PARTI TAMÁS

DOI: 10.66103/KK.2025.4.13

Üzenet a palackban

A.M. Turing és a fordított Turing-teszt

Hogyan beszélhetünk arról, aminek a megértéséhez még nem állnak rendelkezésünkre a szükséges fogalmak, logikai struktúrák és narratívák? Hogyan beszéljen a jövőről, és annak jelennel való összefüggéseiről az, aki megfigyel és leír valami újat? Turing kérdésként fogalmazza meg az átadni kívánt információt, amit azután úgy helyez el a szövegben, hogy a közlésre szánt információ ki nem zárható logikai következtetés eredményeként jusson el az olvasóhoz. Többek között ez a metodológiai pontosság különbözteti meg őt a jósoktól.

Tárgyszavak:
fordított Turing-teszt, digitalizáció, számítógép, ChatGTP, Mesterséges Intelligencia

How can we articulate phenomena for which the requisite concepts, logical structures, and narratives are not yet available? How can an observer who seeks to describe something novel speak of the future and its relation to the present? Turing addresses this challenge by casting the information he wishes to convey in the form of a question. He situates it within the text in such a way that the intended insight emerges for the reader as the outcome of an inescapable logical inference. This methodological precision is, among other things, what sets him apart from mere soothsayers.

Keywords:
reverse Turing test, digitization, computer, ChatGTP, Artificial Intelligence

I. Teszt és "anti"- teszt

Alan M. Turing az 1950-ben, "Computing Machinery and Intelligence[1]" címen publikált tanulmányában - amelyben a "Gondolkodhat – e egy gép?" kérdést

1 A. M. Turing: I. - Computing Machinery And Intelligence. Mind, 1950. október. 433-460. o. [továbbiakban: *Turing,* Computer]

is feltette - felvetette egy teszt gondolatát. Ez a teszt a Turing – teszt, aminek ma már többféle változata is létezik. Ugyanakkor arra következtethetünk, hogy Turing ebben az írásában óvatos, logikailag ki nem zárható következtetésként a fordított turing teszt gondolatát is megfogalmazta, amely szerint egy digitális számítógép nem csupán kommunikálni lesz képes az emberrel, hanem bírálóként is részt vehet a Turing – tesztben. Turing mindezt egy olyan korban fogalmazta meg amely kor még híján volt azoknak a fogalmaknak, és a hozzájuk kapcsolódó logikai struktúráknak és narratíváknak, amelyek lehetővé tették volna, hogy megértsék azt amiről ír. Turing írásának ez az üzenete a századvég emberének, a XX. századból a XXI. századba lépő embernek szól. Ahogy maga is írja, "*[m]indazonáltal úgy gondolom, hogy a század végére a szavak használata és az általános művelt vélemény annyira megváltozik, hogy anélkül beszélhetünk a gépek gondolkodásáról, hogy ellentmondásra számíthatnánk.*"[2],

Nos Turing annak idején azonnal el is vetette a "Gondolkodhat -e egy gép?" kérdést[3], a rá való hivatkozások azonban általában mégis úgy interpretálják a cikket, hogy Turing itt kifejezetten ezzel a kérdéssel foglalkozik. A cikket elolvasva azonban kiderül, hogy Turing itt alapvetően nem ezzel a kérdéssel foglalkozik, hanem egészen pontosan az emberi gondolkodás egy speciális vetületének, a meggyőzésnek, illetve ennek kapcsán a megtévesztésnek a gépi szimulálásában rejlő lehetőségeivel.

Turing a hivatkozott[4] írásában nem csupán a "Gondolkodhat – e egy gép" kérdést vetette el, hanem egyúttal lemondott a „gép" és a „gondolkodás" fogalmainak meghatározásáról is. Ez érthető, hiszen a gép és az ember összehasonlíthatósága terén ma sem vagyunk jobb helyzetben, tekintettel arra, hogy az emberi gondolkodásra, az emberi tudatra és az emberi intelligenciára vonatkozóan jelenleg sem rendelkezünk egzakt meghatározásokkal.

A Turing - teszt alapját egy "imitációs játék"[5] képezi. A játékot hárman játsszák. A játékban három ember vesz részt, egy férfi "(A)", egy nő "(B)" és egy kikérdező/bíró (the interrogator) "(C)", aki lehet férfi vagy nő. A kikérdező a másik kettőtől

2 *Turing,* Computer. 8. o.

3 "Az eredeti kérdés, „Gondolkodnak-e a gépek?", véleményem szerint túl értelmetlen ahhoz, hogy megvitatásra méltó legyen." - M.Turing, 1950., 8.o. Turing mindezt azzal indokolta, hogy amennyiben a meghatározásokat úgy alakítanánk, hogy azok amennyire csak lehetséges, tükrözzék e szavak szokásos használatát, tehát ha a „gép" és a „gondolkodás" szavak jelentését pusztán azok általános használatából próbálnánk levezetni, akkor nehéz lenne elkerülni a következtetést, hogy a feltett kérdésre a választ egy statisztikai felmérésben kell keresnünk, amit Turing képtelenségnek tartott. Ezért az eredeti kérdést két másik, ahhoz szorosan kapcsolódó olyan kérdéssel helyettesítette, amelyek a "gép" és a "gondolkodás" szavak értelmezéséhez képest egyértelműbben kifejezhetők.

4 Vö. *Turing,* Computer.

5 Ezt a játékot gyakran használják "B" játékos kihagyásával "viva voce" néven, hogy kiderítsék, valaki valóban megértett-e valamit, vagy csak „papagájként tanulta meg" (*Turing,* Computer. 444-445. o.) Vö. még: *Bender, Emily M. – Gebru, Timnit – McMillen-Major, Angelina – Shmitchell, Shmargaret*, On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? https://faculty.washington.edu/ebender/papers/Stochastic_Parrots.pdf (letöltve 2021.07.29. napján)

elkülönített szobában tartózkodik, a játék célja pedig az, hogy a kikérdező "(C)" az A-nak és B-nek írásban, távírón vagy közvetítőn keresztül feltett kérdésekre adott választok alapján meghatározza azt, hogy a másik kettő közül ki a férfi és ki a nő. A játék során A megpróbálja C - t téves azonosításra késztetni, míg a harmadik játékos B célja, hogy segítse a kérdezőt[6].

Turing valójában arra volt kíváncsi, hogy m*i történik akkor, ha egy gép veszi át "A" szerepét ebben a játékban, és hogy ebben az esetben vajon a kihallgató ugyanolyan gyakran fog-e rosszul dönteni a játék során, mint amikor egy férfi és egy nő játszik*[7]. Ezzel már el is jutunk a Turing - teszt lényegéhez, amelyben alapesetben szintén három alany szerepel, egy ember C a bíró (interrogátor) szerepében, egy gép A tesztalanyként és egy másik ember B kontrollként. A bíró szöveges csatornán[8] kommunikál és nem tudja ki a gép és ki az ember. Feladata az, hogy azonosítsa melyik kommunikációs partnere a gép és melyik az ember.

A teszt alanyainak száma tehát minimum három (A, B és C, ezek közül C tölti be a bíró (interrogator) szerepét, de versenykörnyezetben több bíró és több chatbot is részt vehet. A játék *valamennyi résztvevője képes kell, hogy legyen olyan kommunikációra, amelyet a többi résztvevő megért*, és ehhez képest másodlagos tényező az, hogy ki milyen szerepet játszik a játékban (C férfi és nő is lehet), a Turing - tesztben pedig már a résztvevők neme is közömbös.

Az A, B és C kölcsönhatásain alapuló *narratíva* az imitációs játékszituáció története, amelyben A és B elrejti a kilétét, C-nek pedig ki kell derítenie, hogy ki kicsoda. Vagyis egy gép A és egy ember B azonos körülmények között próbál meggyőzni a személyazonosságáról egy harmadik résztvevőt C-t, aki szintén ember. A tesztnek ez a dramaturgiája az a kontextus, amelyben a megfigyelés és mérés "történetet" kap, amelyen keresztül értelmezhetővé válik.

II. Az üzenet

Turing ezután feltette a következő kérdést: *"Fordítsuk figyelmünket egy bizonyos (konkrét) C digitális*[9]*számítógépre. Igaz-e, hogy ha ezt a számítógépet megfelelő tárkapacitással látjuk el, kellőképpen növeljük a működési sebességét, és ellátjuk egy megfelelő programmal (megfelelően programozzuk), akkor C képes lesz*

6 *Turing,* Computer. 433. o.
7 *Turing,* Computer. 433. o.
8 Ma már a Turing - teszt audiovizuális változatai is léteznek.
9 *Turing,* Computer. 436. o.

kielégítően eljátszani az imitációs játékban "A" szerepét, miközben "B" szerepét egy ember (man) tölti be?"[10]

A narratíva változatlan, továbbra is az A, B és C által létrehozott játékszituációról szól, ezért is tűnik fel, hogy Turing az itt hivatkozott számítógépet C - vel jelöli, ugyanúgy mint azt a játékost aki a narratíva szerint a bíró (interrogator) szerepét tölti be. Felmerül tehát a kérdés, vajon ez csak véletlen egybeesés? Turing figyelmetlen volt és nem vette észre, hogy ezzel a szöveg nyelvtani értelmezésben zavart okoz, logikai síkon pedig egy újabb irányt nyit az irása logikai értelmezése előtt? Vagy épp ez utóbbi volt a célja a C megjelölés új kontextusban való felhasználásával? Vagyis fel akarta hívni a figyelmet a fordított Turing – teszt lehetőségére, ahol a Bíró szerepét is egy gép tölti be, és a gép próbálja eldönteni, hogy kommunikációs partnerei közül ki az ember és ki a gép. Úgy is fogalmazhatunk, hogy míg a Turing tesztben az ember a bíró, a gép a vizsgázó, addig a fordított Turing tesztben a gép a bíró és egy másik gép a vizsgázó. A mérce azonban továbbra is az ember.

Gondolhatnánk, hogy amikor Turing az imitációs játékproblémát a Turing-tesztté átfogalmazva felcseréli a szerepeket, és a géppel veteti át A szerepét, rábírva azt arra, hogy a beszélgetés során adja ki magát embernek, pusztán csak arra volt kíváncsi, hogy egy gép sikeresen el tudja-e játszani az imitációs játékban az A emberi játékos szerepét, míg a másik szerepet egy valódi ember játssza. A fordított Turing-teszt pedig (ahol a gép a bíró, aki megpróbálja azonosítani az embereket) sokkal később jelenik meg, az 1990-es/2000-es években a CAPTCHA tesztek kontextusában. Gondolhatnánk továbbá, hogy Turingnak nem volt célja a kérdező személyének a géppel való azonosítása, hiszen nem is javasolja azt, és amit látunk, az valójában Turing szavainak véletlen kétértelműsége. Csupán látszólagos átfedés, ami abból adódik, hogy Turing zavaró módon „újrahasznosította" a „C" betűt. – Vagyis Turing figyelmetlen volt.

A magam részéről azonban - különös tekintettel a cikk témájára, és figyelembe véve azt is, hogy Turing kódfejtő matematikus volt - nem hiszem, hogy figyelmetlen lett volna, vagy ne javított volna ki minden, a szándékának megfelelő értelmezést zavaró tényezőt. Inkább úgy gondolom, hogy Turing szándékosan alkalmazta a számítógépre a C megjelölést[11], amivel egy másik lehetséges irányt adott az írása logikai értelmezésének. Mégpedig egy olyan irányt, ami már 1950-ben közvetlenül rámutatott a fordított Turing - teszt lehetőségére, vagyis arra,

10 *Turing,* Computer. 441. o. "Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?"

11 A szöveg nyelvtani és a logikai értelmezés alapján nincs okunk különbséget tenni (C) és C között.

hogy a számítógép, nem csupán tesztalanyként, hanem bírálóként is részt vehet majd a tesztben.

Mindenesetre Turing szövegéből itt már nem következik a bíró típusa (ember vagy gép, itt már nem is említi a bírót), azonban azzal, hogy a gépet ugyanitt C-vel jelöli a szöveg alapján már nem zárható ki, hogy a C (az interrogator, a bíró) szerepét is gép töltse be, annál is inkább mert az idézett mondatban Turing kifejezetten arról beszél, hogy a „konkrét" C számítógép képes lehet az ember, vagyis A szerepét eljátszani, miközben B továbbra is ember, vagyis a gép emberi szerepbe is behelyettesíthető. (Emlékezzünk, az eredeti imitációs játékban C = ember (bíró), A = férfi (későbbi változatban: gép), B = nő (később: ember). Vagyis Turing itt azt a hipotézist veti fel, hogy C számítógép részére A szerep értelmezhető, ami logikailag megnyitja a lehetőséget arra, hogy C számítógép C szerepre is alkalmas legyen, és így gép bírálja el a másik gépet, ezzel pedig Turin lényegében elvezet minket a fordított Turing – teszt megfogalmazásához.

Persze a CAPTCHA[12] tesztek alkalmazása óta a fordított turing teszt már nem újdonság számunkra, hiszen a CHAPTCHA tesztek esetében is a gép bírálja el, hogy ki tekinthető humán, vagy gépi kommunikációs partnernek. Az azonban ma is figyelemre méltó, hogy Turing 1950-ben milyen pontosan feltéképezte mindezt, és hogy milyen finom és elegáns logikai megoldással hozta ezt a tudomásunkra. Egy nyelv által kifejezett „logikai kapu[13]" megalkotása révén tulajdonképpen anélkül fogalmazta meg a fordított turing teszt lényegét, hogy azt megnevezte volna.

III. A jelen

Turing számított rá[14], hogy a "Gondolkodhat-e egy gép?" kérdést legkésőbb a XX. és XXI század fordulóján már valóban fel lehet tenni.[15] Nos a XXI. század első negyedének a végén a ChatGPT5 egyik változata egyenesen a Thinking megjelölést viseli.

12 CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) – teljesen automatizált Turing teszt a számítógép és az ember megkülönböztetésére.

13 A logikai kapu technikailag egy digitális áramkör, amely egy vagy több bináris bemeneti jel (0 vagy 1) alapján logikai műveleteket végez, és egyetlen kimeneti jelet produkál. Ezek az alapvető digitális egységek alkotják a számítógépek, memóriák és más digitális rendszerek hardverbázisát, lehetővé téve a bináris adatok összetett manipulációját és az információ feldolgozását.

14 *Turing,* Computer. 441. o. Úgy gondolom, hogy körülbelül ötven év múlva lehetséges lesz olyan számítógépeket programozni, amelyeknek a tárolókapacitása körülbelül 109, és amelyek olyan jól játszanak az utánzás játékát, hogy egy átlagos kérdezőnek öt percnyi kérdezés után legfeljebb 70 százalék esélye lesz a helyes azonosításra.

15 *Turing,* Computer. 441. o.

Az MI adaptációja Humán – Gép kommunikáción alapuló jelenség, amelynek során a kommunikációban résztvevő felek, az MI és az ember összehasonlíthatóvá válnak, és ennek során közöttük hasonlóságok és különbségek egyaránt felfedezhetők. Amíg azonban az MI felépítését és működésének elveit pontosan ismerjük, ugyanezt az emberi agyról, az emberi gondolkodásról, az emberi intelligenciáról korántsem mondhatjuk – el, vagyis az ezekről alkotott tudásunk nem teljes, ami lényeges akadálya annak, hogy a gépi és a humán folyamatokat és jelenségeket valóban egymáshoz tudjuk "mérni". Turing korábban elvetett kérdésének megválaszolásához tehát még mindig hiányzik az emberi gondolkodás fogalmának egzakt meghatározása. Az azonban ennek ellenére is nyilvánvaló, hogy humán – MI relációban legfeljebb hasonlóságról beszélhetünk, nem pedig azonosságról, vagyis továbbra is legfeljebb az emberi intelligencia egy modelljével nem pedig humán értelemben vett intelligenciával van dolgunk.

A mai multimodális generatív MI ágensek (modellek) már nem csupán az ember által alkotott szöveg, írott és beszélt változatának, hanem az emberi hangnak és az emberi mozgásnak az egyre tökéletesebb modellezésére képesek, ezért kiválóan alkalmasak arra, hogy a turing tesztet sikeresen teljesítve megtévesszék a velük kommunikáló humán alanyokat. De nem csupán őket.

A Corvinus egyetem kutatóinak egy, a közelmúltban publikált kutatása azt igazolja, hogy az MI ágensek nem teljesítenek jól a gépi partner felismerésében. Vagyis a mesterséges intelligencia (AI) és az ember elkülönítése nemcsak a hús-vér bírálóknak okoz problémát[16]. Vagyis egyre nehezebbé válik a botok és a valós felhasználók megkülönböztetése, ami nemcsak a közösségimédia-felületeken jelent komoly kockázatokat, hanem a technológia terjedésével minden társadalmi rendszerben, így a gazdaságban és a politikában is. A kutatók szerint az lenne ideális, ha az MI jobb lenne annak felismerésében, hogy emberrel vagy géppel beszélget-e, mint abban, hogy embernek tettesse magát. Ennek pedig azért van kiemelt jelentősége, mert az ember és a mesterséges intelligencia párhuzamos, egymással kölcsönhatásban álló evolúciójában a biztonságunk érdekében szükség van egy olyan, ezzel az evolúcióval együtt fejlődő technológiára/eljárásra amely világosan képes különbséget tenni a gép és az ember között. Ez a megközelítés pedig – ahogy Hajnal fogalmaz - a mesterséges intelligenciával szemben biztonsági elvként elvárja, hogy az MI mindig legyen jobb az embernél annak felismerésében, hogy MI-vel vagy emberrel van – e dolga.

16 A legfejlettebb AI is megbukott a fordított Turing-teszten: nem tudja eldönteni, hogy emberrel vagy géppel beszél - Qubit Podcast, https://qubit.hu/2025/07/31/a-legfejlettebb-ai-is-megbukott-a-forditott-turing-teszten-nem-tudja-eldonteni-hogy-emberrel-vagy-geppel-beszel, https://omny.fm/shows/qubit-podcast/a-legfejlettebb-ai-is-megbukott-a-ford-tott-turing-teszten-nem-tudja-eld-nteni-hogy-emberrel-vagy-gppel-besz-l Elhangzik a Qubit podcastjában, Hajnal Zsófia, a Budapesti Corvinus Egyetem doktorandusza és a HUN-REN KRTK Világgazdasági Intézet junior kutatója, valamint Manran Zhu és Vásárhelyi Orsolya.

A kísérlet eredményeképpen az volt megállapítható, hogy a gépi bíráló az esetek döntő részében (gépek közötti kommunikációban nyolc esetből hétszer) emberként azonosította a gépi kommunikációs partnerét, vagyis nem volt képes megkülönböztetni az embert a géptől, ami a szóban forgó narratíva szempontjából aggasztó eredmény.

IV. A sikeres vizsgázó, aki vizsgáztatóként megbukott önmaga előtt

Megállapíthatjuk, hogy az MI olyan sikeres az emberi kimenetek modellezésében, hogy a különbséget már egy másik fejlett MI sem képes felismerni és akkor is átmegy a Turing teszten, ha a bíráló szerepét egy hozzá hasonló gép látja el. Mondhatni az MI csak mint bíráló bukott meg a Turing – teszten, ezzel szemben vizsgázóként annyira sikeres, hogy gépi minősége vonatkozásában nem csak az embert, hanem gépi "önmagát" is képes megtéveszteni. Ebben a megközelítésben pedig, általában a gépi intelligencia oldaláról közelítve olyan, mint egy sikeres vizsgázó, aki a saját pályáján képes önmagát is megtéveszteni, és így vizsgáztatóként megbukik önmaga előtt. Mindez a megtévesztésnek olyan dimenzióit nyitja meg, amelyekre nem vagyunk felkészülve. Ezt pedig érdemes alaposan végiggondolni …

Parti Tamás, Fiume, 2025. augusztus 26.