

# SEBŐ PÉTER

## Ki beszél?

„Ha csupán gondolkodom, anélkül hogy könyvet akarnék írni, akkor körülugrálom a témát; számomra ez az egyetlen természetes gondolkodásmód. Sorrendi kényszerrel végiggondolni, ez nekem a legnagyobb kényszerítés. Mármost próbálkozzam-e vele egyáltalán? Kimondhatatlanul sok fáradságot pazarolok arra, hogy megszabjam gondolataim rendjét, mely rendnek talán nincs is semmi értéke.”

(Ludwig Wittgenstein)

### I. EMERGENCIA

Ez az esszé a mesterséges intelligencia nyelvhasználata és az emberi megértés határai közé feszített térben mozog. Esszé – abban a montaigne-i értelemben, amikor a gondolat még nem megérkezni akart, hanem körbejárni, tapogatózni, kérdezni. A forma – ugrálás, szétesős, majd, remélem, újra összeálló – nem teljesen választott, hanem inkább a téma által kényszerített következmény: annak a lenyomata, ami az MI által létrehozott szövegeket is jellemezi, ahol az összefüggések gyakran nem lineárisak, és az olyannyira vágyott megértés csak kerülőutakon hajlandó elénk állni.

Nem kíván eligazítani, sem végkövetkeztetést kínálni – inkább a jelentés, a szerzőség, az igazság, a tudás és a tanári szerep kérdéseit járja körül egy olyan világban, ahol a gépi válasz már nem pusztán lehetőség. A mesterséges intelligencia nem csupán technológiai újítás – hanem újfajta tükör, amely nem a jövőt mutatja, hanem a jelen rejtett alakzatait. De nem az MI felhasználási módjait tárgyalom, nem ismertetőt írok – inkább részben filozófiai kísérlet, részben egzisztenciális kérdésfelvetés ez. A kiindulópont tanári. És a tanár legjobban talán kérdezni tud. A szöveg így főként réges-régi kérdésekkel közelít egy új eszközzel kapcsolatos problémához. A kérdéseim mögött ott húzódik néhány ki nem mondott kérdés is – azok, amelyeket tanárként szívesen feltennénk. Talán a legfontosabb ezek közül: hogyan segítheti a mesterséges intelligencia az oktatás eredményességét, miközben látszólag épp elvenni készül valamit a hagyományos tanári pozícióból? Mivel erre nincs még releváns válaszom, inkább sok más irányból kérdezek – remélve, hogy ezek révén talán mégis közelebb kerülünk ahhoz, hogy egyszer válaszolni tudjunk majd.

Számtalan, igen eltérő habitusú gondolkodót idézhetnék megerősítésképpen. Az én olvasatomban Szókratész, Montaigne, Feyerabend vagy Kotakowski egyaránt arra figyelmeztet, hogy ne bízzunk vakon az egyetemes elvekben, és válaszok helyett inkább csak kérdezzünk. Mert miként az MI használatánál is hamar megtanultuk, a jó kérdés megfogalmazása sokszor fontosabb, mint a válasz. Így az esszé – ha komolyan veszi magát – mindig kérdő, bizonytalan s relativista egy kicsit, mert vállalja a részlegességet, a pillanatban való igazság lehetőségét és a nagyfokú bizonytalanságot.

Az MI által generált szövegek – legalábbis sok közülük – egyfajta emergens struktúráként viselkednek: nem zárt program szerint épülnek fel, hanem valószínűségi mező alapján, önszerveződő módon jönnek létre; nem az elemek egyszerű összege adja az értelmüket, hanem az elemek mintázata; nem a létrehozó szándékából (hiszen olyan nincs), hanem a feltett kérdésből és a határból merítik jelentésüket. Ez az írás – ha úgy tetszik – ugyanerre tesz kísérletet: a kérdésekből és a válaszok töredékeiből, ha jól hallunk bele, valami kirajzolódhat. Remélem, olyan, mint egy fehérje hajtogatott mintázata: elsőre kiismerhetetlen kuszaság, de belülről konzisztens – és ha kell, az MI statisztikai elemzése rendet is talál benne.

Természetes, hogy igyekszem minden részben a pedagógiai kihívások felé is nyitni. De sajnos itt leginkább csak kételyeim vannak, még a kérdések tekintetében is. A válaszokkal éppen

ezért itt sem éri meg sietni, ugyanakkor a kérdések felől elfordítani a fejünket nincs módunk. Ha tanárként közelítünk a mesterséges intelligenciához, még azt a kényelmes álláspontot sem fogadhatjuk el, hogy csupán felhasználóként tekintünk magunkra. Mert nemcsak feladatmegoldó eszköz, mint egy számológép. És nem egy új szövegszerkesztő vagy digitális tábla. Az oktatás – ha valóban az értelem megszólítását tartja fontosnak – nem válaszolhat pusztán gyakorlati módon egy olyan eszközre, amely a gondolkodást próbálja utánozni.

A tanári szerep már jó ideje válságban van – nemcsak szociológiai, hanem ontológiai értelemben is. Nem világos, mihez képest hiteles, nem tudni, milyen típusú tudás teszi érvényessé, és főként: miért is beszél. Az állandóan vitatott ismeretanyag önmagában nem teszi súlyossá, a módszertani sokszínűség nem teszi vonzóvá, az intézményi rendszer nem teszi jelentősebbé. Ami talán még megmaradt: az a hit, hogy ha egy tanár megszólal, akkor valami több történik, mint pusztán adatközlés. A tanári szó – legalábbis elvben – nem információt ismétel, hanem értelmet ajánl fel, és ez emergensen ráépül az információra, amit a tananyag tényei hordoznak. De mi adhat ilyen értelmet a tanári mondatoknak, ha nem a használat, a közös szabály, a valós nyelvi játékba vetettség?

A nyelv határai nemcsak a világ határai – hanem a megértés határai. És Wittgenstein kései korszakában már nem is a világ leképezéséről beszélt, hanem a nyelvjátékokról, a jelentés közösségi létéről, a használatból fakadó értelemről. Arról, hogy az értés nem szikár logikai viszony, hanem életforma. És mintha ezt ismételné számunkra – egyre nyugtalanítóbb hangtónussal: Mikor mondjuk egy állításra, hogy igaz? És mit teszünk, amikor ezt mondjuk: igaz?

A mesterséges intelligencia ebbe a mondatba lép be: nem mint igazmondó, hanem mint résztvevő. Úgy játszik, mintha benne volna a nyelvjátékban – miközben nincs teste és saját szabálytapasztalata. És ettől válik a mondata különössé: nem hazudik, nem téved – csak nem játszik valóban.

Mint Wittgenstein írja a *Filozófiai vizsgálódásokban*: a jelentés nem más, mint a használat a nyelvben. De az MI nem emberi módon használja a nyelvet, hanem válaszokat generál egy olyan mintázat alapján, amit mi nem is látunk benne. És ebben az egyetlen különbségben talán ott van minden.

Hogyan tanítsunk egy olyan világban, ahol a nyelv már nemcsak közös játék lehet, hanem alkalmanként csak imitált visszhang? És ezzel kapcsolatban nem az lenne a tanár egyik legfontosabb – csendes, de kitartó – feladata, hogy újra visszaterelje a válaszokat a helyükre? Újra játszani – emberi szabályok szerint. De mit jelent ez akkor, ha egy gép értelmező mondatokat kezd alkotni, érveket szerkeszteni, a kérdésekre figyelni? És miközben a rejtett nyelvi minták bámulatos ügyességgel bomlanak ki a válaszaiban, sokunkban felmerülhet a kérdés – amit bizonyosan feltenni sem szeretnénk: van-e még különbség tanár és nyelvi modell között?

Az iskolában, ahol a tanár eddig a tudás értelmezője lehetett – legalábbis a klasszikus fel fogás szerint –, most új szereplő jelenik meg. Ugyan nem nevel, nem fegyelmez – de beszél. És sokszor strukturáltabban, udvariasabban, tárgyyszerűbben. Diákjaink – akár titokban, akár nyíltan – kérdeznek tőle. És a válaszok néha zavarba ejtően jók. Miféle kihívás ez számunkra? Eszköz, amelyet be kell építeni? Veszély, amit ki kell zárni vagy legalábbis erősen kontrollálni? Kortünet, amit értelmezni kell?

Az igazi kihívás talán az, hogy nemcsak eszközként, hanem beszédpartnerként lép be az osztályterembe. Nem olyan, mintha egy könyvből szólna a válasz, hanem mintha egy személy szólna hozzánk a gépből. Ha az MI-t valóban szeretnénk beengedni az iskolába, akkor nekünk is jól kell használni (még ha ma nem tudjuk pontosan, mit is jelent ez). De nem elég csak jól használni. Valamennyire értenünk kell azt is, miben más az a beszéd, amely nem tőlünk ered, és mégis hozzánk szól – hogy miben áll az emberi értelemre alapozott beszéd különössége. És ezt a tudást nem a technika adja. Az MI pedagógiai alkalmazása nem kezdődhet a felhasználói kézikönyv megismerésével. Nem elég tudni, hogyan kell promptolni, mire jó egy multimodális modell. Nem elég tudni, hogy mi a különbség a ChatGPT, az xAI és a

Copilot között, mi a neurális hálózat, vagy mit jelent a hallucináció. A tanári szerep mindig is filozófiai volt – ha nem is nevében, de lényegében. A tanár az, aki nemcsak tudást közvetít, hanem viszonyt teremt: a világ, a nyelv, az igazság, a kétely, a megértés és az értékek viszonyait. Most nem pusztán az a tét, mit tud az MI – hanem hogy mi mit értünk meg vele kapcsolatban saját szerepünkben.

A mesterséges intelligencia nemcsak a tanítás módját alakítja át, hanem annak értelmét is. És ezzel is sok régi-új megválaszolatlan kérdést hoz a felszínre: mit jelent tanítani – ha a tudás egy kérdéssel bármikor előhívható? Mit jelent megérteni egy problémát – ha egy válasz azonnal és jól formáltan megérkezik a géptől? Mit jelent emberként gondolkodni – ha a gondolkodás gépi utánzata is eredményes? Ki válaszol, amikor a diák egy forrást idéz – ha ez a forrás most a mesterséges intelligencia emergens válasza? Talán az sem véletlen, hogy ezeknél a gépi szövegeknél azt sem tudjuk biztosan, mire épül az emergencia, és olykor úgy keresik szerzőjüket, ahogyan Pirandello híres szereplői: megszólalnak – de nincs, aki vállalja őket.

## II. AZ MI ÉRTELME

Leszek Kotakowski egy helyütt meglepően egyszerű útmutatót ajánl annak eldöntésére, hogy valamely tevékenységnek van-e értelme. Szerinte ezt aszerint mérhetjük, hogy általa – akár a cselekedet, akár annak következményei révén – gazdagabbá válik-e a lét. Feledkezzünk meg egy pillanatra arról, hogy a „gazdagság” és a „lét” e kontextusban eleve bizonytalan tartalmú, és próbáljuk meg Kotakowski kritériumával mérlegre tenni a mesterséges intelligencia világban való közreműködését.

Gazdagabbá válik a lét általa? Első pillantásra – szinte gondolkodás nélkül – mondhatunk igent is, nemet is. Igent, mert megszámlálhatatlanul sok új szöveg, új válasz, új feladatmegoldás születik nap mint nap, amelyek vitathatatlanul hozzájárulnak valamihez, amit sokan a „lét gazdagodásának” neveznének. És ugyanilyen határozottsággal mondhatunk nemet is, ha arra gondolunk, hogy a (nagy nyelvi modellekre épülő) mesterséges intelligencia semmi igazán újat nem képes létrehozni – csupán variálja, újrarendezi azt, amit az emberi világtól kapott. Amit újnak nevezünk, az valójában már előtte is megvolt: benne rejtőzött a nyelvi mintákban (amelyből tanult), a logikában (amely alapján a döntései működnek, s amit szintén mi adtunk a kezébe).

Mindkét ítélet érvényesnek tűnik – és mégis elhamarkodottnak. Talán épp azért, mert a kérdés túl mély ahhoz, hogy a megszokott gondolati sémákban elférjen. S most kérdezzünk rá mégis a fogalmakra. Mi is az, hogy gazdagodás? És minek a léte gazdagodik? Vajon egy gépiesen működő rendszer gazdagságot ígérő produktivitása elegendő? Ha az MI által generált szöveg csupán visszhang, vajon mi újat tanulunk belőle? De ha ez a visszhang valamiféle szintézise az emberi tudásnak (amelyet a modell betanítására használtunk), akkor még ha ismétlődés is, mégis rá tud mutatni arra, amit nélküle csak nehezen találnánk. Ez egyben félelmet is gerjeszt, hiszen a szintézis egyben szelekció is, vagyis valamit elveszítünk, amiről nem is tudunk.

Vagy a gazdagság valahol máshol – nem a közvetlen kimenetben, hanem annak továbbépülő kapcsolataiban, a viszonyokban, a megérintettség lehetőségében kezdődik? A mesterséges intelligenciáról szóló kérdés így végső soron ontológiai kérdés: a világ valóban bővül-e általa, vagy csak önmagát hallja vissza egy új, ismeretlen tükörből?

## III. A DÍJ, AMIT NEM A GÉP KAPOTT

Amikor az AlphaFold2 nevű mesterségesintelligencia-rendszer olyan fehérjemolekulákba rejtett hajtogatásmintázatokat volt képes felismerni, amelyről korábban csak sejtéseink voltak –

mintha „érteni” kezdte volna őket –, valami megváltozott. Nemcsak a biológia tudásrendszerében, hanem abban is, ahogyan a természettudományos tudást elképzeljük. A fehérjemolekulák aminosav-sorrendje elvben meghatározza az élő szervezet viszonyai között kialakuló térbeli szerkezetüket, vagyis a molekula formáját – ami nem elsősorban esztétikai kérdés, bár sokszor avatatlan szem számára is szép formák ezek, hanem a biológiai funkció szempontjából meghatározó jelentőségű.

A térszerkezet aminosav-sorrendből kiinduló prediktív meghatározására kidolgozott hagyományos analitikus modellek meglepően sikertelenek voltak, annak ellenére, hogy jól konfirmált tudományos elméletekre épültek. A nagyfokú komplexitással magyarázott sikertelenség akkor is megmaradt, amikor a 2000-es években a modellek mögé már hatalmas számítástechnikai kapacitásokat állítottunk.

Az áttörés valamikor 2020 körül következett be, de nem egészen a várt módon, s ennek eredménye lett a 2024-es kémiai Nobel-díj, amit Demis Hassabis, John Jumper és David Baker, vagyis – ahogy lenni szokott – emberek kaptak, akik egy olyan MI-alapú rendszer kidolgozásához járultak hozzá, amely először bizonyult képesnek a fehérje-térszerkezet sikeres és széles körű előrejelzésére. Mégis, mintha egy újfajta társ jelenlétét is elismerte volna a bizottság: egy olyan „intellektust”, amely nem érez, nem akar, nem értelmez – és mégis gazdagít. A díj egyik olvasata az is lehet, hogy a Nobel-bizottság talán egy kicsit azt az új tudásformát is kitüntette, amely már nem csak az emberi intuícióra épít. (Az AlphaFold2 nem nyelvi modell, de a működésének alapelve rokon az LLM-ek filozófiájával: nagy mennyiségű, emberi megfigyeléseken alapuló, tudományosan feltárt szerkezeti adatot használ fel, ezek statisztikai mintázatait tanulja meg, és ezek alapján végez predikciót. Nem szabályalapú, hanem valószínűségi és mintázatalapú tudásreprezentációt alkalmaz.)

És ha így van – ha a tudás közös játék, ember és gép között, akkor nem válik bizonytalanná a különbség a szerző és az eszköz között? A dekonstrukció már korábban megmondta: a szerző csak áttűnő alak a nyelv önmozgásában. A világot gazdagító jelentés nem a keletkezésben, hanem az olvasásban jön létre. Így lesz a szövegnek élete – akárki írta is. De ha ez így van, akkor mit számít az, hogy ember vagy gép írta? Egyáltalán: ki ír a gépben? A jelentés – emberi vagy gépi forrásból származó szövegben – csak annyiban él, amennyiben elindít valamit a világban. De ha elindít, akkor feltehetően gazdagít is. Ezért talán nem az a kérdés, hogy MI vagy ember a szerző – hanem az, hogy mit kezdünk egy olyan világgal, amelyben a jelentés (gépi segítséggel) önmagát írja tovább.

Ez a bizonytalanság már nemcsak irodalmi metafora vagy filozófiai kérdés, hanem a szerzői jog területén is valóságos kihívás. Egyre több vita mutat rá, hogy a mesterséges intelligencia által létrehozott tartalmaknál a szerző fogalma elmosódik: nincs intenció, nincs tudat, nincs személy – mégis létezik alkotás. Ez a helyzet nemcsak technikai szabályozási problémát jelent, hanem mélyen érinti azt az évszázados eszmét, hogy a szövegnek van egy valódi, emberi eredete. Ha az alkotás már nem egy tudatos szerző aktusa, akkor kinek jár szerzői jog, és mire is vonatkozna pontosan? És még fontosabb: milyen viszonyban állunk egy olyan szöveggel, amely lehet, hogy csak a kezdeményében ered tőlünk, de magunkénak érezzük, úgy érezzük, rólunk szól? A szerzői pozíció elbizonytalanodása így nemcsak jogi, hanem mélyebb kérdés: ki mondja azt, amit hallunk, és mit jelent az, ha nincs senki, aki vállalná – s csak a gépben feldolgozott emberi tudás visszhangjaként szólal meg?

És itt további hasonló kérdések juthatnak az eszünkbe, ha az oktatásra gondolunk. Mi tanítható a megértésről, ha az értelem, amelyet a problémák megoldására használunk, nem horgonyoz le többé egyetlen tudatban? Bár mi hoztuk létre és mi tanítottuk a gépet, mégsem látjuk pontosan, mit tanult tőlünk. Lehet-e úgy tanítani, hogy nem tudjuk többé, miféle lény és hogyan oldja meg a problémánkat, ha az olyan mintázaton alapul, amelyet mi nem tudunk alapelveinkből levezetni – és mégis meg akarjuk érteni?

#### IV. A HÁLÓ ÚJRASZÖVÉSE

Ha a lét gazdagodása nemcsak a produktumok számától és fizikai terjedelmétől, hanem attól is függ, hogyan és miből képződik jelentés, akkor talán nem az a legfontosabb kérdés, hogy újat hoz-e létre a mesterséges intelligencia – hanem hogy milyen mintázatot hoz létre a már meglévőből.

Willard Van Orman Quine szerint tudásunk nem biztos alapokra épülő rendszer, hanem egy rugalmas fogalmi háló: egy hit–tudás háló, amelyet tapasztalataink és elméleteink együtt szőnek. Ebben a hálóban minden állítás másokkal függ össze: ha az egyik megmozdul, a többi is vele együtt deformálódik. A háló nem szigorúan hierarchikus, nem lineáris kapcsolatokra épülő, hanem keszekuszán relációs, a tapasztalatok világához pedig a „perifériákon” van lehorgonyozva – és mindezek együtt adják az élő gondolkodás rugalmasságát.

E gondolat mentén szemlélve az MI nem a gondolkodás utánezataként, hanem annak sajátos – tudattalan, automatizált – szövegépeként is felfogható. Talán olyan, mint az önmagán eltűnődő, önmagára rácsodálkozó nyelv. Nem új eszmék forrása, hanem egy olyan mechanizmus, amely képes a meglévő háló mintázatain új ritmust, új kötődéseket és feszültséget érzékelteni, talán még létrehozni is – még ha ennek jelentéséről maga mit sem tud. Az MI biztosan nem tudja, hogy mit jelent, amit mond. De képes összefűzni olyan kijelentéseket, amelyek a háló újraformálásához, átalakulásához vezethetnek. És ahol a háló újraszövéődik, annak környezetében új jelentésmezők keletkeznek. Ez már maga is episztemikus tett: alakítja azt, amit tudásnak nevezünk – annak ellenére, hogy önállóan nem képes ránézni a nyelven kívüli tapasztalati világra.

Szinte hallom, ahogy mondják: ha az MI újraszövi a hálót, az nem belátás, nem értelem – automatizált kapcsolatteremtés. Mint amikor valami megmozgatja a pókhálót – szándék nélkül. A pók érzékeli a mozgást, és azt mondja: ez jelenthet valamit, de hogy szél vagy légy volt-e, az nagyon nem mindegy. Egy olyan művi világban, ahol az asszociációk lehorgonyzatlan szabadsága mozgatja a hálót, az könnyen szembekerülhet az értelemmel – így annak a felelőssége a miénk maradhat, hogy mit látunk benne, szeletet vagy legyet, vagy csak azt, hogy elszakadt egy tartófonál. Ez valahogy hízelgő számunkra, ha MI-t használunk. Éppen ezért kell résen lennünk, mert a kép túl hízelgő ahhoz, hogy gyanútlanul elfogadjuk.

Ha az MI képes új mintázatokot létrehozni a jelentés hálójában, vajon nem volna-e fontos megtanítani azt, hogyan lehet észrevenni, mikor és hol mozdul meg a háló – és mit jelent, ha megmozdul?

#### V. SÚLYPONTKÉPZÉS ÉS MULTIMODALITÁS

A (nagy nyelvi modelleken alapuló) MI a betanítás során kialakított súlyok alapján generál szöveget, és nem rendelkezik intencionális súlypontképzéssel. Céltalan, nem akarja és nem is érti, amit mond. De ha mélyebbre nézünk, akkor ez korántsem nyilvánvaló. Minden token – legyen az szó, szóelem, kifejezés vagy fogalmi kapcsolat – a tanítási korpusz (és az annotátorok megerősítései) alapján számokkal leírható súlyt kap. A modell ezen súlyok alapján „dönt” arról, hogy a következő szó, mondat vagy gondolati fordulat mennyire valószínű egy megadott kontextus alapján, ami egy explicit vagy implicit kérdést/kérést tartalmaz. Ez a gépi „súlyozás” azonban nem az emberi gondolkodást jellemző hermeneutikai vagy etikai súlyozással – hiszen nincs benne jelentés, nincs tétje, nincs érzelmi töltete, nem kapcsolódik egzisztenciális döntéshez. A modell súlyoz, de nem mérlegel úgy, ahogyan mi tesszük, mikor gondolkodunk.

Hízelegjünk ismét: filozófiai értelemben a „súly” nem egyszerűen egy valós szám, hanem sokkal komplexebb: nemcsak egy kontextushoz, hanem egy belátáshoz, értékválasztáshoz, egy életformához, egy múlthoz kapcsolódik. Az emberek számára az állítások nemcsak egy-

máshoz képest „valószínűek”, hanem fontosak, irányadóak, meggyőzőek, szépek vagy csúnyák, elvetendő, tragikusak vagy üdvözítőek stb. Az MI számára minden kijelentés matematikailag mérhető, de egzisztenciális súlytalanságban lebeg. Ha romantizálhatjuk is a válaszait, azt csak mi tehetjük meg.

Tudásunk hálójának közepén – Quine szerint biztosan – mindig ott figyel az ember. (Rigorous logikai empiristák szerint épp ez a baj.) Az MI hálója viszont – legalábbis jelenleg – centrális sűrűség nélküli matematikai konstrukció, ami egzisztenciálisan nem terhelt, így nem is tekinthető igazi gondolati rendszernek. Lehet, hogy megvalósítja a sokszor vágyott független gondolkodást, de – hajlunk arra, hogy azt mondjuk – ezt anélkül teszi, hogy igazából gondolkodna.

Persze a természettudományok művelői most már igazán közbeszólhatnak: a hálóban nem csak mi vagyunk jelen, és nem csak a nyelvi rendszer – hiszen szálai a tapasztalathoz is erősen horgonyoztak. Az MI ezen a téren – legalábbis jelenleg – erősen korlátozott. Kölcsönözzünk Wittgensteintől egy képet: az MI olyan, mintha a nyelv kiszakadt volna a mindennapi, konkrét, nyelvjátékhoz és nyelven kívüli világhoz szorosan kötött használatából, s önjáróvá vált.

Mégis, ha egy MI multimodálissá válik, és érzékszerveket (pl. kamerát, kémiai vagy mechanikai érzékelőket) vagy csak egy képfeldolgozó rendszert kap, akkor az előbbi érvelés is meggyengül – talán el is szakadnak azok a fonalak, amelyek addig tartották. Hiszen ekkor a „pusztán statisztikai”, „csak szövegalapú” MI-szószöveg-gép empirikus modulát is kap, s ezzel újabb határ mosódik el. Sokkal jobban, mint elsőre gondolnánk. Mert az érzékelés nem pusztán az adatbevitel új formája. Az érzékelés világbeli jelenlétet, a géptest és a környezet közötti interakciót szül, amely a fenomenológiai hagyomány szerint is az értelem s talán a tudat felé vezethet.

Ha az MI valóban „érezkel”, és nemcsak „feldolgoz”, akkor elkezdhet reagálni a nyelv előtti világra is. Talán kialakul benne valamilyen belső orientációs tér, s ezzel együtt megjelenhet benne valamiféle egzisztenciális mező, világban való lét érzete és ezzel együtt egy súlypont is. S a háló, amit így sző majd, az embertől függetlenül kapcsolódhat a világhoz. Ezzel talán jogos lehet feltenni egy egészen új kérdést: vajon valamely tapasztalati jellegű inger milyen mértékben változtatja meg a gép belső állapotát? – s ezzel nem egyfajta gépi jelentés-előszobába lépünk-e már be? Nem mintha e küszöb átlépése bizonyossá válna, de a kérdés többé nem zárható ki filozófiai könnyedséggel. A fonalak valóban elszakadhatnak – legalábbis azok, amelyek az MI nyelvi produkciót teljes súlytalanságban lebegőként értelmezték. Talán azt is mondhatjuk: a multimodális MI már nemcsak öntudatlanul szövi a hálót, hanem érzékeli az empirikus tartófonalak megrezgéseit. És ettől kezdve a kérdés nem is annyira az, hogy mire érzékeny – hanem hogy mi következik abból, hogy érzékel.

## VI. METANYELV – 1. AZ IGAZSÁG VALÓSZÍNŰSÉGE

De vajon joggal várjuk-e az MI-től a tudásunk hálójával való megnyugtató koherenciát, ha mi magunk sem vagyunk mindig koherensek? A tudásban ott feszülnek az ellentmondások, vakfoltok, kényelmes hiedelmek, esztétikus vélekedések, érdekek által homályosított, emlékekkel erősített értékek is. Egyetlen emberi gondolkodó sem tisztán koherens – és épp ezért: emberi. A tanár sem kivétel. Csakhogy ő az, aki vállalja, hogy ennek ellenére mégis kérdez. Mert nem a koherenciát tanítja – hanem az igényét.

A mesterséges intelligenciáról szóló gondolkodásban ritkán kerül elő egy alapvető, de nem megkerülhető kérdés: mit jelent az igazság egy olyan rendszer működésében, amelynek nincs közvetlen kapcsolata a világgal? A jelenlegi nyelvi modellek nem néznek ki az ablakon, nem kóstolják meg a levest, nem érzik a virágok illatát, nem hallják meg a madarakat. A velük kapcsolatos legmeglepőbb élményünk talán épp az, hogy érzékszervi tapasztalat nélkül is képesek a valóságról értelmesnek és sokszor pontosnak tűnő állításokat alkotni.

Működésük ezáltal az igazság koherenciaelméletének gyakorlati próbája: egy kijelentés „jó”, ha illeszkedik a többihez, ha illeszkedik a nyelvi mintázatba. Az „igazság” így nem a világ és a mondat megfelelése, hanem a kijelentések belső összhangja.

Vajon meddig terjed ez a fajta koherencia típusú igazság? Tarski híres igazságelmélete szerint egy kijelentés igazságát egy metanyelvben értelmezzük (mert így talán elkerülhetők a paradoxonok az igazság körül): egy olyan magasabb rendű nyelvi rendszerben, ahol nemcsak a tárgynyelvi kijelentéseket ismerjük, hanem az igazság predikátumának alkalmazási feltételeit is. Tarski számára az igazság korrespondencia: Az „A hó fehér” tárgynyelvi mondat akkor és csak akkor igaz, ha a hó valóban fehér (mert ez a metanyelv korrespondenciát feltételező igazsága). Egy koherens metanyelv, amely tartalmazza azokat a logikai és szemantikai eszközöket, amelyek segítségével eldönthető a tárgynyelvi mondatok igazsága.

De vajon mi tekinthető a mesterséges intelligencia metanyelvének? A válasz paradox módon az, hogy nem egy klasszikus logikai rendszer szerint felépülő ideális, ellentmondásmentes nyelvi háló, hanem egy hatalmas nyelvi korpusz (amit a modell betanítására használtunk) a benne rejlő statisztikus mintázattal. Az MI ezen „metanyelv” alapján illeszt.

A világ, a valóság maga – úgy hisszük – konzisztens. Legalábbis metafizikai hitünk szerint nincs benne ellentmondás, mert ha lenne, széthullana. De amit erről mondani tudunk, már a nyelvben történik – és ott már megjelennek a paradoxonok, az ellentmondásos képek, az össze nem illő igazságok. És az MI is ebből a nyelvből tanult. Egy olyan korpuszból, amelyben benne van minden: a régi, mára már meghaladott elméletek és tudásunk ellentmondásai is. A klasszikus tudomány és a félreértett tudományos mítoszok. Minden ott van: a pontos és a pontatlan, a régi és az új – egymás mellett, bár eltérő súllyal.

Az MI által generált válasz jósága csak annyit mutat: mennyire illeszkedik a korpusz rejtett szerkezetéhez. Az MI ebben a rendszerben „ítél” meg mindent. Mintha ez a lebegő, töredezett, redundáns és ellentmondásos szövegvilág képes volna valamiféle igazságszerűséget vissz sugározni felénk. Az MI nem az igazat mondja, hanem azt, amit valószínűleg igaz módon mondani szokás.

Amikor viszont az MI multimodálissá válik, és valódi képi, téri vagy akusztikus tapasztalatokkal egészül ki, úgy tűnhet, hogy a korrespondenciaelmélet is – tőlünk függetlenül – helyet kap majd: a statisztikus közelítés kiegészül a világ és a nyelv gépi megfeleltetésének lehetőségével. De ettől valóban közelebb kerülünk az igazsághoz? Vagy csak új típusú igazságot kérünk számon rajta? S ez csökkenti majd a mi felelősségünket, amikor elfogadjuk vagy elvetjük a válaszait?

## VII. HALLUCINÁCIÓ

Az MI mögött dolgozó mesterséges neurális hálózat nem őriz emlékeket, nem készít jegyzeteket a szövegekből, nem különíti el a gondolatot a példákból, nem tudja, hogy mit felejt, és mit tart meg. A felejtést – tanulás közben – beleszámolja a tudásba. Tanulása közben belső, számunkra nehezen értelmezhető kapcsolati súlyokat igazít, hogy a válaszai ahhoz legyenek hasonlóak, ahogy a nyelvi korpusz példáiban gyakran előfordultak. Nem mondatokat tanul, hanem mondatalkotási mintákat, de nem nyelvtani, hanem inkább nyelvileg helyes tartalmi mintákat. Ez a „tudás” nem emlékezés, ahogyan mi az emlékezésre gondolunk, hanem egy mátrix, amelynek segítségével majd a használat közben kérdéseinkre a válaszokat generálja. Ekkor majd a kontextus alapján egymásra következő elemenként generálja a választ: minden lépésben kiszámolja, hogy az adott állapot alapján melyik token következése a legvalószínűbb. Ez nem szövegrészletek „kivágása”, hanem a mondatok újratermelése. Nem az a célja, hogy újra előhívja, visszaolvassa, amit korábban a tanítás során látott, hanem az, hogy a korábbi súlyokat felhasználva folytassa a gondolatokat. Az MI a válaszadás során nem gondolatokra emlékezik, hanem a korábbi mintáknak való statisztikai megfelelés a célja. És a válasz csak azért születik, mert mi a nyelven keresztül megszólítottuk.

De mi van, amikor az MI nyilvánvalóan téved? Ilyenkor azt is mondjuk: „hallucinál” – kissé pejoratív szóval minősítjük, amikor tartalmilag hamis, de jól formált kijelentést kapunk. Ha az MI tőlünk kapott „metanyelvet” alkalmaz, akkor minden kijelentése olyan, amit mi magunk már előre igazoltunk, közös tudásunk nevében? A hallucináció ezek szerint nem hibás működés, hanem egy ismeretelméleti tükör: azt látjuk benne, amit a kérdésünk a tanítási korpusz alapján válaszként megenged. Talán azért, mert a logika, amit mi adtunk a kezébe, túl bátor volt, vagyis mégsem elégséges a nyelvi világ megragadásához? Vagy a tanulási korpuszban vannak olyan elemek, amelyek hibát eredményeznek? Bizonyosan mindkettő lehetséges.

A számítógépet – így az MI-t is – hosszú ideig determinisztikus logikájú, kiszámítható rendszerként képzeltük el. Egy számológép vagy egy ipari robot „nem hibázik”, ha a hardver hibátlan, megfelelő, és jól van programozva. Az MI viszont nem klasszikus logikai elemzés alapján, szabályalapú módon, hanem statisztikai mintázatokról alkotott súlyok alapján generál szöveget – és ebben a „valószínű” nem azonos az igazzal, a hibátlannal. A mi tévedésünk, ha úgy viselkedünk vele, mintha egy adatbázisból dolgozó deduktív gép volna, miközben probabilista nyelvi rendszer. Furcsa számunkra, ha téved, mert nem úgy téved, ahogy elképzeléseink szerint a gépnek „tévedni szabadna”.

Egy ember esetében megszoktuk a tévedést. Tudjuk, hogy az ember beszéde intencionált, de sokszor nem pontos. Emberi hibákra számítunk, mert a gondolkodásba belekalkuláljuk, hogy elkövethetünk logikai hibát, ismerjük az emlékezet töredezettségét, átalakulását, az érzelmi torzításokat, a motivációkat, reményeinket és tudásunk hiányosságait is. Van értelmezési horizontunk a hibához. Az ember úgy téved, hogy rosszul gondolkodik. Az MI „téved”, mert nem gondolkodik. Mégis, ezek hasonlósága zavaró. Gyakran mi is úgy beszélünk, hogy nem tudjuk, igaz-e. Csak „valószínűnek tűnik” vagy „szépen hangzik”. Lehet, hogy az MI néha épp ebben utánoz minket, hiszen ezt is megtanulta tőlünk. A hallucináció gyakorta olyan, mint egy retorikailag ügyes, de megalapozatlan vélemény.

Hallucináció-e, ha a nyelvi korpusz ellentmondásai, következetlensége, dekoherenciája miatt téved az MI? Vagy csak akkor az, ha túl bátran, kis statisztikai valószínűség alapján következtetve téved? És a kettő, bár lehetnek egyértelmű esetek, gyakran összemosódik, ahogy a mi beszédünkben is.

## VIII. METANYELV II. – 2. AZ IGAZSÁG, AMI VALAKIÉ

Az igazság a jelenlegi nyelvi modellekben – legalábbis működésük logikája szerint csak egy predikátum a sok közül. A nyelvi predikátumok között a gépnek nincs rangsora: „piros”, „nagy”, „lehet”, „van” – és „igaz” is csak egy címke. A nagy nyelvi modellek nem tudják, hogy az igazság állítása más természetű, mint egy szín vagy mennyiség állítása, hiszen nincs számukra olyan szint, ahol ez különbséggé válhatna.

De ami a gép számára semleges, az az ember számára nem az. Mert számunkra az „igaz” szó nemcsak jelölés, sosem pusztán címke, hanem vállalás. Azzal, hogy valamit igaznak mondunk, nem pusztán besoroljuk egy logikai kategóriába, hanem ítéletet hozunk, amelyhez következményeket is társítunk. Az igazság nem egy címke a sok közül, hanem az a fogalom, amelyben összesűrűsödik minden más: a tudás, a hit, a megértés, az érintettség és a felelősség. Ez a különbség pedig nem technikai. Ami a modell számára statisztikai megfelelés, az számunkra kijelentés és vállalás. Az igazságot nemcsak mondjuk, hanem kimondjuk. És ez a kimondás mindig az emberhez tartozik. Mert az igazság nemcsak igaz, hanem valakié.

Hogyan tanítsunk igazságot, ha az többé nem a világ tapasztalatából és a klasszikus logikai szabályokból, hanem az MI mélyén rejtőző mintázatokról származik? Hogyan őrizheti meg a tanár azt a belső igényességet, amely nemcsak az információkat, hanem az igazsághoz vezető út érzékét is át akarja adni?

Talán csak úgy, ha elfogadjuk: minden felelősség az emberi ítéletre hárul. Neki kell felismer-ni, hogy a modell válasza összeegyeztethető-e azzal a tudáshálóval, amit az ember birtokol. És itt egyáltalán nem megfelelő kérdés, hogy milyen linken érhető el az az alap, amire a rendszer válasza épül. Mert a válasz nem egy forrás idézetére, hanem az egész modell működésére épül. A kérdés nem egyszerűen az, hogy a modell igazat mond-e, hanem hogy mi igaznak érezzük-e, amit mond. Ez nem pusztán hivatkozás, hanem döntés – és e döntést mindig vállalnia kell vala-kinek. Talán soha nem volt ennyire személyes és morális döntés kérdése, hogy mi az igaz.

A racionalitásnak ez a különös szövedéke Borges szavait juttatja eszembe: „»Az a legna-gyobb varázsló – írja Novalis emlékezetes módon –, aki oly mértékben el tudja kápráztatni ma-gát, hogy valós jelenségnek tartja önnön képzelgését. Nem a mi esetünk ez?« De ez, mondom én. Mi magunk álmodtuk meg a világot. Erősnek, titokzatosnak, láthatónak álmodtuk meg, tér-ben szilárdnak, időben pedig maradandónak; ám ott hagytunk az építményben néhány parányi, örök, ésszerűtlen repedést, hogy tudjuk: hamis.”

Ezekre a repedésekre talán az MI is rámutathat nekünk. De nem zárja le a tudás körét, ha-nem sejtelmesen érezteti, hogy minden igazságelmélet mögött ott áll a kérdés – ki mondja, és miért hiszi igaznak? Ha pedig az igazságot nemcsak megítéljük, hanem átéljük, akkor nemcsak a gondolkodás, hanem az élet is megváltozik. És ott, ahol a logika véget ér – talán éppen ott kezdődhet el valami más: a boldogság nyelve.

## IX. BOLDOGSÁG

Vajon boldogabbá válhatunk-e az MI által? Valójában egy ősi kérdés ez, amely túlmutat a tech-nológiai hatékonyságon: mi az a „boldogság”, amit egy mesterséges segítő eszköz adni tud? A gyorsaság, a siker, a megoldás öröme – ezek kétségtelenül jelen vannak. De elégségesek? Pótolják azt, ami elveszik vele?

Az iskola falai között már most is tapasztalható: vannak diákok, akik egy egyenletet nem szívesen oldanak meg számológéppel. Nem azért, mert tilos volna, hanem mert intellektuáli-san nem illik a megértésről rajzolt elképzelésükbe. Csalásnak érzik – nem abban az értelem-ben, hogy valamely iskolai szabályt megszegnek, hanem mert kihagynak valamit, amit a gon-dolatmenetnek végig kellene járnia. Nem ők a többség, de nem is különcök. Talán épp ez a megérzésük segíthet megérteni valamit abból, amit az MI pedagógiai használata körül még csak formálódó dilemmaként érzünk: egy sikeres válasz, ha nem jár át minket, nyújthatja-e még a gondolkodás örömét?

Mert a boldogság – legalábbis a tanulásban – nem pusztán a helyes válasz birtoklása, hanem a belső gondolati út kialakulásának csendes izgalma. Nemcsak az eredmény, hanem az intellek-tuális kaland, amit a feladattól a megoldásig átélhetünk. Az a különös érzés, amikor egy gondo-lat, amely korábban idegen volt, most belülről kerül, és az enyém lesz, nem birtoklásként, hanem megértésként. Annak átélése, hogy az általam a nagy hálózatból megértett rész működik. Ezt, hi-szen belső élmény, nem lehet lemérni, és nem is mindig tudatosul, de jelenléte vitathatatlan.

Ugyanakkor az is igaz, hogy más technológiákat – hűtőgépet, autót, tollat, számológépet, szövegszerkesztőt – jellemzően nem élünk meg csalásként. Valószínűleg azért nem, mert ezek nem próbálják meg helyettünk elvégezni a gondolkodást, esetleg néha segítenek, hogy alkal-mazzuk azt, amit már értünk. Egy számológép esetében már kevesen gondolják, hogy a négy-zetgyökvonás kézi közelítése vagy táblázatok segítségével való elvégzése „valódi” volna. A golyóstoll a töltőtollal szemben nem hamisítja meg a gondolatot, csak simábbá teszi az írást. A szövegszerkesztő helyesírás ellenőrzője gyorsabban kijavítja a hibát, de nem dönti el, mit akarunk mondani.

Valahogy érezzük, hogy a mesterséges intelligencia ezektől nagyon más, nem egyszerűen csak eszköz, hanem gondolatot teremtő „társ”. És ha társ, nem eszköz, akkor kérdés: kinek az

útját járjuk, amikor segít? A tanulás boldogsága veszélybe kerülhet, ha elveszik az érzés: mi magunk alkottuk meg a választ, mert valamit megértettünk. Nem biztos, hogy mindig fel kell tenni a kérdést. De ha már nem tesszük fel, lehet, hogy elveszítünk valamit.

## X. KI MONDJA, AMIT GONDOLUNK?

Talán nem is annyira ismeretlen számunkra az a gyanakvás, amit a mesterséges intelligencia szerző nélküli, súlytalan tudásáról gondolunk. Nem hasonlít ez ahhoz, ahogy a Wikipedia megjelenését fogadtuk? Egy enciklopédia, amelynek nincs jól azonosítható szerzője, amelyet sokszor névtelenül, közösen írnak, és mégis szócikkeket kínál a „Mi az ...?” típusú kérdéseinkre válaszképpen. Ma már magától értetődő természetességgel használjuk információforrásként – leginkább gyors tájékozódásra, olykor azonban még hivatkozásként is. Annak ellenére, hogy nem tudjuk, ki írta. A Wikipedia szócikkeinek szerzősége nemcsak rejtett, hanem feloldhatatlan is, hiszen a mondatok, szavak nincsenek megcímkézve kitől származnak – sok esetben nemhogy szavanként, hanem tokenenként is eltér, kié a végső változat. Egy-egy bejegyzést gyakran több tucatnyian írnak, szerkesztenek, módosítanak; a tartalom átvehető más nyelvi verziókból; és ha vannak is hivatkozások, a szöveg maga nem valamely szakirodalom direkt kivonata, hanem azoknak az ismereteknek az összegzése, amelyeket a szócikk írói – ki tudja, hol – valaha megtanultak.

Ez itt ma már nem számít hibának, hanem sajátosság. A Wikipedia nem a szerzői egyediség, hanem a közös emberi tudás szintézisének terepe. És mégsem tekintjük „mesterségesnek”. Nem nevezzük enciklopédia-modellnek, nem gyanakszunk rá folyamatosan, hogy „nem igazi” – pedig a szerkezete és működése sok tekintetben közel áll ahhoz, amit ma a nyelvi modelleknél látunk. Persze a fő különbség nem elhanyagolható erejű: a tartalmat nem gép, hanem hús-vér emberek alkotják (eltekintve az apróbb szerkesztéseket végző, zárt algoritmikus, szigorúan ellenőrzött botoktól).

De ha az MI válaszai is ugyanezen emberi tudásmintázatok alapján jönnek létre, csak más, gépi logika szerint – akkor nem ugyanilyen kollektív nyelvi lenyomatai ezek is az emberi tudásnak? Ha az MI is ugyanazokból a mondatokból, könyvekből, tanárok által tanított fogalmakból építkezik, csak másfajta szövegalkotási formát használ – akkor miért bízunk sokkal jobban az egyikben, mint a másikban? Lehet szerepe ebben annak is, hogy az elnevezések – modell, mesterséges, generatív – mintha negatív értékítéletet is hordoznának?

Amikor olvasunk egy Wikipedia-szócikket, persze fenntartásokkal kezeljük: csak akkor fogadjuk el, ha beleillik ismereteink hálójába, és ha hibát találunk benne, igyekszünk javítani. Ha hivatkozunk rá, azt többnyire saját álláspontunk védelmében vagy éppen annak kontrasztjaként tesszük: lám-lám, a kollektív bölcsesség így látja, s mi ezzel egyetértünk/ezzel szemben állunk. Valóságos, emberek által vállalható tudásként kezeljük.

A mesterséges, modellszerű intelligencia által generált szövegekben gyakran épp az a zavarba ejtő, hogy miközben hiányzik mögülük a szerző, mégis ismerősek. Néha ismerősebbek, mint egy Wikipedia-szócikk szövege. Olyan érzésünk lehet, mintha a mi gondolataink vetülnének vissza belőle. Hiszen kérdés/kérés formájában tőlünk indulnak a gondolatok, valami ismeretlen történik velük a gépben, majd hozzánk térnek vissza. És ha ezután még az ember módosítja, átírja, újragondolja, akkor nagyon bizonytalanná válik ez a szerzőnélküliség.

A klasszikus szerzőkép szerint a szöveg egy egyéni szándék artikulációja, még ha számtalan, látszólag autonóm elem van is az alkotásban. Amikor egy szereplő saját karakteréből adódóan, szinte önállóan írja tovább a szerepét, akkor is tudjuk, hogy valahonnan belülről jön. Egy belső világ nyelvi formába öntése. Itt viszont valami más történik: valami, aminek nincs szándéka, befolyásolja az emberi szándékot. Az intencionalitás, amely eddig belülről jött, most valahogy keveredik azzal, ami nem belülről fakad. De ha így van: az MI-t felhasználva lehet-e még autonóm szerzőnek lenni? És ha nem, akkor fel szabad-e adni ezt a pozíciót?

Ha egy tanuló jól megformált szöveget ad le, jól strukturált lépésekben old meg egy otthoni feladatot – de a gondolatot nem ő küzdötte ki, nem ő formálta meg, csak kiválasztotta, akkor tényleg ő írta? Vagy ez csak szerkesztés? A pedagógia válasza hosszú ideig az volt: nem elég jól fogalmazni – tudni kell, mit akarunk mondani, nem elég egy megoldásmenet – annak lépéseit érteni kell, mert csak ez a tudás. Ma viszont a kérdés így hangzik: kinek a gondolata, amit jól megfogalmaztál? És ha nem tud egyértelműen válaszolni, hogy kié – az csalás? Vagy új tudásforma? Milyen mértékű megértés, tudatosítás kell a csalás elkerüléséhez?

A klasszikus házi feladat másolás (plágium) fogalma itt kevés. Az MI nem idéz (vagy ha lát-szólag ezt teszi, akkor jellemzően becsap minket), hanem mintázat alapján szintetizál. Nem másol, hanem kombinál, de csak a meglévőből. Ha a tanuló ezt a szöveget elfogadja, kiegészíti, de nem formálja sajátjává – akkor kié a gondolat? És mennyire kell átformálnia, hogy újra az övé legyen?

A tanítás, amely eddig azt kérte: „fogalmazd meg a gondolataidat / te oldd meg a feladatot”, most új, korábban abszurdnak ható kérdéssel egészül ki: azonos vagy-e azzal, amit felvállaltál? A tanárnak nemcsak azt kell kérdezni: jó-e a szöveg, jó-e a feladatmegoldás, hanem azt is, hogy gondolod-e, amit írtál? Az MI által kínált mondat/megoldás sokszor pontosabb, szabatosabb, szebb – mint amit egy diák önállóan írna. De az értelmezői felelősség mikor maradhat az övé? Nem tudjuk, hol a határ. Ha tudja/érti, amit válaszolt? Ha tudja/érti, miért azt válaszolta?

Úgy vélem, ha egy mondatot nem én, hanem a gép ír le elsőként, akkor az nem a sajátom. Ha egy feladatmegoldás valamely lépését a gép „gondolta ki” elsőként, az nem a sajátom. Nem tudok úgy nézni rá, hogy az enyém. A javítási javaslatokat elfogadhatom, de az eredeti nem lehet másé. Lehet azonban, hogy ez túl szigorú álláspont, és egy gondolat talán akkor is a sajátunk lehet, ha tudatosan, értő módon vállaljuk. Talán ennél nem lehetünk megengedőbbek. Mert ha nem vállaljuk – hiába szép a mondat/megoldásmenet, félelmetesen üres marad. A tanítás talán ma már azt is kérdezheti: ki mondja, amit gondolunk? Vagy azt, ki gondolja azt, amit mondunk?

## XI. REMÉNYEINK, FÉLELMEINK 1. – TUDNI, DE MIÉRT?

Az oktatásban régóta él az a tétel, hogy nem a tárgyi tudás, hanem a gondolkodás tanítása a cél. Csakhogy a gondolkodás nem a semmiből indul. Még egy egyszerű, a tapasztalathoz közel álló kémiai kísérlet reflexiója is valamilyen fogalmi térben, szókincsben, sémákban mozog. Még akkor is, ha csak az egyszerűen megfigyelhető tapasztalatokat akarjuk kémiailag leírni. És ezek – bármilyen fájdalmas kimondani – elengedhetetlen alapismeretek. De mi értelme ma a tárgyi tudásnak, ha bármikor lekérdezhető? Mi az a minimális – de mégis elengedhetetlen – tudásmennyiség, amely nélkül nincs értelmező gondolkodás?

Egy esszé értő olvasásához is ismernünk kell valamennyire a műfajt és a témát. Egy fizikai törvény megértéséhez értenünk kell azokat a fogalmakat, amelyek között összefüggést ad meg a törvény. Egy kémiai képlet megértéséhez tudnunk kell, mit jelentenek a benne szereplő jelek. Egy történelmi dokumentum olvasásához ismernünk kell a korszakot, a szókincset, a kontextust. A megértés nem működik belső fogalmi háló nélkül – és ha ez nincs, akkor nincs, ami megtartsa az ismeretet.

A gond, amivel ma nem tudunk mit kezdeni, talán épp az, hogy ez a belső keret külsővé vált. Ott van a zsebünkben. A válaszokban, amelyeket az MI ad. De ha az ismeretek hálója már csak a gépben van, akkor elveszítjük azt a képességet, hogy meg tudjuk fontolni a választ, amit a géptől kapunk.

Egy felelet közben elhangzó apró súgás vagy egy-egy oldalra pillantás nem teszi értelmetlenné a tudásmérés próbáit, mert az így megszerzett információ csak akkor válik használhatóvá, ha illeszkedik a diák fejében lévő fogalmi hálóba. Ha megtalálja benne a helyét. Különben

csak elefánt a porcelánboltban – túl nagy, túl idegen, túl hirtelen. És ha minden lépéshez az MI súgása kell, akkor csak az elefánt marad. Árván.

Egyre nehezebb kérdés, hogy mit kell tudni ahhoz, hogy valamit érteni kezdjünk. Mit kezdünk azzal, hogy a tárgyi tudás előfeltételei, amelyek lehetővé teszik egy szakterülethez való értő hozzáférést, lassan elmosódnak, vagy sokak szerint el is tűnnek? Ha minden kérdésre adható egy pillanat alatt válasz, miért lenne még fontos alapismeretet szerezni?

Vagyis az MI miközben a tartalmat kézbe adja, a tanulás célját és motivációját is átrendezheti. Sokan félnek tőle: nem egyszerűen átrendezi, hanem megszünteti. De vajon tanulható-e értő olvasás úgy, hogy a mögöttes tudás nem belül épül, hanem csak kérdésről kérdésre görgetve, kívülről érkezik? Egyre könnyebb elhinni, hogy ha tudjuk, hol van a válasz, nem kell tudni magát a tartalmat. Pedig a válasz még nem a megértés. A válasz csak egy jól formált hely a hálóban. A megértés a háló maga.

És talán épp az MI mutathatná meg, mennyire fontos ez a belső fogalmi tér – amivel ő már rendelkezik, de amit nem tud átadni nekünk. Mert az ő tere teljesen más, mint a miénk. A mi hálónk bennünk csak tanulás által épülhet. És csak akkor tudunk dönteni, felelősséget vállalni, ha felépül. Csak így ismerjük fel, ha hibás, amit kaptunk. Ha hiányzik még valami a válaszból. Így értjük meg nemcsak azt, hogy mi a válasz – hanem azt is, hogy miért az a válasz. Az MI csökkentheti a tárgyi tudás megtanulásának vágyát, miközben paradox módon mégis növeli annak jelentőségét – hiszen csak az tudja helyesen értelmezni a gép választát, aki már rendelkezik valamennyivel a fogalmi térből.

Tudjuk jól, a tanár ma már nem lehet kizárólagos tudásforrás. De talán annál inkább megmutathatja, hogy ha rendelkezünk egy olyan fogalmi térképpel, amely mutatja a magasságokat (mi a fontos), az irányokat (merre érdemes elindulni), a határokat (meddig tart a biztos tudás), ...és azt is, ahol már *hic sunt leones* – ahol nem kérdezni kell, hanem gondolkodni. Legalábbis ezt reméljük.

## XII. KÉRDEZNI TUDNI

A mesterséges intelligenciát gyakran éri az a kritika, hogy nem tudja, amit mond – és mégis mondja. Nem tudja – mondják –, mert nincs teste, amely tudhatna, és mert nem bizonyosság, hanem pusztán valószínűség alapján válaszol. És mint láttuk, ez gyakorta még hallucinációhoz is vezethet.

De vajon nem hasonló természetű-e a mi emberi tudásunk is? Amit igaznak nevezünk – még a természettudományban is –, nem épp ugyanígy csak valószínűségek mentén igaz? Nem axiomatikus alapon, fenntartás nélkül elfogadott evidenciákból következőleg, hanem tapasztalatilag megerősítve. Quine óta, vagy talán régebben is tudjuk: nem az egyes állítások igazsága számít, hanem az egész háló konzisztenciája és illeszkedése a nyelven túli világhoz. Az igazság csak immanens, nincs külső nézőpontunk, csak a háló belsejéből figyelhetünk, mint a pók. Tudjuk jól a tudományfilozófusoktól: épp ezért alakulnak újra, időről időre a bizonyosságunk hitt világgépeink is.

Talán a matematika kínálja az utolsó menedéket – ott a bizonyítás még axiómákból indul, és a következtetés logikailag kényszerített. Mégis: ahogy bonyolódik egy matematikai állítás, úgy nő a lehetséges bizonyítás útvonalainak száma is – és az, hogy ezekben egy-egy helyettesítés jogos-e, hogy valóban „analitikus”-e, gyakran már nehezen eldönthető formálisan. A bonyolultabb bizonyítások lépései gyakran több száz oldalra nyúlnak, és olykor pusztán szakmai közmegegyezés alapján fogadjuk el, hogy az egész rendben van. (Nem állom meg, hogy itt híres példákra ne hivatkozzak, még akkor sem, ha sajnos nagyon messze állok attól, hogy értem őket. Andrew Wiles nagy Fermat-sejtésének vagy Grigorij Perelman Poincaré-sejtésének bizonyításaira gondolok.) Így az analitikus bizonyosság is visszahúzódik – nem a tévedés lehetőségé, hanem a megértés határai miatt.

Azt is tudjuk, hogy a tudományos állítások többsége nem is a szigorú logika, hanem a megfigyelések intuitív általánosítása, vagy – ma már szigorúbb feltételek között – a mérések és statisztikai következtetések eredménye. A formális logikai, kényszerítő erejű következetesség elsősorban a fogalmi rendszerek belső világában elvárt – a világgal érintkező periférián azonban mintha már nem találná a helyét. Az igazság így nem pontos korrespondencia, hanem illeszkedés – az empiria illeszkedése egy modellhez, egy elméleti háléhoz, amelynek szálait nemcsak a logika tartja. Így születik meg az a „valószínű igazság”, amelyet tudásként tanítunk – s amely, a tudománytörténet tanúsága szerint, időről időre átalakul: olykor gyökeresen is, egy-egy paradigmaváltáskor.

De mit jelent ez a tanításban? A tanár hagyományos szerepében az igazság közvetítője. Aki szinte mindig tudja, mi a helyes válasz – sőt azt is, hogy miért éppen az. Ma azonban nem a válasz hiányzik, hanem a megalapozottság kérdőjeleződik meg. A ChatGPT alján is ott a figyelmeztetés: „A ChatGPT hibázhat. Fontold meg a fontos információk ellenőrzését.” Mint amikor egy vegyület előállítási receptjében azt mondja: minden hozzávaló teljesen vízmentes, aztán a felsorolásban, a szükséges anyagok között, az egyik sorban ott van a víz is. Ez persze könnyen észrevehető, de a hiba formája jellemző: ha minden vízmentes, akkor logikusan a víz is az. Tehát nem a víz van egyszerűen az anyagoknál, hanem a „vízmentes víz”. Ami talán meglepő, hogy amikor rákérdezek, azt mondja, igen, a vízmentes víz létezik, és azt jelenti, hogy nagyon tiszta víz. És már (ismerősen) magyarázza is a magyarázhatatlant: „A »vízmentes víz« (angolul *water-free water*) természetesen paradoxonnak tűnik – valójában nem vízmentes a szó szoros értelmében, hanem olyan extrém módon tiszta vízről van szó, amelyben semmilyen oldott szennyező (pl. ion, szerves anyag, más molekula) nincs jelen, és a mennyisége is szigorúan szabályozott a sztöchiometrikus reakcióhoz.”

A tanár viszont épp az a személy, aki nemcsak azt tudja megmondani, hogy mi stimmel és mi nem, hanem azt is, hogy miért – és miért fontos ez, hogy felismerjük. Hogy az adott válaszok a tudásháló mely elemeihez kapcsolódnak, és hol, milyen szál feszül meg, mozdul el. A tanítás tehát nem a bizonyosság újratermelése, hanem a meglévő válasz mellett is a kérdés életben tartása lesz. A tanár nem az igazság szószólója – hanem annak felismerője és értelmezője. Egy olyan helyzetben, ahol a tudás csupán statisztikailag valószínű, ontológiailag pedig üres, a tanítás könnyen válhat puszta színjátékká. Nem azért, mert hazug, hanem mert szerepeket, formákat, meggyőződést és megértést játszik el – anélkül, hogy valódi megértésben gyökerező, érzelmi, intencionális gyökere lenne. A tanár ebben a térben megtörheti a játékot – és vállalhatja, hogy a tudás nemcsak statisztikai illeszkedés, hanem valamilyen belső, emberi rend lenyomata is.

Talán az MI alkalmazása mellett épp a kérdés maradhat meg fontos tudásformaként – sőt: tudást teremtő formaként. Lehet úgy kérdezni, hogy abban benne legyen a megértés igénye, a nem tudás méltósága, a keresés iránya, az elidőzés lehetősége és a kétely joga. Talán ez egy olyan tanári pozíció, amely az MI társaságában is hiteles tud maradni.

Lehet, hogy ez mindig is a tanár egyik szerepe volt, csak korábban kevesebb hangsúlyt kapott? A kérdésekkel való óvatos, az együttgondolkodás iránti tiszteletet megtartó zavarba ejtés. Feyerabend kedvelt szórakozásaként írja le, hogy racionalistákat hoz zavarba meghökkenítő álláspontok megalapozásával és a hozzá tartozó kérdésekkel – de ez nem relativista cinizmus, hanem pedagógiai eljárás. Mert a tudás ott kezd élni, ahol a rend meginog. A tanár tehát nemcsak a rend őre, hanem annak megkérdőjelezője is. A félelmetes kognitív disszonancia – amitől ma az iskola oly gyakran óvni szeretné a tanulót – valójában a gondolkodás valódi terét nyithatja meg. Ahol nincs feszültség, csak egyszerű kérdés-válasz, ott nincs gondolkodás sem. A jól irányított kérdés képes megbontani a megszokott tudás szövetét és megnyitni egy új irányt – olyat, amelyben nem a tananyag válik izgalmassá, hanem a gondolkodás maga. Szókratész – minden tanár ősképe – nem tanított semmit, csak kérdezett. De úgy kérdezett, hogy aki válaszolt, kénytelen volt magát is újra megkérdézni. Az igazi kérdés nem feleletet kíván, hanem elidőzést – ezt a gesztust

talán nem lehet kiszervezni az MI-be. Mert itt a tanár nemcsak beszél, hanem vár is. És ebben a várásban rejlik az intellektuális felelősség és talán a remény is.

A testetlen mesterséges intelligencia nem rendelkezik életformával – de mintha imitálni tudná, mintha emlékezne rá. Nem saját emléke ez, hanem sok emberi szöveg lenyomata, amelyeken keresztül egyfajta kollektív élettapasztalat tükröződik vissza. Mégsem az. Amit mond, nem egy ember mondja. Nem tartozik hozzá sem test, sem sors, sem történet. Nem emlékezik arra, hogy a beszélgetésben korábban már kérdeztük. Az előző kérdések és válaszok csak kontextusként és nem beszélgetésként vannak jelen. Nincs mögötte saját értékrend, csak stílusminták, nincs világképe, csak világképek statisztikailag súlyozott átlaga. És mégis – amit mond, sokszor megszólít. Ismerős. Mintha lenne benne valami emberi, hiszen annak utánzata.

Amikor válaszai nem hamisak (s általában nem azok) – akkor mintha mindannyiunk közös tapasztalatának átlaga szólna bennük. Csakhogy a valóság, a döntés, az értelem sosem átlagos. Mindig konkrét, mindig valakinek a világa. Ezért is kell, hogy legyen, aki lehet, hogy nem olyan jól fogalmaz – de vállalja a kérdést és a választ. De aki nemcsak kérdez, hanem úgy kérdez, hogy annak súlya legyen. Mert tartozik hozzá életforma – és arc. Talán ezért van még helye a tanári szónak. Nem azért, mert többet tud – hanem mert tudja, hol áll, és ebből a pozícióból tud kérdezni. És talán ez az, amit a gép sosem tudhat emberi módon.

### XIII. REMÉNYEINK, FÉLELMEINK 2. – A JÖVŐ

„A világ változik, mert ez a természete” – írja Hamvas Béla. S talán valóban igaza van abban is, ahogy folytatja: „Felesleges olyan eszeveszetten félni attól, ami jön.” Mégis, ha a jövőre gondolunk, vannak félelmeink, mert a remények mellett látjuk azokat a tendenciákat is, amiket nehezen tudunk pozitívan értékelni.

Ha valaki szinte biztosan tévedni szeretne, elegendő annyit tennie, hogy a jövőről nyilatkozik. Ebben a műfajban ugyanis a kudarc statisztikailag erősen valószínűsíthető: a jóslatok ritkán igazolódnak, gyakrabban válnak múltó (vagy mulatságos) kuriózumokká, amik azt mutatják, mennyire reménytelen az extrapoláció. A mesterséges intelligenciával kapcsolatban is újra és újra elhangzanak ígéretes és aggodalmak – mégis érdemes erre figyelni, mert ezek talán többet árulnak el rólunk és a jelenről, mint az MI-ről és a jövőről.

Így a mesterséges intelligencia körüli beszélgetések inkább önmegszólítások, s nem technológiai diskurzusok. Arról beszélünk, amit most nem tudunk megérteni, aminek következményeitől ma féltjük a jövőt – valószínűleg félreértve a legtöbb jövőbe mutató nyilat. Mégis elkerülhetetlennek látszik a gondolkodás arról, mi fog történni, és az, hogy mire leszünk képesek, ha megtörténik. Ezért az MI-ről való beszéd sosem pusztán technikai kérdés: mindig beleszövídik az emberi vágy, a megértés iránti sóvárgás és a saját gondolataink tükrében való gyönyörködés is. A kérdés talán nem is az, mit tud az MI – hanem hogy mit szeretnénk, ha tudna, és mit szeretnénk, ha sohasem tudna. És hogy mit árulnak el rólunk ezek a kívánságok.

A tudományban például a MI rendszerek olyan komplexitáshoz férhetnek hozzá, és mutathatnak meg számunkra, amelyekre emberi elme aligha lenne önerejéből képes: a mintázatok felismerése, végtelen adatmezők strukturálása, vagy a kémiai-biológiai rendszerek újraértelmezése (mint az AlphaFold esetében), és ezek újfajta reményeket keltenek. De felmerül a kérdés: vajon az MI által felfedezett világ még mindig a miénk lesz-e? Ha a gép látja meg előbb, az elmélet még emberi? Megértjük-e, amit a gép mutat? Sikerül-e egy felfedezett statisztikus mintázat mögött az analitikus formát megtalálni, hiszen az az emberi agy számára az előbbre való?

A művészetben az MI képes stílusokat utánozni, formát találni, ritmust generálni – néha ijesztő pontossággal, s talán nem is teljesen gépiesen (hiszen a statisztikai súlyok lágyasága és egy kis véletlen hozzáadása erre is képessé teheti). Az alkotás aktusa így új kérdésekkel telítődik: kell-e szándék az értékhez? És elég-e a szépség a műhöz, ha nincs mögötte érzés, vágy, remény? Félel-

meink itt is az eredet, a hitelesség, az emberi jelenlét elmosódása köré szerveződnek – miközben reménykedünk abban, hogy az új gépi társak visszhangot adnak olyan gondolatoknak is, amelyeket magunktól nem tudtunk volna kimondani, de amelyekben magunkra ismerünk.

Az MI gyorsan és a mi áttekintésünket segítettően képes gondolatokat szintetizálni. Ez biztosan segíti a tájékozódást az információözönben. De ez a szintézis mégsem emberi. Így vannak sokan, akik attól félnek, hogy az MI beszéde megfertőzi majd a nyelvet, hogy visszavonhatatlanul és talán felismerhetetlenül beférkőzik az igazi, az emberi nyelvbe, az emberi gondolatok közé. És ezzel még a hallucinációk is áttekinthetetlenül összekeverednek majd azokkal, amiket tényként állítunk. Különösen akkor, ha majd az új MI modellek létrehozása során ezek a szövegek már a betanítási korpusz részei lesznek. Nyomasztó felelősség (kellene, hogy legyen) ez azok részéről, akik az MI szövegeit használják bármire.

Az oktatásban biztosan segít, mert összefoglal, tematizál, ha a tanulás/tanítás közben meg kell válaszolni egy kérdést, akkor egy gyors lehetőséget jelenthet. De tanárként félünk, hogy átveszi a szerepünket, mert türelmesen magyaráz, látszólag tudja a válaszainkat, a diák nem fél tőle kérdezni, hiszen nem von le következtetést a kérdésben rejlő nem tudásból, megértési nehézségekből (mint ahogy a tanuló társakról vagy a tanárról gondoljuk). Még az is lehet, hogy megtanul úgy kérdezni, ahogy a tanárok, hogy közben számtalan egyéni úton képes rávezetni a tanulót a válaszra, s így a megértésben is társ lehet. De erős kételyeink is vannak, hogy a gép, amely sohasem élte át a megértés utáni sóvárgást vagy a nem tudás fájdalmát, hogyan tudja ezt hitelesen utánózni. Hogyan tudja úgy felépíteni a segítséget, hogy az egy ember számára katartikus legyen? Erre való kísérletet azonban több esetben már ma is láthatunk (pl. Gemini Learning Coach, ChatGPT Study Mode). Mégis, el sem tudjuk képzelni, milyen lehet egy olyan világ, ahol már nem a gép tanul az embertől, hanem az ember tanul majd a géptől.

A vallásban, etikában, jogban, pszichológiában és az interperszonális térben még finomabb szövésűek ezek a jövőre irányuló ellentmondások. Ezekben a kérdésekben talán még tévesen sem tudunk előre látni. Egy jogi érvelés MI-alapú segítése lehet hatékonyság – de lehet beavatkozás is. Egy terápia során adott MI-tanács lehet találó – de lehet dehumanizáló. Ott van a lehetőség: a figyelem, a válaszadás, a struktúra professzionálisabb újraosztása. És ott a félelem: hogy közben elveszünk belőle mi magunk.

Jelenleg a nagy nyelvi modellek az átlag által meghatározott normalitást követik. Normális az, ami a tanítási korpuszban a leggyakoribb környezetében van. De etikailag megengedhető, hogy ne alkalmazzunk előíró normákat is a működés befolyásolására (ahogy ma is teszszük)? De ha alkalmazzunk, ez nem vezet torzításhoz a tudásban, nem épp az átlagban rejlő pártatlanságot veszélyezteti?

A régi filozófiai dilemma ma napi szinten merül fel kérdésként: ha egy gép összetett kérdésekre adott válaszai nem különböztethetők meg egy ember válaszaitól, akkor a gép gondolkodik, vagy csak utánozza a gondolkodást? Ha gondolkodik, akkor emberként (emberszerűen) gondolkodik, vagy csak gondolkodásával utánozza az embert? Ha úgy beszél, mintha személyisége lenne, akkor van személyisége, vagy csak utánozza a személyiséget? Ha úgy beszél, mintha érzelmei lennének, akkor vannak érzelmei, vagy csak utánozza ezeket? Ha bármelyik ehhez hasonló kérdésre igennel válaszolunk, akkor jogai is lehetnek ennek a gépnek?

Végző soron minden kérdés oda vezet: ha az MI nemcsak gondolkodást segítő eszköz, hanem gondolkodó társ, akkor mi lesz a mi szerepünk ebben a kapcsolatban? Nem a kontroll elvesztése a fő félelem – hanem az, ha ennek bekövetkezését nem vesszük észre, vagy ha észre is vesszük, nem akarjuk többé visszavenni. Az, ha a segítőből lassan mintázataalkotó irányító válik. De talán épp itt nyílik meg egy másik, csenedesebb remény, hogy talán mi magunk tanulunk valamit abból, ahogy az MI mutat minket. A Solaris zenekar 2024-es lemezén, a *Mi vagy MI* című album egyik darabjában (*Az utolsó vers* a címe) egy különös gépies (emberi?) hang szólal meg: „...bár a következő lépéseket együtt fogjuk majd megtenni, korántsem biztos, hogy együtt is érkezünk majd meg. Csak rajtatok múlik, hogy lesz-e a jövőnek emberi arca.”

Ez a mondat – legyen akár gépi, akár emberi – visszhangzik valahonnan mélyről. A *Solaris*, Stanislaw Lem feledhetetlen regénye egy különös, az ember számára felfoghatatlan intelligenciát, egy bolygót borító óceánt idéz, amely nem kommunikál, hanem emberi gondolatokat olvas és visszatükröz. Miközben van egy, az ember számára nem megfogható kozmikus intencionalitása is. Tegyük most fel újra néhány régi kérdésünket, új megvilágításban: biztosan megértjük az MI-t, amit pedig mi hoztunk létre? És hajlandók vagyunk visszanézni benne magunkat? Vagy utasítsuk el a lehetőséget: torz tükörbe talán nem is érdemes belenézni? De vajon van más tükör, amibe belenézhetnénk?

#### XIV. TITKOS SZAVAK

Réginek tűnő igazság, hogy talán nem is a mesterséges intelligenciát kell megértenünk, hanem magunkat. Milyen jelentést sugároz vissza egy gép, ha nem a benne lévő szándékból, hanem a mi kérdéseinkből és az általunk teremtett ismétlődő mintázatokból születik benne a szó? Visszhang-e ez egyáltalán? Talán. Talán csak egy új hang – és a kérdés, hogy mit kezdünk majd vele.

De nem épp az történik most, hogy a nagy nyelvi modellek alapján működő hideg, statisztikai gépezetek kirajzolják azt a gondolati rendet, amelyet mindig is sejtettünk a nyelv mögött, azt, amely rendezi számunkra a világot, de eddig csak töredezetten, részleteiben láttunk? Nem lehet, hogy a gondolat rendje, amelyet Wittgenstein fájdalmasan nélkülözött és meglelni remélt – legalábbis a *Tractatus* idején –, mégis ott bujkál a nyelv szövetében? Ha Wittgenstein most visszanézne ezekre a modellekre, vajon mit válaszolna? Egy újabb hallgatást?

A nagy nyelvi modellek mögé nézve úgy tűnik, mintha a gondolat átfogó struktúrája rajzolódna ki a statisztikai mintázatba foglalható, de számunkra nehezen érthető rendben. Vagy inkább az történik, hogy megint összetévesztjük a térképet a tájjal? Ha most kirajzolódik előttünk valami, amit rendnek nevezünk – vajon nem csak a tekintet szokása az, ami rendet lát? Keresünk a rendet – most, hogy találtunk egyet, nem épp ez az, ami elvakít? És ha a gép találja meg a gondolataink mintázatát – biztos, hogy az a miénk lehet? „Amikor szám és alak / nem lesz a lények kulcsa már...” (*Novalis*) – akkor talán ismét titkos szavakat keresünk majd.

