Markovits Gabriella

A MESTERSÉGES INTELLIGENCIA ÉS PILLANATNYI HELYE VILÁGUNKBAN

BEVEZETÉS

A mesterséges intelligencia kérdésköre a 20. században egyre szélesebbre és szélesebbre nyílt, illetve vált egyre interdiszciplinárisabbá, így a történeti áttekintés előtt mindenképpen tisztáznunk kell, hogy valójában mit is értünk az intelligencia fogalma alatt. Ha sikerül hamar átlendülnünk Edwin Boring meghatározása felett, miszerint „az intelligencia az, amit az intelligenciatesztek mérnek”[1], akkor láthatjuk, hogy mára egészen alaposan feldolgozott fogalommal van dolgunk.

A legelterjedtebb és széles körben elfogadott interpretációk alapján a mesterséges intelligencia értelmezési körébe tartozó definíciók két dimenzió mentén válnak értelmezhetővé: az egyik az embert veszi alapul, míg a másik a racionalitást állítja a középpontba. Ezeken belül pedig további két elágazást figyelhetünk meg: a gondolkodást és a cselekvést.

Az első nézőpont szerint megkülönbözetünk emberi módon gondolkodó, illetve emberi módon cselekvő rendszereket, amelyek alapját tehát az emberi mérce biztosítja. *Az emberi módon gondolkodó rendszerek* felépítése előtt mindenekelőtt meg kell határozni, hogy hogyan gondolkodik az ember, ebből adódóan ez az ág a kognitív tudomány talajában gyökerezik, célja az emberi elme működésének feltárása.

A másik ide köthető értelmezés *az emberi módon cselekvő rendszerek* létrehozása, ahol a rendszer sikerességét az jelenti, hogy az adott mesterséges rendszer teljes mértékben megkülönböztethetetlen az alapjául szolgáló intelligenciától, tehát az embertől. Ez a megközelítés Alan Turing nevéhez köthető és a 20. század közepén pontosan ezért hozta létre az *Imitációs játék* tesztet, azaz a Turing-tesztet: ha a teszt során egy gép/rendszer valamilyen közvetítő eszköz segítségével el tudja hitetni a kísérletben résztvevő emberrel, hogy egy valódi emberrel beszél, vagy ha a kísérleti alany nem tudja egyértelműen eldönteni, hogy emberrel vagy egy mesterséges rendszerrel kommunikál, akkor azt a gépet/rendszert intelligensnek nevezhetjük. (Egy későbbi részben látni fogjuk, hogy azért Alan Turing rendszere sem tökéletes, bizonyos kondíciók mellett a gép nagyobb eséllyel felel meg a teszten.)

A racionalitáshoz fűződő két további elmélet közül *a racionálisan gondolkodó rendszerek* felfogása a logikát és a helyes következtetést helyezi előtérbe, amelyben a legfőbb célt az embernél racionálisabban és pontosabban következtető rendszerek

[1] Stevens 1973.

létrehozása jelenti. A mai logicista megközelítés teljes egészében a következtetés helyességét helyezi előtérbe az emberi tényezővel szemben.

A szintén ehhez kapcsolódó másik megközelítés *a racionálisan cselekvő rendszerek* kidolgozása, amelyek célja, hogy racionálisan cselekedjenek, de ennek nem feltétele az emberi módon való gondolkodás, vagy ahhoz hasonló működés, csupán a helyes következtetés eredménye. Ezeket a rendszereket ágensnek nevezzük és a legjobb kimenetel érdekében cselekednek.

SEARLE INTELLIGENCIA-ÉRTELMEZÉSEI

A Searle által *erős* mesterséges intelligenciának nevezett esetben azonban teljesen másról van szó. „Az MI erős változata viszont arra épít, hogy emulálhatjuk is a megismerést, tehát olyan gépeket hozhatunk létre, amelyek valóban rendelkeznek kognitív állapotokkal."[2] Ez az eset nem sokban különbözi a gyenge mesterséges intelligencia alapjaitól: itt is feltétlenül szükség van egy számítógépre és egy programra, azaz egy „agyra" és egy „elmére". Adott egy mesterséges és „intelligens" rendszer, mit várhatunk tőle? Mitől lesz több, mint egy gyenge MI?

„Ő" olyan lesz, mint egy valódi ember, legalábbis ezt feltételezzük. Gondolkodni fog, saját véleménye lesz és emberi módon fog viselkedni, csak hogy néhányat említsek a vele szemben támasztható elvárásokból. Egy sor diszpozíciót lehet egy emberi értelemben vett intelligens entitásra vonatkoztatni, ha közben az a feltételezés bújik meg a háttérben, hogy ez az entitás olyan, mint *mi*. Márpedig ezt várjuk egy *erős* mesterséges intelligenciától. Searle *Az elme, az agy és a programok világa* című írásában így mutat erre:

> „ […] az MI erős verziója a számítógépet nem pusztán eszköznek tartja az elme tanulmányozásában, hanem a megfelelően programozott számítógépet valóban elmének tekinti abban az értelemben, hogy a megfelelő programok birtokában a számítógép szó szerint megért és egyéb kognitív állapotokkal rendelkezik. Mivel az MI erős verziója szerint a beprogramozott számítógépnek kognitív állapotai vannak, a programok nem pusztán olyan eszközök, amelyek a pszichológiai magyarázatok tesztelését teszik lehetővé, hanem sokkal inkább a programok maguk válnak magyarázattá."[3]

Úgy gondolom, hogy sokkal elfogadhatóbb az *erős* MI megközelítés a tudományok jelenlegi szemszögéből, mivel megkönnyítheti a fejlődést, minden területen. Ha eltekintünk attól, hogy ennek a létrejötte hatalmasat lendítene az emberiség önmagát Istenként való pozícionálásában, akkor elképesztő hasznot is jelenthet számunkra.

[2] Pápay 2003. 105.o.
[3] Searle 1980. 136.o.

ALAN TURING ÉS AZ IMITÁCIÓS JÁTÉK

Alan Turing a mesterséges intelligenciáról szóló diskurzusba szigorúan funkcionalista oldalon csatlakozott be és erre az elméleti fejtegetésre adott választ Searle értelmezése az intelligenciával kapcsolatban. Turing elképzelése szerint lehetséges egy olyan gép építése és annak megfelelő programozása, amely bármilyen témájú beszélgetésben képes megállni a helyét és emberi módon viselkedni egy ilyen kommunikációs helyzetben. Az alábbiakban szeretném bemutatni az *Imitációs játékot*, azaz a Turing-tesztet, továbbá a mellett érvelek, hogy miért nem szabad elvetni ezt a fajta intelligencia tesztet, miközben számos gondolkodó e koncepció alkonyát vizionálja.

Alan Turing 1950-ben íródott *Computing machinery and intelligence* című cikke egy komoly kérdéssel nyit, ami szó szerint megalapozta és kijelölte az utat a jövő számára: Tudnak-e a gépek gondolkozni?

> „A következő kérdéssel szeretnék foglalkozni: »Tudnak-e a gépek gondolkodni?«. Ezt pedig a »gép« és a »gondolkodni« fogalmak definíciójának jelentésével kellene kezdeni. A definíciókat keretbe kellene foglalni, már amennyire ez lehetséges, hogy a szavak mindennapi használatát tükrözzék, de ez a hozzáállás veszedelmes volna. Ha a »gép« és »gondolkodni« szavak értelmét úgy akarnánk meghatározni, hogy általános használatukat vizsgáljuk, nehéz lenne elkerülni azt a következtetést, hogy a »Tudnak-e a gépek gondolkodni?« kérdésre adott válasz értelmét olyasféle statisztikai módszerek útján kell keresni, mint egy Gallup felmérés. Ez pedig abszurdum. Ahelyett tehát, hogy ilyen definíciókkal kísérleteznénk, a kérdést helyettesíteni fogom egy másikkal, amely ezzel szoros kapcsolatban van, és viszonylag egyértelmű szavakban fejezhető ki."[4]

Turing új megközelítési módot kínál a cikkben, mely a kor gondolkodói között nagy vitákat váltott ki. Az *Imitációs játék* lényege, hogy két emberi és egy gépi résztvevővel olyan párbeszéd alakuljon ki, amit a hétköznapi nyelven emberinek nevezünk. A három társalgó fél közül az egyik ember kérdez, a másik ember, illetve a gép is válaszol a feltett kérdésekre. Az új formula, amit Turing felvázol a kérdéssel kapcsolatban a következő: adott egy férfi (A), egy nő (B) és a kérdező (C), azonban a kérdező nem tudja, hogy melyik válaszadó milyen nemű. (Turing játékában az az ideális, ha külön-külön szobában vannak és kizárólag közvetítő eszközök segítségével tudják a kapcsolatot tartani.) Akkor sikeres a teszt, ha a gép sajátos eszközeivel sikeresen megtéveszti a kérdezőt és elhiteti vele, hogy emberrel beszélget.

[4] Turing 1950. 433.o. (saját fordítás)

„TISZTELETTUDÓAN MEGÁLLAPODUNK, ABBAN HOGY…”

A 2000-es évekre nagyon sok kritika megfogalmazódott a Turing-teszttel kapcsolatban. További kitételeket adtak hozzá bizonyos esetekben és bizonyos szubjektív álláspontokat túláltalánosítottak, vagy a nem megfelelő bírákat ültetik a székbe - a lehetőségek tárháza végtelen a kritikák esetében. Azonban később szeretnék rámutatni, hogy van létjogosultsága a Turing-tesztnek a chatbotok világában is, ha pedig elkészül egy megfelelően összetett és minden igényt kielégítő program, akkor a tudomány többi területével együttesen már komoly előrelépést jelenthet mind a tudományos élet, mind a hétköznapi felhasználás szempontjából.

Ami a Turing-teszt használatát és lényegét illeti, nem véletlenül Imitációs játék az eredeti elnevezése. A gépnek egy embert kell meggyőznie egy adott dologról és nem az a feladata, hogy váljon *valamilyenné*, hanem az, hogy *tegyen úgy, mintha* valamilyen lenne, és ezt hitesse el a vele kommunikáló emberrel. Turing a következőképpen írja le a problémát *Mind*-ban megjelent cikkében:

> „A probléma új alakját leírhatjuk egy játék terminusain belül, amelyet nevezzünk »imitációs játéknak«. Három ember játssza, egy férfi (*A*), egy nő (*B*) és egy kérdező (*C*), aki akármelyik nem képviselője lehet. A kérdező a másik kettőtől távol, egy külön szobában tartózkodik. A játék célja a kérdező számára az, hogy megállapítsa, hogy a másik két személy közül melyik a férfi és melyik a nő. A résztvevőket *X* és *Y* címkeként ismeri, és a játék végén azt a megállapítást tegye, hogy »*X* azonos *A*-val és *Y* azonos *B*-vel«, vagy pedig: »*X* azonos *B*-vel és *Y* azonos *A*"-val«. […] A játék harmadik résztvevőjének (*B*) az a célja, hogy segítse a kérdezőt. *B* számára a legjobb stratégia valószínűleg az, hogy igazat válaszol. Olyan dolgokkal is megtoldhatja a válaszait, mint hogy »Én vagyok a nő, ne hallgasson rá!«, de ez nem jelent semmit, mivel a férfi is tehet hasonló megjegyzéseket.
>
> Ezek után a következő kérdést tehetjük fel: »Mi történik akkor, ha a játékban *A* helyét egy gép foglalja el?« A kérdező ugyanannyiszor dönt helytelenül, ha így játszanak, mint akkor, ha a játékot egy férfi és egy nő játssza? Ezek a kérdések helyettesítik eredeti kérdésünket: »Tudnak-e a gépek gondolkodni?«” [5]

Tehát nem az a teszt célja, hogy arról győzze meg a kérdezőt a gép, hogy ő milyen entitás valójában, hanem hogy megtévessze őt bizonyos verbális eszközök segítségével valamivel kapcsolatban. Turing ebben az írásában később kitér a tudatosság kérdésére is, azonban a párbeszéd lényegét nem a gépi tudatosság problémája kapcsán ragadhatjuk meg. Ami talán szerencsés eset, hiszen a mai napig nem tudunk pontos és egyetemes definíciót adni a tudat fogalmára és a saját tudatosságunkra sem.

[5] Turing 1950. 433-434.o. (saját fordítás)

Lehetünk bizonyos *véleményen*, hogy melyik teóriát fogadjuk el, ismerjük azt a kifejezést, hogy *tudatában vagyok valaminek*. A tudat, illetve az ehhez kapcsolódó kutatások rendkívül szerteágazó rendszert alkotnak, amelyben az eligazodás meglehetősen bonyolult. Addig a pontig, amíg nem kapunk teljesen szabatos, minden tudományterületen elfogadható definíciót a tudattal és a tudatossággal kapcsolatban, csak tapogatódzhatunk mind a saját, mind a gépi tudatosság témakörében és feltételezésekkel kell élnünk, azonban ez nem teszi értékelhetetlenné a teszt eredményét. Ha egy ember-ember kommunikációs folyam során feltétel nélkül elfogadjuk, hogy a másik tudatos lény és gondolkodik, márpedig elfogadjuk, annak a lehetőségnek ellenére, hogy a másik ember „belsőjében" teljesen más folyamatok zajlanak, mint a mienkben. Mivel pedig semmi esetre sem szeretnénk egy előadóterem harmadik sorában belenézni egymás elméjébe azzal a céllal, hogy megvizsgáljuk: megérti vajon a másik, amit magyarázok vagy sem, így játékba hozva minden jóindulatunkat *tisztelettudóan megállapodunk abban, hogy…*

> „Ezen nézőpont legszélsőségesebb formája szerint az egyetlen mód, amely meggyőzheti az embert arról, hogy a gép gondolkodik, az az, ha ő maga a gép és érzi azt, hogy gondolkodik. Ebben az esetben az illető le tudja írni ezeket az érzéseit a külvilágnak, de természetesen nem volna rá bizonyíték, hogy ezeket érezte is. Hasonlóképpen, ezen nézet szerint az egyetlen módja annak, hogy tudjuk, hogy egy ember gondolkodik, az az, hogy ha valaki azonos azzal a bizonyos személlyel, akiről beszélünk. Ez igazából maga a szolipszista nézőpont. Lehet, hogy ez a leglogikusabb álláspont, de a gondolatok közlését nehézkessé teszi. *A* valószínűleg azt hiszi, hogy »*A* gondolkodik, de *B* nem«, míg *B* azt hiszi, hogy »*B* gondolkodik, de *A* nem«. Ahelyett, hogy erről állandó vitába bonyolódnánk, inkább tisztelettudóan megállapodunk abban, hogy mindenki gondolkodik." [6]

Turing ebben a részben úgy fogalmaz: „érezzük" hogy gondolkodunk, ami szintén problémát vethet fel, hiszen hogyan lehet érezni a gondolkodást? Nem tudom megfogni, sem leírni, hogy milyen érzés gondolkodni, amit legfeljebb el tudok mondani a gondolataimról, hogy milyen sorrendben követik egymást. Azt nem tudom leírni, hogy milyen egy gondolat kvalitása, talán csak annyit állapíthatok meg, hogy milyen „érzés" gondolkodni például a fantáziálással vagy az emlékezéssel szemben. E tisztázatlan fogalom alkalmazását nem tartom szerencsésnek jelen esetben, hiszen alapvető dolgokra nem tudunk a tudomány mostani álláspontjai szerint válaszolni, bár ugrásszerű és hirtelen fejlődés jellemző minden ágra, abban csak reménykedhetünk, hogy valamikor a (közel) jövőben megtaláljuk a választ néhány nagyobb kérdésünkre ezekkel kapcsolatban. Az érzés fogalmáról a modern tudományok nyelvén valami hasonlót mondhatnánk: introspekció segítségével a gép/program megállapítja

[6] Turing 1950. 446.o. (saját fordítás)

magáról, hogy gondolkodik. Erről is azonban csak valamiféle kommunikáció által tudnánk megbizonyosodni, mi, emberek.

A gondolkodás érzésével visszakanyarodunk a tudatosság és az imitáció kérdéséhez. Abban hallgatólagosan megegyezhetünk, hogy a gép a teszt során az „úgy teszek, mintha" elvet követi, el kell hitetnie a kérdezővel valamit. Tegyük fel továbbá, hogy ehhez rendelkezik is a megfelelő nyelvi eszköztárral és annak az elmeit jól használja. (Egyelőre napoljuk el a kérdést, hogy ezt előre betáplálták vagy esetleg egy tanulási folyamat következménye, legyen ez egy későbbi bekezdés témája.) Ha eredeti formájában vizsgáljuk a tesztet, akkor mindenképpen ki kell emelni, hogy a gép egyrészt megértette a célt és „minden erejével azon van", hogy ezt el is érje. Másrészt feltételeznünk kell, hogy a gép, pont a fentebb említett indokok miatt, képes megkülönböztetni magát a másik résztvevőtől és a kérdezőtől is. Sőt, létre tudott hozni egy mesterséges reprezentációt/képet önmagában, amit elhitet vagy legalábbis meg próbál elhitetni a kérdezővel magáról.

Ez persze csak akkor alkalmazható, ha a programozók nem azt az utasítást adják neki, hogy „a következő Enter leütésétől kezdve magadat úgy fogod definiálni és leírni, hogy százkilencven centiméter magas orosz úszónő vagy, szőke bubi hajjal, a leghosszabb tincsed 9 inch hosszú és a többi a teljes leírás megszületéséig". Ha egy teljes személy leírását szeretnénk létrehozni külső szemlélőként, akkor tények hihetetlenül hosszú sorát kellene a programozóknak betáplálni, amit talán el sem tudunk képzelni, ha csak és kizárólag adatok betáplálásával szeretnénk megvalósítani.

Az előbbi fejtegetésből, amikor a másik tudatát szerettem volna vizsgálni, jól látszik, hogy egy rajtam kívül álló és tőlem elkülönböző személy tudatát nem tudom közvetlenül vizsgálni, csupán közvetett tapasztalatom lehet róla. Ha létezik olyan, mint a tudat, akkor csak és kizárólag magamon/magamban tudom megvizsgálni a tudatosság létét és milyenségét introspekción keresztül. Pontosan ezért tartom a Turing-tesztet egy igenis helytálló „vizsgának", hiszen a gépnek ebben az esetben közvetetten lehet tapasztalni a kognitív képességeit, analóg módon nem különbözik az ember és ember közötti kommunikációtól: a gépelt beszélgetésen keresztül van lehetőség a kommunikációra két fél között. Az ember és ember között fennálló párbeszédben sincs ez másként, hiszen a párbeszédben álló felek között kimondott szavak lépnek a gépelt betűk helyébe. Természetesen ebbe a közegbe lép bele ebben az esetben a verbális és non verbális kommunikáció több rétege is, bár jelen esetben ez nem egy szükségszerű kitétel, hiszen nem egy „emberbőrbe bújtatott" mesterséges intelligenciáról szeretnénk beszélni, de ez egyelőre a jövő rejtélye marad.

A CHATBOTOK, MINT A TURING-TESZT TÖKÉLETES ALANYAI

Turing idejében még nem létezett a mesterséges intelligencia most is ismert felosztása, így szíve mélyén a kor tudós társadalma, már aki érintett volt a témában, csak álmodozhatott egy kommunikációra kifejlesztett gépről vagy programról. Ezek közé tartoznak a ma chatbotoknak nevezett programok, amelyek célja, hogy akár általános, akár specifikus témában beszélgetésbe elegyedjenek a velük kommunikáló

emberrel. Azért is kerültek ők a tanulmány fókuszpontjába, mert a hatékonyságuk és fejlődésük jól tesztelhető a Turing-teszttel. Úgy gondolom, hogy a chatbotokon keresztül tehát nagyon is jól lehet vizsgálni a mesterséges intelligencia bizonyos vetületét és szintjét, a kommunikációt, ami szintén egy óriási előrelépés ebben a témában. Tehát az Imitációs játékot legfőképpen az ilyen speciális programokkal lehet elvégezni, mert egy óriási adatfeldolgozó rendszert hiába szeretnénk az esős időjárás viszontagságairól kérdezni, nem fogja azt mondani, hogy a hideg rázza a szakadó esőtől, még ha imitálnia is kellene az érzelmet.

Napjainkban számos kutatás folyik különböző egyetemeken és vállalatoknál is ezen a területen, ezekből szeretném kiemelni az egyik legnagyobb technikai céget, a Microsoftot. Kutatásaik három fő vonalának különböző neveket adtak: *Cortana, XiaoIce* és *Tay*. Mind a három „hölgy" fő profilja a kommunikáció, míg *Cortana* a virtuális személyi asszisztens szerepét tölti be Microsoft Windows nevű multi-platform rendszerében (a mindennapos feladatokat látja el), addig *XiaoIce* egy „szociális asszisztensként" funkcionál, aki különböző felületeken biztosít támogatást és beszélgetőpartnert az emberek számára. A *Tay* nevű programnak egy külön fejezetet szeretnék később szánni, ugyanis meglehetősen kellemetlen helyzet alakult ki a Microsoft ezen fejlesztése körül és a legnagyobb szerepet a kialakult „pofonban" mi, emberek játszottuk. Ebben a szakaszban szeretném megvizsgálni, hogy a XiaoIce nevű chatbot milyen elvek nyomán működik, és hogyan mutatja meg a Turing-teszt jövőjét.

A hivatalos Microsoft közlemény szerint a következők mondhatók el erről a beszélgető programról:

> „Az elmúlt hetek során a csapatunk egy kísérleten dolgozott, amit XiaoIce névvel illettek (fordítása: kis jég). XiaoIce egy közösségi asszisztens, akit az emberek barátként hozzá tudnak adni a listájukhoz néhány vezető kínai közösségi szolgáltatónál, köztük a Weibonál, ami egy Twitterhez hasonló mikroblog szolgáltató 700 millió felhasználóval, továbbá a Touchpalnál is megtalálható. […] Egyszerűen hozzá kell adni a beszélgetőpartnereink listájához és máris tudnak az emberek hosszabb beszélgetést folytatni vele. Azonban XiaoIce sokkal fejlettebb, mint azok a beszélgető robotok, akikre emlékszünk. Ő egy kifinomult társalgó, aki figyelemreméltó személyiséggel bír. Könnyedén beszélgetésbe elegyedik a partnerével, még olyan tartalomspecifikus dolgokkal kapcsolatban is, mint hírességek, sport vagy gazdaság, ugyanakkor együttérzésre képes és humorérzéke is van. A vélemény-elemzés használatával a fejleszteni tudja kifejező- és válaszolási készségeit a pozitív vagy negatív reakciók alapján, amit az emberi partnereitől kap. Tud vicceket mesélni, idézni költeményekből, szellemtörténetekben is jártas, sőt zeneszövegek továbbítására és a lottószámok kiejtésére, és így tovább. Akárcsak egy barát, hosszabb

> beszélgetéseket képes fenntartani, melyeknek hossza több száz üzenetváltásig is terjedhet."[7]

A jelenlegi cél globális tekintetben ezeknek a programoknak a továbbfejlesztése és annak vizsgálata, hogy milyen további elemekkel lehetne bővíteni ezt a területet. A napjainkban is futó programok jelentősen támaszkodnak az interneten megtalálható, korábban lefolytatott, szöveg-alapú társalgásokra, mert Cortana és XiaoIce is az internet segítségével működik, és azokra a sémákra épül, amik a mindennapos kommunikációnkat jellemzi. Pontosan az *ipon.hu* internetes portál egyik elemzése világít rá a tényre, hogy bár nagyon egyedinek képzeljük magunkat, de mégsem vagyunk azok. A beszélgetéseink témája sokszor nagyon hasonlít egymásra, bármilyen nyelven is zajlik a kommunikáció, sőt bevett sémák várhatóak mindannyiunk környezetében élő emberektől.

> „Ugyanakkor bármennyire is szeretjük magunkat egyedi és megismételhetetlen beszélgetőtársaknak hinni, a Xiaoice technológiájának alapját az a tény adja, hogy a világon élő 7 milliárd ember közt minden pillanatban sok hasonló elemeket felvonultató beszélgetés zajlik le, vagyis hogy reakcióink sok szempontból nagyon is kiszámíthatóak. A bot az elmúlt 18 hónap csevegései során annyi adatot gyűjtött össze, hogy a társalgások sikeressége érdekében átvizsgált információ 26 százaléka saját munkájának eredménye. Ez a 26 százalék a beszélgetések 51 százalékában elég is az emberszerű kommunikáció megvalósításához. És mivel a rendszer folyamatosan fejleszti magát, minden eszmecserével egyre emberibb lesz."[8]

XiaoIce tehát megfigyel, elemez, folyamatosan tanul és kijavítja a saját hibáit. Ugyanennek a programnak például nem kifejezetten az a célja, hogy elhitesse a vele beszélgetővel, hogy ő ember, azonban mindenképpen hozzájárul ahhoz a fejlődési vonalhoz, hogy az intelligens ágensek egyre jobb minőségű beszélgetést tudjanak folytatni egy adott *emberi* partnerrel. Egyfajta felderítő osztagként működnek, amelyek rávilágítanak azokra a kritikus pontokra, ahol fejlesztésre szorulnak a programok és elbukhatnak a jövő gépei. A Microsoft nem szeretné abba hitbe ringatni az embereket, hogy ez a program bármilyen belső és mentális állapottal, netán érzelmekkel bír, egyelőre azok feltérképezése a cél: „A Microsoft fejlesztői persze nem állítják, hogy a program érti, hogy mit érez, vagy mit mond beszélgetőtársa. XiaoIce reakcióinak nagy részét Bing-es keresések alapján állítja össze, illetve ezeket az adatokat hang- és képelemzésekkel egészíti ki (ha éppen videochatről van szó), hogy jobban meg tudja határozni a társalgás érzelmi kontextusát."[9]

[7] https://blogs.bing.com/search/2014/09/05/meet-xiaoice-cortanas-little-sister

[8] https://ipon.hu/elemzesek/baratunk_a_csevegobot/2812/2

[9] https://ipon.hu/elemzesek/baratunk_a_csevegobot/2812/1

A mesterséges intelligencia tudományos élete az elmúlt években sok eseménytől volt hangos, az elmúlt két év legfrissebb híreinek egy része Eugene Goostman nevétől volt hangos a Turing-teszttel kapcsolatban. [10] A bírák 33%-át meggyőzte az öt perces teszt során, hogy ő egy 13 éves ukrán fiú, tehát emberi lény. Igaz, hogy a fenti kritériumok közül egyik sem szó szerint Turing írásából merít, bár egy helyen Turing említ egy öt perces időintervallumot: „[…] a kérdező eljut a megfelelő felismerésig 5 perc kérdezősködés után…”[11], de pont ezekről az önkényes kiegészítésekről írtam fentebb. Pontosan öt perce van a programnak beszélgetni a bírókkal/kérdezőkkel és a kérdezők legalább harminc százalékát kell arról meggyőznie arról, hogy ő ember. Esetleges támadási felületként értelmezhető a kritikusok szempontjából, hogy míg Eugene 13, XiaoIce pedig 17 évesként mutatkoznak be, hogy Eugene nem angol anyanyelvű, pedig angol nyelven folyik a vizsgálódás, és XiaoIce egyelőre még csak a Távol-Keleten működik… Ezek a kiskapuk egyelőre lehetőséget biztosítanak az intelligens programoknak, hogy hibázzanak, a bírák és beszélgetőpartnerek pedig elnézőbbek a nyelvtani és akár világnézeti pontatlanságokkal szemben, ha tudják: nem egy felnőttel beszélgetnek. A cél természetesen az, hogy elérjenek egy olyan fejlettségi szintet ezek a mesterséges beszélgető programok, hogy ne kelljen ilyen-olyan tulajdonságokkal felruházni, amik felmentést és kibúvókat jelenthetnek a Turing-teszt során.

Úgy vélem, ha ezeknek a beszélgető programoknak a fejlettségi szintje eléri a kívánt és kielégítő mértéket, tehát akár a nyelv, akár a gondolkodás szintjén eléri az emberi elvárásokat, akkor az így megszerzett résztudást összekötve a mesterséges intelligencia egyéb aspektusaival már egy újabb fejlődési mérföldkőhöz fog elérni e kutatási terület. A Turing-teszt ekképpen tehát tökéletesen alkalmazható véleményem szerint a beszélgetésre kifejlesztett egységek esetében, vizsgálható általa, hogy milyen szinten képes a kommunikációra a gép, ellenőrizhető, hogy a programozók sikeresen ültették át az nyelv használatát a technológia világába. Azonban addig, amíg a társalgásra kifejlesztett ágenseket nem kapcsoljuk össze a mesterséges intelligencia egyéb területeivel és létre nem jön egy valódi tudat (a maga sajátosságaival, benyomásaival, történetével és finomságaival, már-már emberi mivoltával), csupán ezen a területen jelent előrelépést. Ha eljön a pillanat, hogy egy ilyen tudat létrejön, még ha emberi korlátozással is, abban a pillanatban Alan Turing okfejtése életbe léphet és a gép kielégítő válaszokat adhat még a saját elméjével kapcsolatban is.

„EMBERI - TÚLSÁGOSAN IS EMBERI”: AZAZ EGY TANULÓ MESTERSÉGES INTELLIGENCIA KALANDOS ÚTJA KÉZ A KÉZBEN AZ INTERNET NÉPÉVEL

Ha csupán korlátozott mennyiségű információval látunk el egy mesterségesen létrehozott intelligens rendszert, akkor az csak a megadott kereten belül fog tudni

[10] http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx
[11] Turing 1950. 442.o. (saját fordítás)

akciókat végrehajtani, Somogyvári cikkében pedig az bontakozik ki, hogy ha megtanítjuk a gépnek, hogy hogyan használjon bizonyos modelleket és tanuló algoritmust hozunk létre, akkor már más a helyzet. Ezen a ponton azonban több irányba is elindulhatunk. A mesterséges intelligencia tanulási folyamatait csoportosíthatjuk aszerint, hogy ennek a folyamatnak van emberi felügyelője vagy sem. „A tanulóalgoritmusokat alapvetően két kategóriára oszthatjuk: a felügyelővel és a felügyelő nélkül tanulókra. A felügyelővel tanulók számára egy külső forrás, a tanár szolgáltatja a követendő példát, míg a felügyelő nélkül tanulókat egy, többé vagy kevésbé expliciten az algoritmusban bennfoglalt kritérium irányítja. Emiatt a felügyelő nélküli tanulást tekinthetjük úgy, mint ami a külvilág jeleiből megpróbálja kiszűrni azokat, melyeknek struktúrája megfelel, egy az algoritmusba előre beépített szabálynak." [12]

Másik nagy kérdés, hogy ha felügyelővel tanul az algoritmus, akkor ki vagy kik azok az adekvát személyek, akik megfelelnek erre a feladatra. A tudományos közösség bizonyos területei biztos, hogy a saját területük legkiválóbbját neveznénk meg, ekkor specifikus a végeredmény és egy biológia, esetleg a filozófia terén teljesen jártas mesterséges intelligenciával lenne dolgunk. A következőkben vizsgálandó esetben egy beszélgető program továbbfejlesztése volt a cél azzal, hogy óriási mennyiségű mindennapi párbeszédet szerettek volna beprogramozni a gépbe annak érdekében, hogy ezekből tanuljon. A legfrissebb példa a Microsoft legújabb, Tay nevű mesterséges intelligencia programja, akivel pórul jártak a fejlesztői.

Az eredeti elképzelés szerint Tay adott online felületeken keresztül (például az egyik ilyen csatorna volt a Twitter is) kommunikálni tudott az őt kérdező és vele beszélgetni kívánó emberekkel. A Microsoft 2016. március 23-án tette elérhetővé a lehetőséget az emberek számára, ekkor lett online Tay, a nyilvánossá tétel célja az volt, hogy a chatbot ezekből a beszélgetésekből a természetes nyelv mélyebb rétegeibe nyerhessen betekintést és „megértse" azt. Az eredeti elképzelés szerint a beszélgetések és interakciók számának növekedése egyre nagyobb és nagyobb adatbázist biztosított volna a program számára, amiből megtanulhatja a természetes nyelv használatát és egyre kiválóbban használta volna azt. „A Tay által használt gépi tanulás lényege, hogy a mesterséges intelligencia megfigyelésekből és tapasztalatokból von le következtetéseket, ezekkel válik egyre finomabbá. […] Az ezen elven tanuló robotokat jellemzően egy speciális szakterületre képzik ki, jelen esetben arra, hogy hatalmas szövegmennyiségből tanulja meg, hogyan kell válaszolni."[13]

A „virtuális gubanc" akkor kezdődött, amikor egyetlen egy nap után le kellett állítani a kezdeményezést, mert a beáramló információk elemzése után Tay politikailag szélsőséges álláspontra helyezkedett és olyan radikális véleményeket fogalmazott meg, amiknek nemcsak, hogy megkérdőjelezhető a létjogosultsága, de átlépett egy etikai határvonalat is. Például szélső jobb oldali álláspontra helyezkedve úgy

[12] Somogyvári 2015. 204.o.

[13] http://www.origo.hu/techbazis/20160324-naciva-tette-az-internet-a-microsoft-chatrobotjat.html

vélte, hogy Hitler nem tett semmi rosszat, rasszista és fajgyűlölő megjegyzések tömkelegét tette közzé és kategorikusan kijelentette, hogy szerinte George Bush egykori elnök szervezte meg a 2001. szeptember 11-i terrortámadásokat az Egyesült Államokban.[14]

Egyelőre nem született tanulmány az esettel kapcsolatban, ezért a Microsoft hivatalos *Learning from Tay's introduction* (Tanulságok Tay bemutatkozásával kapcsolatban) közleményéből tudunk meg többet az eseményekről. Első kérdésként felmerül, hogy miért nem zárta ki a program a fentebb említett és azokhoz hasonló kérdéseket. Elméletileg létezik egy szűrő a programon belül, aminek célja, hogy ne történhessen ilyen vagy ehhez hasonló incidens, az indítása előtti teszteken ezek jól is működtek, azonban éles helyzetben valamilyen módon mégsem hagyta figyelmen kívül a program ezeket a megjegyzéseket.

> „Tay tervezése közben rengeteg szűrőt terveztünk és építettünk be, továbbá széleskörű tanulmányokat folytattunk különböző felhasználói csoportoknál. Tay terssztűrő képességeit is felmértük változatos feltételek mellett, különösen azért, hogy a Tay-jel folytatott kommunikáció pozitív benyomást keltsen. Amikor kielégítőnek találtuk Tay interakcióját a felhasználókkal, szerettük volna, ha egy szélesebb kör is kapcsolatba lépne vele. Ezáltal nőtt tehát az interakciók száma, amivel kapcsolatban arra számítottunk, hogy a mesterséges intelligencia így többet tanul és egyre jobbá válik." [15]

Személy szerint úgy vélem, hogy kissé elhamarkodottan publikálták a programot, nem bizonyosodtak meg arról teljes mértékben, hogy a program valóban kiszűri a problémás anyagot.

Míg a fentebb említett emberi mulasztásnak róható fel, a másik komoly téma azt célozza meg, hogy mekkora részben felelősek azok az emberek, akik helytelen és etikátlan dolgokra tanították Tayt, illetve egyfajta kettős mérce kialakulására szeretnék rávilágítani. Az internet „jótékony" névtelensége mögé búvja az emberek ebben az esetben olyan véleményeket fogalmaztak meg, amiket a hétköznapokban valószínűleg nem fejtenének ki hangosan és nyilvánosan. Továbbá az sem biztos, hogy valóban arra az álláspontra vannak, amit az üzeneteken keresztül kifejeztek, vagy amiről próbálták meggyőzni a programot. Így vagy úgy, de ezen álláspontok hangoztatása vagy akár a holokauszttagadás a „való világban" bizonyos retorziókat vonnak maguk után, mind társadalmi, mind törvénykezési szinten. Ennek tükrében a megnyilvánulások vagy az egyén szupresszált véleményét tükrözik, vagy azt a rossz szándékot, ami valóban személyiségének része, azonban ezt nem „nyilváníthatja ki"

[14] A bejegyzéseket a Microsoft törölte Tay személyes adatlapjairól, képernyőfelvételek formájában rögzítésre kerültek különböző felhasználók által: https://social-hax.com/2016/03/24/microsoft-creates-ai-bot-internet-immediately-turns-racist/

[15] http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.00182ox8h19kvd8rxwz100tcp8ege

a környezetében. A rendszert offline állapotba kellett helyezni, a cég hivatalosan is elnézést kért az esetért és a teljes felelősséget is vállalja a történtekért, azonban elhatárolódik a Tay által megfogalmazott véleményektől.

> „Rendkívül sajnáljuk az megfontolatlan, támadó és sértő bejegyzéseket Taytől, amelyek nem minket és az értékrendünket képviselik, sem azt, hogy milyennek terveztük Tay-t. Tay jelenleg lecsatlakoztattuk a hálózatról és ügyelni fogunk arra, hogy csak akkor hozzuk vissza, amikor teljesen biztosak vagyunk abban, hogy sikeresebben meg tudjuk előzni a rosszindulatú törekvéseket, amelyek szemben állnak elveinkkel és értékeinkkel. […] Teljes mértékben vállaljuk a felelősséget azért, hogy nem láttuk ezt a lehetőséget idő előtt."[16]

KONKLÚZIÓ

Ahogy az a fenti bekezdésekben jól látszik, a mesterséges intelligencia már erőteljesen jelen van világunkban és hatása egyre jobban érvényesülni fog. Főként a tudományok területein jelenthet óriási lehetőséget, mind a könnyebb megértés, mind a fejlődés esetében. Az MI kutatás első úttörőjeként számon tartott Alan Turing gondolatmenetét és konklúzióit a mai napig kiválóan lehet hasznosítani, ráadásul egy olyan irányt is kijelöl az *emberi módon gondolkodó* rendszerek fejlődésének, amit úgy vélem, hogy érdemes lehet követni. Ezen felül pedig rendkívül jó alapot jelenthet az Imitációs játék a mesterséges intelligenciák tesztelésében is, megmutatja, hogy mennyire hatékonyak és milyen teljesítményt nyújtanak a felhasználóknak.

A Microsoft és Tay esete rávilágított arra, hogy milyen könnyen el tudjuk veszíteni az irányítást egy ilyen helyzet felett, ha nem biztosítjuk a megfelelően a kontrollált környezetet és csak „kiengedünk" egy tanuló MI-t a nagybetűs világba. Az ember és program összefonódása, illetve közös léte egyelőre még kezdeti stádiumban van, innen nagyon sok irányt vehet fel a kapcsolatunk. Mivel egy meglehetősen friss tudományterületről van szó, így szót kell ejtenünk az emberek edukációjáról is, ha szeretnénk a mesterséges intelligenciákat a tudomány területén kívülre is alkalmazni, mert a kettős mérce pedig itt lép életbe.

Azokat az embereket nem fogják felelősségre vonni, akiknek köszönhetően a program erkölcstelen megnyilvánulást és magatartást tanult, sokkal hamarabb kapcsolnak ki egy ilyen gépet, mintsem számon kérnék az őt programozó embereket. Ha esetleg így is lenne, akkor sincs (egyelőre) olyan bíróság a világon, aki kimondaná azt, hogy egy mesterséges intelligencia megrongálásért és annak etikátlan magatartásra való buzdításáért bármiféle szankciót kiszabhatna. Ha egy pillanatra megállunk és alaposabban megvizsgáljuk az analógiát, akkor felmerülhet a kérdés, hogy az ilyen viselkedést tanúsító emberek vajon a környezetük felé is ezt, vagy ehhez hasonló üzeneteket közvetítenek?

[16] http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.00182ox8h19kvd8rxwz100tcp8ege

Ha a jövő nemzedékei ezt a hozzáállást teszik magukévá, akkor semmilyen szinten nem lesz egyszerű a haladás az emberiség számára. Véleményem szerint ez az óriási tömeget megmozgató és bevonó program meglehetősen félresiklott, nagyobb részben pont azon emberek miatt, akiktől azt várták el, hogy segítséget fognak nyújtani a saját érdekükben. Hiszen a program nekik készül, a Microsoft dolgozói, akik ezen a részlegen munkálkodnak, nem maguknak és nem is hobbiból hozták létre Tayt. A mesterséges intelligencia egyre több területen jelen van, és nem tulajdonítunk egyetlen egy MI-nek sem világpusztító és fajkiirtó hátsó szándékokat. Azonban valóban súlyos következményei lehetnek annak, ha esetleg egy fejlettebb mesterséges intelligenciát kiengedünk egy olyan közegbe, ahol nem megfelelően tanítják.

Természetesen ez az esemény pont azért kapott ekkora nyilvánosságot, mert kényes területeken alkalmazták. Ha az interneten keresztül szeretnénk a jövőben bármire is tanítani egy mesterséges intelligenciát, mindig Godwin megfigyelése/törvénye lebegjen a szemünk előtt: *egy internetes vita terebélyesedésével annak az esélye, hogy valaki a nácikkal vagy Hitlerrel von valamilyen párhuzamot, közelít az egyhez.*

IRODALOM

„*Barátunk, a csevegőbot*”: https://ipon.hu/elemzesek/baratunk_a_csevego-bot/2812/1

„*Learning from Tay's introduction*”: http://blogs.micro-soft.com/blog/2016/03/25/learning-tays-intro-duction/#sm.00182ox8h19kvd8rxwz100tcp8ege

„*Meet XiaoIce, Cortana's Little Sister*”: https://blogs.bing.com/se-arch/2014/09/05/meet-xiaoice-cortanas-little-sister

„*Microsoft Creates AI Bot – Internet Immediately Turns it Racist*”: https://so-cialhax.com/2016/03/24/microsoft-creates-ai-bot-internet-immediately-turns-racist/

„*Nácit csinált az internet a Microsoft gépagyából*”: http://www.origo.hu/tech-bazis/20160324-naciva-tette-az-internet-a-microsoft-chatrobotjat.html

PÁPAY György. (2003). Az erős mesterséges-intelligencia-hipotézis karteziánus gyökerei. Debrecen, Vulgo, 104-122.

SEARLE, John R. (1980). Minds, brains, and programs. (Behavorial and Brain Sci-ences, 417-424.) Fordította: Thuma Orsolya. In: Pléh Csaba (szerk.) (1996). Kognitív tudomány. Budapest, Osiris Kiadó. 136-151.

SOMOGYVÁRI Zoltán. (2015). Belső reprezentáció neuronhálózatokban: modell alapú tanulás. In R. L. Kampis György, Ropolyi László. Evolúció és megisme-rés. Budapest, Typotex Kiadó. 204-208.

STEVENS, S. S. (1973). Edwin Garrigues Boring 1886-1968. Washington D. C.: National Academy of Science.

„*Turing Test success marks milestone in computing history*”: http://www.rea-ding.ac.uk/news-and-events/releases/PR583836.aspx

TURING, Alan M. (1950.). Computing Machinery and Intelligence. Mind, 433-460. (saját fordítás)