

MI: a misztikumon túl

AI: over mysticism

KILIÁN Imre

Deutsche Telekom Systems Solutions Hungary Kft, Pécs
7621 Pécs, Nagy Lajos király útja 11.
tel: [+49 69 9731792500](tel:+49699731792500)
www.deuschetelekomitsolutions.hu
Imre-Zoltan.Kilian@t-systems.com

Abstract

The emerging AI technologies, based on large language models, are exposing many social, legal, and ethical questions too. We think that these should and can be answered within the corresponding disciplines. Unfortunately, the technology still suffers from childhood diseases. The price of their tolerant and flexible answering ability is their inability to perform strict logical inferences, and their nature to produce sometimes quite senseless answers – hallucinations. To eliminate these, the technology is continuously under refinement – yet we think, the hybrid combination of the oldschool rule-driven technology can provide an ultimate solution.

Összefoglaló

A nagy nyelvi modelleken alapuló MI technológiák rengeteg társadalmi, jogi és etikai kérdést felvetnek, amelyeket jórészt a megfelelő tudományterületeken belül kell és lehet megoldani. Sajnos ezeken túl még komoly gyermekbetegségektől is szenvednek. Az efféle rendszerek toleráns válaszadási képességének ára a következtetési képesség hiánya és a hallucinációra, vagyis a teljesen légből kapott válaszokra való hajlam. Ezek kiküszöbölésére a technológia finomításával is lehet válaszolni, igazán hatékony megoldást viszont inkább a korábbi, szabályvezérelt technológiák bevonásával egy hibrid megoldás adhat.

Kulcsszavak

mesterséges intelligencia, nagy nyelvi modellek, csevegőprogramok

1. BEVEZETÉS

A Turing próbán, mint műszaki-tudományos kihíváson a ChatGPT tavaly őszi megjelenésével lényegileg átléptünk. Azóta viszont folyamatos MI lázban ég a világ. Ezernyi fórumon, ezernyi színvonalon tárgyalják a témát: Mi lesz, ha a MI végleg átveszi az uralmat az emberiség felett? Foglalnánk-e időpontot a robotpszichológushoz? Betegségünkben bejelentkeznénk-e a chatdokihoz, netán elfogadnánk-e a mesterségesen intelligens gépbíró ítéletét perpatvar idején?

El tudna-e végezni egy MI egy egyetemi programot? Le tudná-e tenni a vizsgákat, majd megszerezné-e az oklevelet?

Végül egy nyilvánvalóan megválaszolt – költői – kérdés: Tanulnánk-e a géptanártól? Mert tesszük. Lényegileg Gutenberg, a könyvnyomtatás, de talán már az írás feltalálása óta – a lényeg azóta sem, csak a technológia és a körítés változott. Kicsit provokatívabb azonban a kérdés másik oldala: elfogadnánk-e a MI által kiosztott egyetemi vizsgajegyet?

A kérdések sokféle irányból megközelíthetők. A jelen írásban az etikai kérdésekre is inkább műszaki irányból próbálunk választ adni. Ennek során a szerző a modern csevegőrobotokat a saját, klasszikus, szabály alapú technológiával készített ITSy-Bitsy megoldásával [5] is összeveti.

2. MENNYIRE INTELLIGENS A MI?

Az intelligenciafokot – főleg a pszichológusok – különféle kísérleteken és különféle mérőszámokon (IQ, EQ, stb) keresztül mérik. Informatikusként ezen kísérletek végigvitelére nemigen van mód, de

kíváncsian várom és várjuk a kollégák eredményeit. Előrebocsátom, hogy félek tőle, hogy meztelen a király, vagyis a MI súlyosan levizsgázza.

Egyetemi oktatóként úgy érzem: a csevegőprogramok minden bizonnyal képesek lehetnek bizonyos egyetemi tantárgyak vizsgakövetelményeit teljesíteni – esetleg némi külön felkészítés után – ha a megfelelő kötelező olvasmányokat elolvassák. Elsősorban a leíró jellegű tantárgyakra gondolhatunk – pl. a híresen hírhedt orvosi anatómia szigorlatra, vagy más társadalomtudományi, netán jogi vagy közgazdasági tárgyak vizsgáira.

A MI intelligenciájának egyfajta objektív műszaki mérőszáma lehet a neuronsejtek (vagy a neuronkapcsolatok) száma. Ez a szám nem kicsi. Az IBM Watson modellre épülő szoftver kódátíró alkalmazásában a kapcsolatok (nem 0 súlytényezők) száma 20G ($20 \cdot 10^9$), de hallani T nagyságrendű (10^{12}) modellekről is. [3]

Ezt a számot talán a természetes neuronok és szinapszisok számával célszerű összehasonlítani. Az emberi idegrendszer becslések szerint kb. 10^{11} neuront, és 10^{14} szinapszist tartalmaz, vagyis egy idegsejt átlagban 1000 másik idegsejttel kapcsolódik. [4] A gépi intelligencia tehát kb. 1%-nál jár. Ez nem kevés, hozzátevé, hogy az emberi idegrendszernek csak tört része foglalkozik a beszédfeldolgozással és a gondolkodással, komoly része különböző fizikai és élettani funkciókat lát el.

3. TERVEZŐK, JOGÁSZOK ÉS DÖNTÉSHOZÓK KÖZÖTT

Szögezzünk le egy fontos dolgot. A MI-t emberek fejlesztik, emberek építenek bele különböző lehetőségeket, és ezeket a lehetőségeket más emberek használják jó vagy rossz célokra. A legbékésebb eszközöket is lehet rossz célra használni: egy fél téglá vagy a székely viccek tízcentis bicskája (a székely személyi igazolvány) teljesen elegendő lehet emberi élet kioltására is...

A mai MI megoldások adattömeget alakítanak át egy másik, békés adattömegbe. A tervezők döntenek el, hogy mit mibe alakít át, úgy, hogy épp afféle adattömegeket tanítanak vele. Aztán az eredmény adattömeget valaki támadásnak érezheti, pl. azért, mert valami valótlanságot, vagy csak egyszerűen sértőt, netán mítoszrombolót állít... Bárkiről: a felhasználóról, egy nemzeti hősről, egy vallásalapítóról, stb.

Ugyanúgy a tervezők döntenek el, hogy az eredmény átlép-e ténylegesen a *saját metaszintjén*, több lesz-e egy pusztá adattömegnél, vagyis végrehajtható-e, esetleg emberi ellenőrzés nélkül is. A tervezők döntenek el, hogy egy csevegés eredményeképpen öntevékenyen elindul-e valami más program, netán magának a csevegésnek az eredményprogramja, és az elindított program jótékony hatású, vagy esetleg károkozásra is képes lehet.

Tervezők döntenek el azt is, hogy egy eszközt alapvetően pusztításra terveznek-e, vagy békés célra, mindenesetre a legtöbb békés eszközt is lehetséges rossz célokra is használni.

Tervezési kérdés az is, hogy van-e vészleállító gomb, főkapcsoló, vészfék egy berendezésbe építve, amin keresztül bárki bármikor, vagy esetleg csak felkent személyek, de mégis leállíthatják a szerkezetet.

Végül, de nem utolsó sorban tervezési, de nagyrészt jogi kérdés is az, hogy egy pusztító eszközhöz ki hogyan, milyen jogosítványokkal fér hozzá, ki mikor, milyen feltételek mellett tudja a békés eszközt vagy akár a pusztító eszközt birtokolni, élesíteni, kioldani, működtetni, ill. a működtetés milyen jogi következményekkel, retorziókkal vagy mással jár.

Természetesen sokkal bonyolultabb erkölcsi, etikai, netán jogi kérdések is felmerülhetnek. Ki a felelős azokért a tartalmakért, amelyeket a MI feldolgoz, tárol, és esetleg vissza is ad? Ki a felelős, ha a MI-t – diszkriminatív, rasszista, militáns szövegekkel etette, maga is azzá válik?

Talán a legfontosabb ezzel összefüggő kérdés a *döntési hatáskör* kérdése. A ChatGPT vagy más társ-csevegőprogramok a csevegésen (a válaszgeneráláson) túl ma még mást nem tesznek, de még ezekre vonatkoztatva is: kié a végső döntés? A generatív MI alkalmazásokban megkerülük-e, rövidre zárják-e az emberi felelősség, a döntés emberi nyugtázásának kérdését?

4. MENNYIRE INTELLIGENS A MI?

Az alább leírtak sok olyan tapasztalatra épülnek, amit a ChatGPT-vel szereztem ugyan, de feltehetőleg a többi, nagy nyelvi modelleket használó csevegőprogramra is igaz, mert a működés módjukban nincs igazán nagy különbség. [1]

Ezek a programok mind a *mélytanulmányos technológiára* épülnek. A mélytanulás igen összetett, sokváltozós (és eredményében is esetleg sokdimenziós) függvények példák alapján történő megtanulására alkalmas. Fontos megjegyezni, hogy ezeket a függvényeket a technológia pontról-pontra leköveti, ha nincs hajszaóra megegyező tanító minta, akkor viszont többé-kevésbé hasonló mintákat keres, netán néhány hasonló részminták részeredményéből is össze képes rakni a végeredményt.

A csevegőprogramok *valószínűségi alapon* működnek. Már az egyszerű nyelvi modellek is bizonyos nyelvi szekvenciák (együttes szóelőfordulások) valószínűségét tárolják, a csevegőprogramok ezekből számolják ki egy teljes mondat vagy akár egy párbeszéd valószínűségét, és a legvalószínűbb választ bocsátják ki eredményként. A nyelvi modellek nagysága csupán a modellben tárolt egymás után következő nyelvi elemek számára utal – a kezdeti, pár szavas *környezeti ablak* immár ezres nagyságrendre nőtt.

Mindamellet a csevegőprogramok – hiányzó információ esetén – a legvalószínűbbnek vélt választ adják vissza. Amennyiben ez a válasz a hétköznapi tapasztalatnak ellentmond, *hallucinációnak* is nevezik.

A hiányzó információkkal kapcsolatban azonban egy *filozófiai kérdés* is felmerül (a *negatív mindentudás* vagy omniscience kérdése), amit az episztemikus modális logikákban is érintenek. Tudhatja-e valaki, hogy mit nem tud. Honnan tudhatja a MI, hogy a helyes válasz a „nem tudom”? (Csevegőprogramok biztosíthatnak lehetőséget a válasz valószínűségére vonatkozólag alsó küszöbérték előírására is).

5. CSEVEGŐPROGRAMOK KÉPESSÉGEI

Egyetemi környezetben ismerhetjük a jó képességű, ám a vizsgán készületlenül megjelenő diákok helyzetét. A kérdésről fogalma sincs egy ilyen diáknak, de az olvasottsága és a szakmai ismeretei elegendőek ahhoz, hogy az oktatói kérdésekre legalább valószínűsíthetően (és formailag) helyes válaszokat adjon.

A csevegőprogramok szintén ehhez hasonlóan dolgoznak. Ha az oktató szövegtestben szerepelt egy a feltett kérdéshez komolyan hasonló kérdés-válasz pár, ill. egy efféle párbeszéd-fordulat, akkor a válasz feltehetőleg korrekt lesz. Ha nem, akkor hasonló szövegrészletek alapján rakja össze a hallucináló választ. Így fordulhat elő, hogy pl. a „Mit tudsz Petőfi Sándor haláláról?” kérdésre a ChatGPT által adott válasz:

"....A forradalom és szabadságharc leverése után, 1849. július 31-én Petőfi Sándor elindult a harctérrel szomszédos Terézvárosból, hogy megkeresse a harctéri változatban ott tartózkodó feleségét, Júlia Szendreyt, és megpróbálja elhagyni az országot...."

(Talán vmi 1956-as szövegdarabot házasíthatott össze a '48-as környezettel?)

A valószínűségi működésmód egyben azt is jelenti, hogy - legalábbis a jelen helyzetben - nincs mód hajszálpontos megoldásra, vagyis nincs mód a program *homályos tudását* pl. adatbázis lekérdezéssel kiegészíteni.

Egy csevegőprogram képtelen az *elvonatkoztatásra*. A tudása hasonlít az emberi tudáshoz – a tárolt tények tárolásával és valószínűségi előkeresésével – és mindez független bármilyen elvonatkoztatási szinttől.

6. TÚLEGYSZERŰSÍTÉS. GONDOLKODJUNK ÍTÉLETLOGIKÁBAN!

A csevegőprogramok talán legfontosabb *hiányossága*: a *gondolkodás*. Változatos kérdésfeltevésekkel igazolható, hogy még egyszerűbb következtetési képességet igénylő feladatokra is képtelenek.

Méző László klasszikusában [2] az analógiára való emberi képtelenséget említi annak bizonyítékaként, hogy az emberi agy másképp működik, mint a számítógép. (Persze nem mindenki, de mégis sokunk esetén így van ez.) Efféle zsákutcákba nem nehéz a MI csevegőprogramokat sem bekergetni – ami nyilvánvalóan nem hallucináció, hanem tényleges hiányosság.

Kérdés: Ha innen 250 km-re van Budapest, onnan 260 km-re van Bécs, akkor milyen messzi van innen Bécs?

ChatGPT: Ha Budapesttől 250 km-re vagy, és Bécs 260 km-re van Budapesttől, akkor Budapest és Bécs közötti távolság összege a két távolság hozzáadásával számolható ki.

A válasz nyelvhelyességén túllépve, már a távolságok összeadása sem feltétlenül helyes... amennyiben a három helyiség nem egy egyenesen, sőt, a két végpont nem átellenes oldalon helyezkedik el.

Az alkalmazott módszer viszont tisztán hamis eredményre vezet egy másik – analóg - tranzitív reláció kiszámításakor...

Kérdés: A nagyapám nagyapja hányadik ősöm?

ChatGPT: A nagyapád nagyapja a harmadik ősöd. Az ősök számozásában az első ősöd te magad vagy, a második ősöd a szüleid, a harmadik ősöd a nagyszüleid, és a negyedik ősöd a nagyapád nagyapja.

A matematikában a magasabb szintű logikai nyelvek használatának a legfőbb célja: tömör logikai összefüggések felírása, amivel esetek ezrei, sőt, gyakran analóg esetek is leírhatók. A tömör logikai szabályok felírásának legfontosabb hátulütője: esetleg nem létezik rá teljes megoldó algoritmus – vagyis a logikai rendszer *eldönthetetlen*. Az eldönthetlenséggel szembemenő stratégia a *reifikáció* – (talán kiegyenesítésnek vagy kilapításnak fordíthatnánk), amikor az ismert esetekre külön-külön felírjuk és/vagy alkalmazzuk a logikai szabályokat. Egy reifikált tudásra építve ismert esetekben, ismert körülmények között otthonosan mozog az emberi elme, és gond nélkül megtalálja a megoldásokat, de az analógiákat nem ismeri fel, és hasonló helyzeteket vagy egyáltalán nem, vagy csak nagy nehézségek árán tud kezelni vagy megoldani.

7. SZABÁLYVEZÉRELT CSEVEGŐPROGRAMOK

Itt kell szót ejteni a csevegőprogramok szabályvezérelt technológiájáról. A szerző által készített ITSy-Bitsy csevegőprogram [5] végigjárja a vonatkozó módszertani lépéseket: a szóalak-elemzés angol nyelvre meglehetősen primitív feladat. Ugyanígy, viszonylag kellemes dolog a kötött szórendű angol nyelvre klasszikus elemzőprogramot készíteni. Ami ezután jön, az már nyelvfüggetlen: az elemzési fa átalakítása logikai nyelvvé, majd a logikai nyelv kiértékelése – részben egy kiterjedt *tudásállomány*, egy *ontológia* segítségével, részben háttérbeli konkrét adatbázisok, vagy egyéb célalkalmazások (pl. útkereső program) lekérdezése.

Egy szabályvezérelt csevegőprogram *nem toleráns*. Nem valószínűségi alapon dolgozik, ezért csakis azokat a szerkezeteket ismeri fel, amit beprogramoztak. Nem fog esetleg elfogadható válaszokat létrehozni sem a gépelési hibák, sem nyelvtani hibák, sem esetlegesen hiányzó információdarabok esetén.

Viszont *nem is hallucinál* – a létrehozott válaszok általában hajszálpontosak, és minthogy nem valószínűségi alapon működik, így nem is lehetséges, hogy egymástól független részinformációkat kapcsoljon egyetlen gondolatba.

Álljon itt egy összehasonlító példa, amire az ITSy-Bitsy jobban teljesít a ChatGPT-nél. A „Hány gerinces barátja van Micimackónak?” kérdésre a szabályvezérelt megoldás szépen megkeresi Micimackó barátait a Százholdas Pagonyban, majd eldönti hogy hányan gerincesek (Nyuszi egyes barátaitól és üzletfeleitől eltekintve mindenki), végül pedig megszámlálja őket.

Kérdés: How many vertebrate friends does Winnie-the-Pooh have in Hundred Aker Wood? (sic!)?

ITSy-Bitsy: Winnie-the-Pooh has 7 vertebrate friends in Hundred Aker Wood.

Ugyanerre a (de immár magyarul feltett) kérdésre a ChatGPT válasza (rövidítve):

ChatGPT: „Micimackó barátai a Százholdas Pagonyban:

1. Micimackó, ...

...

4. Nyanya Kanga, akinek a fia Rudi

5. Rudi, Kanga fia

6. Dácsi, Kanga és Rudi fia

... összesítve tehát megállapítható, hogy Micimackónak csak a Tigris barátja gerinces.”

Az eredmény magyarázata – eltekintve a Nyanya Kanga, Rudi, Dácsi és egyéb hallucinációktól: feltehetőleg semmilyen információt nem olvasott a ChatGPT sem arról, hogy a könyv szereplői gerincesek volnának, sem arról, hogy az egyes állatfajták gerincesek-e. Tigrisről talán bevillanhatott, hogy ő ragadozó, és a két fogalmat pedig összetéveszthette.

8. ÖSSZEFOGLALÁS

A MI megoldások széles körű alkalmazásba vétele rengeteg társadalmi, jogi és felelősségi problémát vet fel. Ezek azonban túlnyomórészt tisztázhatóak a megfelelő szakemberek bevonásával, és a már ma is folyó diskurzusok kiértékelésével, a kételyek eloszlatására pedig már ma is megvannak az alkalmas technológiák. A megrendelők döntenek el, hogy ezeket a tervezők beletervezik-e, ill. a megvalósítást végző szakemberek tényleg beleépítik-e egy MI rendszerbe.

Nem hiszem, hogy a MI tiltása bármire is jó megoldás volna – (szerintem az atombombát kellene betiltani.) Persze a MI létrehozásában, alkalmazásában, forrásanyagában keletkező szerep- és felelősségi körök szétválasztása mindenképpen fontos.

A manapság használatos nagy nyelvi modellek technológiája (LLM) sok szempontból érdekes – pl. az öntanulási képességük miatt, viszont egyelőre olyan gyermekbetegségektől terhes ami feltétlenül kijavítást követel. A gép a neurális hálós technológiának köszönhetően jól felismeri a bemenő mondatok (feltett kérdések) nyelvtani szerkezetét, és azokhoz többé kevésbé korrekt válaszokat is tud adni, de a kibocsátott válaszokban sosem lehetünk teljesen biztosak. A nagy nyelvi modellek ugyanis egyrészt valószínűségi alapon dolgoznak: a legvalószínűbbnek ítélt értelmezést és választ bocsátják ki magukból. Másrészt a tudásanyag reifikálásával a tudáselemeket közvetlenül képezik le a neurális hálóban, ezért sem analógiás, sem másféle összekapcsolt logikai következtetési képességük nincsen. Csak azokat az információkat tudja, amiket a tanítás során elolvasott, és azokat sem 100% biztosan.

Az említett kívánnivalók egyik javítási lehetősége: ha a gyártók egyfajta hibrid megoldást követnének: egy párhuzamos, szabályvezérelt csevegőprogram biztosan pontos és tömör választ ad – ennek megíúsulása esetén használható a nagy nyelvi modelleken alapuló, mélytanulósos megoldás.

9. HIVATKOZÁSOK

- [1] IBM Unveils watsonx Generative AI Capabilities to Accelerate Mainframe Application Modernization (<https://newsroom.ibm.com/2023-08-22-IBM-Unveils-watsonx-Generative-AI-Capabilities-to-Accelerate-Mainframe-Application-Modernization-22-Aug-2023>)
- [2] Mérő L.: Észjárások. *Tericum Kiadó, Budapest, 1997.*
- [3] Kayt Sukel: Fact Sheet. Neurotransmission: The Synapse. (<https://dana.org/wp-content/uploads/2019/05/fact-sheet-neurotransmission-synapse-baw-2020.pdf>. 06-Sep-2023)
- [4] AI Impacts: Scale of the Human Brain (<https://aiimpacts.org/scale-of-the-human-brain/> 06-Szept-2023)
- [5] Kilián I.: Csevegőrobotok. Az ITSy-Bitsy modell. *Erdélyi Magyar Műszaki Tudományos Társaság, SzámOkt 2020. konferencia kiadványa, Kolozsvár, pp. 110–117, 2020.*