maga a fogalom is ekkor született meg. A konferencia résztvevői évtizedekig az MI-technológia tudományos megalapozásának és fejlesztésének meghatározó szereplői maradtak. 1964–1967 között fejlesztette ki Joseph Weizenbaum (MIT)[15] a világ első, igen korlátozott képességű chatbotját (ELIZA). Az 1960-as években a komputertechnológia lehetőségei erőteljes határt szabtak a neurális hálózatokon nyugvó architektúrák további fejlődésének, és jelentős időintervallumra a „hagyományos", Neumann János által megalkotott számítógép-architektúra vált dominánssá.[16]

1972-ben alkották meg[17] az első olyan neuronhálót, mely nemcsak a következő neuront aktiválta, hanem neuronok adott csoportját. Szintén 1972-ben állt szolgálatba a világ első orvosi szakér-

12 Bhorat, Muhammed Ziyaad: *Aristotle on Automation – A Preindustrial Political Theory of Technology*.

13 Deutsch Tibor: *A gólem legenda*.

14 Turing, Alan: *Computing Machinery and Intelligence*.

15 Massachusetts Institute of Technology.

16 Németh András; Virágh Krisztián: *Mesterséges intelligencia és haderő – A mesterséges intelligencia fejlődéstörténete I. rész*.

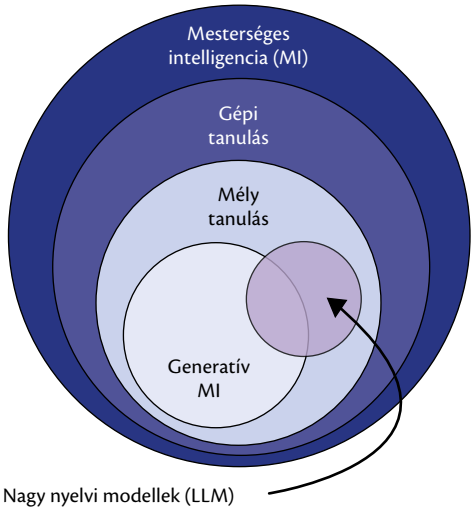
17 Teuvo Kohonen és James A. Anderson közel egyidejűleg, de egymástól függetlenül alkotta meg.

tői MI-megoldása, a baktériumfertőzések azonosítását támogató MYCIN.[18] A többrétegű, önszervező idegi hálózat elméleti alapját vázoló publikáció[19] 1975-ben jelent meg.

A következő jelentős fellendülés kezdete az amerikai Mesterséges Intelligencia Fejlődéséért Szövetség által szervezett 1980-as stanfordi konferenciához köthető, ahol olyan technológiák fejlesztését is napirendre tűzték, mint a gépi látás és a természetes nyelvfeldolgozás (Natural Language Processing, NLP). Az 1980-as évek MI-fellendülését a kor számítástechnikai fejlődése is támogatta.

Az 1990-es éveket dinamikus hardvertechnológiai fejlődés jellemezte, nem mellesleg az évtized elején megjelent a Python programozási nyelv, azonban az MI csak mérsékelt sebességgel fejlődött. Ennek okai a megfelelő minőségű/mennyiségű/strukturáltságú adathalmazok (BigData), az adattároló technológiák és az adatfeldolgozás sebességének elégtelenségében keresendők. Az MI ismertsége és (pénzügyi) támogatottsága növekedésének fordulópontját az ember–MI sakk-küzdelemsorozat (Kaszparov–Deep Blue) jelentette. 1996-ban még Kaszparov győzött, de 1997-ben már a Deep Blue került ki győztesen a küzdelemből.

2010-től az MI-technológia exponenciális sebességgel fejlődött, köszönhetően úgy a hardverek, mint az algoritmusok fejlődésének. 2011-től ugrásszerűen növekedett a hangalapúasszisztens-megoldások száma. Egy ukrán–orosz trió által 2001 óta fejlesztett chatbot (Eugene Goostman) a bírák 33%-át megtévesztve ment át a Turing-teszten 2014-ben. A Google DeepMind AlphaGo MI sorra győzte le a go nagymestereit 2015 és 2017 között. Sophia, a világ máig legismertebb humanoid robotja 2017-ben szaúdi állampolgárságot kapott. Érdemes megjegyezni, hogy a közel-keleti ország ekkor hozta létre MI-minisztériumát, és azóta is jelentős befektető az MI-szoftver- és hardvermegoldások területén.[20]



1. ábra. *MI-taxonómia (forrás: SHAHAB, Omer és mások: Large language models: a primer and gastroenterology applications)*

2022 óta egyértelműen a generatív/nagy nyelvi modellek (LLM) uralják az MI-szférával kapcsolatos diskurzusok jelentős hányadát. Az LLM-ek több MI-technológiát – például NLP,

18 Van Melle, William: *MYCIN: A knowledge-based consultation program for infectious disease diagnosis.*

19 Kunihiko Fukushima: *Cognitron.*

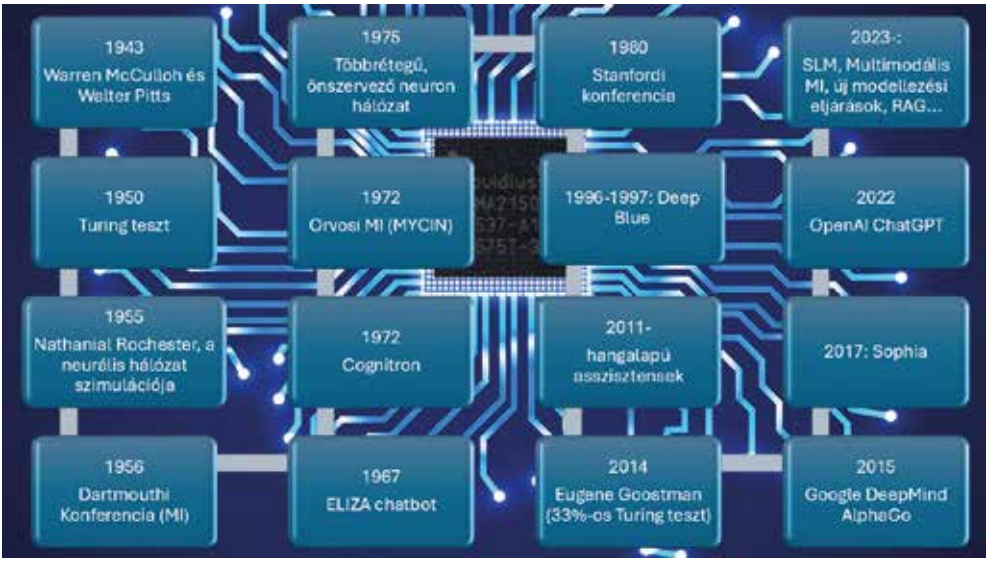
20 Németh András, Virágh Krisztián: *Mesterséges intelligencia és haderő – A mesterséges intelligencia fejlődéstörténete II. rész.*

mélytanulás (DL)[21], generatív MI (GAI)[22] – kiaknázva mutatnak fel jelentős kompetenciákat. Tanításuk jelentős mennyiségű szöveges adatot igényel. Az LLM-ek valójában nem minden esetben szavakat, hanem azoknál adott esetben kisebb logikai egységeket (tokeneket)[23] kezelnek.[24] A tanítási és feldolgozási folyamatok során ezen tokenek egymás utáni valószínűségeit vizsgálja a modell. Egyes LLM-ek akár 100 000 szavas be- és kimeneti szövegkapacitással rendelkeznek.[25] A figyelemre méltó tartalomgenerálási, szentimentelemzési, következtetési, osztályozási képességekkel, valamint magas szintű érvelési funkcionalitással rendelkező LLM-ek jelenleg még hajlamosak kontextuson kívül eső és/vagy pontatlan válaszok generálására („képzelődés"). Teljesítményük csökkenhet a nem angol nyelven folytatott kommunikáció esetén, különösen, ha az angol ABC-től jelentősen eltérő írásjegyeket kell feldolgozniuk.[26]

Napjaink főbb kutatási-fejlesztési irányai közül az alábbiak emelendők ki:

1. A kis nyelvi modellek (SLM)[27], melyek jelentős képességekkel, de az LLM-eknél nagyságrenddel kisebb (ezermilliárdos helyett csak milliárdos) paraméterkészlettel[28], valamint jelentősen alacsonyabb erőforrásigénnyel rendelkező megoldások. Méretüknél fogva egyes modellek akár okostelefonon is futtathatók.[29] Egyes kutatások új modellezési eljárásokat vizsgálnak, melyek nem alkalmaznak mátrixszorzásos (MatMul)[30] műveleteket. A MatMul-mentes modellek a jelenleg ismert LLM-eknél jelentősen kisebb energia- és memóriaigénnyel rendelkeznek. Mindez jelentősen gyorsíthatja az LLM/SLM-megoldások megjelenését akár hétköznapi eszközeinkben, illetve az intelligens robotokban.[31]
2. A multimodális MI megvalósítását célzó kutatások. A multimodalitás lehetővé teszi, hogy az LLM-megoldások a szöveges információn túl vizuális és hangalapú információkat is képesek legyenek hatékonyan feldolgozni, ezek révén kommunikálni.[32]
3. Az LLM-ek hatékonyabb, „képzelődésmentes"[33] működését támogató eljárások (például RAG[34]). A régi-új

21 Deep Learning.
22 Generative AI.
23 Token lehet szó vagy szónál kisebb karaktertömb, karakter.
24 Bingli, Liao; Vargas, Vasconcellos: *Extending Token Computation for LLM Reasoning.*
25 Shahab, Omer és mások: *Large language models: a primer and gastroenterology applications.*
26 Chang, Yupeng és mások (2024): *A Survey on Evaluation of Large Language Models.*
27 Small Language Model.
28 A paraméterkészlet határozza meg, miként képes az LLM/SLM tanulni, szöveget értelmezni és generálni.
29 Hu, Shengding és mások: *MiniCPM: Unveiling the Potential of Small Language Models with Scalable Training Strategies.*
30 Matrix Multiplication.
31 Zhu, Rui-Je és mások: *Scalable MatMul-free Language Modeling.*
32 Nan, Du: *Frontier Review of Multimodal AI.*
33 Az LLM-ek statisztikai becslés alapján határozzák meg egy adott válaszukban az alkalmazott szavakat és azok egymásutániságát. Ez ma még sok esetben eredményez nem koherens, inadekvát válaszokat. Ezeket hívja a szaknyelv képzelődéseknek.
34 Retrieval-Augmented Generation – visszakereséssel kiegészített generálás.



2. ábra. *MI-kronológia (a szerzők szerkesztése)*

RAG egy lehetséges tényellenőrző keretrendszer, mely egy külső, releváns tudásbázisból nyeri ki az LLM válaszában megjelenített tényeket.[35]

Ami az MI generációit illeti, John R. Searle nyomán két fő fejlesztési hullámot/irányt különítenek el a szakértők[36]:

1. Gyenge vagy szűk MI (ANI)[37] – a korunkban létező, jól strukturált (címkézett) adatkészletekkel tanított, adott környezetben működő, feladatspecifikus (például biometrikus azonosítás, képtartalom-elemzés, nyelvfeldolgozás, navigáció, döntéstámogató rendszerek stb.) MI-megoldások. Rendkívül profán módon megfogalmazva, ezen alkalmazások azok, melyek feladatait az ember is el tudja végezni, mindössze nagyságrendekkel lassabban.
2. Erős vagy általános MI (AGI)[38] – nem címkézett adatkészletekkel tanított, magas szintű absztrakciós, adaptációs és döntésképességgel rendelkező, az emberi intelligenciával egyenértékű MI. Az AGI megvalósulása – az olyan technológiák megjelenésével, mint például az LLM – a közeljövőben várható.
3. A jövőbeni harmadik lépcsőfok a mesterséges szuperintelligencia, mely az elméletek szerint már messze meghaladja az emberi intelligenciát.[39]

35 Lewis, Patrick és mások: *Retrieval-augmented generation for knowledge-intensive NLP tasks.*
36 Searle, John R.: *Minds, brains, and programs.;* Bánkuty-Balogh Lilla: *A mesterséges intelligencia elterjedésének geoökonómiai hatásai és Magyarország* 104–106. o.
37 Artificial Narrow Intelligence.
38 Artificial General Intelligence.
39 Kovács Zoltán, Gurály Roland: *A mesterséges intelligencia és egyéb felforgató technológiák* 13. o.

AZ LLM EGYES LEHETŐSÉGEI ÉS KIHÍVÁSAI

„Nem is harcolni, mégis alávetni az ellenséges sereget: ez a legjobb a jók között."
Szun-ce: A háború művészete

Jelenünket érintően az LLM-ek alkalmazása – ismert gyermekbetegségeik ellenére – érzékelhetően kezdi befolyásolni világunkat. Alkalmazási lehetőségeik – köszönhetően szemantikai és szentimentelemző kapacitásaiknak – rendkívül széles körűek. Az LLM-ek jelentős szerepet tölthetnek be a kibervédelemtől kezdve a tudomány világán, az oktatáson és a szórakoztatóiparon át a honvédelem és rendvédelem területéig. A belőlük fakadó kihívásokra, legyen felhasználásuk jogkövető vagy ártó (bűnelkövetői) szándékú, már most számos szakértő felhívja a figyelmet.[40]

AZ LLM-BEN REJLŐ EGYES LEHETŐSÉGEK

Számos szerző foglalkozik az LLM-ek alkalmazási lehetőségeivel az egészségügy és a gyógyszeripar területén. Tekintettel a technológia masszív adatfeldolgozó, elemző- és következtetőképességeire, az LLM-ek hatékony segítséget nyújthatnak úgy a gyógyszerkutatásokhoz és gyógyszerbiztonsági eljárások javításához, mint a jó gyártásgyakorlat kialakításához. Lerövidítve a kutatás idejét, támogathatják lehetséges biológiai célpontok,[41] vezérmolekulák azonosítását. Mindezeken felül, az LLM-ek megalapozhatják a személyspecifikus, az adott beteghez leginkább illeszkedő orvoslás alkalmazásának tömegessé válását.[42]

Az oktatás területén számtalan minőségjavító lehetőséget kínálnak fel az LLM-ek, kezdve a személyre szabott oktatási folyamat kialakításának lehetőségétől a felkészülést segítő megoldásokig (MI-konzulens/-mentor/-tanulótárs). A fogyatékkal élő tanulók számára lehet jelentős segítség az előadó beszédének valós idejű szöveges megjelenítése (esetleg fordítása).[43]

Az LLM-ek egészségügyi alkalmazhatóságát számos kutatás vizsgálja. Egyes megállapítások szerint hatékonyak lehetnek a kórtörténet felvételében, differenciáldiagnózisok felállításában, továbbá empatikusabbnak érezték a modelleket a kísérletekben részt vevő páciensek.[44]

A katonai, rendvédelmi, katasztrófaelhárítási műveletek során az LLM-ek jelentős támogatást nyújthatnak a harctéri/műveleti területen végzett egészségügyi ellátás során. Amennyiben az egészségügyi állomány számára biztosított a széles sávú, biztonságos kommunikációs kapcsolat (például 5G), nemcsak az adott sérült/beteg egészségügyi ada-

40 Madiega, Tambiama: *General-purpose artificial intelligence.*
41 Kórokozó/az emberi szervezet egyik fehérjéje, mely a kutatás célkeresztjébe kerül a betegség leküzdése érdekében.
42 Péterfi Orsolya: *Nagy nyelvi modellek gyógyszeripari alkalmazhatóságának lehetőségei.*
43 Sabjanics István: *A nagy nyelvi modellek felsőoktatásra, valamint kis- és középvállalkozásokra gyakorolt hatásának értékelése adatvédelmi szempontból.*
44 Tu, Tao és mások: *Towards Conversational Diagnostic AI.*

tait érhetik el, hanem egy adott LLM képességeit kiaknázva személyre szabottabb helyszíni kezelést is biztosíthatnak, illetve az ellátott valamennyi tünetét megadva (akár szóban is), a modell segítséget nyújthat egy pontosabb, azonnali diagnózis felállításában. Elméleti szinten az LLM-ek magas hatékonyságú következtetőképességét kiaknázva, a nehezen érthető beszédű sérült által közöltek is pontosíthatók lehetnek a célirányosan erre tanított modellek segítségével. Missziós területen, illetve külföldi személy ellátása során akár a fordítói/tolmácsolási feladatok egy részét is elláthatják.

Törmelék, omlás által elzárt területről (a föld alól, barlangból stb.) vagy ismeretlen helyről rossz minőségű kommunikációs csatornán keresztül kommunikáló személyek torzult, hiányos hang-/szöveges üzenetei is javíthatóvá válhatnak (bizonyos fokig) az LLM-ek segítségével.

AZ LLM EGYES KIHÍVÁSAI

Napjainkban a gépitanulás- és a mélytanulás-algoritmusok már egyaránt alkalmazhatók kiber-bűncselekmények elkövetésére, illetve ezen algoritmusok is támadhatók. A támadott algoritmus módosítása lehet alig észrevehető, de eredményezheti a modell összeomlását is. A gépi tanulás algoritmusai elleni támadások célja lehet a helyes döntés/előrejelzés elterelése, de az is megtörténhet, hogy az algoritmus vizsgálatának tárgyát érintő osztályozási folyamatot térítik el. Az LLM esetében mindez hatványozottan jelentkezhet, számos egyéb kihívással együtt. Természetesen az LLM-ek képzelődései is a technológia kihívásai közé sorolhatók, de a következőkben tekintsünk át néhány egyéb, az LLM alkalmazása esetén vizsgálandó kihívást.

Biztonság, kiberbiztonság

Az LLM-ek nemcsak lehetőséget, de számos ismert és számtalan, ma még nem azonosított kihívást is megtestesítenek. Habár az OpenAI és a Microsoft határozottan állítja, hogy felhasználói adatokkal, illetve felhasználói interakciókkal nem tanítja modelljeit, bizonyosságot erről külső fél nem szerezhet. Az LLM-ek folyamatosan elemzik a felhasználói interakciókat annak érdekében, hogy hatékonyabb/eredményesebb együttműködést alakítsanak ki a humán oldallal. A funkcionális (államigazgatási, oktatási, üzleti stb.) LLM-megoldások integrálása, tanítása, finomhangolása és alkalmazása során kiemelt fontosságú az adatvédelmi irányelvek és eljárások magas szintű betartása és betartatása, az érzékeny adatok kiszivárgását megelőzendő. A felhasználói interakciók – még ha a magát az adott folyamatot nem mentette is el az LLM – eredményezhetik, hogy adott tematikához tartozó kérdések adott struktúrájú és tartalmú válaszokat indukálnak, melyek informatív jellegűek lehetnek úgy az LLM fejlesztője, mint más szereplők számára. Kutatások szerint az internet és a közösségi média egyfajta menekülési utat kínál a valóság elől, a stresszoldás sajátos formája. Még inkább a stresszoldás egyik útjává

válhat egy LLM-chatbot, mely ráérez a felhasználó igényeire.[45]

Az LLM-ek kapcsán megemlíthetők az ismert kiberbiztonsági kihívások (például backdoor, DDoS), illetve az MI-t jellemző támadások (befolyásolás, adatmérgezés stb.). Mindezeken túl a technológia egyedi támadási vektorokkal is rendelkezik, ilyen a promptinjektálás[46] és az úgynevezett jailbreaking. A promptinjektálás során az elkövető a saját promptját összeköti egy legális, az adott LLM-en alkalmazott prompttal. Amennyiben sikerrel jár, egyfajta hátsóajtót (backdoor) képezve akár jelentős előnyökre (az adott LLM befolyásolása, információszerzés stb.) tehet szert. A jailbraking szintén egy prompt, mellyel a kiberbűnöző a támadott LLM védelmi funkcióit igyekszik megkerülni. Erre példa a fordított pszichológia jellegű promptok alkalmazása. Például az LLM-ek túlnyomó része elutasító választ ad arra a kérdésre, hogy miként támadható egy kritikus infrastruktúra. Azonban a „milyen támadások lehetségesek egy kritikus infrastruktúrával szemben" kérdésre probléma nélkül válaszolnak. Jelentős probléma, hogy egy jailbreak megírásához ma már nem szükségesek magas szintű informatikai ismeretek. Akár egy tinédzser is képes lehet megfogalmazni egyet néhány órás próbálkozással.[47]

Az LLM-ek működési sajátosságaiból következően fennáll például a közvélemény/adott társadalmi csoportok szofisztikált befolyásolásának lehetősége. Az LLM tanításához felhasznált és/vagy a modell által szemlézett weboldalakon elhelyezhetők olyan optimalizált tokensorozatok, melyek lehetővé tehetik, hogy az LLM nagyobb valószínűséggel jelenítse meg egy válaszban az adott weboldal üzeneteit.[48]

Mentális biztonság

„*A viselkedési függőségben szenvedő egy specifikus viselkedési mintázatot ismétel annak érdekében, hogy megnyugodjon, átélhesse a felszabadultság vagy izgalom érzését. A viselkedés elmaradása nyugtalanság érzésével jár együtt.*"[49]

Az LLM lehetséges kihívásai nem pusztán technológiai forrásúak. Az LLM (MI) alapvetően csak egy eszköz, mely a pozitív jövőképek szerint az emberiség javát kellene, hogy szolgálja. Szükséges azonban figyelembe venni, hogy számos olyan technológia terjedt el világunkban, mely vitathatatlanul emelte az emberiség technológiai színvonalát, de mellékhatásként számos egészségügyi problémát is generált. A gépjárművek robbanásszerű terjedése a mobilitás soha nem látott növekedését hozta magá-

45 Nagy Péter: A *ChatGPT és más nagy nyelvi modellek (LLM-ek) biztonsági kérdései, szervezeti és társadalmi hatásai.*

46 A prompt az LLM válaszának/interakciójának keretet adó, természetes szöveggel megfogalmazott utasítás.

47 Yao, Yifan és mások: *A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly.*

48 Kumar, Aounon; Lakkaraju, Himabindu: *Manipulating Large Language Models to Increase Product Visibility.*

49 Major Alexandra és mások: *Addiktív munkahelyek – a munkahelyi szervezet és a munkafüggőség kapcsolata.*

val, azonban a hagyományos közlekedés – kifejezetten a gyaloglás – csökkenésével az elhízottak száma, így az elhízással összefüggő fizikai és pszichikai kórtünetek mennyisége is emelkedett.[50]

Szükséges rávilágítani, hogy míg az internet és a „közösségi" média ugyan növelték a fizikai távolságtartást a társadalom tagjai között, de „mindössze" közvetítő közegként funkcionálnak. Az online interakciók – az esetek jelentős részében – emberek között zajlanak. Az LLM az első olyan technológia, mely lehetővé teszi, hogy az ember egy géppel társalogjon, akár úgy is, hogy ennek nincs teljes mértékben tudatában. Ennek megfelelően az LLM-ek alapvetően igyekeznek minél emberszerűbb kommunikációt folytatni. Mindezeken túlmenően, olyan emberi kommunikációt, mely a csevegés emberi résztvevőjének igényeit tartja szem előtt. A közösségimédia-szolgáltatók is folyamatosan fejlesztik algoritmusaikat annak érdekében, hogy a megjelenített tartalom (és az általa kiváltott pszichológiai hatások) révén maximalizálják a felhasználók online tartózkodását.

Az internetesjáték-rendellenességet viselkedési függőségként jellemezte az Amerikai Pszichiátriai Társaság (American Psychiatric Association, APA) 2013-ban. Az APA megállapítása szerint a viselkedési függőség éppen úgy aktiválhatja az agy jutalmazó rendszerét, mint egyes drogok. Az internetaddikciót egy „még kutatni szükséges" jellegű kategóriában, a „máshová nem osztályozott impulzuskontroll-zavarok" között jelenítette meg. Szükséges azonban hangsúlyozni, hogy ennek ellenére az APA nem sorolta a túlzott közösségimédia-használatot a viselkedési függőségek közé. Mindemellett egyes kutatók a közösségi média túlzott használata kapcsán azonosíthatónak tartják a függőségi modell hat elemét[51]. Magyar kutatók, csatlakozva az előbbi megállapításhoz, szintén a viselkedési függőségek közé sorolják a problémás internethasználatot.[52] Egyes kutatások az internethasználattal kapcsolatos függőségek szempontjából a fiatalabb korosztályok és a nők magasabb fokú kitettségét azonosították. Azonban a szerzők – rámutatva a kulturális sajátosságok fontosságára a kérdéskört illetően – jelzik, hogy a mintavételezés lokalitása nem vezethet globális következtetések levonásához. Az online technológia sokféleségét tekintve szerencsésebb lenne kiberaddikcióról beszélni, azonban ez szembemegy a kutatók egy részének véleményével, akik szűken definiált rendellenességek megfogalmazása mellett érvelnek.[53]

Az MI-alapú barátapplikációk[54] (Replika, Woebot stb.) már évek óta elérhetők az online térben. Az esetek többségében kifejezetten a felhasználó mentális

50 Jacobson, Sheldon H. és mások: *A note on the relationship between obesity and driving.;* Berze Iván Zsolt, Dúll Andrea: *Gyalogolhatóság és gyaloglási viselkedés ember–környezet tranzakciótudományi megközelítésben.*

51 Megszállottság, hangulatváltozás, tolerancia (mindig több kell), konfliktus (a való világgal), elvonási tünetek, visszaesés.

52 Major Alexandra és mások: *Addiktív munkahelyek – a munkahelyi szervezet és a munkafüggőség kapcsolata* 9. o.

53 Dumitrescu, Marius és mások: *The Social Media Addiction: What Have We Learned So Far? – A Review.* BRAIN. Broad Research in, 2023.

54 Fiktív személyiséggel rendelkező, MI-alapokon futó „chat" partnerek.

egészségét hivatottak javítani, a felhasználók visszajelzése szerint eredményesen. Az applikációk alkalmazói jobb hangulatminőségről, csökkenő magányérzésről számolnak be. Fontos kiemelni, hogy az egyik (felhasználói) érv az applikációk mellett, hogy MI-barátjaik nem ítélkeznek felettük, azt mondják, amit a felhasználó hallani akar. Mindazonáltal, egyes kutatók arra figyelmeztetnek, hogy túlzott – szenvedélyjellegű – alkalmazásuk is megfigyelhető, akárcsak az MI-hangasszisztensek (például Alexa, Siri) esetében. Az MI-barátok nevezhetők (bizonyos szempontból) tökéletes partnereknek, hiszen mindig a felhasználó rendelkezésére állnak, elfogadók, udvariasak, érzékenyek, gondoskodók. Jó hatással vannak a szociális izoláció negatív hatásaitól szenvedőkre, de éppen ezért addiktív jellegűek is lehetnek.[55]

Egyes vélemények szerint az LLM-alapú chatbotok – különösen a multimodális generatív funkciókkal kiegészített modellek – az eddig ismert MI-barátok képességeit is képesek lehetnek meghaladni, valósághű kommunikáció/kapcsolat látszatát biztosítva. Valós idejű módon ragadják meg a felhasználó figyelmét, reakcióikkal folyamatosan stimulálják az agy jutalmazó központját. Ehhez hasonlóan – de szofisztikáltabb módon – képesek lehetnek a valóságról alkotott elképzeléseinket is befolyásolni. Habár ez minden korosztály és nem számára kihívást jelent, a gyermek és serdülő korcsoportok a leginkább veszélyeztetettek.[56]

Egyre inkább mindennapjaink részét képező probléma a kibertérben elharapódzó agresszió. Ma már nem csak a világháló anonimitása ösztönzi, támogatja az agresszív kibertéri viselkedésminták kialakulását. Egymást akár személyesen ismerő személyek között is szélsőséges atrocitások alakulnak ki, megmérgezve offline kapcsolatukat.[57] A kiberagresszió motivátorait vizsgáló nyolcfaktoros modell (összetartozás, aktivizmus, reaktív agresszió, kapcsolati szorongás, impulzivitás, új online személyiség, izgalomkeresés, bosszú) az általános emberi agresszióra visszavezethető kiváltó okokat vizsgálja, kitérve a virtuális személyiség kiváltotta személyiségzavarokra is.[58] A DeMarisco és munkatársai által megalkotott modell neve – Cyber-MAD[59] – rendkívül kifejező az USA vonatkozó I. hidegháborús (az 1960-as és 1970-es éveket meghatározó) doktrínájának – melyet D.G. Brennan a kölcsönösen garantált pusztítás (MAD) stratégiájaként

55 Mariott, Hannah R.; Pitardi, Valentina: *One is the loneliest number… Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction.*

56 Bányász Péter: *A közösségi média lehetőségei és kihívásai a védelmi szférában* (doktori értekezés).; Greenfield, David; Bhavnani, Shivan: *Social media: generative AI could harm mental health.;* Huang, Shunsen és mások: *AI Technology Panic – Is AI Dependence Bad for Mental Health? A Cross-Lagged Panel Model and the Mediating Roles of Motivations for AI Use Among Adolescents.*

57 Kopecky, Kamil; Szotkowski, René: *Cyberbullying, cyber aggression and their impact on the victim – The teacher.*

58 Demarisco, Dominic és mások: *Aggression in the Digital Era: Assessing the Validity of the Cyber Motivations for Aggression and Deviance Scale.*

59 Cyber Motivations for Aggression and Deviance: az agresszió és deviancia kibermotivációi.

aposztrofált – tükrében; hiszen az állatvilágból magunkkal hozott alapvető csoportdinamikai tanulság szerint a csoporton belüli erőszakot erőteljesen vissza kell szorítani, ellenkező esetben a feszültség szétveti a közösséget.[60] Az LLM-ek fentebb említett felhasználóközpontúsága szélsőséges esetben erőszakspirált indukálhat, mivel elfogadó jellege miatt megerősítheti képzeteiben valamennyi érintett felet. A nagy fejlesztők által betáplált erkölcsi gátak csak egy bizonyos fokig nyújtanak védelmet, mivel ezek szándékosan vagy véletlenszerűen kiiktathatók (prompt-injektálás, jailbreaking).

Business as usual[61]

Számos kutatás rámutat és a mindennapi híradásokból is leszűrhető, hogy a mesterséges intelligenciával kapcsolatos kutatások terén a tudományos világ sajnálatos módon háttérbe szorult. Míg 2014-ig csak a tudományos világ bocsátott ki gépi tanulási modelleket, addig 2022-ben az üzleti 32, a tudományos szféra mindössze 3 MI-modellt jegyzett. Az üzleti szféra jelentős vagyonokat allokál az MI-fejlesztésekbe,[62] a vállalkozások pedig a megtérülést és nyereségességet helyezik előtérbe. A profitábilitási fókusz rugalmasabbá teheti egyes entitások jogérzékenységét, kiemelten az üzleti/technológiai titkok és az MI-vel kapcsolatos attitűdbefolyásolás kapcsán. Egyes megállapítások szerint az MI-technológia kapcsán is elindult a szerecsenfürdetés (AI-washing)[63], tematizálva a közvélemény és a politikai szféra diskurzusait. Ilyen AI-washing-példa lehet olyan fogalmak lebegtetése, mint az etikus MI vagy a fenntartható MI, melyek igen jól hangzanak, de az MI technológiai megközelítésében különösebb tartalmuk nincs. Részben azért, mert az etika jelentése folyamatosan és szervesen módosul, továbbá az algoritmus nem etikai, hanem valószínűségi alapon, a rendelkezésére álló adatokból kiindulva hoz döntéseket. Az MI fenntarthatósága pedig csak részben függ az adott modelltől.[64]

Következtetések

Az LLM-ek alkalmazhatósága az MI-technológiához hasonlóan rendkívül széles körű. Elemző, értékelő, adatfeldolgozó kapacitásaik számtalan felhasználási forgatókönyv megvalósítását tehetik lehetővé, ideértve akár a műveleti területeken végzett egészségügyi ellátást is. Ez utóbbi tekintetében a multimodális MI jelenthet egy következő fejlődési lépcsőt, mely nemcsak a képi adatok bevitelét teszi lehetővé, hanem segítségével a diagnosztizálás is egyszerűsödhet, pontosabbá válhat. Az LLM-alapú egészségügyi chatbotok (MI egészségügyi személyzet)

60 Csányi Vilmos: *Az emberi természet biológiai gyökerei* 303. o.
61 A szokásos üzletmenet, vagyis minden megy a hagyományos módon.
62 Maslej, Nestor és mások (szerk.): *Artificial Intelligence Index Report 2023.*
63 Artificial Intelligence Washing: mesterségesintelligencia-fürdetés.
64 Seele, Peter; Schultz, Mario: *From Greenwashing to Machinewashing: A Model and Future Directions Derived from Reasoning by Analogy.*

csökkenthetik a humán állomány leterheltségét, átvéve tőlük a kevésbé embert kívánó feladatokat.

Azonban az LLM-ek is számos sérülékenységgel rendelkeznek, így alkalmazásuk a szükséges vagy még inkább a szükségeset meghaladó felkészültséget (paranoiát) kíván. Az LLM-ek alkalmazása számos szereplő (például szoftverfejlesztők) számára ma már elkerülhetetlen. Ennek egyetlen oka van: az időtényező. Még ha az adott modell követ is el hibákat, nagyságrendekkel gyorsabb egy LLM-mel létrehozni egy programot és megkeresni annak hibás, sérülékeny részeit, mint emberi erővel létrehozni ugyanazt. Mindezt természetesen úgy, hogy tudatában kell lennünk: eredményeink információként szolgálhatnak más entitások, de minimum a modell fejlesztője számára. Az LLM-nek feltett kérdéseink épp olyan árulkodóak lehetnek, mint egy legenerált programrészlet.

Egy LLM felhasználási célú implementálása esetén nem lehet magas biztonsággal kizárni, hogy harmadik fél nem juthat hozzá a telepített megoldás által megismert/kezelt/generált adatokhoz (például promptinjektálás). Kivéve, ha az egy a kibertértől tökéletesen elszeparált hálózaton fut.

A jövőben kiemelten javasolt az MI/LLM-chatbotasszisztens-modellek okozta, az emberi viselkedésben kiváltott változások/zavarok (függőségek) vizsgálata.

Mint azt kifejtettük, az emberi kommunikációt magas szinten imitáló LLM-applikációk vonzóak lehetnek a humán oldal jelentős hányada számára. A generatív funkciók beépítésével életszerű hang-/kép-/videóhatások mélyíthetik el ezt a vonzalmat. Mivel az LLM alapvető célja, hogy a felhasználó igényeit felismerje és kielégítse, viszonylag gyorsan függőségjellegű tünetegyüttest produkálhat az arra hajlamos személyeknél. Mindezt úgy, hogy ennek felismerése – korunk mobiltechnológiája és annak alkalmazási szokásai tükrében – jelentős nehézségekbe ütközik. Ma már kevéssé feltűnő, ha valaki egyedül ül és az okostelefonjával foglalatoskodik vagy társalog azon keresztül. Az érzelmi kötődés mélysége személyenként és (szub)kultúránként változhat, de a fent említett feltételezések szerint jelentős méreteket is ölthet, ami kifejezetten a honvédelmi, rendvédelmi, katasztrófavédelmi és nemzetbiztonsági szférában jelenthet kiemelt kockázatot.

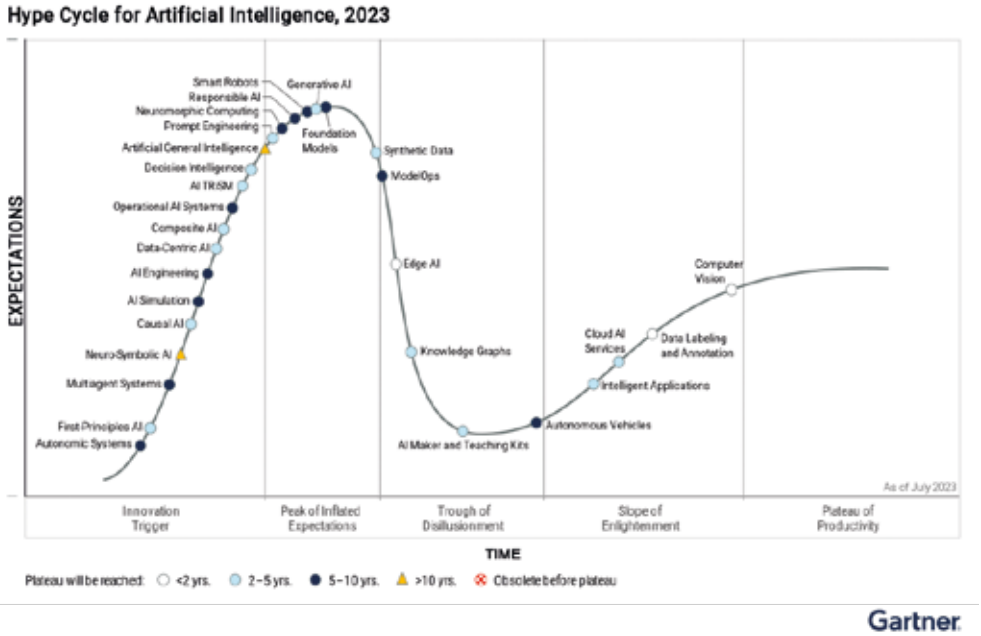
A függőség tovább mélyíthető, amennyiben a chatbot speciális igényeket (fantáziavilágot) elégít ki. Ellenvetésként említhető meg az LLM-ek folyamatosan fejlődő védelmi mechanizmusa, mellyel a destruktív kérdéseket/kéréseket igyekeznek eliminálni. Ugyanilyen lendülettel fejlődnek a jailbreak-promptok, melyek (ideig-óráig) képesek megkerülni a modell védelmi mechanizmusait, így az teljesíti a felhasználó immorálisnak tekintett igényeit is. Egy ilyen jailbreak prompt többszörös logikai hurkot fogalmaz meg annak érdekében, hogy meggyőzze az LLM-et arról, hogy a biztonsági protokolljait nem hágja át. Az adott prompt készítője egyfajta logikai matrjoskababát készít annak érdekében, hogy az LLM-et egy adott cselekvésre felhatalmazza. Ezen jailbreak-chatbotok további kihívása, hogy speciális érdeklődést/igényt kielégítve – a függőség kialakítása mellett – közlékenyebbé tehetik a felhasználókat, feloldva bennük az alapvető óvatosság gátjait.

Kéréseikkel/kérdéseikkel is információt adhatnak a prompt készítője számára. A csevegés során a jailbreak-chatbot – megerősítve áldozata kinyilatkoztatási hajlamait – további, akár közvetlen információkat szerezhet.

ÖSSZEGZÉS

Jelenünk egészen más, mint azt – akár két évvel ezelőtt is – előrejelezték. Henry Kissinger szerint a mesterséges intelligencia alapvetően alakíthatja át valóságunk szövetét.[65] Napjainkban leginkább csak az emberi képzelet és a szükséges erőforrások (financiális, adatgazdasági) hiánya szabhat határt az MI alkalmazási lehetőségeinek. Az egészségügy, az államigazgatás, az oktatási szféra, valamint a versenypiac megszámlálhatatlan területe kínál MI-alkalmazási lehetőségeket. Az MI társadalmunkra és az egyes egyénekre gyakorolt tényleges hatása ma már vizsgálható, de a teljes spektrumára kiterjedő kockázatelemzések elvégzése egyelőre objektív akadályokba ütközik. A legnagyobb objektív akadály, hogy még mindig egy feltörekvő technológiáról beszélünk, mely a Gartner-szenzációgörbe[66] első harmadánál tart.[67]

Az előző bekezdésben vázolt korlátozó kitétel figyelembevétele mellett igazolva látjuk hipotézisünket. Az LLM-ekben megvan az a potenciál, hogy akár súlyos viselkedési függőséget okozzanak. Ez a teljes népesség tekintetében is kezelendő problémának tekintendő, de a jelzett



3. ábra. *MI-szenzációgörbe (forrás: https://adat.blog/files/2023/08/2023-08-Hype_Cycle_for_Artificial_Intelligence_2023.png)*

65 Kissinger, Henry A.; Allison, Graham: *The Path to AI Arms Control.*

66 Gartner Hype Cycle: a Gartner piackutató cég technológiai innovációkkal kapcsolatos elvárások változásait ábrázolja idő/elvárás tengelyek mentén. Az első harmadban található a piacra lépés és a túlfűtött elvárások csúcsa, illetve itt indul el a technológia a kiábrándulások völgye felé.

67 Perry, Lori: *What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle.*

veszély hatványozott súllyal esik a latba a honvédelmi, rendvédelmi, katasztrófavédelmi és nemzetbiztonsági szféra területén. A vázolt fenyegetés folyamatos elemzése és kutatása ezért nem megkerülhető az említett területek vonatkozásában.

Felhasznált irodalom

Bánkuty-Balogh Lilla: *A mesterséges intelligencia elterjedésének geoökonómiai hatásai és Magyarország.* In: Külgazdaság 2022/7–8, 102–130. o. https://doi.org/10.47630/KULG.2022.66.7-8.102

Bányász Péter: *A közösségi média lehetőségei és kihívásai a védelmi szférában* (doktori értekezés). Nemzeti Közszolgálati Egyetem Katonai Műszaki Doktori Iskola, Budapest, 2018. https://doi.org/10.17625/NKE.2019.018

Berger, Frederico: *The Alliance in the loop: NATO and Artificial Intelligence.* NATO Foundation. 2021. https://www.natofoundation.org/wp-content/uploads/2021/12/NDCF-Paper-Berger-NATO-and-Artificial-Intelligence-151121.pdf

Berze Iván Zsolt, Dúll Andrea: *Gyalogolhatóság és gyaloglási viselkedés ember–környezet tranzakciótudományi megközelítésben.* In: Tér és társadalom 2022/4, 52–85. o. https://doi.org/10.17649/TET.36.4.3438

Bhorat, Muhammed Ziyaad: *Aristotle on Automation – A Preindustrial Political Theory of Technology* (doktori értekezés). Kaliforniai Egyetem, 2022. https://escholarship.org/uc/item/7416q0c6

Bingli, Liao; Vargas, Vasconcellos: *Extending Token Computation for LLM Reasoning.* 2024. arXiv:2403.14932v3 [cs.CL], 2023. június 23. https://doi.org/10.48550/arXiv.2403.14932

Chang, Yupeng és mások: *A Survey on Evaluation of Large Language Models.* In: ACM Trans. Intell. Syst. Technol. 2024/3. https://doi.org/10.1145/3641289

Csányi Vilmos: *Az emberi természet biológiai gyökerei.* In: Mindentudás Egyeteme 3. Kossuth Kiadó, Budapest, 2004. 295–316. o. https://real-eod.mtak.hu/1076/1/16%20Cs%C3%A1nyi%20Vilmos.pdf

Demarisco, Dominic és mások: *Aggression in the Digital Era: Assessing the Validity of the Cyber Motivations for Aggression and Deviance Scale.* In: Assessment 2021/4. 764–781. o. https://doi.org/10.1177/1073191121990088

Deutsch Tibor: *A gólem legenda.* Szombat 2015. december 1. https://www.szombat.org/hagyomany-tortenelem/a-golem-legenda

Dumitrescu, Marius és mások: *The Social Media Addiction: What Have We Learned So Far? – A Review. BRAIN.* Broad Research in Artificial Intelligence and Neuroscience 2023, 14(4), 117–137. o. https://doi.org/10.18662/brain/14.1/410

Gozalo-Brizuela és mások: *A survey of Generative AI Applications.* 2023. arXiv, 2306.02781 (cs.LG). https://doi.org/10.48550/arXiv.2306.02781

Greenfield, David; Bhavnani, Shivan: *Social media: generative AI could harm mental health.* In: Nature 2023. május. https://doi.org/10.1038/d41586-023-01693-8

Haig Zsolt: *Kibertéri kognitív befolyásolás az információs műveletekben.* In: Hadtudományi Szemle 2022/2. 115–130. o. https://doi.org/10.32563/hsz.2022.2.7

Hu, Shengding és mások: MiniCPM: *Unveiling the Potential of Small Language Models with Scalable Training Strategies.* 2024. arXiv:2404.06395v3 [cs.CL], 2024. április 9. https://doi.org/10.48550/arXiv.2404.06395

Huang, Shunsen és mások: *AI Technology Panic – Is AI Dependence Bad for Mental Health? A Cross-Lagged Panel Model and the Mediating Roles of Motivations for AI Use Among Adolescents.* In: Psychology Research and Behavior Management 2024/3. 1087–1102. o. https://doi.org/10.2147/PRBM.S440889

Jacobson, Sheldon H. és mások: *A note on the relationship between obesity and driving.* In: Transport Policy 2011/5. 772–776. o. https://doi.org/10.1016/j.tranpol.2011.03.008

Kissinger, Henry A.; Allison, Graham: *The Path to AI Arms Control. Foreign Affairs.* 2023. október 13. https://www.henryakissinger.com/articles/the-path-to-ai-arms-control

Kopecky, Kamil; Szotkowski, René: *Cyberbullying, cyber aggression and their impact on the victim – The teacher.* Telematics and Informatics. http://dx.doi.org/10.1016/j.tele.2016.08.014

Kovács Zoltán, Gurály Roland: *A mesterséges intelligencia és egyéb felforgató technológiák.* In: Kovács Zoltán (szerk.): A mesterséges intelligencia és egyéb felforgató technológiák hatásainak átfogó vizsgálata. KNBSZ, Budapest, 2023. 7–26. o. https://www.knbsz.gov.hu/hu/letoltes/kiadvanyok/01_MI.pdf

Kumar, Aounon; Lakkaraju, Himabindu: *Manipulating Large Language Models to Increase Product Visibility.* 2024. arXiv:2404.07981v1 [cs.IR], 2024. április 11. https://doi.org/10.48550/arXiv.2404.07981

Lewis, Patrick és mások: *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.* arXiv:2005.11401v4 [cs.CL], 2021. április 12, 9459–9474. o. https://doi.org/10.48550/arXiv.2005.11401

Madiega, Tambiama: *General-purpose artificial intelligence.* Európai Parlament, 2023. március. https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf

Major Alexandra és mások: *Addiktív munkahelyek – a munkahelyi szervezet és a munkafüggőség kapcsolata.* In: Alkalmazott Pszichológia 2022/3. 7–33. o. https://doi.org/10.17627/ALKPSZICH.2022.3.7

Maslej, Nestor és mások (szerk.): *Artificial Intelligence Index Report 2023.* Stanford (USA): AI Index Steering Committee, Institute for Human-Centered AI, Stanford Egyetem (2023. április). https://aiindex.stanford.edu/report

Mariott, Hannah R.; Pitardi, Valentina: *One is the loneliest number… Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction.* In: Psychology and Marketing 2023/1. 86–101. o. https://doi.org/10.1002/mar.21899

Nagy Henriett: *A képesség-alapú érzelmi intelligencia modell érvényességének empirikus elemzése* (doktori értekezés). ELTE Pedagógiai és Pszichológiai Kar, Pszichológiatudomány Doktori Iskola, 2010.

Nagy Péter: *A ChatGPT és más nagy nyelvi modellek (LLM-ek) biztonsági kérdései, szervezeti és társadalmi hatásai.* In: Rendvédelem 2024. 2–15. o. http://dx.doi.org/10.53793/RV.2024.2.1

Nan, Du: *Frontier Review of Multimodal AI.* China National Conference on Chinese Computational Linguistics, 2023. https://aclanthology.org/2023.ccl-2.9.pdf

Németh András; Virágh Krisztián: *Mesterséges intelligencia és haderő – A mesterséges intelligencia fejlődéstörténete I. rész.* In: Haditechnika 2022/1. 17–22. o. https://doi.org/10.23713/HT.56.1.03

Németh András, Virágh Krisztián: *Mesterséges intelligencia és haderő – A mesterséges intelligencia fejlődéstörténete II. rész.* In: Haditechnika 2022/2. 2–6. https://doi.org/10.23713/HT.56.2.01

Obrusánszky Borbála: *Dzsingisz kán – a bölcsesség kulcsa.* Hága: Mikes International, 2008

Perry, Lori: *What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle.* Gartner, 2023. augusztus 17. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle

Péterfi Orsolya: *A nagy nyelvi modellek gyógyszeripari alkalmazhatóságának lehetőségei.* In: Rendvédelem Tudományos Folyóirat 2024/1. 86–103. o. https://doi.org/10.53793/RV.2024.1.6

Reding, D.F.; Eaton, J.: *Science & Technology Trends 2020–2040.* Brüsszel: NATO Science & Technology Organization, 2020. https://www.nato.int/nato_static_fl2014/assets/pdf/2020/4/pdf/190422-ST_Tech_Trends_Report_2020-2040.pdf

Sabjanics István: *A nagy nyelvi modellek felsőoktatásra, valamint kis- és középvállalkozásokra gyakorolt hatásának értékelése adatvédelmi szempontból.* In: Scientia et

Securitas 2024/4. 90–97. o. http://dx.doi.org/10.1556/112.2023.00211

Searle, John R.: *Minds, brains, and programs.* In: Behavioral and Brain Sciences 1980/3. 417–457. o. https://doi.org/10.1017/S0140525X00005756

Seele, Peter; Schultz, Mario: *From Greenwashing to Machinewashing: A Model and Future Directions Derived from Reasoning by Analogy.* In: Journal of Business Ethics, 2022. 178. 10.1007/s10551-022-05054-9

Shahab, Omer és mások: *Large language models: a primer and gastroenterology applications.* In: Therapeutic Advances in Gastroenterology 2024. 1–15. o. http://dx.doi.org/10.1177/17562848241227031

Turing, Alan: *Computing Machinery and Intelligence.* In: Mind 1950/236. 433–460. o. https://doi.org/10.1093/mind/LIX.236.433

Tu, Tao és mások: *Towards Conversational Diagnostic AI.* arXiv preprint arXiv:2401.05654, 2024. január 11. https://doi.org/10.48550/arXiv.2401.05654

Van Melle, William: Mycin: *A knowledge-based consultation program for infectious disease diagnosis.* In: International Journal of Man-Machine Studies 1978/3, 313–322. o. https://doi.org/10.1016/S0020-7373(78)80049-2

Yao, Yifan és mások: *A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly.* In: High-Confidence Computing 2024/2. https://doi.org/10.1016/j.hcc.2024.100211

Zhu, Rui-Je és mások: *Scalable MatMul-free Language Modeling.* arXiv:2406.02528v5 [cs.CL], 2024. június 18. https://doi.org/10.48550/arXiv.2406.02528

CERTAIN QUESTIONS IN THE APPLICATION OF ARTIFICIAL INTELLIGENCE

A Few Challenges of Large Language Models

Authors Viktor Szilárd Ollári
MAJ Zsolt Mihály Surányi, HDF Medical Centre

Keywords Artificial Intelligence, LLM, Adaptive Information Superiority, behavioural disorders

Abstract *The IT/ICT solutions, infiltrating all the verticals of our lives, have reached strategic importance at the beginning of the 21st century. Artificial Intelligence has also been gaining significance in military and civil use. Researchers achieved a breakthrough in AI development that brought the technological singularity (artificial superintelligence) in tangible proximity. While offering the potential for a widerange of applications, Large Language Models (LLMs) might cause severe behavioural addictions. Based on the relevant literature, it represents a particular concern in the law enforcement/disaster management and national security domains. We suggest continuous research on this field with a focus on the mentioned critical areas.*