VÉGH JÁNOS

Mi is valójában a mesterséges intelligencia?

What is the Artificial Intelligence, actually?

DOI: HTTPS:// DOI.ORG/10.53793/RV.2024.1.2

Absztrakt

A mesterséges intelligencia (MI) az emberek által végzett, nem rutinszerű tevékenységek ellátására készített számítógépes rendszerek neve, de ismertté szöveggenerálási képessége (LLM) révén vált. Meg kell értenünk alapfogalmait és működési elveit, valamint használatának következményeit energiafelhasználási, fenntarthatósági és környezetszennyezési szempontból is. Az MI lehetőségei messze állnak attól, amit feltételeznek, de veszélyei sem akkorák; feltéve, hogy megértjük, helyesen használjuk és szabályozzuk használatát.

KULCSSZAVAK: MESTERSÉGES INTELLIGENCIA, GENERATÍV, CHATGPT

Abstract

Artificial Intelligence (AI) is the name of computing systems imitating non-routine actions of humans. It became famous due to its text generation ability (Large Language Model, LLM), mainly the ChatGPT application. We need to comprehend its base notions and principles, furthermore energy consumption, sustainability, and environment damaging aspects of its operation. The possibilities of AI are far from the presumed ones, but so are also its dangers; provided that it is used and regulated correctly.

KEYWORDS: ARTIFICIAL INTELLIGENCE, GENERATIV, CHATGPT

Bevezetés

Napjaink egyik varázsszava mindannyiunk számára, függetlenül attól, hogy számítógépes feladatokkal (esetleg ezen belül mesterséges intelligenciával) foglalkozó, társ-tudományi, az alkalmazás iránt érdeklődő felelős szakember, érdeklődő kívülálló (aki vagy lelkesedik érte, vagy fél tőle, vagy csak a békés együttélés érdekében szeretne többet tudni) vagy az „utca embere" vagyunk, a mesterséges intelligencia (MI), ami az angol nyelvű kifejezés (Artificial Intelligence, AI) tükörfordítása. Az MI azért nőtt ki a tisztán tudományos tevékenység köréből, mert egyrészt akár komoly gazdasági előnyöket („újabb ipari forradalom") is ígér, másrészt a természetes nyelvekhez idomított felhasználói felületei (a nagy nyelvi modellek (LLM) felhasználásával) tömeges felhasználást tesznek lehetővé és önálló életre keltek, harmadrészt az MI technikai kivitelezésének zárt módú kezelése magában hordozza a kiszámíthatatlan viselkedés veszélyét és az érdekcsoportok általi befolyásolás lehetőségét. A most éppen mindenhatónak és marketing szempontból abszolút szükségesnek gondolt MI-t a „vezetői összefoglalók" alapján mindegyik vállalat szükségesnek tartja versenyképességének megtartásához, a szó szoros értelmében „bármi áron". Nagy szerep jut a magukat hozzáértőnek valló, bizottságokba szervezett szakértőknek és az irányításra kényszerülő vezetőknek. A kiterjedt terület hiányos szakmai ismerete, az üzleti tőke mohósága, a társadalom és az egyének felfokozott várakozása vagy félelme, nem kellően átgondolt fejlesztésekhez és váratlan következményekhez vezet.

Összefoglaljuk a szükséges ismereteket az MI (és főleg az LLM) alkalmazása iránt érdeklődők és azt gyakorlati alkalmazások céljára felelősséggel bevezetni kívánók számára, akik nem tudnak kiigazodni a rajongók és az ellenzők szélsőséges állításai között. Az MI és az LLM felhasználói számára követhető stratégia, hogy a tanulmány általános értékelési szempontjait alaposan megértik és saját körülményeikre „benchmark" céllal (jól ismert tényekre vonatkozóan) tanulmányt készíttetnek a ChatGPT-vel, majd saját tapasztalatukból (sok munkával) döntik el, hogy ott és úgy alkalmazható-e. Amit nagyon célszerű időnként megismételni, mivel a rendszerek folyamatos változásban vannak. Törekedtünk arra, hogy a tanulmány tartalma a kapcsolódó szakmai területekkel nem ismerős érdeklődő számára is érthető legyen

alapszinten. A speciálisabb ismereteket igénylő részek egy különálló, de elérhető Függelék elnevezésű dokumentumba (URL1) kerültek és adnak többlet információt, a fő szöveghez képest minimális átfedéssel. Jelen tanulmány főként az MI világában eligazodni kívánó és nem azt fejleszteni vagy technikai működését megérteni akaró olvasó számára szól.

Mi is az MI?

Az MI-vel kapcsolatban a tájékozódást nagyban megnehezítik az álhírek, a marketing vélemények, a valótlan tények, a valós tények elferdítése, valamint az elérhető információ (jelentős részben már az MI által manipulált) sokszínűsége. Egyik esetben sem lehet figyelmen kívül hagyni az informáltságot (a személyest, a különböző „véleményvezéreket", a média véleményét, a hivatalból nyilatkozókat, a céges érdekeket képviselőket, a tényszerűséget nélkülöző rajongókat és a megrettenteket), valamint az egyre nagyobb keresletet kihasználó, egyre növekvő számú (és sokszor elszabadult fantáziájú) „szakértőt" is beleértve. Arra, hogy hogyan is működik az MI, a magukat szakértőknek tartó művelői és felhasználói – sőt fejlesztői – is egyre kevésbé tudnak válaszolni; szélsőséges értékelésű hatásai pedig kiterjednek az oktatástól a művészetekig, a gazdaságtól a tudományig; felelőtlen fejlesztői által könnyen elérhetővé tett fejlesztőanyagai következtében egyre több területen jelennek meg és befolyásolják életünket (Kovács 2023).

A tanulmány *nem* MI-ellenes, hanem MI-kritikus. Igyekszik feltárni a területtel kapcsolatos felfokozott technológiai várakozással (valamint éppúgy a teljes kiábrándulással) szemben álló másik oldal érveit is. Megpróbálja – a teljesség igénye nélkül – a nagyon sokszor nem egyértelmű alapfogalmakat és összefüggéseket definiálni, azokat egységes és viszonylag átlátható keretbe helyezni, egymáshoz való viszonyukat tisztázni, továbbá az MI-t csupán alkalmazni kívánó szakember számára érthetően megfogalmazni a társ-tudományainak állításait. A tanulmány rövid és koherens módon tárgyalja a szükséges alapfogalmakat és vezeti be az olvasót az elemeiben ismert fogalmak összefüggő rendszerébe, valamint további tájékozódáshoz ad útmutatást az érdeklődő olvasó számára.

Az MI meghatározása

A természetes intelligenciát (TI) általában az emberi létezéshez (vagy legalább fejlett biológiai rendszerhez kötött (Macpherson et al. 2021), biológiai szövetek bonyolult együttműködésén alapuló magasabb rendű viselkedési módnak tekintik, amelynek viselkedését sem a legmagasabb agyi funkciók, sem az elemi neurális műveletek területén nem ismerjük teljes mértékben (Buzsáki 2019). Úgy tekintjük azonban, hogy a biológiai tanulás (az intelligens működés) lényegi része, hogy a rendszer módosítani tudja saját jövőbeli működését, amihez „jövőbe látás", azaz „*intuíció is szükséges*". Lényeges momentum, hogy nem csak egyszerű mennyiségi tanulásként, inkrementálisan, a korábban már látott viselkedésminták valamelyikét (esetleg azok valamilyen kombinációját) használjuk, hanem egy *új és váratlan helyzetre is jól reagáljunk* (nem-inkrementális, ugrás-szerű tanulás). *Augusto és Gambino (2019)* meghatározása szerint *a kizárólag a korábban látott viselkedésmintákon alapuló rendszer nem intelligens: nem szerez új képességet, hanem változatos módokon próbálkozik*. Ennek alapján meg lehet különböztetni *betanulást*, azaz viselkedésminták elsajátítását, ami esetleg több megtanult viselkedésminta valamilyen módon való kombinálását is jelentheti, de a viselkedés csak a már látott mintákra korlátozódik; és *megtanulást*, ami az addigi képességekhez hozzáadja az új helyzetben szükséges viselkedés képességét (Végh 2024). Lényegében a betanított munkás és szakmunkás, ill. az asszisztens és menedzser/orvos/tanár megkülönböztetésével analóg: másfajta tudásról és másfajta képességről van szó. Más szavakkal kifejezve, ezzel teljesen összhangban van az a meghatározás, amivel a ChatGPT (a fejlesztői által összeállított tudásbázis alapján) megkülönbözteti a gépi (és az ennek alapjául szolgáló mesterséges) és az emberi intelligenciát. Mindegyikre szükség van a megfelelő helyen és összefüggésben. Az MI matematikai vonatkozásainak (és a fogalomnak) jó megfogalmazását tartalmazza a Mi az MI? című könyv (URL2), annak ellenére, hogy negyedszázada íródott.

A kiegészítő dokumentumban bevezetünk további, (kényszerűségből) jelzővel ellátott intelligencia fogalmakat is, amik részben a biológiai és technikai rendszerek különbözőségéből fakadnak, részben a biológia működés félreértése miatt használt probabilisztikus, valamint a technikai megoldások következményeként előálló kvázi-probabilisztikus működésre utalnak. A digitális számítógép használatának mind tömegesebb elterjedésével előállt lehetőség egyre több terület ad az emberi intellektus előállítását célzó próbálkozások számára, bár mint Chu– Prokopenko–Ray (2018) magyarázza: „*szigorúan véve, ma sem értjük, milyen elv alapján számol az élő anyag*". Úgy pedig igen nehéz valamit mesterségesen előállítani, hogy nem tudjuk, az eredeti hogyan működik és – az átgondolatlan technikai megvalósítások miatt – egyre kevésbé tudjuk ellenőrizni a megvalósított másolatot, viszont különféle érdekektől motiválva az internet segítségével tömegesen és korlátozás nélkül használjuk.

Az MI, mint tudományos terület

A számítógépet és az agyat tanulmányozó tudományok egymást nem mindig előnyösen befolyásoló kölcsönhatásban állnak egymással (Macpherson et al. 2021; Hassabis et al. 2017). Mindkét területen használják a társ-tudományok, például a matematika, a fizika, az informatika, az elektronika, a számítógéptudomány, az idegtudományok eredményeit. Mindegyik terület képviselői saját területük elsőrangú szakemberei és ezt feltételezik a vonatkozó társ-tudományok képviselőiről is, de magára a társ-területre vonatkozó tudásuk erősen korlátozott. Minden tudományterület valamilyen közelítéseket használ a természet jelenségeinek leírására és ezeknek a közelítéseknek (sokszor csak implicit) érvényességi köre van, ami már sokszor a társ-tudományok képviselőinek figyelmén kívül marad, így esetlegesen a társ-tudományok eredményeit és következtetéseit olyan területekre is alkalmazni akarják, ahol a közelítések már nem alkalmazhatók. Jellemző példái éppen a számítógéptudomány és a bioinformatika. Az MI ezeknek a társ-tudományoknak egymásra hatásából született; mára (legalábbis méreteit és kutatási intenzitását tekintve) lényegében önálló tudományterületté vált és magán viseli az említett „túl-használat" jellemzőit.

Fogalmi tisztázás

A legtöbben már ott elakadnak, amikor magát az MI fogalmat próbálják megérteni, így a kapcsolódó fogalmak megértése reménytelenné válik. Tekintve, hogy már a TI megfogalmazásakor is bizonytalankodtunk, nem is csoda: azt sem tudjuk pontosan megfogalmazni, hogy mihez, és annak melyik részéhez hasonlítjuk. A bizonytalan megfogalmazású TI és a marketing ihlette MI megfeleltetése egymásnak komoly értelmezési nehézségeket okoz (Kar–Kornblith–Fedorenko 2022). 2023-ban talán a legismertebb MI-vonatkozású megvalósítás, a ChatGPT álláspontja (azaz a rendszer tudásbázisát készítő OpenAI fejlesztők által összeállított szövegkészlet tartalma) ellentmondásos. Egyrészt korrektül fogalmaz önmagáról: „*Én* egy *gépi intelligencia vagyok*, *amely mesterséges intelligencia alapokon működik*", azaz egy lényegében tisztán matematikai módszer műszaki megvalósítása. Másrészt viszont, ha technikai rendszerként általa okozott károkról (energiafelhasználás és környezetszennyezés) kérdezik, akkor megtagadja és csupán egy szoftvernek minősíti önmagát, azaz csupán a szoftvert futtató hardver okoz kárt; miért használnak hardvert.

Az MI fogalma

Az MI Fang–Su–Xiao (URL3) által adott definíciója az emberi intelligencia közelebbről nem definiált fogalmához vezet vissza: „Az MI abban különbözik más berendezésektől, amelyeket emberi tevékenységek javítására vagy helyettesítésére terveztek, hogy *nem csak rutin vagy ismétlődő tevékenységek megoldásán alapszik, hanem az intelligencia-szintű emberi viselkedésen is.*" Ez utóbbi meghatározás – lényegében helyesen – kizárja fogalomköréből az egyszerű (főként visszacsatoláson alapuló) szabályozó és vezérlő köröket, az egyszerű osztályozási és mintafelismerési eseteket, a bonyolult döntési fák és optimalizálási eseteket, a statisztikai számítások és mintázatok keresése alapján hozott döntéseket, az idősorok vizsgálata alapján megalapozott előrejelzések készítését stb., viszont fontosnak tartja az emberi viselkedés imitálását. Megfontolandó lenne Augusto és Gambino (2019) meghatározását használni az MI-re is: „az intelligencia egyrészt lehetővé teszi külső ingerekre megfelelő válasz adását, másrészt *további képességek* megszerzését".

Már csak az emberi tevékenységek sokrétűsége is megnehezíti azok imitálásának csoportba sorolását. A megvalósítások technikai és matematikai háttere, valamint társadalmi/technológiai hatásai és népszerűsége/elterjedtsége miatt *az MI-nek lényegében nincs egységes fogalma és kezelése*. Ehhez még hozzájárul a megvalósításban és felhasználásban (főként gazdaságilag és politikailag) érdekeltek marketingje, különböző hiedelmek és várakozások terjesztésében nagyon, tényszerű ismeretek megszerzésében sokkal kevésbé aktív egyének és csoportok tevékenysége, továbbá a fejlesztés gazdasági hasznából részesedni kívánók felelőtlen magatartása és az „érdekes, szórakoztató és hasznos" technológiát „ingyen" használni kívánók érdeklődése. Ennek megfelelően maga a fogalom is gyorsan változik, és a változáshoz a szakmai-technológiai háttérnek van a legkevesebb köze. A technológiai várakozás gyors felfutása idején kiadott 2018-as Európai Unió (EU) meghatározás még főként az automatizálást (elsősorban érzékelés, egyszerű számítások és beavatkozások; osztályozás; matematikai algoritmusok gyors végrehajtása) értette alatta, de jövő idejű megfogalmazással, a hívők erősen csoda-váró hangulatában irreális távlati célokat tűzött ki. Ráadásul (a marketing ihletésű vezetői összefoglalók szintjén) összekavarodnak a csoda-váró rajongói vágyálmok, az ide vonatkozó matematikai módszerek használata és fejlesztése az ezeket megvalósító eszközök használatával. A fogalmat 2022-ben az EU az eredeti jelentéstől teljesen elrugaszkodva, gyakorlatilag a ChatGPT csevegőprogram testére szabta: (tanításhoz) *bemenő adatokat használó és generált tartalom kimenetet előállító, egy bizonyos listában szereplő technológiákkal*

*előállított szoftver*ként határozta meg (ennek alapján a négy év előtti megfogalmazás, az automatizálás, már nem tartozik az MI fogalmába; biztosan nem felel meg az MI eredeti modellezési szándékának sem; továbbá erősen kérdéses, hogy maga az antropomorf ambíciójú MI belefér-e a testére szabott kategóriába). Az európai álláspont nyomán született 2019-es Magyarország Mesterséges Intelligencia Stratégiája (URL4) is „*fizikai eszközigény* nélkül" képzeli el az MI fejlesztését és használatát (és természetesen nem látta előre az említett fogalomváltozást).

Az álláspontok szinte egyetlen közös vonása, hogy a legelső számítógépes korszak túlzó és nagyobbrészt megalapozottság nélküli, időközben már többszörösen csalódáshoz vezető elvárásai alapján tűzik ki céljaikat. Egyszerűen nem akarják tudomásul venni a valóságot tartalmazó szakértői jelentéseket: „*Hat évtizeddel később a kitűzött kognitív képességek még mindig elérhetetlennek tűnnek*" (Jordan 2019). Kétségtelen tény, hogy az évtizedekkel ezelőtti állapothoz képest a matematikai módszerek, továbbá a számítási technológia és erőforrások használata óriási mértékben fejlődött, de ettől a kitűzött cél megalapozottság nélküli vágyálom maradt. Maga a terület azonban a több rendelkezésre álló hardver és adat alapján ambiciózusabb és olyan elvárásokat támaszt, mint legutóbb a „dot.com" buborékkal kapcsolatban; hasonló eredményesség várható. Ezt a várakozást az sem befolyásolja, hogy a manapság legnépszerűbb csevegő alkalmazások (készítői által írott adatkészlet) tisztán „nem"-mel válaszolnak arra a kérdésre, hogy tudnak-e gondolkodni, önálló felfedezéseket tenni, dolgokat/eseményeket előre jelezni, a természetes intelligencia számára megoldhatatlannak bizonyult feladatokat megoldani. A rajongók irreális fantázia-álmai önállóvá váltak és nem hagyják befolyásolni magukat a keserű tények által. Ebben a képben nem jut hely sem a technológia tömeges felhasználásával okozott károknak, környezeti ártalmaknak, energiafelhasználásnak, sem a tömeges társadalmi felhasználás által okozott károknak és veszélyeknek. *Figyelemre méltó, hogy a stratégiákat összeállító és a technológia használatát szabályozni kívánó „szakértők", politikusok és jogászok milyen kevéssé ismerik a technológia elméleti és gyakorlati tulajdonságait, továbbá használatuk következményeit. Megelégszenek az érdekelt cégek által készített, a vezetőknek szánt hangos marketing anyagok ismeretével. Még az sem tűnik fel, hogy – az utólagos felelősségre vonástól félve – a fejlesztő cégek egyre inkább követelik a terület fejlesztésének és használatának szabályozását, azaz saját önállóságuk csorbítását.*

Az MI megítélése

Ez a bizonytalanság nem véletlen. Az EU által elfogadott jelenlegi (2023) MI szabályozás (URL5) elismeri, hogy *nincs általánosan elfogadott meghatározás (jellemző módon gépi intelligenciáról beszél, de MI-t mond).* Hivatkozás nélkül, de egyetért Lighthill (1972) (URL7) fél évszázados kutatási jelentés megállapításával, hogy *fogalmilag sem lehet egységes MI-ről beszélni. Nem csak működési módját és tulajdonságait nem ismerjük, de még a fogalmát sem tudjuk pontosan meghatározni. Bevezetéséről (megvalósíthatóságáról és hatásairól) sem készült tanulmány. R*észleges tiltásának dokumentumában (URL6) is csak az szerepel, hogy a (szakmai) *„előadó a mesterséges intelligencia-rendszer nemzetközileg elismert fogalom meghatározásának használatát szorgalmazza*" (figyelmen kívül hagyva, hogy pár sorral odébb állította, hogy nincs olyan).

A betiltás és a hiányzó fogalmi meghatározás mellett legalábbis furcsán hat, hogy "A mesterséges intelligencia egy kiforrott és használatra kész technológia, amely felhasználható az ipari folyamatok során keletkező, egyre növekvő adatmennyiség feldolgozására." (Ami a vezetői összefoglalóban található marketing szöveg; magának az MI-nek, pontosabban az azt belülről ismerő fejlesztőknek más a véleménye: „egyáltalán nem tükrözi a jelenlegi állapotot a mesterséges intelligencia területén". Ez az állítás sikeresen összemossa a vágyálmokat, a különböző matematikai módszereket megvalósító rendszereket, a különböző technikákat, a különböző alkalmazási területeket; a néhány valóban sikeres módszer, technika és terület álcája mögé rejtve a valójában fogalmilag sem ide tartozókat.) Erre vonatkozóan a magyar stratégia (URL4) helyesen állapítja meg, hogy az „általános MI kutatása fejletlen és bizonytalan" (de az akkor is igaz marad, ha használatát leszűkítjük egy-egy területre). K*ét különböző intelligenciáról van szó: a gépi intelligenciát kutatjuk és az emberi intelligenciát próbáljuk elérni*, lásd még az MI véleményét is a Függelékben (URL1). *Csupán közelítőleg lehet az MI fogalmait a TI hasonló fogalmaira leképezni (Kar–Kornblith–Fedorenko 2022).* A Jordan (2019) által javasolt sorrendfordítás (AI helyett IA, Intelligence Augmentation) egyáltalán nem csupán szójáték. A *számítógépes rendszer óriási ismeretanyaghoz fér hozzá, több adatot tud tárolni, gyorsabban működik, mint az ember; jól kiegészíti és segíti az emberi intelligenciát (természetesen feltéve, hogy helyesen valósították meg), de nem pótolja azt*. Rutin feladatokra és a leggyakrabban előforduló esetekben akár teljes értékű megoldásként használható, de mindenképpen a TI felügyelete alatt.

A működési mód teljes félreértésének (vagy inkább az erre vonatkozó tudatlanságnak) a következménye, hogy a félrevezetett társadalom elitnek mondott rétege olyan követelményeket támaszt az MI-vel kapcsolatban, mint átláthatóság, kiszámíthatóság, elszámoltathatóság, korrektség, bizonyos elvárások és értékek tiszteletben tartása. Amire egy tapasztalati eloszlásokon és matematikai valószínűségeken alapuló rendszer még bonyolult (és TI alapú) „MI algoritmusok" használatával sem képes, és amit csak tetéz a megismerhetőségre (adatkészletek, algoritmusok és technikai eszközök) vonatkozó, céges érdekek védelmére szolgáló korlátozás.

Bár a 70-es években, a már kicsit haladottabb számítógéptudomány korszakában (a fizikai Nobel-díjas) R. P. Feynman (2018) felhívta a figyelmet arra, hogy jobb lenne a „haladó algoritmusok" (advanced algorithms) kifejezést használni a területre. Kicsit részletesebben talán „az ember működésének valamely aspektusához hasonló viselkedést legalább közelítőleg imitálni próbáló matematikai algoritmusok" lenne rá egy jó összefoglaló név. Ezeket az elméleti módszereket pedig csak konkrét (matematikai, informatikai, elektronikai és újabban biológiai) megoldások használatával lehetséges megvalósítani, tulajdonságaik megértéséhez pedig a használt technológia tulajdonságait kell figyelembe venni. Ezen megoldások jelentős korlátozásokat szabnak mind a megvalósítás, mind a felhasználás számára; sőt, a megvalósítás döntő módon befolyásolja a tapasztalható tulajdonságokat.

Az elnevezés helytelen használata annyira elterjedt, hogy a robbanásszerű felfutás idején vezető elméleti tudós is kifakadt ellene: *„Ne hívjunk mindent MI-nek"* továbbá *„valójában miért nem intelligensek a mai rendszerek"* (Pretz 2021). *„A rosszul megválasztott elnevezés megakadályozza a vonatkozó intellektuális és felhasználási problémák megértését"* (Jordan 2019). A kapcsolat fordítva is igaz: az MI módszereinek biológiai relevanciája legalábbis kérdéses (Macpherson 2021). Mindennek ellenére, az elterjedésnek immár hagyományteremtő ereje van, pedig a valódi ok, amint egy újságíró a Google fejlesztői konferenciájával kapcsolatban megfogalmazta (Jordan 2019): *„az MI gyorsan olyan marketing kifejezéssé vált, ami csak annyit jelent 'számítógépek által végzett számítógépes tevékenység'".*

Adatfeldolgozás MI módra

Technikai kivitelüket illetően, a rendszerek jellemzően elektronikus eszközök (változó mennyiségű és nem mindig helyesen értelmezett biológiai imitációval). A jellemző eszköz a számítógép, illetve az azon futó program; de vannak speciális (számítógép-szerű, jellemzően nagyrészt analóg) cél-berendezések is. Amint a Függelék tárgyalja (URL1), a számítási folyamatokat lehet vezérelni adatokkal, amelyek megkeresik a feldolgozásukhoz szükséges utasításokat; közvetlenül utasításokat adhatunk egy-egy adat(csoport) feldolgozására, vagy egy esemény hozza működésbe az előbbiek valamelyikét.

Biológiai lényként az ember is adatvezérelten működik (legalábbis olyan értelemben nincs vezérlőegysége, mint a számítógépnek). Például a szemünkkel látott élmények olyan adatokat generálnak, amelyek bekerülnek neuronjaink hálózatába, ott a rendkívüli sűrűséggel összekapcsolt neuronok hálózatán egy jól meghatározott pályán végig haladva átalakulnak, és cselekvésben vagy emlékezésben nyilvánulnak meg. Lényegében ezt a funkcionalitást szeretnénk megvalósítani technikai elemek (jellemzően számítógép) felhasználásával: pl. a robotok adatból (vezérlés vagy érzékelés) cselekvést valósítanak meg. A mesterséges neurális hálózatok lényegében logikailag ezt teszik, technikai megvalósításuk (ezért működésük és tulajdonságaik is) azonban lényegesen különbözik.

A számítógépek elemi programutasítások sorozatának végrehajtásán alapulnak; egy-egy bonyolultabb feladat megoldására számos elemi utasítást végrehajtó programot kell írni és annak futtatása eredményezi a megoldást. Egy ilyen lehetséges program annak imitálása, hogy a megkapott adatokból – lényegében eléggé általánosan megfogalmazott feladatot végrehajtó program (egy algoritmus) felhasználásával – látszólag programozás nélkül állítjuk elő a kívánt eredményt. Az ilyen (általunk ál-MI-nek nevezett) megoldásban ugyan az adatok vezérlik a feldolgozást, és az „MI algoritmus" módosításán keresztül (az általános megfogalmazást bizonyos kérdések esetében speciálissá téve) módosítani tudjuk a számítógépes programot, de az minden módosítástól bonyolultabbá válik és működése egyre közelebb kerül ahhoz, amit közvetlen programírással (TI alapján) el tudnánk érni. (Vegyük észre, hogy három szinten is elrejtjük a TI-t. Az adatkészletet TI készíti (a nyers szövegeket és a szövegösszefüggéseket is); az algoritmusokat TI írja; az újabb keletű, ún. „prompt engineering" szakma pedig immár nyíltan a TI beavatkozása: ha még az előző két elemmel nem működik kielégítően, akkor a paraméterek „finomhangolásával" a TI beavatkozik, hogy (arra az egy kérdésre) minél tökéletesebb válasz szülessen.) Az algoritmus-vezérelt működésnek mind a megvalósítást, mind a működést, mind annak eredményességét illetően komoly korlátai vannak. Az egyik alapvető probléma az ún. skálázhatóság: a felületes hasonlóság alapján készített („játék"-szintű) modell egyáltalán nem működik valós méretű problémák megoldásakor.

Turing-teszt

Alan Turing javasolta az intelligencia elfogadhatóság mérésére azt a tesztet (bár inkább odavetett ötletnek, mint rendszeresen használható tesztnek szánta), hogy a mesterséges rendszer akkor mondható intelligensnek, ha egy ember egy rövid társalgás során nem tudja eldönteni, hogy emberrel vagy géppel lépett kapcsolatba. Azaz, a Turing által eléggé nagyvonalúan megfogalmazott emberi szereplő intelligenciájától, a kommunikáció eszközeitől, a feltett kérdések körétől, az elérhető adatkészletek körétől és időszerűségétől, a csevegés időtartamától, a megengedett „gondolkodási" időtől stb. is függ a teljesítés. A kérdésfeltevés lényegében értelmetlen, lévén, hogy két teljesen eltérő, csupán nevében egyező intelligenciát hasonlítunk össze. A Függelék szerint (URL1) az ún. generatív eloszlásfüggvénnyel dolgozó rendszerek definíció-szerűen átmennének, viszont az emberek többsége nem feltétlenül menne át a Turing-teszten.

A Turing-teszt sikeres teljesítésének demonstrálására irányuló lobbizás igen erőteljes. Az érdeklődést felkeltő publikus *„A ChatGPT teljesítette a Turing-tesztet"* (Biever 2023) (ezzel szemben áll a ChatGPT önértékelése, lásd a Függelékben (URL1)) szalagcím alapján a már csak korlátozottan elérhető közlemény címe *„Könnyű intelligencia tesztek, amelyeket a ChatGPT nem tud teljesíteni"*, a közleményben pedig az áll: *„a GPT-4 és más alapú rendszerek valószínűleg átmennének a népszerű Turing-teszten olyan értelemben, hogy az emberek nagy részét bolonddá teszik, legalábbis egy rövid csevegés során"*. A szöveggenerálásra elég jó hasonlat: *egy jó minőségű márkás árucikk helyett gyártsunk olyan utánzatot, amelyik rövid időn keresztül meg tudja téveszteni azokat az embereket, akik nem ismerik az eredetit*. Aminek természetesen csak akkor van értelme, ha el tudjuk adni eredetiként (a párhuzam a ChatGPT felhasználásával nem véletlen). Munkavállalók MI-vel való helyettesítésére is igaz.

Az értékelés szerint az MI alapú rendszerek nem felülmúlják az embereket, hanem a használt benchmark képességei végesek (és a rendszereket „felkészítik" a tesztek teljesítésére, lásd a 2+2=? feladatot). A „contamination" jelensége (előbb-utóbb felbukkannak a helyes válaszok is a betanításhoz használt adatkészletben) révén a felhasznált nagy mennyiségű adat között megtalálja a helyes választ. *Feltéve, hogy az adatkészletben megtalálható és ott csak helyes válaszok szerepelnek*. Ha nem nyert, „újra húzhat". Ha nem sikerült, az nem újsághír. Aztán a fejlesztők tökéletesítik a rendszert, hogy azt az egy benchmarkot teljesítse.

Az MI története

Az MI szinte az emberiséggel egyidős és kötődik az emberi fantáziához. Az ember nem csak fantázialényeket teremtett, hanem valódi (ember alkotta) teremtmények működését is antropomorfizálta: saját (megtapasztalt és többé-kevésbé megértett) működéséhez hasonlította. Hogy manapság számítógép és elektronika a nyersanyag, az már az első MI „forradalom" idején is magától értetődött. A második idején nyilvánvalóvá vált, hogy a „haladó algoritmusok" nem elegendőek („expert knowledge" és megfelelő interfész is kell), a mostani harmadik „aranykort" pedig elsősorban az elképesztően nagy mennyiségű adat felhasználása motiválta és dominálja; széles körű használatát a számítógépes nyelvészet valóban imponáló fejlődése tette lehetővé. Vegyük észre az alábbiakban: minden eddigi megvalósításban csupán arról van szó, hogy valamilyen trükkel elbújtatjuk a TI eredményeit, és megpróbáljuk a gyanútlan felhasználóval elhitetni, hogy az általa látott eredmény az MI terméke. Végül is, *az emberi intelligencia mesterséges megvalósítása azt jelenti, hogy a mesterséges szerkezetnek antropomorf viselkedést és kinézetet adunk azzal a céllal, hogy a többi embert megtévesszük.*

Nem-számítógépes MI

Az EU (URL14) 2018-ban így definiálta az MI-t: *„A mesterséges intelligencia intelligens viselkedésre utaló rendszereket takar, amelyek konkrét célok eléréséhez elemzik környezetüket és – bizonyos mértékű autonómiával – intézkedéseket hajtanak végre*". Eme meghatározás szerint sok száz év óta, naponta használt rendszer például a vízszint szabályzó, aminek konkrét célja a vízszint állandó értéken tartása és a környezet folyamatos elemzése során a vízszint csökkenésének észlelésekor autonóm módon intézkedik: kinyit egy szelepet és vizet enged a tartályba, majd a vízszint helyreállítása után autonóm módon elzárja a szelepet. A TI a szelepbe van építve. Az ókori vízórák „vezérlőszerkezete" egy felelős rabszolga vagy pap volt (beépített TI).

Több szempontból is jó példa az MI-vel kapcsolatban az a mechanikus szerkezet, amelyiket a zseniális mérnök, Kempelen Farkas 1769-ben mutatott be „sakkozógépként". Sokkal bonyolultabb tevékenységet (de lényegében továbbra is egyetlen célt): gépi megvalósítású sakkpartnert hozott létre. Lényegében az emberi test és az emberi agy részleges modellezése volt a cél – mechanikusan, és valójában a TI bevonásával. Az antropomorf jelleget sakkfigurák mozgatása, a bábu fejének és kezének mozgatása stb. valósították meg; a lényeget pedig, a tudásbázist a TI szolgáltatta: szó szerint elbújtatta a TI-t a látszólagosan

MI megoldásban. Abban a korban már bonyolult óraművek (és általuk mozgatott figurák) léteztek, de nem lehetett velük kommunikálni (nem voltak interaktívak); lényegében ezt az igényt elégítette ki a sakkozógép.

Mai terminológiával úgy mondhatnánk, hogy a sakkozó asztal és a török figura alkották a felhasználói interfészt, a működtető mechanika az igénybe vett infrastruktúrát, a tudásbázist (avagy adatkészletet) pedig a TI (emberi sakkmester) szolgáltatta, közvetlenül, valós időben. Erre sem igaz, hogy az MI legyőzte volna a TI-t: egy jól sakkozó ember győzte le a neves embert, illúziót keltően használva a bonyolult infrastruktúrát. A sakkozó automata új elemként (igaz, nagyon korlátozott módon) *kommunikál* az emberi felhasználóval (Kempelen Farkas egyéb tevékenységei – pl. a beszélőgép alapján – feltételezhetjük, hogy ez volt a valódi érdeklődése). Még akár azt is mondhatjuk, hogy ilyen értelemben az általa készített automata átment a Turing-teszten: az emberek túlnyomó többségét megtévesztette. Itt már jól látható az embernek az az igénye, hogy elfogadja a mesterséges teremtmény korlátozott valódiságát; azzal együtt, hogy látja annak „bábu" jellegét, szögletes mozdulatait, a vizuális és akusztikus kommunikáció hiányát stb. Tudomásul veszi, hogy az illúziónak korlátai vannak.

Illúziót kelteni természetesen szavakkal és képekkel is lehet; feltéve, hogy a felhasználónak szándékában áll elhinni az illúziót; lényegében ezen alapszik a mesék és legendák többsége. Felhasználóinak nagy része számára az MI alapú termékek használata olyan érzést biztosít, mint az „Aliz kalandjai Csodaországban" (Alice's Adventures in Wonderland, 1862) közismert szatirikus fantázia-mese vagy az „Óz, a csodák csodája" (The Wizard of Oz, 1939) főhősének az az élmény, hogy egy olyan fantázia-világba kerül, ahol *az állatok és a fantázia-teremtmények antropomorfizáltak*, továbbá a környezet és az események a valós és a szürreális keveréke. Vegyük észre, hogy ezek a teremtmények kommunikálnak (azaz beszélnek és viselkednek), továbbá – bár furcsán – gondolkodnak.

Egy másik irodalmi párhuzam lehet a híres regény „Don Quijote de la Mancha" két világa: az elképzelt lovagi és a valóságos világ. Ahol szintén keverednek a valós és szürreális elemek: az ábrándképben szereplő óriások a valós világban szélmalmok és az ábrándképben rohamozó hős lovag valójában egy ábrándkép által bolonddá tett, jó szándékú, naiv ember. Itt a társadalom bolonddá tett része játssza Don Quijote szerepét és az MI-ból az „intelligencia" szóhoz kötődő ábrándkép a lovagregények légkörét. Az ember, a szélmalom és a rohamozás is valóságos, csak szerepük látszik másnak a két világban. Saját világában mindegyik konzisztens; a baj (ott is) akkor következik be, amikor a képzelt és a valóságos világ keveredik (komoly célra, munka-segítőként használjuk). Jelen tanulmány szerzőjének jut Sancho Pansa szerepe: igyekszik a szegény, bolonddá tett társadalmat a valóság talajára visszarángatni.

Számítógépes MI

A ChatGPT-t létrehozó OpenAI világosan megkülönbözteti a kizárólagosan elméleti erőforrásokat használó „mesterséges intelligencia" és a technikai erőforrásokat használó „gépi intelligencia" fogalmakat. Mai felhasználói, stratégiakidolgozói, szakértői stb. viszont nem tudnak ilyen megkülönböztetésről. Ezért is fontos lenne a fogalmak tisztázása. A két fogalom összetévesztéséből származik a magyar MI stratégiában (URL4) is tetten érhető nagyon naiv álláspont, miszerint *„A mesterséges intelligencia egy globális technológia, határok és fizikai eszközigény nélkül"*, valamint a – különösen a globális energiaválságban szenvedő korunkban – fájó tévhit, hogy *„Ezúttal azonban nincs a fejlődésnek természeti erőforrás-igénye"*. Aminek maga a Stratégia szövege is számos ponton ellentmond: *„Cél a kutatási és általános fejlesztési feladatok elvégzéséhez szükséges erőforrások központi infrastruktúrával való kiszolgálása, a létrejövő infrastruktúra igény szerinti folyamatos fejlesztése, hosszú távú fenntartása."* Amennyiben csupán elméletileg gondolkozunk különböző módszerekről, akkor valóban csak gondolkodó fők által kifejlesztett matematikai módszerekről kell beszélnünk; ez az MI. Ha azonban azokat technikai eszközökkel (jellemzően számítógépes programokkal) valósítjuk meg, akkor a képbe bevonul a technikai megvalósítás a természeti törvényekből következő korlátaival együtt, ami a *gépi intelligencia*.

Nem lehet az óvatossági és hihetőségi szempontokat sem figyelembe venni, valamint hibákat és sebezhetőségeket sem megérteni az elméleti és technológiai háttér tárgyalása nélkül; különösen, mivel a technológiai fejlődés iránya egyre kiszámíthatatlanabb és kockázatosabb. Figyelembe kell venni energetikai és környezetszennyezési szempontokat, mert az EU 2023-as álláspontja szerint (URL6) is: *„A mesterséges intelligencia-technológiák és adatközpontok nagy szénlábnyommal rendelkeznek a megnövekedett számítási energiafogyasztás, valamint a tárolt adatok mennyisége és a keletkezett hő-, elektromos és elektronikus hulladék mennyisége által előidézett magas energiaköltségek miatt, így még nagyobb lesz a szennyezés."* Pedig az már az MI pár év előtti felfutása előtt is magas volt (Jones 2018).

Sok elméleti tanulmány lehetőségként említi, hogy *az MI csökkentheti bizonyos tevékenységek lábnyomát, annak említése nélkül, hogy az ennek elérésére szolgáló technológia használata mennyivel többet növel rajta*. Egyetlen, az újabb keletű gigantomániás szellemben épített rendszer betanítása több tíz tonna szén-dioxid kibocsátásával egyenértékű (URL8), a GPT-3 több mint ötszázzal. Az ilyen rendszerek betanításának

energiaigénye gigawatt-óra nagyságú és hosszú hetekig/hónapokig tart (Luccioni, Viguier és Ligozat 2022-es cikkükben 118 napot említenek) (URL8). Nem elegendőek a merész célkitűzések és ambíciók; a technológia pontos megértésére és megvalósítására van szükség, a csodavárás nem stratégia. Esetleg a kidolgozók figyelmét fel lehetne hívni pl. arra, hogy a közgazdaságtan szerint a haszon eléréséhez befektetés és termelési költség is szükséges; vagy legalább arra, hogy ingyen ebéd nincs.

A technikai és biológiai számítási módok alapvetően eltérnek; a biológiai működés számítógépes imitálásának elve önmagában több nagyságrendi számítási hatásfokcsökkenést okoz (Végh 2021), ami a teljesítmény fokozására használt megoldások következtében mára odáig vezetett, hogy az igazán fontos problémák (pl. globális felmelegedés, személyre szabott genetikai gyógyítás (Conklin–Kumar 2023)) megoldásának lehetővé tételéhez a számítási hatásfoknak kb. 11 nagyságrendet kellene javulnia. Az egyértelműen az MI által gerjesztett számítási teljesítményigény exponenciálisan növekszik (Mehonic–Kenyon 2022), és a technikai megvalósítás legelemibb követelményeit sem tekintő matematikai eljárások (pl. mélyrétegű tanulás) fejlesztése egyértelműen felelőssé tehető érte. A szükséges energiaigény nagyságrendje már rövid távon is a Földön megtermelhető energia összes mennyiségének nagyságrendjét közelíti (Bailey 2022). Emellett az MI által előidézett szénkibocsátás, vízfelhasználás, környezetszennyezés, disszipáció stb. növekedése jó úton van ahhoz, hogy az emberiség fennmaradásának egyik legnagyobb akadályát képezze (URL8). Lényegében erre utal, hogy „*a gépi tanulás olyan modellt használ, ami költséges, alacsony hatékonyságú és fenntarthatatlan*" (Bailey 2022).

Nagy bátorságra (vagy botorságra) vall egész iparágakat (lényegében jövőnket) MI-alapúvá alakítani, mégpedig kötelező jelleggel, szakmai megalapozottság nélkül, politikai akaratként megfogalmazva. Ami azzal együtt, hogy egyre csökkenő számítási hatásfokkal, egyre növekvő energiapazarlással, vízfelhasználással, környezetszennyezéssel és szénkibocsátással alkalmazzuk, elgondolkodtató; különösen, hogy az adatmennyiség racionalizálása fel sem merül.

Az MI hullámai

Az MI neve lényegében azokat az elvárásokat tükrözi, amiket első kutatói az 50-es években feltételeztek képességeinek határairól: az emberi intelligencia gépi megvalósítása, főként gondolkodás, *magas szintű kognitív képességek modellezése volt a cél* (human-imitating) (Jordan 2019), a társ-tudományok eredményeinek felhasználásával. A vonatkozó társ-tudományok azonban csak alacsonyabb (nem kognitív) szinten foglalkoztak jelfeldolgozásssal és döntéshozatallal. Rögtön az elektronikus számítógép megalkotása után megszülettek a túlzó elképzelések a „*gondolkozó, tanuló és kreatív gépek*" építésére vonatkozóan, bár megalkotói *pszichológiai modellezésre és nem mérnöki eszközökkel létrehozott berendezések készítésére szánták* (a más tudományterületen történő „túlhasználat" újabb példája). Még arra vonatkozóan is eléggé korán megfogalmazódtak kérdések (Turkle 1991), hogy az MI valóban olyan jellegű-e, mint amilyet az emberi agy is használ problémák megoldására.

A megjelenő új technológiákkal kapcsolatos várakozás a tapasztalat szerint nagyjából ugyanúgy változik; a technológiai várakozás görbéje, avagy hype néven ismert. Megjelenésük után óriási a várakozás, egyre több terület fedezi fel azokat és alkalmazza saját területére. Idővel viszont leáll az újabb területek mohó belépése, továbbá a régebben elkezdett alkalmazási területek nem érik el az elvárt (nem feltétlenül reális) sikereket, és a csúcs után következik a kiábrándulás időszaka. Ezután következik a megértés és letisztulás, hogy voltaképpen mire jó és mire nem jó, továbbá az egyes területeken mire képes az új technológia, és ezután valamilyen szinten stabilizálódik használata.

Az MI, mint technológia részben ezzel egyező, részben ilyen értelemben szokatlan viselkedésű. Több hullámban is megszületett (minden szakember generáció újra felfedezte és az éppen rendelkezésre álló technológiával nagyrészt függetlenül megalkotta), de egyes hullámai a hype általános viselkedésének megfelelnek. Az MI hullámaiban azonosíthatjuk a „felfokozott várakozás" csúcsait, amiket törvényszerűen követnek a csalódás hullámvölgyei. Természetesen eltérő technikai és társadalmi körülmények között, az előző hullámból származó „konszolidált használat" alapvonala fölött.

Az első hullám lényegében egyrészt a biológiai és elektronikai működés vélt hasonlósága és a logikai működés gépesítése igényének megvalósításából fakadt, de hamar kiderültek a megvalósítások technikai, sőt elvi korlátai. A második hullám kifulladását a társ-tudományoknak tulajdonították (nem elég fejlett elektronika, nem eléggé jó matematikai módszerek, az agy működésének nem kellő szintű ismerete stb.). A mostani hullám már teljesen más hátteret talált: minden területen jelentős fejlődés következett be. Az általános számítógépesítés miatt megnőtt a számítógépes feldolgozás elterjedtsége és szerepe. Mivel kezelhetetlenül naggyá duzzadt a számítógépekkel termelt és általuk feldolgozandó adatmennyiség, és egyrészt nem is áll rendelkezésre megfelelő mennyiségű és szakértelmű programozó szakember, másrészt igen drága a szakértelmük, felmerült az ötlet, hogy *matematikai módszereken alapuló, megfelelően általános*

szoftver használatával, egyedi programozás nélkül dolgozzák fel az adatokat.

Az általános feldolgozási módszerek működési hatásfoka mindig lényegesen kisebb, mint a speciálisaké; az MI esetében ez akár nagyságrendeket is rontott a feldolgozás hatásfokán. A (főként matematikus) szakértők azonban nem különböztették meg az energiafogyasztással és szén lábnyommal valóban nem rendelkező MI-t, és az ezeket a matematikai módszereket alkalmazó gépi intelligenciát, így az ezzel járó problémákat (amelynek nagy részét a fejlett nyelvi modellek generálták) csak a gépi intelligencia tömeges alkalmazása hozta felszínre.

A várható tömeges felhasználást természetes nyelvet használó interfészek kifejlesztésével akarták segíteni. Ennek segítésére fejlesztették a nyelvi modelleket, amelyek használata később önállósította magát és mára a bajok fő okává vált. Továbbá az internet elterjedtsége miatt jelentősen megnőtt eme adatok elérhetősége és a társadalom (beleértve a társ-tudományokat is) várakozása az MI elvárt jövőbeli teljesítményére vonatkozóan. Ne feledjük el a kockázati tőke profit várakozását sem.

Az MI mostani hulláma kifulladásának 2023 közepétől vagyunk tanúi; az előző két nagy hype alapján, törvényszerűen. A felfokozott várakozás visszaütött. Az „intelligencia" szóhoz fűződő képzetek, párosulva a szakterületek legalábbis merész jövőképeivel, irreális elvárásokat, reményeket és félelmeket generáltak; éppúgy, mint az előző két hullám idején. Ironikus módon, éppen a használat megkönnyítésére fejlesztett nyelvi interfész tette nyilvánvalóvá, hogy az elvárások irreálisak. Mára már a napi sajtó is foglalkozik a felfokozott várakozásból eredő csalódottsággal. A kezdeményezők, támogatók, javaslattevők, gyártók, fejlesztők stb. egymással versengve igyekeznek kihátrálni az MI mögül; azért ügyelve, hogy ne veszítsék el arcukat. Egy dolog nem változott az MI körül: *nem értik valójában mi is, mire való, hogyan működik, mire jó és mire nem jó az MI.*

A gyakorlati (pénzügyi, tudományos, technológiai stb.) hasznot remélt gazdasági szereplők csalódottak. Az ilyen szempontokat nem tekintő, csupán szórakozni vágyó vagy házi feladatának megoldási kötelezettségét teljesíteni akaró stb. nagyközönséget azonban még (vagy egyre inkább?) hatása alatt tartja a varázs, és az MI alkalmazásába befektető gazdasági szereplők futnak a pénzük után. A technikailag megvalósítható *gépi intelligencia* (bármilyen nagy mértékben tökéletesített eszközökkel megvalósítva is) *csupán nevében mutat hasonlóságot az emberi intelligenciához; ezért a kutatások nagy része megvalósíthatatlan célt tűz ki. „A gépi tanulásban óriási adatmennyiség és feldolgozó kapacitás használata sem nem hatékony, sem nem jövőbiztos. Az MI-nak sokkal okosabbá és nagyságrendekkel hatékonyabbá kell válni" (Webber 2022).*

A nyilvános és nagymértékű kiábrándulás oka lényegében a (z LLM alapú) ChatGPT, de a fogalmi tisztázatlanság miatt az egész MI került gyanúba. Nem kellene azonban a fürdővízzel együtt a gyermeket is kiönteni. A betiltás után az EU drasztikusan (bár megértés hiányában, igen rosszul) módosította az MI-vel kapcsolatos álláspontját, ami nyilván maga után vonja Magyarország Mesterséges Intelligencia Stratégiájának (URL4) hasonló értelmű módosítását. A technológiai várakozás görbéjének szokásos menete szerint néhány hónap múlva következhet(ett volna) a konszolidált használat időszaka. Nagy kérdés, hogy a fejlesztők önként vállalt önkorlátozása, a felhasználók fokozódó óvatossága és a hatóságok addig meghozott szabályozásai után hogyan folytatódhat a történet. Annak, hogy az említettek megértenék, hogy mit fejlesztenek, használnak és szabályoznak, semmi jele.

Az MI kategóriái

Bár ma már semmi nyoma annak, hogy emlékeznének rá, 1959 óta létezik (Samuel 1959) az MI-vel kapcsolatos módszerek matematikai alapú felosztása, amelynek használata sok bajt megelőzhetett volna. Ebben a szerző a gépi tanulási módszereket lényegében három kategóriára osztja. A *determinisztikus függvényosztály* jól meghatározott eredményt ad, mint pl. a szorzótábla vagy egy játékban következő legjobb lépés, vagy szükséges vízszint. A *probabilisztikus függvényosztály* valamely jól ismert tulajdonságú eloszlás egy elemét adja eredményül, pl. egy szabályozott mennyiség egy bizonyos érték körül, egy bizonyos szórással rendelkező eredményt ad, vagy egy idősor várható következő eredményét megalapozott matematikai módszerekkel megbecsüli. A *generatív függvényosztály* látszólag hasonló a probabilisztikushoz: ott is valamiféle eloszlás alapján adunk eredményt, de az eloszlás tulajdonságait féligmeddig a felhasználó és a felhasználási körülmények szabják meg, azaz *lényegében semmit nem tudunk róla előre mondani.*

A jelenleg problémákat okozó ChatGPT generatív modellt használ (ezt jelenti nevében a 'G'). A matematikában kevésbé ismerősök számára, némiképp egyszerűsített magyarázatként: a három függvénnyel, három különböző módon határozzuk meg a 2+2=? feladat eredményét. A feladat megoldását a *determinisztikus* esetben a szokásos módon, egzakt matematika használatával modellezzük. Az eredmény többszöri próbálkozás esetén is ugyanaz, jól tesztelhető.

Probabilisztikus esetben gondoljunk pl. a liszt csomagolására. A 2 kg-os csomagolás esetében szinte biztosak lehetünk abban, hogy a mérlegre téve nem 2000 g lesz a mutatott érték a mérlegen, és biztosak lehetünk

abban, hogy sok mérés esetén átlagosan megkapjuk a 2000 g értéket, valamilyen kis szórással. *Ismerjük az eloszlást* és annak tulajdonságait, továbbá elfogadjuk, hogy az esetek néhány %-ában mondjuk csak 1997 g-os zacskót kaptunk. Mind az 1997 g-os, mind a 2002 g-os zacskó esetén az mondjuk, hogy vettünk egy 2 kg-os csomagolású lisztet. Az eredmény minden egyes próbálkozás esetében más; csak a próbálkozások eredményének sokaságáról tudunk statisztikus megállapítást tenni; csak ilyen értelemben tesztelhető.

A generatív eloszlás és annak Turing-tesztje

Már a ChatGPT-re gondolva, részletezzük a generatív eloszlás néhány tulajdonságát. A fogalom megértéséhez gondoljunk a 2+2=? feladat kérdőíves közvélemény-kutatással (vagy éppen egy közösségi fórumon feltett kérdésre adott válaszok összegzésével) történő megoldására.

Első lépésként megkérdezzük az embereket, hogy véleményük szerint mennyi a helyes eredmény. Mondjuk az emberek 30%-ának válaszában a helyes érték 4, 35% szerint 3, 25% szerint 5 és 10% szerint 6. Más közvélemény-kutatás és más megkérdezettek véleménye alapján, más tapasztalati eloszlást kapunk. Akárhogyan is, kapunk egy generatív (tapasztalati) eloszlást, amelyik félkész állapotban tartalmazza a feladat megoldását. *Nem ismerjük előre magát az eloszlást, de meg kell adnunk, hogyan készítünk belőle választ*. Van ugyan átlagértékünk is, meg szórás is, meg az első függvény módszere alapján ismerhetjük a helyes értéket is; de mi lesz a feladat eredménye?

Az eloszlás átlagértékét nem adhatjuk meg eredményként, mert az eredmény csak egész szám lehet. Ezért adjuk meg a kérdőíven feltett kérdésre kapott válaszok valamelyikét (vegyük észre: feltételezzük, hogy a helyes válasz is közöttük van). Első próbálkozásként vegyük a leggyakoribb választ: 2+2=3. Az így megadott válasz ugyanezzel a módszerrel, egy másik felmérés szerint, más érték is lehet, a mi programunk szerint viszont a mi kérdőíves betanításunkkal létrehozott adatkészlet használatával a helyes válasz mindörökké 3. Azaz, programunk minden esetben hibás választ ad. Az algoritmus ugyan jó, csak rosszul választottuk meg a betanításhoz használt adatkészletet.

A *probabilisztikus* modellhez való hasonlóság alapján mondhatjuk azt is, hogy válasszuk az adatkészlet valamelyik elemét válasznak. Sem a kérdezőbiztosnak, sem a generatív modellnek, sem a válaszadónak nem kell tudnia, hogy erre a kérdésre csak egy helyes válasz van. Végül is, egy „intelligens" társadalomban többféle helyes válasz is létezhet, sőt, a társadalom valamelyik intelligens tagja is adhat a körülményektől vagy az időtől függően más válaszokat ugyanarra a kérdésre. Azaz, valamilyen módszerrel, alkalomszerűen kijelöljük a kérdőíven megadott valamelyik választ, és azt adjuk meg válaszul a feltett kérdésre. A válasz a TI egy képviselőjétől érkezett, tehát a kérdező a TI által adott választ kap; azaz olyat, amilyet egy találomra kiválasztott embertől is kaphatna. A program tehát pontosan ugyanúgy viselkedik, mint ahogyan találomra kiválasztott emberek viselkednének. Teljesíti a Turing-tesztet: nem tudjuk megmondani, hogy a válasz közvetlenül egy embertől vagy a válaszokat összesítő programtól érkezett.

Nem mellesleg: ha a helyes érték is a lehetséges válaszok között van, a mi programunk már nem nulla valószínűséggel ad helyes választ. Azaz, valakinek rögtön az első esetben, másnak akkor sem, ha tízszer megkérdezi. Sőt, matematikailag korrekt értelemben bejelenthetjük, hogy a mi programunk 30% valószínűséggel helyes eredményt ad, pontosan ugyanolyan arányban, mint a TI. Azaz sikerült gépi módon előállítanunk az intelligenciát, ami *egyetlen* feladatban teljes körűen, Turing értelmében imitálni tudja az emberi intelligenciát. Igaz, hogy csak egyetlen feladat megoldására, de ha hozzáadjuk a 3+2=? feladat megoldásának kérdőívét is, akkor már két feladatra is kiterjed. További kérdőívek hozzáadásával előbb-utóbb eljutunk oda, hogy a „feladatok széles körében" tudja imitálni az emberi intelligenciát. Lényegi különbség azonban nincs.

Ha elég buzgók vagyunk, akkor a generatív függvényt akár „kreatívvá" is tehetjük. Az ugyanis akár véletlen is lehet, hogy a helyes válaszok a 3, 4, 5, 6 számok közül kerülhetnek ki (igaza van, ez az adott kérdőív kitöltőitől függ: nem felelet választós). Miért ne lehetne akár 2 vagy 7 a helyes válasz? (némelyek ezt az „intuíció" fogalommal tévesztik össze, mások egyszerűen kitalálásnak, hallucinációnak tekintik.) Azt ugyan először szomorúan vesszük tudomásul, hogy a 3+3=? feladat kérdőíve még továbbra sem szerepel programunk tudásában, de a kreativitás engedélyezésével már erre a kérdésre is kapunk választ. Végül is, látjuk, hogy két számról, azok összeadásáról van szó, meg két kérdőíves adatkészletről, továbbá a kreativitás is megengedett. Így az utóbbi kérdésre is rávághatjuk azonnal a választ. Még akár az is előfordulhat, hogy helyes választ ad a programunk, legfeljebb a találati pontosság kisebb lesz. Viszont nem kell beismernünk, hogy csak bizonyos számok összeadásáról van ismeretünk (legfeljebb azt, hogy csak összeadásról). Baj csak akkor van, ha a kérdező tudja a helyes választ. Különben abban a meggyőződésben kel fel a képernyő elől, hogy programunk mindenre tudja a választ (legalábbis úgy csinál). Ha egy nem-tanítható változatról van szó, akkor ez a verzió megbukott. (Bár az előbbiek szerint ez nem biztosan derül ki: ha a készítők szerint 30%-os pontossággal működik, akkor a

nyilvánvalóan hibás válasz ellenére is el kell fogadnunk, hogy maga az MI rendszer jó. Sőt, még az emberi tévedést is imitálni tudja).

Eddig „MI algoritmus"-unk kizárólag az eloszlás kezelésére szorítkozott. Ha azt látjuk, hogy sokan tesznek fel ilyen jellegű kérdéseket, akkor készítünk egy „MI algoritmus fejlesztést". Kiegészítjük az algoritmust azzal, hogy készítünk egy összeadó szubrutint (a determinisztikus függvény szerint), és ha két egész szám összeadását kéri a felhasználó, akkor megértjük a két számot, elvégezzük az összeadást, megadjuk az eredményt, és az újabb verzió már átmegy a vizsgán (teljesíti a benchmarkot). Sőt, tökéletesen működik. Mármint két szám összeadását tökéletesen végzi (viszont az MI használatával ágyúval lövünk verébre). A javulás nyilvánvaló: ha többnyire ilyen feladatokat adunk, akkor jelentősen javul a rendszer pontossága. Igaz, hogy a TI két ponton is belenyúlt a generatív módon működő MI rendszer működésébe: felismerte, hogy milyen célra akarjuk használni az MI rendszert, és annak algoritmusába beépítette a TI által, a determinisztikus módszer alapján működő összeadó szubrutint. Aztán a bemutatón azt mondjuk, hogy mindezt az MI csinálja, hiszen javítottunk az MI algoritmuson.

Az adatkészlettel kapcsolatos probléma egyúttal rámutat a tanulás egyik nehézségére is. Ha folyamatosan tanuló rendszert használunk, amelyik a kapott válaszokat beépíti adatkészletébe, akkor a tudatlan vagy maliciózus felhasználók válaszai vagy akár a gépi tanulás idő- és sorrend viszonyai (lásd Függelék alapján (URL1)), az adatkészleten alapuló tapasztalati (generatív) eloszlás függvény folyamatosan változik, és egy idő után még jó eloszlásfüggvénnyel indítva is elkezdhet rossz válaszokat adni, éppen a tanulás következtében. Esetleg éppen a bemutatón siklik ki (lásd URL9). A generatív eloszlásfüggvény tulajdonságait nem ismerjük előre (ellentétben a probabilisztikussal); azok a „kérdőíveken" megadott válaszoktól függenek, tehát esetlegesek. Időben változhatnak, manipulálhatók is.

Vegyük észre szomorú tanulságként: generatív eloszláson alapuló MI programunk:

- csak olyan eredményt tud adni, amilyet adatként már látott betanítása során
- ha mégis újat alkot, abban nincs sok köszönet
- még ha ismeri is a helyes választ, nem biztosan azt adja meg
- ha pontosabbá akarjuk tenni, TI bevonására van szükség
- még a betanítása során már látott feladatot sem biztosan tudja megoldani
- fogalma sincs arról, mit is csinál: számokkal és valószínűségekkel dolgozik.

Aki az MI generatív módszerét használja, az vállalja a fentieket és magára vessen az eredményért.

A generatív eloszlások alkalmazása médiára

Eloszlásokat generálni természetesen nem csak számokból és kérdőívek alapján lehet. Betűk, szavak, mondatpanelek; mozdulatsorok elemei; zenei hangok; ecsetvonások; hangulati elemek megjelenése: bármi, amit matematikailag többé-kevésbé jól le tudunk írni. (A képek/hangok manipulálása és hamisítása különösen veszélyes. Nem csak egy választási kampányban lehet el nem hangzott mondatokat valamelyik jelölt szájába adni vagy meg nem történt rendezvényen felvett képeket készíteni. Kérdésessé teszi az eddig bizonyító erejű hang és/vagy képfelvételek felhasználhatóságát is. A hangszín, a hangsúly, a hanghordozás, az akcentus, a gesztusok, a mimika stb. mind leírhatók matematikailag. Egy esetleg évek múlva sorra kerülő tárgyaláson nem csak a bíróság, hanem maga a szereplő is elbizonytalanodik: a megtévesztés a halványuló emlékkel kombinálva tökéletes illúziót kelthet.) Az MI jelenleg legtöbb vitát kiváltó és sok veszéllyel fenyegető területe az a természetes nyelvi interfész, amelyet eredetileg a számítógépekkel való kapcsolattartás megkönnyítésére fejlesztettek, de ami azóta önálló életre kelt. Egy számítógéppel kommunikálni önmagában is nagy kihívás és az említett Turing-tesztnek is lényeges része. Részben ezek voltak az okai a természetes nyelvi interfészek fejlesztésére irányuló erőfeszítéseknek, amelyek az eredetihez képest teljesen más irányba fejlődnek. Ma már főként szöveggenerálásra irányulnak az erőfeszítések, főként arra, hogy valóságosnak hassanak (azaz minél jobban megtévesszék az MI rendszer felhasználóját), sokkal kevesebb gondot fordítva arra, hogy a tartalomnak valós értelme legyen. Bár ezt nem igazán szokták az adatvezéreltnek beállított MI rendszerek működésének ismertetésekor hangsúlyozni, a rendszerek adatait feldolgozandó, kézileg (TI által készített) algoritmusokat is be lehet és kell tenni; ezek lesznek az „MI-algoritmusok". Hogy ezek mennyit számítanak, arra jó példa, hogy a *GPT-4 és a GPT-3.5 még éppúgy a 2021 szeptemberi adatkészlettel dolgozik, mint a GPT-3; ennek ellenére lényegesen „intelligensebbnek" tűnik. Igen, az algoritmusába rengeteg új TI került. De még így is kell hozzá „prompt engineering"*.

A nyelvek természetes módon tagozódnak hangokra, szavakra, kifejezésekre; vannak felismerhető szabályok, hangulati- és stíluselemek. Könnyen lehet statisztikát készíteni, hogy milyen hangok vagy szavak fordulnak elő együtt nagy valószínűséggel, és annak alapján csupán valószínűségeket használva, értelmesnek tűnő (mert gyakran így használt) pl. szókombinációkat létrehozni.

Ami önmagában igen nagy számú lehetőséget jelent (pl. az angolban a kb. 40,000 szóból előállított, csupán kétszavas kombinációkkal milliárdnyi lehetőséget kellene vizsgálnunk), amiből azonban bizonyos lehetőségek kizárásával (erre szolgálnak azok a bizonyos nyelvi modellek) észszerű nagyságú kombinációkra korlátozzuk a lehetőségeket. Általában azonban még ezekre sem tudunk egyszerű matematikai algoritmusokat megadni. Ilyenkor próbálunk meg neurális hálózatokat használni a feladat megoldására.

A neurális hálózatok abból az elképzelésből indulnak ki, hogy természetes neuronjaink hálózata matematikai algoritmus nélkül tud megoldani feladatokat, így hasonlónak gondolt mesterséges neuronhálózatokat hoznak létre, amelyekben az elemek egymással kommunikálnak, rétegeket képeznek, működési paramétereiket állítják (erről jó és érthető összefoglalót ad Wolfram (2023)). Hogy miért pont úgy, ahogyan, arra mindig akad egy olyan feladat, amelyre azt lehet mondani, hogy „mert így működik". Másik feladatra már nem (lásd például a 2+2=? feladatot), így egy szűk körű intelligenciát lehet megvalósítani. Ha egy nyelvi feladatra szélesebb körű „intelligenciát" akarunk létrehozni, akkor igen nagy számú (nyelvi modellek esetében milliárdos) mintát kell használnunk a betanításhoz; a megvalósítás hosszadalmas (akár több hónapos) és a siker esetleges. A nyelvi modellek már önmagukban kizárják a lehetetlen kombinációkat, a gépi megvalósítások pedig az ún. gyorsítók használatával tovább csökkentik a megoldások előállításához szükséges időt, a megoldás pontosságát csökkentve. A kereskedelmi szolgáltatók pedig a leggyakoribb kérdésekre adandó válaszokra kész szókombinációkat vagy akár mondatpaneleket állítanak össze, amiket korábban (még a felhasználói kérdés elhangzása előtt) olyanra formálnak, amilyet a TI is használna. Az elvből következően az MI válaszainak összeállításakor semmi olyan alkotóelem nem kerül a válaszba, ami korábban nem szerepelt a betanításhoz használt szövegekben, és annak nagy része már a kérdés elhangzásakor össze volt állítva. A mondatok és kifejezések összeállítása előfordulási gyakoriság alapján történik, de az ún. MI algoritmusok is szerephez jutnak. A fejlesztés arra irányul, hogy a válasz formailag minél valósághűbb legyen; szerencsés esetekben a használati gyakoriságok valamennyire tartalmi összefüggés látszatát is kelthetik. Az egyhangúság csökkentésére változnak a szavak és a szókapcsolatok, a hiányzó tényeket pedig „gyakran együtt használt" alapon másik (látszólag akár oda illő) ténnyel pótolják.

Javasolt osztályozás

Amint fentebb láttuk, a generatív eloszlások használatakor „orosz rulett"-et játszunk: maga a fejlesztő sem tudja megmondani, mi lesz a feltett kérdésre a válasz. Az okát abban találtuk meg, hogy *matematikailag nem ismert tulajdonságú eloszlást használtunk, a válasz kiszámítási algoritmusát az eloszlás ismerete nélkül kell elkészítenünk, és valószínűségi elemek használatával adjuk meg az aktuális választ*. Ez a hatás tisztán elméleti felhasználásban, a csupán matematikát használó rendszerekben is jelentkezik. A technikai megvalósítás további bizonytalansági tényezőket hoz be (lásd Függelékben (URL1)), amelyek kvázi-véletlenszerű működést eredményeznek, gyakorlatilag teljesen átláthatatlanná téve az ún. MI-rendszerek működését; fejlesztők és felhasználók, tanulni és szórakozni vágyók számára egyaránt; beleértve a rendszert munka céljára használni kívánókat is. Felidézve, hogy az MI fejlesztés eredeti célja az emberi viselkedés modellezése volt, az látjuk, hogy a mostani fejlesztés teljesen félreszaladt: inkább irányítani akarja az emberi viselkedést, mintsem hasonlítani rá. A káosz akkora, hogy a kognitív tudomány (ami az eredeti célközönség volt) az AI rövidítést az „Alien Intelligence" (földön kívüli intelligencia) jelentéssel használja (Frank 2023). Már szó nincs arról, hogy emberi intelligenciát más technikai megoldással hozzunk létre.

A vágyálmok alapján lényegében három kategóriára lehet osztani a megvalósításokat. A legtágabbról, az emberi intelligencia teljes körét megvalósító (sőt, a fanatikusok szerint: felülmúló) technikai rendszerekről legfeljebb a realitáshoz egyáltalán nem kötődő rajongók beszélnek. A legszűkebb, az emberi intelligencia megnyilvánulásának egyetlen, vagy csak nagyon kevés vonatkozását megvalósító rendszerek (rutinfeladatok folyamatos végzése, szoros szabálykövetés, lehetséges esetek felmérése, osztálybasorolás, nagy mennyiségű számítás gyors elvégzése) akár az embernél jóval gyorsabbak lehetnek, hibátlanul dolgoznak, nem fáradnak. A középső, az előbbinél jóval ambiciózusabb rendszerek többre törekszenek, de képességeik és eredményeik egyáltalán nem állnak arányban kitűzött céljaikkal (a jelen történések leginkább az ebbe a kategóriába tartozó rendszerekre vonatkoznak). Nem véletlenül, ugyanaz történik, mint amit az MI első hullámát lezáró jelentés (URL7) említ: a kevéssé sikeres és perspektivikus területek igyekeznek a sikeres és perspektivikus területek álcája mögé bújni, és az „egységes MI" kép alkalmazásáért harcolni. Ami törekvés a „vezetői összefoglaló" szintjén sikeres is. Ezeknek a területeknek érdeke, hogy a működés „fekete doboz" (URL10) jellegű legyen: ha tudjuk, mi és hogyan történik, el tudjuk dönteni, hogy az jó-e nekünk. A közvetlen üzleti érdekek indokolják, hogy a fejlesztett feldolgozási algoritmusok, a válaszadáshoz használt adatkészletek és a felhasználó által kapott válaszok előállításának módja mind céges titoknak számítanak.

A felhasználási területek alapján lehet talán a legrosszabbul és legkevésbé megalapozottan osztályokra bontani az MI-t, és csupán a tiltás fokára vonatkozó kategóriákba sorolni az alkalmazásokat, mint ahogyan az EU jelenleg tette (URL11). Egy jellemző példa, hogy a „magas kockázatú" csoportba sorolja a „veszélyes viselkedést ösztönző hangtámogatást használó játékokat", de a „minimális kockázatú' csoportba az „MI-kompatibilis videojátékokat". Közben van még két kategória. Emberek sorsára vonatkozó döntést hozni MI alapján a kormányoknak tilos, pl. magáncégeknek szabad. Teljesen kivonja a szabályozás alól a katonai, belügyi stb. alkalmazásokat. Valószínűleg amúgy sem lenne rá hatásköre. Az a tény, hogy az oktatás és ismeretterjesztés – az EU bizottsági javaslat ellenére – nem tartozik abba a kategóriába, amelyikben az MI eszközeivel valótlanságok, téves következtetések, tévinformációk terjesztése stb. tiltott, komoly kockázatot jelent; önmagáért beszél. Sokkal értelmesebb lenne a működés elve alapján részekre bontani az MI területeit, mint a felhasználási terület szerint, a működési módtól függetlenül (lásd Függelékben (URL1)).

Használhatóság

Az MI használhatóságát sok mítosz övezi. Mint említettük, lényegében azt az illúziót igyekszik kelteni, hogy emberi partner; a fejlesztés nem működésének javítására, hanem az illúziókeltés tökéletesítésére irányul. Emberi tulajdonságokkal úgy közelíthető, hogy egyrészt igen nagyképű (mindenhez hozzászól és részletesen kifejti véleményét, igyekszik szakértőnek látszani); gátlás nélkül kitalál nem ismert (csupán nyelvtani értelemmel rendelkező) tényeket, és azokat a valódi tényekkel összeötvözi (bár tagadja, hogy ezt tenné); nem tudja megadni a forrásokat, hogy honnét származik az információ; ismételt és konkrétabb kérdések esetén elbizonytalanodik. Valójában úgy állítja össze „mondanivalóját", ahogyan azt feltételezése szerint a kérdező hallani szeretné. Aki valóban innét szeretne tájékozódni, azt akár teljesen félrevezetheti; ráadásul a változatosság érdekében a látszólag változatlan körülmények között (szándékoltan véletlenszerű, szándéktalanul technikailag kvázi-véletlenszerű vagy a használat idejétől és körülményeitől függő mértékben) akár homlokegyenest eltérő válaszokat adhat. Ha az érdeklődő saját maga által jól ismert tényekre vonatkozó kérdésekkel teszteli az MI tudását, két-három kérdéssel kideríthetí, hogy az imponáló válaszkészség mögött előre elkészített sablon nyelvtani elemek állnak, és a válaszoló nem ismeri az általa használt szavak és mondatelemek jelentését, csupán a gyakori használat alapján tippel.

A működés mögött egy korlátozott (bár elképesztően nagy) méretű adatkészlet és annak készítőinek emberi intelligenciája áll, az MI algoritmus pedig olyan, amilyet a készítés időpontjában a készítők a kérdés feltevésének körülményeire előre láttak. Amint valami hiba kiderül, a fejlesztők (TI) javítanak az algoritmuson és a javított algoritmus már teljesíti a célul kitűzött tesztet. Úgy tűnik, a skálázhatóság megértése sem erőssége az MI hívőknek. Természetesen ma mindenki, mindenre az MI-ben találja meg a csodaszert, és „minél nagyobb annál jobb" elven épült rendszereket akar használni. A fejlesztése (valószínűleg nem véletlenül) a szoftver fejlesztésnek ahhoz a módjához hasonlít, amikor először kitűzzük a teljesítendő feladatot (milyen benchmark teljesítése), aztán addig gyötörjük az algoritmust és az adatokat, amíg a tesztprogram sikeresen le nem fut. Ami természetesen csak annyit bizonyít, hogy az illető tesztet sikerrel vette a program. Egy másik teszt természetesen másik hibák kijavítását teheti szükségessé. Ha az nem sikerül, csak az első benchmark sikeres teljesítését tesszük közzé. Természetesen folyik a rakéta-ellenrakéta fejlesztés: a ChatGPT3.5 által még nem teljesíthető teszteket a ChatGPT4 már teljesíti, hiszen ismert a hibák oka. Közben azonban megszületik a másként tesztelő benchmark, amely már a ChatGPT4 számára is nehéznek bizonyul (URL12). Figyelni kell az értékelések megfogalmazására és az értékelők mögött álló (többnyire üzleti) érdekekre. Orvosi vizsgakérdések megválaszolásában a *GPT-4 lényegesen jobban vizsgázott, mint a GPT-3.5,* de még mindig alatta marad *egy átlagos emberi orvostanhallgató* teljesítményének (URL12). Ami nem áll ellentétben a GPT-4 technikai bejelentésével (URL13), miszerint felülmúlja *az emberek többségének* vizsgaképességét. Az *emberek többsége* valóban rosszul teljesítene az *orvostanhallgatók* tudásának mérésére szolgáló teszten. Bemutatunk azonban két olyan esetet, ahol a követelményszint magasabb és a benchmarkot nem lehet manipulálni. Az elsőben a természet fejleszti a benchmarkot, a másodikban a technológia.

A COVID-19 pandémia kezelése

Sajnos a közelmúltban előfordult az az esemény, ami „élesben" tesztelte az MI képességeit. Ilyen szempontból nagyon jó időzítéssel, 2019-ben, a hype (technológiai várakozás) erősen felfutó periódusában, a matematikai módszerek és az elektronikai eszközök teljességének birtokában, a módszerek és eszközök használatában jártas nagy számú kutató jelenlétében, a COVID-19 pandémia előidézte azt a társadalmi szükségletet, hogy az MI is harcba szálljon a járvány megfelelő kezelésének érdekében. Ehhez a kormányoktól megkapták a szükséges anyagi- és emberi erőforrásokat is.

Illusztratív példaként szolgál ama számos eszköz esete, amelyeket a COVID-19 járvány diagnosztizálására és előrejelzésére fejlesztettek ki, és amelyek megvalósítására 2019. és 2020. során jelentős erőforrásokat biztosítottak. Ezek analízise világos következtetéseket hozott. A nagy tekintélyű The Alan Turing Institute jelentése (von Borzyskowski et al. 2020) az UK Data Science and Artificial Intelligence közösségnek a COVID-19 járvány elleni küzdelemre adott válaszát ismerteti. Következtetése, hogy bár számos projektben széles skálán történtek fejlesztések, az eredmények nem igazán voltak használhatók, jócskán teret hagyva a jövőbeli járványok sokkal jobb előrejelzésének. Ugyanerre jutott a British Medical Journal-ban megjelent tanulmány, amely a COVID-2019 diagnózisára és prognózis céljára kifejlesztett előrejelző eszközök mindegyikét alkalmatlannak találta klinikai használatra. Mindössze két olyat talált, amelyik további vizsgálatra érdemes. Egy, a Nature Machine Intelligence folyóiratban megjelent tanulmány 415 olyan gépi tanulás alapú eszközt vizsgált, amelyek mellkasröntgen és CT felvételek alapján próbáltak a páciensek esélyeire következtetni. A következtetés megint csak nem biztató: egyik sem találtatott alkalmasnak klinikai használatra.

Úgy tűnik, ez az éles teszt, amiben a résztvevők minden lehetőséget megkaptak (a nagy figyelmet és a kiugrás/bizonyítás lehetőségét is), továbbá szó szerint az életük volt a tét, egyáltalán nem igazolta az EU 2018-as reményeit, hogy „a mesterséges intelligencia segít a világ legnagyobb kihívásainak megoldásában: a krónikus betegségek kezelésében”. Hasonló eredményre jutott a National Institute of Health is: „*Valójában, még ha számos alkalmazás célozta is meg a COVID-19 előrejelzését és diagnosztizálását, közülük csak néhány elég érett arra, hogy klinikai körülmények között alkalmazni lehessen*” (Comito–Pizzuti 2022).

Elektronikus tervezés

Érdekes és stílusos ötlet hibás elvű LLM-et használni a hibás elvű hardver gyorsítóegységek tervezésére (Hemsoth Pricket 2023). Az MI/LMM fejlesztésében érdekelt cégek által támogatott technikai újságírók véleménye szerint is, *még nem sokat látni abból, amit az ilyen elven működő elektronikus fejlesztő rendszerektől elvárnak*. Az eddigi tapasztalatok szerint nem boldogulnak az összetett feladatokkal, nem tudják elérni a zárt forrású (TI által készített) rendszerek által elért teljesítőképességet. A gyorsító rendszerekben használt megoldások ugyanis eltávolítják a szükséges információ egy részét, a látszólagos gyorsabb működés érdekében félbevágják a számítást (lásd Függelékben (URL1)). Ennek következtében a tervezőrendszer éppúgy bugyután viselkedik, mint a csevegő robotok: az általuk tervezett rendszerek is bugyután viselkednek, vagy egyáltalán nem működnek. Továbbá ezek a problémák csak a TI által készített „demonstrációs célú” darabok elkészítésével csökkenthetők, ami másképpen fogalmazva azt jelenti, hogy ha betanításhoz használt adatkészletben nem szerepel a tervezendő „Intellectual Property” (IP), akkor az MI nem fogja azt megtervezni. *Újat nem tud alkotni, csak egy (vagy több) régit megtalálni és abból kicsit átfogalmazva, egy „generált” IP-t készíteni.* Hogy az működőképes-e, az attól függ. Az MI nem tudja, mi a terv célja. Csupán készít valamit „tartalmilag és formailag”, probabilisztikus értelemben, ahhoz hasonlót.

Összefoglalás

A mesterséges intelligencia néven ismertté vált számítógépes technológia egyre inkább bevonul életünkbe. Működése és hatásai egyre átláthatatlanabbá válnak, miközben fogalma is egyre tisztázatlanabb. Olyan számítógép-alapú technológia, amelynek főbb tulajdonságait megérteni csak akkor lehet, ha technikai működésének elveit legalább madártávlatból ismerjük. A tanulmány ebben próbál segíteni. Nem csodaszer, de megfelelő óvatossággal akár segítségünkre is lehet, bár manapság a fejlesztések nem erre irányulnak. Sem a *hívők sokszor tapasztalt (a marketinges „vezetői összefoglaló” anyagokon alapuló) csodavárása, sem a világvége hangulat nem reális. „Az LLM technológia használata kikerülhetetlen, ezért annak tiltása nem megoldás"; A ChatGPT és más LLM rendszerek meggyőző, de sokszor hibás szöveget állítanak elő; használatuk torzíthatja a tényeket és tévinformációkat terjeszthet” (van Dis et al. 2023). Meg kell tanulnunk együtt élni vele: mértékkel, utólagos emberi ellenőrzés mellett, az emberi tevékenységek rutin jellegű részének gyors elvégzésére használni, de semmiképpen nem helyette.*

Irodalomjegyzék

Augusto, E.–Gambino, F. (2019) Can NMDA Spikes Dictate Computations of Local Networks and Behavior? *Frontiers in Molecular Neuroscience*, Vol. 12/238. doi:10.3389/fnmol.2019.00238 https://www.frontiersin.org/articles/10.3389/fnmol.2019.00238/full [Letöltve: 2024.02.16.].

Bailey, B. (2022) AI Power Consumption Exploding. *Semiconductor Engineering*, August 15th, 2022. https://semiengineering.com/ai-power-consumption-exploding/ [Letöltve: 2024.02.16.].

Biever, C. (2023) ChatGPT broke the Turing test – the race is on for new ways to assess AI. *Nature*, 25 July 2023.

https://www.nature.com/articles/d41586-023-02361-7.pdf [Letöltve: 2024.02.16.].

Buzsáki, Gy. (2019) *The Brain from Inside Out*. Oxford University Press. ISBN: 978-0-19-090538-5.

Chu, D.–Prokopenko, M.–Ray, J. C. J. (2018) Computation by natural systems. *Interface Focus*, 19. October 2018. doi: 10.1098/rsfs.2018.0058 https://royalsocietypublishing.org/doi/10.1098/rsfs.2018.0058 [Letöltve: 2024.02.16.].

Comito, C.–Pizzuti, C. (2022) Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. *Artifical Intelligence in Medicine*, Vol. 128:102286. doi: 10.1016/j.artmed.2022.102286 https://www.sciencedirect.com/science/article/pii/S0933365722000513?via%3Dihub [Letöltve: 2024.02.16.].

Conklin, A. A.–Kumar, S. (2023) Solving the big computing problems in the twenty-first century. *Nature Electronics*, Vol. 6. pp. 464-466. https://www.nature.com/articles/s41928-023-00985-1.epdf [Letöltve: 2024.02.16.].

Feynman, R. P. (2018) *Feynman Lectures on Computation*. CRC Press. ISBN 9780429500442.

Frank, M.C. (2023) Baby steps in evaluating the capacities of large language models. *Nature Reviews Psychology*, 2. pp. 451–452. https://doi.org/10.1038/s44159-023-00211-x [Letöltve: 2024.02.16.].

Hassabis, D.–Kumaran, D.–Summerfield, C.–Botvinick, M. (2017) Neuroscience-Inspired Artificial Intelligence. *Neuron*, 95(2). pp. 245-258. doi: 10.1016/j.neuron.2017.06.011 https://pubmed.ncbi.nlm.nih.gov/28728020/ [Letöltve: 2024.02.16.].

Hemsoth Pricket, N. (2023) What happens when LLMS design AI accelerators? *TheNextPlatform*. https://www.nextplatform.com/2023/09/25/what-happens-when-llms-design-ai-accelerators/ [Letöltve: 2024.02.16.].

Jones, N. (2018) How to stop data centres from gobbling up the world's electricity. *Nature*, 13 September 2021. https://www.nature.com/articles/d41586-018-06610-y [Letöltve: 2024.02.16.].

Jordan, M. I. (2019) Artificial Intelligence – The Revolution Hasn't Happened Yet. *Harvard Data Science Review*, Issue 1.1, Summer 2019. doi:10.1162/99608f92.f06c6e61 https://hdsr.mitpress.mit.edu/pub/wot7mkc1/release/10 [Letöltve: 2024.02.16.].

Kar, K.–Kornblith, S.–Fedorenko, E. (2022) Interpretability of artificial neural network models in artificial intelligence versus neuroscience. *Nature Machine Intelligence*, 4. pp. 1065-1067. doi:10.1038/s42256-022-00592-3 https://arxiv.org/abs/2206.03951 [Letöltve: 2024.02.16.].

Kovács, Z. (szerk.) (2023) *A mesterséges intelligencia és egyéb felforgató technológiák hatásainak átfogó vizsgálata*. Katonai Nemzetbiztonsági Szolgálat, 2023. https://www.knbsz.gov.hu/hu/letoltes/kiadvanyok/01_MI.pdf [Letöltve: 2024.02.16.].

Macpherson, T.–Churchland, A.–Sejnowski, T.–DiCarlo, J.–Kamitani, Y.–Takahashi, H.– Hikida, T. (2021) Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research. *Neural Network*, Vol. 144. pp. 603-613. https://doi.org/10.1016/j.neunet.2021.09.018 [Letöltve: 2024.02.16.].

Mehonic, A.–Kenyon, A. J. (2022) Brain-inspired computing needs a master plan. *Nature*, Vol. 604/1. pp. 255-260. doi:10.1038/s41586-021-04362-w https://www.nature.com/articles/s41586-021-04362-w [Letöltve: 2024.02.16.].

Pretz, K. (2021) Stop Calling Everything AI, Machine-Learning Pioneer Says. *IEEE Spectrum*. https://spectrum.ieee.org/the-institute/ieee-member-news/stop-calling-everything-ai-machinelearning-pioneer-says [Letöltve: 2024.02.16.].

Samuel, A. R. (1959) Some studies in machine learning using the game of checkers. *IBM Journal Res. Dev.* 44(1959), pp. 1210–1229.

Turkle, S. (1991) Dangerous Thoughts . . . And Machines With Big Ideas. *The New York Times*, March 17, 1991, Section 7, Page 1. https://www.nytimes.com/1991/03/17/books/dangerous-thoughts-and-machines-with-big-ideas.html [Letöltve: 2024.02.16.].

URL1: *Függelék* https://github.com/jvegh/TimeAwareComputing.github.io/blob/main/ChatGPT.pdf [Letöltve: 2024.03.02.].

URL2: (2000) *Mi a MI? Avagy a mesterséges intelligencia témaköre, célkitűzései*. https://doksi.net/hu/get.php?lid=484 [Letöltve: 2024.03.02.].

URL3: Fang, J.–Su, H.–Xiao, Y. (2018) *Will Artificial Intelligence Surpass Human Intelligence?* http://dx.doi.org/10.2139/ssrn.3173876 [Letöltve: 2024.02.16.].

URL4: (2019) Magyarország Mesterséges Intelligencia Stratégiája. https://digitalisjoletprogram.hu/files/2f/32/2f32f239878a4559b6541e46277d6e88.pdf [Letöltve: 2024.02.16.].

URL5: General-purpose artificial intelligence. https://epthinktank.eu/2023/03/31/general-purpose-artificial-intelligence/ [Letöltve: 2024.02.16.].

URL6: Jelentés a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról szóló

európai parlamenti és tanácsi rendeletre irányuló javaslatról. https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_HU.html [Letöltve: 2024.02.16.].

URL7: Lighthill, J. (1972) *Artificial Intelligence: A General Survey*. https://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm [Letöltve: 2024.02.16.].

URL8: Luccioni, A. S.–Viguier, S.–Ligozat A-N (2022) *Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model*. https://arxiv.org/abs/2211.02001 [Letöltve: 2024.02.16.].

URL9: Germšek, M. (2023) How Overhyped is ChatGPT in 2023? The Truth about its Hype Explained. https://shape-labs.com/articles/how-overhyped-is-chatGPT [Letöltve: 2024.02.16.].

URL10: (2023) ChatGPT is a black box: how AI research can break it open. *Nature*, 619, pp. 671–672. https://www.nature.com/articles/d41586-023-02366-2 [Letöltve: 2024.02.16.].

URL11: Európai Bizottság (2023) A mesterséges intelligenciáról szóló törvény. https://digital-strategy.ec.europa.eu/hu/policies/regulatory-framework-ai [Letöltve: 2024.02.16.].

URL12: Rosol, M.–Gąsior, J. S.–Laba J.–Korzeniewski, K.–Mlynczak, M. (2023) *Evaluation of the performance of GPT-3.5 and GPT-4 on the Medical Final Examination*. https://www.medrxiv.org/content/10.1101/2023.06.04.23290939v1.full.pdf [Letöltve: 2024.02.16.].

URL13: Achiam, J. et al (2023) *GPT-4 Technical Report*. https://arxiv.org/abs/2303.08774 [Letöltve: 2024.02.16.].

URL14: Európai Bizottság (2018) A közös európai adattér kialakítása felé. https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:52018DC0237 [Letöltve: 2024.02.16.].

van Dis, E. A. M.–Bollen, J.–Zuidema, W.–van Rooij, R.–Bockting, C. L. (2023) ChatGPT: five priorities for research. Conversational AI is a game-changer for science. Here's how to respond. *Nature*, Vol. 614. pp. 224-226. https://www.nature.com/articles/d41586-023-00288-7 [Letöltve: 2024.02.16.].

Végh, J.–Berki, Á. J. (2021) Why learning and machine learning are different. *Advances in Artificial Intelligence and Machine Learning*, 1(2). pp. 136-154. doi:10.54364/AAIML.2021.1109 https://www.oajaiml.com/uploads/archivepdf/77931109.pdf [Letöltve: 2024.02.16.].

Végh, J. (2021) Revising the Classic Computing Paradigm and Its Technological Implementations. *Informatics*, 8/4(2021). 71. doi:10.3390/informatics8040071 https://www.mdpi.com/2227-9709/8/4/71 [Letöltve: 2024.02.16.].

von Borzyskowski, I.–Mazumder, A.–Mateen, B.–Wooldridge, M. (eds.) (2020) *Data Science and AI in the age of COVID-19*. Reflections on the response of the UK's data science and AI community to the COVID-19 pandemic. The Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2021-06/data-science-and-ai-in-the-age-of-covid_full-report_2.pdf [Letöltve: 2024.02.16.].

Webber, F. (2022) Third AI Winter ahead? Why OpenAI, Google & Co are heading towards a dead-end. *Cortical.io.* July 18, 2022. https://www.cortical.io/blog/third-ai-winter-ahead-why-openai-google-co-are-heading-towards-a-dead-end/ [Letöltve: 2024.02.16.].

Wolfram, S. (2023) *What Is ChatGPT Doing … and Why Does It Work?* Wolfram Media, Inc. ISBN-13: 978-1-57955-081-3 (paperback), ISBN-13: 978-1-57955-082-0 (eBook). https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/ [Letöltve: 2024.02.16.].