

REFLEX

Neuman Péter

NAÍV GONDOLATOK GÖDEL NEMTELJESSÉGI TÉTELEI ÉS A MESTERSÉGES INTELLIGENCIA LEHETETLENSÉGÉNEK KAPCSOLATÁRÓL

„Akik úgy gondolják, hogy a fizika törvényei társadalmi konvenciók csupán, próbáljanak meg kilépni ezekből a konvenciókból a 21. emeleti lakásom ablakán át” – jegyzi meg szellemesen a fizikus Alan Sokal (1996a)¹. Valami hasonlót mondhatnánk azoknak, akik bizonyítani kívánják, hogy mesterséges intelligencia nem létezhet. Hívják fel mondjuk az Amazon vagy valamelyik jobb bank automata ügyfélszolgálatát!

Mesterséges intelligencia ugyanis láthatóan van, sokunk számára a mindennapok része. Ez azonban nem akadályozta meg a különböző hátérrel, de filozófiai indíttatással rendelkező elemzőket abban, hogy megfogalmazzanak egy no-go tételt² a mesterséges intelligencia lehetetlenségével kapcsolatban. Az igazsághoz hozzátartozik, hogy Gödel és a mesterséges intelligencia (MI) kapcsolatáról már jóval azelőtt elkezdődött a vita, hogy megjelentek volna működő MI megoldások. J.R. Lucas 1961-ben megjelent cikkében (Lucas, 1961) arra a következtetésre jut, hogy számítógépekkel nem modellezhető az emberi elme. Állításának bizonyítására Gödel tételét idézi, miközben bizonyítottnak tekinti, hogy a mesterséges elme olyan formális rendszer, amelyre ez a tétel vonatkozik, hiszen szükségszerűen tartalmazza a természetes számok aritmetikáját. Lucas nem foglalkozik azzal, hogy vajon az MI-nek muszáj-e modellezni az emberi agyat, ahhoz, hogy működjön, és intelligenciának nevezhessük.

A kételyeket megfogalmazók szerint a modern matematikai logika talán legjelentősebb Gödel nevéhez fűződő tétele(i) bizonyítják, hogy nem várható mindenki ízlését, igényét, definícióját kielégítő olyan mesterséges intelligencia, amely felveszi a versenyt (bármit is jelentsen ez) az emberi aggyal. Világos, hogy az állítást már csak azért is nehéz cáfolni, mert az MI definícióknak, melyek olykor egymást kizáróak, se szeri se száma. A definíciók részletes tárgyalása jócskán túlmutat ezen cikk keretein, ezért a következőkben MI alatt olyan digitális számítógépek segítségével működtetett rendszereket értek majd, amelyeket a közvélekedés mesterséges intelligenciának tekint.

Az ötletet, tehát, hogy a mesterséges intelligenciának elvi korlátai vannak Kurt Gödel a múlt század egyik legjelentősebb matematikusának felfedezése adta. Ha a kételyek valódiak, nem a mesterséges intelligenciába vetett hit lesz az egyetlen „projekt”, amelynek Gödel kijózanító tételei vetnek véget. Az 1931-ben publikált úgynevezett nemteljességi tételek David Hilbert nagyívű programjának a bukását jelentették annak bizonyításával, hogy alkotható lenne a matematika minden ismert eredményét magában foglaló ellentmondásmentes és minden igaz állítást magába foglaló axiomatikus rendszer.

¹ Sokal 1996-ban *A határok áttörése: arccal a kvantumgravitáció transzformatív hermeneutikája felé* címmel jelentetett meg egy teljesen értelmetlen cikket a Social Text című tudományos folyóiratban (Sokal, 1996a). Néhány héttel később Egy fizikus kísérletei a kulturális tanulmányokkal című írásában, melyből a fenti idézet is származik leleplezte magát egy másik lapban.

² Az elméleti fizikában no-go tételeknek nevezzük azokat, amelyek valamilyen fizikai rendszer/szituáció lehetetlenségét állítják és bizonyítják.

Gödel állítása kicsit pontosabban így szól: Minden olyan véges módon meghatározott konzisztens formális rendszer, amelyik elég gazdag ahhoz, hogy tartalmazza az aritmetikát szükségszerűen nem teljes és eldönthetetlen, abban az értelemben, hogy vannak benne állítások, amelyek igazságáról nem tudunk meggyőződni.

Habár a Gödel tétel központi kérdése, vagyis, hogy a Hilbert program sikerét lehetővé tevő axiomatikus rendszer alkotható-e már a 20. század első harmadában benne volt a levegőben³(ahogy a relativitáselmélet is már a 19. század legvégétől fogva), a bizonyítás nagyszerűsége mellett maga a kérdésre adott nemleges válasz is hosszú évtizedeken keresztül hatott a legkülönbözőbb területeken dolgozó természet- és társadalomtudósok gondolkodására. Nem csupán a bizonyítás, hanem az állítás minden rétegének legalább részleges megértése is idő és energia igényes feladat, amelyre kevesebben vállalkoznak, mint amennyien Gödelről beszélnek vagy írnak. Sokan amolyan univerzális törvényként fogták, fogják fel ma is, és vonják le belőle a számukra fontos következtetéseket függetlenül attól, hogy ezen következtetések jogosak vagy nem. A félreértések többsége abból származik, hogy a tételt idézők nem minden esetben törődnek azzal, hogy mire is vonatkozik ez a törvény, nevezetesen a természetes számok aritmetikáját magába foglaló formális matematikai rendszerekre, és alkalmazni kívánják a tételt például a jogrendszerekre. Elképzelhető, hogy nem lehetséges ellentmondásmentes kontinentális jogrendszert alkotni, viszont ez nem vezethető le Gödel tételéből.

A félreértések, téves értelmezések mellett azonban valós és jogos következtetések is felfedezhetők a Gödel tételek utóéletében, még akkor is, ha ezekkel nem mindenkinek kell feltétlenül egyetérteni.

A kimagasló műveltségű magyar fizikus és bencés szerzetes, Jáki Szaniszló (Jáki, 1992) az emberi megismerés határainak létezését látja igazolva Gödel által, amikor azt írja, hogy a matematikai fizika konstrukcióinak mélye mindig rejtve marad előttünk, más szóval mindig lesznek korlátai a felfedezések pontosságának (sic). Egy másik helyen Jáki kissé rosszállóan jegyzi meg (Jáki, 2006), hogy míg a fizikusok mintegy harminc évig figyelemre sem méltatták az ő Gödel tétellel kapcsolatos megállapítását, Stephen Hawkingnak 2002-ben, Dirac születésnapján Cambridge-ben tartott előadásán elhangzottakat kritika nélkül elfogadták, vagyis azt, hogy a nemteljességi tételek következménye, hogy nem találták meg a theory of everything-et, a mindenség elméletét. Igaz, Hawking jó hírt hozott, Gödelre hárította a felelősséget egy olyan projekt vélelmezett bukása miatt, amely sokak több évtizedes munkájának kudarcát jelentette. Jáki ezzel szemben szerénységre intette fizikus kollégáit azzal, hogy figyelmeztette őket, a világ teljes megismerése nem várható a fizikától.

Érdeemes észrevenni, hogy amíg Jáki az emberi megismerés határait látja bele a Gödel tételekbe, addig Roger Penrose (1995) épp ellenkezőleg, az ember gondolkodásnak a gépi gondolkodással szembeni magasabb rendűségeként értelmezi azt, hogy amíg a gépek, a formális rendszerből, a szigorúan a logika által diktált következtetési szabályok béklyójából képtelenek kilépni, az emberi gondolat szabadon szárnyalhat, kikerülve így a nemteljesség

³ Az 1930-as königsbergi konferencián például Gödel már ismertette tételét. A tétel tartalma helyett inkább a bizonyítás volt újszerű a jelenlévők, így Neumann János számára is, aki, maga is foglalkozott a témával. (Dawson, 1984)

Szkilláját, illetve az inkonzisztencia Kharübdiszét. Penrose kritikusi ugyanakkor kétségbe vonják ezt az álláspontot azzal, hogy Gödel nem számítógépekről, hanem formális rendszerről beszél, amely az emberi gondolkodásnak is keretül szolgál. Ebből a keretből elvileg ki lehet törni az embernek éppúgy, mint a gépnek, csak fel kell adni a formális rendszerben való konzisztens lavírozás követelményét.

Jáki figyelmeztetése annak ellenére nem vált népszerűvé, hogy a fizikusok ekkorra már hozzászórtak ahhoz, hogy a megismerés nem lehet teljes. Gondoljunk például a határozatlansági relációra, a kvantummechanikai Neumann-Dirac féle méréselméletre, a kaosz elméletre, vagy éppen csak arra a tényre, hogy a kísérletek szükségszerűen pontatlanok. Nem szokatlan tehát, hogy a fizikai kutatás falakba ütközik. Ezeket a falakat azonban a fizikusok találták meg, és a fizikai kutatások eredményeképpen jutunk el hozzájuk.⁴ A Gödel-féle megismerési korlátot a fizikán belül nem tudjuk kezelni, mélyebb szinten, a logika szintjén épül be a fizikai elméletbe, kívülről jönnek, a fizikus semmit sem tehet eltüntetése érdekében. Bizonyos szerzők (Breuer, 2001; Barlow, 2011; Penrose, 1994) a gödeli korlát és a fizika kapcsolatának elemzésekor együtt tárgyalják a különféle okból és különböző szinteken megjelenő – pl. matematikai és fizikai - „lehetetlenségeket” (Barlow, 2011). Penrose a mesterséges intelligenciával kapcsolatos érveiben (Penrose, 1994) is összeköti a gödeli és a kvantummechanikai határozatlanságokat.

A megismerés akadályai a felhasználó, a tudós szemszögéből rokonságban vannak egymással. Kérdés, hogy érdemes-e, lehetséges-e, és ha igen, akkor hogyan lehetséges és érdemes ezeket ugyanazzal a módszerrel vizsgálni.

Ha a Gödel-tétel akadályozza az MI működését, vagy azt vesszük észre, hogy az időjárás előrejelzés pontossága korlátokba ütközik, az egy-egy projekt kudarcát jelenti. A kvantummechanikai határozatlanságokkal kapcsolatban nem ennyire egyértelmű a helyzet, hiszen ezek nélkül nem létezne kvantummechanika, így például kvantum számítógép vagy atomreaktor sem. Nem egyértelmű tehát, hogy a megismerési korlátok, a lehetetlenségek negatív hatással vannak a fejlődésre. Ennek illusztrálására álljon itt egy elméleti fizikából kölcsönzött példa.

A fizikai megismerés egyik régóta ismert korlátja az, hogy amíg a fizika törvényei csaknem kizárólag zárt rendszerekre érvényesek, a valóságban csak nyílt rendszereket látunk. Azt gondoljuk, hogy a zárt rendszerre meghatározott törvények jó közelítései a valódi, nyitott rendszer bélieknek, és az gyakran igaz is, de nem mindig. Különösen azokban az esetekben nem, amikor a zárt rendszerre levezetett szabályok eredményei nem számok, melyektől való eltérés kerekítési hiba, hanem kvalitatívak, esetleg ontológiai következménnyel járnak. Egy elemi részecske létezése is múlhat azon, hogy a fenti közelítés mennyire jó. A zárt rendszerben érvényes szabályok több esetben vezetnek nemkívánatos, intuícióval ellentétes eredményre. Ilyen például az a régóra ismert tény, hogy például a klasszikus mechanikában kauzalitás nem értelmezhető. Bertrand Russell már 1913-ban észrevette ezt (Russell, 1913). A zárt rendszerek problémájának kezelésében segítségünkre lehetnek az úgynevezett

⁴ Nem annyira egyértelmű a helyzet a kaotikus jelenségek esetében, hiszen itt a megismerés korlátait az elméletben megjelenő nemlineáris differenciálegyenletek okozzák, melyek – ha tetszik – fizikán kívüliek. Ugyanakkor a jelenségnek létezik fizikai interpretációja, ami Gödel esetében hiányzik.

effektív elméletek, melyek bizonyos bonyolult és igényes matematikával meghatározott kényszerfeltételek, peremfeltételek bevezetésével kiegészítve a zárt rendszerre megalkotott törvényekből a való világra érvényes eredményre juthatnak. Az effektív elméletek nem csupán a térbeli koordináták tartományát szűkíthetjük le a a vizsgálandó rendszernek megfelelően, hasonló módszerrel kezelhető például az a probléma is, amikor egy fizikai rendszer viselkedését például csak kis energián vizsgáljuk. Ilyen értelemben effektív elmélet a 20. századi részecskefizika több elmélete, például a Weinberg Salam modell is. Polonyi (Polonyi, 2014) egy klasszikus mechanikai effektív elméletben mutatta meg, hogy az eredetileg hiányzó időirány megjelenik, lehetőséget adva arra, hogy kauzalitást egyáltalán értelmezzünk.

Gödel tételének bizonyításakor megalkotott egy ún. Gödel mondatot, mely nem volt bizonyítható és cáfolható sem az adott formális rendszerben, és belátta, hogy nem konstruálható olyan rendszer, amelyben nem találunk ilyen mondatot. A Gödelre hivatkozó MI-szkeptikusok ezt használják fel állításuk igazolására. Ugyanakkor az MI Gödel mondaton kívüli tartományában nincs ilyen problémás mondat. Ha a fizikai effektív elméletek mintájára úgy tekintünk a működő MI automatákra, mint amelyek a Gödel mondaton kívüli részét használják csak a világnak, úgy az MI automata nem kerül olyan helyzetbe, hogy tanácstalanul kelljen megállnia egy logikai lépés elvégzésekor. Miután a Gödel mondatok száma, bár végtelen, elenyésző az összes mondatéhoz képest, az esetek döntő többségében a gondot észre sem vesszük. Hogy ez így van, azt a természetes számok aritmetikájának mindennapi tökéletes működése is igazolja. A kérdés az, hogy be tudjuk-e úgy állítani az MI automatát, hogy elkerülje a Gödel mondatot. A fizika ismert effektív elméleteiben ez rendszerint nem nehéz feladat, mert vagy a konfigurációs vagy a fázistérnek egy jól körülhatárolható összefüggő részében dolgozunk, például egy adott érték alatti vagy feletti energiával rendelkező részecskéket vizsgálunk vagy egy véges tartományát a konfigurációs térnek. Csak olyan folyamatokat vizsgálunk, amelyek az effektív rendszeren belüliek. Az effektív fizikai rendszerekkel felvázolt analógia talán azért sem haszontalan, mert végső soron az MI automata is egy fizikai rendszer, melyben a mi effektív tartományunkat úgy definiáljuk, hogy kikerülje a Gödel mondatot. Kérdés, hogy meg tudjuk-e határozni, hogy mi lesz ez a tartomány. Gödel mondat egy formális rendszeren belül valószínűleg végtelen sok van, de ez önmagában nem jelent gondot.

De szükség van-e a Gödel mondatok elkerülésére ahhoz, hogy az automata működjön, értsd: hasznot hajtson számunkra. Mi történik, ha az elkerülés nem sikerül, és beleütközünk egy problémás Gödel mondatba? Annál rosszabb biztosan nem, hogy az MI automata nem lesz képes annak a problémának a megoldására, amelyet várunk tőle és leáll, esetleg téves választ ad. Nehéz lenne azt állítani, hogy az emberi aggyal ilyesmi nem fordul elő egy-egy probléma megoldása során. Dummett (Dummett, 1963; Molnár, 2020) a nyelvfilozófiai használatelméletének megközelítésével vizsgálja a Gödel tétel filozófiai következményeit. Ha elfogadjuk, hogy egy szó vagy mondat jelentését a használata és csakis a használata tudja meghatározni, továbbá azt, hogy a Gödelnél olyan jelentős szerepet játszó igazság modelfüggő, az MI automata – melynek fizikai állapotai a használatban lévő formális rendszer modelljét jelölik ki - viselkedése Gödel mondat szerű „eldönthetetlen” esetekben mit kell, hogy jelentsen?

A fentiekben azt próbáltam végig gondolni, hogy Gödel tétele megakadályozza-e azt, hogy hasonlóan kezeljük az emberi és a mesterséges intelligenciát. Egyfajta Turing tesztnek alávétve őket, látunk-e olyanfajta viselkedést az egyikben, amelyik teljesen hiányzik a másikban. Abban az esetben, ha fekete dobozként tekintünk az MI automatára és a vele együtt vizsgált emberi elmére, azt látjuk, hogy mind a kettő valamilyen, a szokványostól eltérő választ ad bizonyos kérdésekre, nem válaszol, vagy egymást kizáró válaszokat ad. Lehet, hogy ezek a pontok a két intelligencia esetében nem ugyanott vannak, de a külső szemlélő nem képes alapvető különbséget tenni a gép és az ember, fogalmazzunk így, tanácstalansága között. Amellett érvelek tehát, hogy pusztán Gödel tételéből nem következik a cikk elején említett no-go tétel.

Irodalom

- Barrow, J. D. (2011): Gödel and physics. In (Baaz el. al), *Kurt Gödel and the Foundations of Mathematics: Horizons of Truth* (255–276). Cambridge University Press.
- Breuer, T. (2001): Von Neumann, Gödel and Quantum Incompleteness. In (Stöltzner, M), *John von Neumann and the Foundations of Quantum Physics* (75–82). Springer Netherlands.
- Dawson J. W. (1984): *The Reception of Gödel's Incompleteness Theorems. In Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984, 253-271.
- Dummett, M. (1963): *The Philosophical Significance of Gödel's Theorem* In (Dummett, M.) *Truth and other enigmas*. Harvard University Press.
- Jaki, S. L. (1992): *The relevance of physics*. Scottish Academic Press.
- Lucas, J.R. (1961): *Minds, machines and Gödel*. *Philosophy*, 36, 112–127.
- Molnár, Z. G. (2020): *Michael Dummett: A Gödel-tétel filozófiai jelentősége*. In (Molnár Zoltán Gábor), *Russelltől Gödelig. Esszék a matematika filozófiai problémáiról*. TypoTeX. [előkészületben].
- Penrose, R. (1994): *Shadows of the Mind*. Oxford University Press.
- Russell, B. (1913): *In the notion of cause*. *Proceedings of the Aristotelian Society, New Series*, 13, 1-26.
- Sokal, Alan D. (1996a): *Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity*. *Social Text*, 46–47. 217–252.
- Sokal, A. D. (June 5, 1996b): *A Physicist Experiments with Cultural Studies*. *Lingua Franca*. https://physics.nyu.edu/faculty/sokal/lingua_franca_v4/lingua_franca_v4.html