

Glavanits Judit

FOGYASZTÓBARÁT MESTERSÉGES INTELLIGENCIA

- a velünk élő AI egyes aktuális kérdései

FOGYASZTÓBARÁT MESTERSÉGES INTELLIGENCIA

- A VELÜNK ÉLŐ AI EGYES AKTUÁLIS KÉRDÉSEI

SZERKESZTETTE:
GLAVANITS JUDIT

GYŐR, 2023.

© Glavanits Judit, G. Karácsony Gergely, Kecskés Gábor,
Solymár Károly Balázs, Szabó Tamás

Szerkesztő:
Glavanits Judit

Szakmai lektor:
Kelemen Roland

Kézirat lezárva: 2023. augusztus

ISBN 978-615-6491-37-4

Kiadó: UNIVERSITAS-Győr Nonprofit Kft.

TARTALOMJEGYZÉK

- 7 A szerkesztő előszava

- 9 SOLYMÁR KÁROLY BALÁZS: A fogalomhasználat egyes kérdései a mesterséges intelligenciáról szóló európai rendelettervezetben

- 29 G. KARÁCSONY GERGELY: Ki védi meg a virtuális személyeket?
– a kiterjesztett ember alapjogi védelme a metaverzumban és azon túl

- 45 KECSKÉS GÁBOR: A mesterséges intelligencia az ENSZ Fenntartható Fejlődési Céljai szolgálatában

- 63 SZABÓ TAMÁS: A pénzügyi szférában alkalmazott döntéstámogató rendszerek átlátható és fogyasztóbarát felhasználása

- 85 GLAVANITS JUDIT: Tudatos és érzelmes mesterséges intelligencia

A szerkesztő előszava

Az Igazságügyi Minisztérium jogászképzés színvonalát emelő programjai keretében a Széchenyi István Egyetem Állam- és Jogtudományi Karának Nemzetközi- és Európai Jogi Tanszéke számos kutatáson és kutatási eredmény publikálásán van túl, jelen kötet mégis egyfajta „szíve csücske” a szerkesztőnek (reméljük a többi könyv nem sértődik meg ezen a súlyos diszkrimináción).

A 2022-ben elnyert pályázat témája: „Fenntartható és fogyasztóbarát mesterséges intelligencia”, amely cím tükrözi a kutatók hozzáállását az új technológiához. Hogyan lehet a felhasználók, a fogyasztók számára olyan szabályozási környezetet teremteni, amelyben nem csak az innovációnak nem szabunk indokolatlan határt, de a fejlesztések az emberiség hosszú távú jövőjét helyezik előtérbe. A jelen kötet szerzőiből álló kutatócsoport a Széchenyi István Egyetemen működő Tech-Augmented Legal Environment (TALE) kutatócsoport keretein belül keresték a választ a fenti kérdésre, ki-ki a saját érdeklődési köre szűkebb alkérdéseire fókuszálva.

A címben található „aktuális” jelző utal arra, hogy aki ma mesterséges intelligenciával (továbbiakban: AI, *artificial intelligence*) vagy annak jogi szabályozásával foglalkozik, a szokásosnál is gyorsabb avulásra számíthat kutatási eredményei tekintetében. A technológia és az AI-alapú fejlesztések exponenciális fejlődésével a jogalkotó aligha tud lépést tartani, ugyanakkor a felhasználók technológiába vetett bizalma növekedni látszik, főleg a fiatalabb generáció látszik igen bátran alkalmazni és mindennapi életébe beépíteni akár a néhány hónapos technológiai újdonságokat is.

A kutatócsoport 2023. november 17-én rendezett tudományos konferenciája a chatGPT felsőoktatási és egyéb gyakorlati felhasználásáról szólt, ahol a résztvevők nagy száma és végzettségük-érdeklődésük sokfélesége egyértelműen jelezte, hogy az AI-alapú technológiák esetében a multidiszciplináris megközelítés elengedhetetlen a téma jobb megértéséhez. A jövőben tehát a kutatók számára nem csak az jelenthet kihívást, hogy a meglévő fogalmi eszköztárba hogyan illesszék be az új és új fejlesztéseket, hanem hogy más tudományterületek szereplőivel is szót értsenek, sőt, közösen gondolkodjanak egy-egy innováció jogi, gazdasági, társadalmi aspektusain – a műszaki megoldásokon jóval túlmutatóan.

A kötet szerzői jogászok, de a megközelítésük túlmutat a jogszabály betűjén, az általuk felvetett gyakorlati kérdések pedig más tudományterületek számára is tartogatnak kérdéseket: a virtuális valóságban épített énkör, a fejlesztők fejében lévő elképzelések és a jogi fogalmak egymásra illesztésének nehézségei,

a fenntarthatósági szempontok érvényesítése már a tervezés szakaszában, valamint a marketing-szempontok és automatizáció határainak tisztázása a feladat, hogy majd ha felmerül bennünk: vajon gondolkodnak-e, éreznek-e helyettünk a gépek, az elméleti és gyakorlati szakemberek közös választ tudjanak adni, bármi is legyen a kérdés, ami az AI alkalmazásával kapcsolatban felmerül.

A kötet szerzői és a szerkesztő őszintén remélik, hogy az alábbi 5 tanulmány gondolatokat ébreszt az AI fogyasztóbarát és fenntartható felhasználásával kapcsolatban. Bár a győri jogi karon folyó kutatásunkból most csak egyfajta pillanatképet villantunk a Kedves Olvasó számára, megígérjük, hogy érdeklődésünk a téma iránt nem fog alább hagyni a közeljövőben, és amint új, tudományos vizsgálatra érdemes kérdések merülnek fel bennünk, nem félünk majd újra billentyűzetet ragadni...

Győr, 2023. november 20.

Dr. habil. Glavanits Judit
kutatásvezető

SOLYMÁR KÁROLY BALÁZS

A FOGALOMHASZNÁLAT EGYES
KÉRDÉSEI A MESTERSÉGES
INTELLIGENCIÁRÓL SZÓLÓ
EURÓPAI RENDELETTERVEZETBEN

A fogalomhasználat egyes kérdései a mesterséges intelligenciáról szóló európai rendelettervezetben

I. ÁLLÍTSUK MEG AZ INNOVÁCIÓT, AMÍG KITALÁLJUK A SZABÁLYOZÁST?

2023. március 22-én vezető technológiai vállalatok tulajdonosai, a Mesterséges Intelligencia területén korszakalkotó eredményeket elért kutatók, valamint társadalomtudósok nyílt levélben^[1] követelték a GPT-4-nél fejlettebb MI rendszerek tanításának felfüggesztését legalább 6 hónapra. Nem általában az MI fejlesztések leállítását követelték tehát, csak az „egyre nagyobb, kiszámíthatatlan, fekete doboz jellegű modellekét egyre növekvő képességekkel”. A nyílt levél nem használja egyik, az utóbbi időben bevett kategória megjelölést sem: sem az *általános célú mesterséges intelligencia (general purpose artificial intelligence)*, sem az *alapmodell (foundation model)* kifejezéseket. Használja ehelyett az *advanced AI* kifejezést, valamint körülírja a jelenséget, olyan módon jellemezve, hogy annak működését „senki, még az alkotóik sem tudják megérteni, előrejelezni vagy megbízhatóan kontrollálni”. A dokumentum ugyancsak hivatkozik az *általános mesterséges intelligencia (artificial general intelligence, AGI)* fogalmára, amely *nem* egy már létező technológia, hanem egy olyan cél, amely felé a jelenleg sokat tárgyalt és vitatott kutatások haladnak. A ChatGPT-t fejlesztő OpenAI az AGI rendszereket úgy határozza meg, mint „*nagymértékben autonóm rendszereket, amelyek túlteljesítik az embert a legtöbb gazdasági értékkel bíró munkában.*”^[2]

2023. március 31-én az olaszországi adatvédelmi hatóság adatvédelmi okokra hivatkozva azonnali hatállyal betiltotta a ChatGPT használatát, és felhívta az OpenAI-t s a GDPR szabályainak megfelelő működés bemutatására. Az aggályok tisztázását követően április 28-án aztán újra elérhetővé vált a rendszer^[3]. A hírben nem a betiltást emelném ki, hanem a gyors újbóli elérhetőséget. Húzhatta volna a hatóság a folyamatot, újabb és újabb hiánypótlást kérve, de nem tette. Az eset jól szemlélteti, hogy ez a technológia egyszerre kelt félelmet, vet fel

[1] Forrás: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (letöltés dátuma: 2023. május 12.)

[2] Forrás: <https://openai.com/about> (letöltés dátuma: 2023. május 12.)

[3] Forrás: <https://apnews.com/article/chatgpt-openai-data-privacy-italy-c0764f0e77273c429564fdb0c475a707> (letöltés dátuma: 2023. május 12.)

aggályokat, kockázatokat, és egyszerre vonzó, izgalmas, nélkülözhetetlen. Bár a szolgáltatással való játék, képességeinek megtapasztalása önmagában is nagyon jó szórakozás, jóval többről van azonban szó: az alapmodellekre építve az ún. downstream fejlesztők üzleti vállalkozásokat indítanak, vagy alakítanak át meglévőket, az alapmodell betiltása vagy felfüggesztése azonnali gazdasági érdeksérelmet okoz.

Mi készíthette erre az állásfoglalásra azokat, akik közül sokan a Mesterséges Intelligencia kutatások legfontosabb alapjait tették le, vagy az OpenAI-t megalapító Elon Muskot? Miért került nem csupán a szakmai közönség, hanem az átlagemberek érdeklődésének középpontjába az MI 2022 decemberében? Hogyan lett hirtelen óriási tétje az Európai Bizottság által előterjesztett, a mesterséges intelligencia szabályozására szánt rendelettervezetnek (a továbbiakban: Rendelet)?

II. INNOVÁCIÓMENEDZSMENT ÉS JOGVÉDELEM

A mesterséges intelligencia felhasználásának előnyei rendkívül széleskörűek és egyértelműek. Jelentős gazdasági és társadalmi előnnyel kecsegtet számos ágazatban, beleértve az energiapiacot, a mezőgazdaságot, a környezetvédelmet, az egészségügyet, a pénzügyi piacot és a közügyeket (Lee, 2022). Különösen hasznos lehet az előrejelzések javításában, a műveletek és az erőforrások elosztásának optimalizálásában, valamint a szolgáltatások személyre szabásában (Madiaga, Chahri, 2022).

Nem szabad ugyanakkor, hogy a napjainkban tapasztalt és a jövőben várható gazdasági előnyök elhomályosítsák látásunkat, fel kell mérnünk a mesterséges intelligencia fejlesztésével és felhasználásával járó, vélt vagy valós etikai és jogi kockázatokat. Az MI-rendszerek mind önmagukban, mind valamilyen termék részeként megsérthetik az olyan alapvető jogokat, mint a magánélethez való jog és a diszkrimináció tilalma, befolyásolhatják és manipulálhatják az emberek döntéseit anélkül, hogy azt a döntéshozó érzékelné, akadályozva ezzel az önrendelkezést, vagy veszélyeztethetik az emberek biztonságát.

Az iparági önszabályozó eszközök, amelyeket egyébként is egy-egy bigtech multinacionális vállalat ural, nem alkalmasak arra, hogy a mesterséges intelligencia káros hatásait kezeljék, így megkezdődött a jogi szabályozás eszköztárának alaposabb vizsgálata. Az MI-rendszerek fejlesztésének és felhasználásának kockázatai egyre inkább a figyelem középpontjába kerülnek, nemzetközi szervezetek (UNESCO, Európai Tanács, ENSZ) sürgetik, hogy a nemzetek veze-

tői ne csak olyan politikákat fogadjanak el, amelyek ösztönözhetik a hasznos innovációt, hanem olyan intézkedéseket is, amelyek megvédik az egyént és a társadalmat a mesterséges intelligencia veszélyeitől (Smuba, 2021).

Jelenleg egy olyan - remélhetőleg csak átmeneti - időszakon megyünk keresztül, amikor a hatályos normák nem képesek megfelelően szabályozni a mesterséges intelligencia fejlesztését, felhasználását a társadalom egészére és az egyes egyénekre gyakorolt negatív hatásainak kezelése érdekében, miközben az új normák megalkotása még várat magára (Lee, 2022). Ebben jelenthet globális jelentőségű változást a Rendelet, amely a GDPR-szabályozáshoz hasonlóan világszerte óriási hullámokat vet, és alapvető hatással lehet a technológiai fejlesztések jövőjére.

A mesterséges intelligencia által generált kihívásokra adott átfogó, koordinált jogalkotói válasszal való késlekedés rendkívül káros és beláthatatlan következménnyel járhat az egyénre és a társadalomra nézve, mivel a mesterséges intelligencia évről évre egyre jobban átszövi a mindennapjainkat az okos eszközökön, a közösségi média platformokon, az oktatáson, az orvosláson, a közlekedésen, az üzleti és a közhatalmi döntéseken keresztül.

A normák megalkotása során kiemelt figyelmet kell fordítani arra, hogy ne szabjunk indokolatlan akadályokat a mesterséges intelligencia fejlődése elé, mindeközben alapelveként tekintsünk (i) az MI-rendszerek átláthatóságára, (ii) a rendszer egész életciklusa alatt folyamatos, beavatkozásra képes emberi felügyelet képességére, (iii) a fejlesztők és a felhasználók felelősségének kérdésére az MI-rendszer által okozott káros hatásokért, (iv) az adatvédelemre és (v) az alapjogok megfelelő tiszteletben tartására. A mesterséges intelligencia szabályozásnak fontos célként kell tekintenie a MI-rendszerekbe vetett bizalom növelésére, mivel a bizalom hiánya akadályozza a MI-rendszerek elterjedését, ezzel csökkentve a mesterséges intelligenciában meglévő potenciális előnyök kiaknázását (Smuba, 2021).

III. AZ EURÓPAI MEGKÖZELÍTÉS

Az Európai Unió a 2010-es évektől kezdett érdeklődni az MI területe iránt, először gazdasági-versenyképességi szempontból, majd 2018 óta egyre inkább a kockázat, az emberi jogok, és általában a veszélyek kerültek az érdeklődés középpontjába.

Zódi (2021) meglátása szerint ennek oka a nagy platformokhoz kapcsolódó botrányokban és az ebből fakadó általános bizalmatlanságban keresendő, va-

lamint abban, hogy Európában mindig jóval nagyobb súlya volt az értékalapú megközelítésnek, szemben az USA erősen szabadpiaci megközelítésével.

A tanulmány az Európai Bizottság Rendelettervezetét egy szűk, de annál fontosabb aspektusból vizsgálja: eleget tesz-e a tervezet terminológiája egy folyamatosan változó technológia szabályozási megragadásával kapcsolatos elvárásoknak? Kellően rugalmas fogalommeghatározásokkal dolgozik-e, ugyanakkor kellőképpen világos-e a szabályozás tárgyi hatálya.

A dolgozat hipotézise, hogy a szabályozási tárgy megragadása során egyszerűre szükséges követni a technológiasemlegesség elvét, és nagyon jól célzottan megragadni azt, amit szabályozni kívánunk. Ennek során a terminológiában, a jogszabály nyelvezetében kerülni szükséges a műszaki kifejezések olyan jellegű használatát, amely azon a hiten alapul, hogy a műszaki-technológiai-tudományos fogalmak pontosan és jól határolják le a tárgyat. Állításom az, hogy a „miért, milyen jellemzőkre tekintettel, mit akarok szabályozni” kérdésre nem-technikai kifejezésekkel is jól leírható választ kell tudnunk adni. Az algoritmusokra, matematikai eljárásokra koncentráló fogalomhasználat könnyen idejét múlttá, meghaladottá válik, és az sem várható el, hogy minden új „buzzword” a jogszabály módosítását igényelje. E szabályozástechnikai kérdéskörön túl a dolgozat tartalmi állásfoglalást is tesz a „mennyire legyen tág a szabályozás hatálya” kérdésben: a technikainak, elméletinek tűnő vizsgálódás haszna ugyanis nagyon is gyakorlati: dollármilliárdok múlnak egy-egy kifejezésen, ha drámaiban akarunk fogalmazni, Európa globális versenyképessége.

IV. EGY „MI NYÁR”, AMIT SOKAN ŐSZNEK HITTEK

Az MI kutatások történetét felívelések, „MI nyarak” és „MI telek” jellemzik. Utóbbiak azon időszakokra utalnak, amelyek során mind a kutatók – és a finanszírozók – , mind a nagyközönség érdeklődése jelentősen visszaesett, csalódottság, kiábrándultság jelent meg a technológiával kapcsolatban. A váltakozás dinamikáját talán leginkább a fent említett általános mesterséges intelligencia elérésében való hit, és az arról való lemondás adta. Amellett, hogy az ún. „*narrow AI*” területén egyre komolyabb eredményeket sikerült elérni, tehát az emberi intelligencia egyes részterületein jól teljesítő rendszerek jelentek meg, újra meg újra kudarcba fulladt egy, az emberrel a versenyt felvenni képes, általánosan intelligens rendszer kifejlesztése.

Megvalósult az emberi arcok szoftver általi felismerése, a gépi fordítás, vagy a bőrrák gépi látáson alapuló korai azonosítása, ugyanakkor nem születtek olyan alkalmazások, amelyek akár csak távolról is hasonlítottak volna a Mesterséges Intelligencia Szent Gráljaként kezelt AGI-ra. 2020 januárjában a BBC vetette fel – számos vezető AI kutatót megszólaltatva – , hogy talán közeledik a következő MI-tél^[4]. Rá egy hónapra megjelent az Európai Bizottság Fehér Könyve a Mesterséges Intelligenciáról, amely felvázolta egy jövőbeni MI-szabályozás kereteit, majd 2021 áprilisában az „AI Act” tervezete. A jogalkotási folyamat kellős közepén, 2022. november 30-án a nagyközönség számára is elérhetővé vált a Chat-GPT, amelynek fényében hirtelen jóval nagyobb jelentőséget kapott a rendelettervezet, egyben rákényszerítve az európai uniós testületeket egyes új, vagy újnak vélt technológiákkal kapcsolatos határozott állásfoglalásra a jogalkotási folyamatban. A GPT-3.5 nyelvi modellt használó, szöveges instrukciók megértésére, és szöveges tartalmak generálására optimalizált alkalmazás tudása megdöbbenetete mind a kutatókat, mind a nagyközönséget. Azok számára, akik korábban legyintettek arra a felvetésre, hogy a Mesterséges Intelligencia képes lesz elvenni ügyvédek, reklámszöveg-írók vagy sajtófigyelést végzők munkáját, hirtelen elemi erővel mutatkozott meg a technológia ereje, amely számos pozitív hatás mellett óriási kockázatot jelent az alapvető jogokra nézve.

V. ÚJ TECHNOLÓGIAI FOGALMAK A HAGYOMÁNYOS MI-FOGALOMRENDSZERBEN

Milyen módon tudjuk elhelyezni az ANI-AGI fogalomrendszerben a „*general purpose AI*” vagy „*foundation models*” fogalmat? A tudományos közmegegyezés szerint ezek a rendszerek nem minősülnek AGI-nak, ugyanakkor a technológia messze túlmutat az ANI rendszerek képességein, és alapvető különbségeket mutat a rendszer betanítása tekintetében is. Az általános célú felhasználás már elnevezésében is arra utal, amit egy AGI rendszertől várnánk, ugyanakkor azt az elvárást, hogy egy AGI rendszer valamennyi gazdaságilag értékes feladatban teljesítsen jobban, mint az ember, nem teljesítik. A „*foundation model*” kifejezés e rendszerek azon jellemzőjét tükrözi, hogy önmagukban nem valamely jól körülhatárolt feladat ellátására szolgálnak, hanem lényegük az, hogy minimális paraméterezés mellett számos, szinte végtelen számú célra

[4] <https://www.bbc.com/news/technology-51064369> (letöltés ideje: 2023. május 12.)

használható rendszer épülhet rájuk. A modelleket alkalmazásprogramozási felületen (*application protocol interface*, API) keresztül teszik elérhetővé más fejlesztők számára, akiket „downstream fejlesztőknek” nevezünk. Érdekes még egy fogalmat tisztáznunk ebben a körben, mégpedig az „LLM” (*Large Language Models*) fogalmát, mivel az általános célú rendszerek közelmúltbeli áttörése elsőként ezen a területen valósult meg a GPT-3 publikálásával 2020 júliusában az OpenAI fejlesztésének eredményeként. A GPT-3 tehát a nagy nyelvi modellek körébe tartozó „*foundation model*”. Az OpenAI által kidolgozotthoz hasonló generatív nyelvi modellek célja, hogy az emberi nyelvhasználatnak megfelelő szöveget generáljanak mesterséges neuronhálózatok bevonásával. A modell tanításához egy óriási adatbázist vesznek igénybe, amelyből a modell megtanulja előrejelezni egy mondatban a következő szót, az ún. felügyelet nélküli tanulás (*unsupervised learning*) módszerét alkalmazva, a szöveg statisztikai mintáit, struktúráit feltárva, „megtanulva”. A következő fázisban a finomhangolás történik, már a felügyelt tanulás módszerének segítségével. Természetesen nem ez volt az első nagy nyelvi modell, de ez hozott olyan áttörést elsősorban a paraméterek számának^[5] és a tanítási adatbázis méretének köszönhetően, amely a közelmúlt vitáit elindította. A GPT-3 képességei és sokoldalú felhasználása nemcsak a „laikusokat” lepte meg, hanem a modell fejlesztőit is. Az általános célú MI tehát egy lépés az AGI távoli ideája felé, olyan lépés, amely újra hihetővé tette, hogy egyszer elérhető lesz a feladatok széles körében az emberrel azonos, vagy magasabb szinten teljesítő gépi intelligencia. Ez a lépés akkora ugrásnak tűnik, hogy különös hangsúlyt kap a felkészülés és a védekezés kérdése.

VI. A RENDELETTERVEZET HATÁLYA AZ EURÓPAI BIZOTTSÁG JAVASLATÁBAN

A jogi szabályozással kapcsolatos alapvető elvárás a tárgyi hatály egyértelműsége, amely nem magától értetődő az új technológiák szabályozása esetében, ahol a szabályozás tárgya maga is újszerű, a jogalkotó számára is bizonytalan körvonalú, másrészt folyamatos mozgásban van.

A definíciónak ennek megfelelően kellően egyértelműnek, jól körülhatároltnak kell lennie, hiszen enélkül a jogbiztonság hiányzik, másrészt kellőkép-

[5] Míg a GPT-2 „csak” 1,5 milliárd paraméterrel dolgozott, addig a GPT-3 ennek a 116-szorosával, 175 milliárd paraméterrel. A GPT-3 tanító adatbázisa egy 45 TB méretű szöveges állomány volt.

pen rugalmasnak ahhoz, hogy a technológiai változásokhoz alkalmazkodni tudjon. A korábbi definíciós kísérletekről általában elmondható, hogy a tágabb megközelítést követték, vagy egyáltalán meg sem kísérelték az MI definiálását (OECD, Európa Tanács, UNESCO) (Renda, Engler, 2023). Egy európai uniós rendeleti szabályozás, amely nagyon komoly költségvonzattal járó kötelezettségeket kíván telepíteni a kötelezettekre, nem teheti meg, hogy nem alkot nagyon világosan értelmezhető MI-definíciót. Tekintsük át, hogy a Rendelettervezetben szereplő eredeti javaslat hogyan változott az Európai Parlament által javasolt módosításig.

Fontos kiemelni, hogy a Rendelettervezet nem az MI, hanem az MI-rendszer fogalom meghatározását adja meg. A Rendelettervezet indokolása ennek kapcsán hangsúlyozza, hogy „*a lehető legnagyobb mértékben technológiasemleges és időtálló kíván lenni, figyelembe véve a mesterséges intelligenciával kapcsolatban gyorsan bekövetkező technológiai és piaci fejleményeket.*”^[6] A definíció sajátossága, hogy azt az I. melléklettel szükséges összeolvasni, amely a mesterséges intelligencia fejlesztésére vonatkozó megközelítések és technikák részletes listáját tartalmazza^[7]. Az I. melléklet folyamatos frissítésére a Bizottság felhatalmazást kap a Rendelettervezet 4. cikkében. A 3. cikk 1. pontja szerint *mesterségesintelligencia-rendszer (MI-rendszer): „olyan szoftver, amelyet az I. mellékletben felsorolt technikák és megközelítések közül egy vagy több alkalmazásával fejlesztettek, és amely az ember által meghatározott célkitűzések adott csoportja tekintetében olyan kimeneteket, például tartalmat, előrejelzéseket, ajánlásokat vagy döntéseket képes generálni, amelyek befolyásolják azt a környezetet, amellyel kölcsönhatásba lépnek;”* Az első látásra gondosnak tűnő definíció számos kritikát kapott mind a szakmai érdekképviseleti és tudományos műhelyek véleményezése során, mind a jogalkotás aktorai részéről. Egyesek túl tágnak, mások túl szűknek, megint mások homályosnak, a kötelezettek által nehezen értelmezhetőnek jellemezték a meghatározást. Grady

[6] Európai Bizottsági javaslat az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról (a továbbiakban a hivatkozások alatt: Javaslat), 5.2.1 pontja

[7] Javaslat I. mellékletében foglaltak szerint, a mesterséges intelligenciával kapcsolatos technikák és megközelítések a Javaslat 3. cikk 1. pontjának megfelelően

- a) Gépi tanulási megközelítések, ideértve a felügyelt, a felügyelet nélküli és a megerősítő tanulást, a módszerek széles skálájának, többek között a mélytanulásnak az alkalmazásával;
- b) Logikai és tudásalapú megközelítések, beleértve a tudás megjelenítését, az induktív (logikai) programozást, a tudásbázisokat, a következtetőmotorokat, a(z) (szimbolikus) érvelést és a szakértői rendszereket;
- c) Statisztikai megközelítések, Bayes-féle becslés, keresési és optimalizálási módszerek.

(2023) szerint a technológiasemlegesség követelménye nem azt jelenti, hogy különböző technológiákra pontosan ugyanazok a szabályok vonatkoznak. Ha valamelyik technológia speciális problémákat vet fel a szabályozó számára, természetesen lehetősége van erre specifikusan reagálni. A Rendelet tervezete az I. mellékletben számos olyan megközelítést, technológiát vonna be az MI-rendszer fogalmkörébe, amelyeket egyrészt senki nem sorolna oda, másrészt nem vetettek és vetnek fel jelenleg sem speciális szabályozási megközelítést. A szerző a hatály erős leszűkítését javasolja, a gépi tanulás technológián belül is egy szűkebb csoportra, a nem interpretálható (*uninterpretable*, máshol *noninterpretable*) gépi tanulásra. Az ebbe a körbe tartozó matematikai eljárások/megközelítések, mint a mesterséges neurális hálók, vagy az ún. *véletlen erdők* lehetetlenné teszik az inputok alapján az outputok egyértelmű levezetését. A tanulmány egy lehetséges definíciós javaslatlal is él, egyben törlésre javasolja az I. mellékletet: az MI-rendszer „olyan rendszer, amely a szolgáltató vagy a felhasználó számára ismeretlen paraméterek alapján gépi tanulással következtet arra, hogy célkitűzések egy adott csoportját hogyan lehet elérni, és a rendszer által generált kimeneteket, például tartalmat (generatív AI rendszerek), előrejelzéseket, ajánlásokat vagy döntéseket képes generálni, amelyek befolyásolják azt a valós vagy virtuális környezetet, amellyel kölcsönhatásba lépnek.

Álláspontom szerint a szerző megállapítása helyes abban a kérdésben, hogy különböző jellemzőkkel bíró technológiák különböző szabályozási rezsimek alá tartozhatnak, nem tartom azonban helyesnek, hogy konkrét technológiascsoport (gépi tanulás) megnevezését tartalmazza a definíció. Legyen bármennyire is bevett és stabilnak, öröknek hitt a gépi tanulás fogalma, Kurzweil (2001) óta tudjuk, hogy a technológiai fejlődés üteme exponenciális, ezért nagyon nehéz bármi biztosat mondani azzal kapcsolatban, hogy melyik lesz a vezető MI-technológia a következő években – vagy hetekben, hónapokban. Másrészt legyen bármennyire is nagy átfedés a gépi tanulás műszaki-tudományos meghatározásában, mégiscsak bizonytalan körvonalú fogalmakról beszélünk, és nem tűnik jó iránynak bizonytalan elemet behozni egy jogszabályi definícióba.

A kérdést úgy is feltehetjük, hogy vajon ad-e bármilyen többletet a definícióhoz a kifejezés használata? Álláspontom szerint nem. Úgy gondolom, a definíció az „ismeretlen paraméterek” megragadásával, valamint a meghatározás többi elemével olyan módon képes elvonatkoztatni konkrét technológiától, hogy annak lényegi, a szabályozás számára releváns jellemzőjét képes megragadni.

Smuha és munkatársai (2021) álláspontja szerint is nagyon tágra szabta az Európai Bizottság az MI-rendszer fogalmát, gyakorlatilag bármely technológia annak minősülhet. Ez a bizonytalanság számos statisztikai vagy logikai eljárással dolgozó fejlesztő számára jogbizonytalanságot okozhat, amely a szerzők szerint kétféle módon lehetne feloldható. Az egyik lehetőség szerint a jogszabálytervezet címe módosulna olyan módon, hogy nem Mesterséges Intelligenciáról, hanem az algoritmusokról vagy a szoftvekről szólna. A szerzők felvetik, hogy bizonyos kritikus területeken nem az a fontos, hogy milyen technológiát alkalmaznak a fejlesztők, hanem maga az alkalmazás, a fókusz így a technológiáról a szabályozott területre helyeződne át. A másik lehetőség itt is a gépi tanulás irányában történő korlátozás lenne, amellyel kapcsolatban a szerzők azt a kockázatot látják, hogy számos, ugyancsak kockázatot hordozó technológiát téveszthetnek szem elől. Álláspontom szerint az első lehetőség egyértelműen elvethető, mivel a hagyományosnak mondható, kiismert és jól kontrollált technológiák vonatkozásában kifejezetten veszélyes egy, a Rendelettervezet szerinti követelményrendszer bevezetése, amennyiben a III. mellékletben felsorolt területek tekintetében szükséges különleges szabályokat előírni, azt a területekre vonatkozó ágazati szabályokban meg lehet tenni.

A European Digital SME Alliance (2021) álláspontja szerint is túl tág az MI-rendszer definíciója, és konkrétan megjelöli az I. melléklet c) és b) pontjában felsorolt rendszereket, mint amelyek évtizedek óta jelen vannak, bármilyen specifikus szabályozás nélkül. A javaslat azonban nem arra a következtetésre jut, hogy a hatályt valamely technológiára koncentrálni kell, hanem konstatálja, hogy az MI pontos definíciójában a szakértők sem értenek egyet, ezért az alkalmazást kellene szabályozni a technológia helyett. Ha nem így jár el a szabályozás, akkor megkerülhetővé válik, pedig MI-nek nem minősülő technológiákkal is lehet alapjogokat sérteni vagy hordozhatnak hasonlóan komoly kockázatot.

Hasonló érveléssel áll elő Renda és Engler (2023): álláspontjuk szerint egy tág definíció elejét veheti annak, hogy egyébként az alapjogokra és a biztonságra kockázatos technológiák kikerüljenek a hatály alól. A tág definíció elejét veheti annak is, hogy szándékosan kikerülje a fejlesztő a szabályozás hatályát, olyan technológiát választva, amely kívül esik a definíción. Egy túlságosan szűk definíció meglátásuk szerint sokkal több kérdést vet fel továbbá abban a kérdésben, hogy valamely konkrét műszaki megoldás a rendelet hatálya alá tartozik-e vagy sem.

A Future of Life Institute 2021. augusztus 6-án kelt véleményében azt kifogásolta, hogy a javaslat azzal, hogy egyetlen rendeltetés irányából ragadja meg a rendszereket, elmulasztja az általános célú MI-rendszerek (*general purpo-*

se AI) szabályozását. A GPT-3, a DALL-E és a hasonló, közelmúltban megjelent rendszerek számos – gyakorlatilag korlátlan számú – különböző célra használhatók, így hiba lenne azokat egyetlen rendeltetés alapján besorolni. Ehelyett egy olyan szabályegyüttesre lenne szükség, amit rendeltetéstől függetlenül minden MI-rendszer esetén alkalmazni kellene.

VII. A TANÁCS FOGALOMRENDSZERREL KAPCSOLATOS JAVASLATAI

A Tanács által 2022. december 6-án elfogadott Általános Megközelítés szűkített a korábbi definíción, indoklásuk szerint annak érdekében, hogy a „hagyományos szoftverrendszerektől” világos kritériumok mentén lehessen elhatárolni az MI-rendszereket. Az Általános Megközelítés törölte az I. mellékletet, és a 4. cikk már nem annak folyamatos felülvizsgálatára hatalmazza fel a Bizottságot, hanem végrehajtási aktus elfogadására annak érdekében, hogy az általa meghatározott két technológia MI-megközelítés technikáit folyamatosan pontosítsa és frissítse. A Tanács tehát két technológiai megközelítésre szűkítette le Bizottság korábbi listáját, a gépi tanulási, valamint a logikai és tudásalapú megközelítésekre. A módosítás lényegét az Általános Megközelítés úgy írja le, hogy nem MI-rendszer az olyan rendszer, amely „kizárólag természetes személyek által definiált szabályokat használ egy lépéssorozat automatikus végrehajtása érdekében” (Tanács Általános Megközelítés 6. preambulum-bekezdés).

Jól látszik, hogy a szabályozási törekvéseknek ez a bizonytalanság, a „fekete doboz”-jelleg áll a középpontjában. Ott válik el a szabályozás tárgyának figyelemre érdemes MI a többi szoftvermegoldástól, ahol bejön a kiszámíthatatlanság. Úgy is fogalmazhatunk, hogy ahol a gépi problémamegoldás „emberivé” válik: ahol maga a probléma sem fogható meg egyszerűen, és ahol a megoldásához vezető út sem algoritmizálható a szó hagyományos értelmében. A 6a és 6b preambulum-bekezdés, amelyet a Tanács újként illesztett a tervezetbe, részletesen kifejti, mi jellemzi ezeket a technológiákat. A gépi tanulási megközelítés meghatározásakor a 6a bekezdés is kiemeli, hogy ezek a rendszerek olyan módon tanulnak az adatokból annak érdekében, hogy megoldjanak egy problémát, hogy nincs megadva lépésről lépésre egy utasítássorozat, amely elvezet az inputtól az outputig. Fel kell azonban hívni a figyelmet arra a tényre, hogy Tanács továbbra sem csak a gépi tanulási megközelítést érti Mesterséges Intelligencia alatt, a logikai és tudásalapú megközelítések – köztük a szakértői rendszerek is – továbbra is részét képezik a javaslatnak. Álláspontom szerint

a szűkítés nem olyan jelentős mértékű, ahogyan arra a dokumentum utal, legalábbis a megközelítések/technológiák tekintetében. Megjelenik ugyanakkor a definícióban az autonómia fogalma, a rendszernek autonóm jellemzőket kell mutatnia. A szabályozási dokumentumok mögött álló szándék kiolvasható tehát az indokolásokból, a technológiák körének behatárolása azonban ezen feltétel mentén nem egyszerű: számos, nem gépi tanuláson alapuló eljárás „hatásában ugyanúgy tűnhet „kiszámíthatatlannak”, mint egy olyan gépi tanuláson alapuló rendszer, amely önmagát „építi fel” az adatokból.” – jutott Zódi Zsolt (2021) erre a következtetésre a Rendelettervezetről szóló blogbejegyzés-sorozatának első részében.

Azt is hozzátehetjük, hogy valamennyi MI-megközelítés esetében létezik algoritmus, nem szabad azt gondolnunk, hogy a szoftver „azt csinál, amit akar”, a lényeg inkább úgy ragadható meg, hogy az algoritmus más absztrakciós szinten megfogalmazott, mint amit a gépek instruálása esetén megszoktunk. Rejtve marad a működés mibenléte előttünk, meglepő és kiszámíthatatlan a végeredmény annak ellenére, hogy a szoftver utasításokat végrehajtva dolgozik^[8].

Egy technológia szabályozása esetében a legfontosabb elvárás a jövőállóság. Ennek fényében bátor lépésnek tűnik, hogy a Tanács a definíció részévé tesz, és nem végrehajtási aktussal módosítható mellékletbe konkrét technológiákat. Nem nehéz ugyanis elképzelni újonnan megjelenő, a korábbi kategóriákba nem besorolható megközelítéseket, amelyek ma még nem ismertek, vagy amelyekre nem gondol a jogalkotó. Ha egy évvel a chatGPT megjelenése előtt kimaradhatott a szabályozásból az általános célú MI-rendszerek köre, akkor nem lehetünk túl bizakodóak. A tágabb definíció mellett érvelők azzal szoktak érvelni, hogy a nemzetközi definíciós kísérletekhez jobban illeszkedhet egy ilyen meghatározás, elősegítve az európai szabályozási megközelítés sikerét. A tág definíciót kritizálók azt vetik fel, hogy túl sok mindenre terjed ki a szabályozás, a túlszabályozás veszélye áll fenn.

Renda és Engler (2023) azonban felhívja a figyelmet arra, hogy a Rendelet hatályán túl van egy másik szűrő is, mégpedig a kockázat: többletkövetelményeket csupán a nagy kockázatú rendszerekre kíván előírni a Rendelet tervezete. Ezzel a felvetéssel szemben az innováció jövője iránt aggódók azt hozhatják fel, hogy éppen elég, ha a szoftverek túl széles köre csak korlátozott számú ágazatban kerül be a rendelet hatálya alá.

[8] Egy mesterséges neuronhálón alapuló mélytanulás eljárás például, ha rosszul kategorizált, utasítva van arra, hogy pontosítsa a modellt, és ennek érdekében módosítsa a súlyokon. Ez a folyamat azonban ember számára nem követhető, nem írható le egyszerűen és kiszámíthatóan.

Az Általános Megközelítés másik jelentős eleme, hogy bevezeti az általános célú MI-rendszer fogalmát, amelynek hiánya az egyik legkomolyabb kritika volt a bizottsági javaslattal szemben. A tervezet 3. cikkébe beszúrt új (1b) pont értelmében az „általános célú MI-rendszer” olyan MI-rendszer – függetlenül a piaci bevezetéstől vagy a szolgáltatás módjától – ideértve a nyílt forráskódú szoftvereket is – a gyártó/szolgáltató (*provider*) szándéka szerint általánosan alkalmazható funkciók ellátást szolgálja, mint a kép- és beszéd felismerés, a hang és videogenerálás, a mintadetekció, a kérdések megválaszolása, a fordítás és egyébek. Az általános célú MI-rendszerek számos kontextusban, számos más MI-rendszerbe integrálva használhatók.” Az általános célú MI-rendszerekre egy önálló, IA címet vezet be a javaslat, amely ezen rendszereknek a Rendeletnek való megfelelőségét tárgyalja, külön előírásokat állapítva meg. A javaslat arra az esetre vonja szabályozási körébe az általános célú MI-rendszereket, ha azokat nagy kockázatú MI-rendszerként használják vagy komponensei egy nagy kockázatú MI-rendszernek. A Rendelettervezet egyes rendelkezéseit azonban nem közvetlenül írja elő alkalmazandónak a javaslat, hanem egy végrehajtási aktusnak szánja annak meghatározását, hogy milyen módon kell azokat alkalmazni, amelyet átfogó hatásvizsgálatnak és konzultációs mechanizmusnak kell majd megelőznie.

VIII. AZ EURÓPAI PARLAMENT MEGFEJTÉSE

Az Európai Parlament MI-rendszer definíciója kitüntetett figyelmet szentel a nemzetközi harmonizációs igénynek, annak, hogy egy teljesen új definíció hátráltatná a nemzetközi elfogadást. A definíció álláspontom szerint kifejezetten szerencsés irányba mozdul el, és a szabályozási technikát, a megfogalmazást tekintve a technikai-műszaki szakkifejezések helyett az általános jellemzők megragadásával operál. A kompromisszumos módosító javaslatok^[9] értelmében az MI-rendszer olyan „gép-alapú rendszer, amelyet úgy terveztek, hogy különböző szintű autonómiával működjön, és amely képes kifejezett vagy implicit célok érdekében olyan kimeneteket generálni, mint például előrejelzések, ajánlások, vagy döntések, amelyek befolyásolják a fizikai vagy virtuális

[9] Az Európai Parlament Belső Piac és Fogyasztóvédelem Bizottságának, valamint az Állampolgári jogok, bel- és igazságügy bizottságának kompromisszumos módosítási javaslattervezete az Európai Parlament és a Tanács a Mesterséges intelligenciára vonatkozó harmonizált szabályok (a Mesterséges Intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról szóló javaslatához, 137.

környezetet.” Az autonómia fogalom, az implicit (értsd: nem ember által definiált) célok követése olyan jellemzők, amelyeket általában a gépi tanulás fogalmával fedhetnénk le, de ha a jövőben egy másik technológia mutatna ilyen jellemzőket, arra is tökéletesen illene a meghatározás.

Az Európai Parlament javaslattervezete az „általános célú MI-rendszer” fogalma mellett (1d), amely már az Általános Megközelítésben megjelenik, külön definiálja a „*foundation model*” fogalmát. Az általános célú MI-rendszer „olyan MI-rendszer, amely alkalmazások széles körében használható, illetve azokhoz adaptálható, amelyekre nem eleve tervezetten és specifikusan tervezték.” A „*foundation model*” definíciója (1c) a tervezet értelmében: „olyan MI-modell, amelynek tanítása az adatok széles körén történt, a kimenetek sokféleségére tervezték, és különböző feladatok széles körére adaptálható”. Az alapmodellekre vonatkozó előírásokon túl további előírások vonatkoznak a „generatív MI” kategóriába sorolható alapmodell előállítóknak. Ide soroljuk azokat az alapmodelleket, amelyen különböző szintű autonómia mellett tartalmat generálnak, legyen szó komplex szövegről, képről, hangról vagy mozgóképről. A sikeres parlamenti bizottságokat követően az Európai Parlament plenáris ülése várhatóan még 2023. folyamán dönt a javasolt szövegről.

IX. HOGYAN TARTSUK ÉLETBEN AZ INNOVÁCIÓT?

Valamennyi technológiai tárgyú szabályozás indokolása megfogalmazza maga számára azt a célt, hogy a jogvédelem mellett biztosítani kell a technológiai fejlődést, az innovációt. Azt már kevésbé érezni, hogy komolyan bele is gondolnak a kockázatba, amelyet az innováció lassítása, akadályozása hordoz. A kockázat különös jelentőségét az adja, hogy hajlamosak vagyunk túlértékelni az ún. „Brüsszel-effektust”, azt a jelenséget, hogy az európai jogalkotás példaként szolgál a világ számára, és azt várni, hogy egy-egy kockázatokat hordozó technológia fejlődését globálisan megnehezítjük. Valójában szabályozási versenyként is megjelenik a globális technológiafejlesztési verseny, és ha Európa nem teremt barátságos légkört, majd teremt más. A technológiai innováció pedig többé nem egy örvendetes, hanem egy nélkülözhetetlen feltétele a gazdasági sikernek: a gazdasági-hatalmi erőviszonyok első számú alakítója.

A jogászok ünneplik az MI-rendelet tervezetét, amely az első átfogó jogi szabályozás lesz a világon. Vajon a GDPR-szabályozást követően Európa újra irányt adhat majd a világ jogalkotásának, újabb jelét adva a Brüsszel-effektusnak, vagy a vesztébe rohan. Mueller (2021) a Center for Data Innovation elemzésében

a Bizottság előzetes gazdasági hatásvizsgálatát is felhasználva kísérelte meg megállapítani az MI-rendelet által a vállalkozásokra telepített költségeket, és annak hatását a versenyképességre. A jelentés számításai szerint a szabályozás a következő öt évben 31 Mrd euróval terheli majd meg az európai gazdaságot, 20%-kal csökkentve az MI-beruházásokat. Egy európai KKV, amely nagy kockázatú MI-rendszert fejleszt, akár 400.000 eurós megfelelési költséggel is szembe-sülhet, ezzel a profit 40%-os csökkenését okozva. Vajon megvalósítható-e ilyen költségek mellett a Bizottság azon célkitűzése, hogy az európai vállalkozások 75%-a használja a technológiát az évtized végére?

X. ÖSSZEFOGLALÁS

Úgy vélem, ma már nem vitatható, hogy a mesterséges intelligencia, legalábbis annak egy jól körülhatárolható területe sajátos szabályozási megközelítést és kitüntetett figyelmet érdemel, a hagyományos jogintézmények használata ezen a területen nem minősül megfelelő eszköznek, amint arra G. Karácsony Gergely (2021) tanulmányában felhívta a figyelmet. Szabályozni tehát kell, ugyanakkor különös körültekintéssel kell eljárni mind a „mit”, mind a „hogyan”, a konkrét kötelezettségek tekintetében, gondosan mérlegelve a költségeket és a hasznokat.

Jelen tanulmány a „mit” kérdéskörét járta körbe, különös óvatosságra intve a szabályozót, hogy túlságosan tágan határozza meg az MI-rendszer fogalmát. A mesterséges intelligencia egységes fogalom meghatározása nem létezik sem a tudományban, sem a jogban, az azonban biztosra vehető, hogy nem kívánja a jogalkotó valamennyi szoftverterméket a szabályozási körébe vonni.

A szabályozási tervezetekből, az indokolásokból és háttéranyagokból kiolvashatóan azon rendszerek szabályozási igénye jelentkezett elemi erejű igényként, amelyek működése megjósolhatatlan, fekete-doboz jelleget mutat, mellékesen technológiai háttérét a gépi tanulás, azon belül a mélytanulás jelenti. Ha a cél ez, nagyon pontosan kell céloznunk, mert egy plusz sor az I. mellékletben képes megölni az európai versenyképességet. Nem áltathatjuk magunkat azzal sem, hogy önmagában a rendelet hatálya alá tartozás nem jelent azonnal kötelezettséget, hogy csak a III. mellékletben felsorolt területeken fejlesztett rendszerek minősülnek nagy kockázatúnak. Éppen elég, ha ezeken a területeken áll meg az innováció, bőven elég ahhoz, hogy végzetes legyen. A pontos célzáshoz nagyon fontos, hogy ne értsük félre a technológiasemlegesség előírását: nem azt jelenti, hogy minden technológiát ugyanúgy kell szabályozni, hanem azt, hogy valamely

technológiát önkényesen, ok nélkül nem helyezhetünk a másik elé. A tanulmány lényegi megállapítása, hogy a pontosság nem azt jelenti, hogy a tudományos-műszaki fogalmakra utalva próbálom megragadni a lényegét (gépi tanulás, mélytanulás, neurális hálók stb.). Nem algoritmus nevekkel kell próbálkoznunk, hanem a szabályozni kívánt terület lényegi jellemzőinek egyszerű és érthető megragadásával, hasonlóan ahhoz, ahogyan az az Európai Parlament javaslatában szerepel. Nem túlzás azt állítani, hogy még a mesterséges intelligencia kifejezést sem kellene használni. Jogi szempontból értékelhetetlenül homályos fogalom, természetesen a marketing értéke jelentős, és egy jogszabály társadalmasítása esetén ez is fontos szempont. Mindezen szempontok figyelembe vétele még mindig csak félsiker. A közelmúlt felforgató generatív MI technológiái ugyanis egészen biztosan a szabályozás hatálya alá esnek majd, és nem mindegy, hogy a támasztott követelményeknek lehetséges-e a megfelelés, képesek-e a költségeket állni az európai startupok, a várható hasznok arányban állnak-e a költségekkel, és kivonulnak-e a nagyvállalatok az európai piacról, tovább erősítve az Egyesült Államok és Kína hegemoniáját a technológia területén. Ebből az aspektusból is megérdemelne egy új hatásvizsgálatot a rendelettervezet.

IRODALOM

1. Apnews.com: OpenAI: ChatGPT back in Italy after meeting watchdog demands. (Elérhető: <https://apnews.com/article/chatgpt-openai-data-privacy-italy%20b9ab3d12f2b2cfe493237fd2b9675e21>. Letöltés ideje: 2023. május 12.)
2. Bbc.com: Researchers: Are we on the cusp of an 'AI winter'?. (Elérhető: <https://www.bbc.com/news/technology-51064369>. Letöltés ideje: 2023. május 15.)
3. DIGITAL SME (2021): DIGITAL SME Position Paper on the EU AI Act. (Elérhető: <https://www.digitalsme.eu/digital/uploads/DIGITAL-SME-Position-Paper-AI-Act-FINAL-DRAFT-1.pdf>. Letöltés ideje: 2023. május 21.)
4. Futureoflife.org: Pause Giant AI Experiments: An Open Letter. (Elérhető: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Letöltés ideje: 2023. május 12.)
5. Grady, Patrick (2023): The AI Act Should be Technology-Neutral. Center for Data Innovation. (Elérhető: <https://www2.datainnovation.org/2023-ai-act-technology-neutral.pdf>. Letöltés ideje: 2023. május 15.)
6. G. Karácsony Gergely (2021): Innovatív jogalkotói megoldások a mesterséges intelligencia-alapú rendszerek szabályozására. In: Török Bernát, Zódi Zsolt (ed.): A mesterséges intelligencia szabályozási kihívásai. Ludovika Egyetemi Kiadó, Budapest.

7. Lee, Jaemin (2022): Artificial Intelligence and International Law. Springer Singapore, Singapore. DOI: <https://doi.org/10.1007/978-981-19-1496-6>.
8. Madiega, Tambiama – Chahri, Samy (2022): European Parliamentary Research Service. European Parliamentary Research Service. (Elérhető: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf). Letöltés ideje: 2023. május 12.).
9. Mueller, Benjamin (2021): How Much Will the Artificial Intelligence Act Cost Europe?. Center for Data Innovation. (Elérhető: <https://www2.datainnovation.org/2021-aia-costs.pdf>. Letöltés ideje: 2023. május 27.).
10. OpenAI.com: Charter. (Elérhető: <https://openai.com/charter>. Letöltés ideje: 2023. május 12.).
11. Renda, Andrea – Engler, Alex (2023): What's in a name? - Getting the definition of Artificial Intelligence right in the EU's AI Act. CEPS Explainer. (Elérhető: https://www.ceps.eu/wp-content/uploads/2023/02/CEPS-Explainer-2023-02_Definition-of-AI-right-in-the-EUs-AI-Act.pdf. Letöltés ideje: 2023. május 12.).
12. Smuha, Nathalie - Ahmed-Rengers, Emma - Harkens, Adam - Li, Wenlong - MacLaren, James - Piselli, Riccardo - Yeung, Karen (2021): How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act. Legal, Ethical & Accountable Digital Society (LEADS) Lab, University of Birmingham. (Elérhető: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3899991. Letöltés ideje: 2023. május 10.). DOI: <http://dx.doi.org/10.2139/ssrn.3899991>.
13. Smuha, Nathalie (2021): From a 'Race to AI' to a 'Race to AI Regulation' - Regulatory Competition for Artificial Intelligence. In: Law, Innovation & Technology. Vol. 13 Iss. 1, 2021., KU Leuven Faculty of Law, Leuven, Belgium. (Elérhető: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3501410. Letöltés ideje: 2023. május 10.). DOI: <http://dx.doi.org/10.2139/ssrn.3501410>.
14. Zódi Zsolt (2021): Az Európai Bizottság Mesterséges Intelligencia Kódexének tervezete. In: Gazdaság és Jog. 2021/5. szám.

JOGFORRÁSOK:

1. A Tanács általános megközelítése az Európai Parlament és a Tanács a Mesterséges Intelligenciára vonatkozó harmonizált szabályok (a Mesterséges Intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról szóló rendelettervezetéről.
2. Az Európai Parlament Belső Piac és Fogyasztóvédelem Bizottságának, valamint az Állampolgári jogok, bel- és igazságügy bizottságának kompromisszumos módosítási javaslattervezete az Európai Parlament és a Tanács a Mesterséges intelligenciára vonatkozó harmonizált szabályok (a Mesterséges Intelligenciáról szóló

- jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról szóló javaslatához.
3. Európai Bizottsági javaslat az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról. Javaslat száma: COM/2021/206.
 4. Fehér könyv a mesterséges intelligenciáról: a kiválóság és a bizalom európai megközelítése. Európai Bizottság.

G. KARÁCSONY GERGELY

KI VÉDI MEG A VIRTUÁLIS
SZEMÉLYEKET?

– A KITERJESZTETT EMBER ALAPJOGI
VÉDELME A METAVERZUMBAN ÉS
AZON TÚL

Ki védi meg a virtuális személyeket? – a kiterjesztett ember alapjogi védelme a metaverzumban és azon túl

I. BEVEZETÉS – A VIRTUÁLIS TEREK SZABÁLYAI

Az elmúlt néhány év talán legnagyobb technológiai szenzációja a szabadon használható virtuális terek, közkeletű nevén a metaverzum fejlődése volt. A technológiával szembeni várakozásokat jól mutatja, hogy a világ egyik legnagyobb informatikai szolgáltatója, a Facebook anyavállalata is felvette a Meta nevet. Ezek a felületek személyi számítógépen vagy mobil eszközön, legteljesebb mértékben pedig virtuálisvalóság (VR)-eszközökkel használhatók. Felhasználási lehetőségeik szinte végtelenek: a kikapcsolódás és a fiktív virtuális világok szórakozási célból való létrehozásától kiindulva az oktatási, munkahelyi kollaborációs feladatokon át, a távol élő családtagokkal és barátokkal szervezett virtuális összejövetelekig az élet számos területén bizonyultak jó megoldásnak. A virtuális terekben a felhasználó először létrehozza saját digitális mását: ez az avatar fogja őt megszemélyesíteni a metaverzum többi felhasználója irányában. Az avatart a felhasználó mozgatja, VR-eszközök használata esetén a digitális alteregő leképezi a felhasználó kézmozdulatait, teste és feje együtt fordul a felhasználóval. A metaverzumban a felhasználók élő szóban (mikrofon segítségével), esetleg írásban kommunikálhatnak egymással, a virtuális tér résztvevői a másik személyről csak a felhasználónevét ismerhetik meg általában. Ezek az így létrehozott virtuális világok szabadon bejárhatók, nincs előre megkötve, hogy a felhasználók merre menjenek vagy milyen feladatokat hajtsanak végre. Ezekre a szabad virtuális világokra példa a Meta által fejlesztett, VR-készülékkel használható tér, a Horizon. Kvázi metaverzumként működnek egyes online többfelhasználós játékok terei is, például a Fortnite vagy a Roblox. E tanulmányban azt vizsgáljuk meg, hogy a virtuális terekben létrehozott digitális személyeket, az avatarokat, megilletik-e bizonyos alapvető jogok, részesülnek-e jogi védelemben az emberekhez hasonlóan. Kiinduló feltételezésünk az, hogy ahol emberek (vagy az általuk irányított karakterek) nagy számban vannak jelen, ott hamar szükségessé válik a békés együttélést biztosító szabályrendszer lefektetése. Feltételezzük, hogy bizonyos, személyhez fűződő alapjogok, úgymint

a személyes adatok és a személyiség védelme, valamint a diszkrimináció tilalma, érvényesülést kívánnak a metaverzumban is.

A digitális technológiák gyors térnyerésével együtt megfigyelhető az a tendencia, hogy e technológiák működésének és hatásainak szabályozása egyre többször kerül ki a nemzetállami jogalkotó kezéből, és kerül át iparági szereplők vagy azokat tömörítő csoportok kezébe (Guihot, Matthiew, Suzor, 2020), amelyek iparági sztetenderdek és jó gyakorlatok kialakításával próbálják elejét venni a negatív hatásoknak. A szabályalkotás ilyen fajta decentralizálódása egyrészt a műszaki fejlődés gyors sebességével, másrészt a hagyományos jogalkotók szabályozó szerepének leértékelődésével magyarázható (Villaronga, Golia, 2019). Ezt figyelhetjük meg az egyre inkább elterjedő virtuális terek vonatkozásában is. A metaverzumba belépve egy saját világ tárul a szemünk elé, amelyben szabadon tevékenykedhetünk. E virtuális világnak is megvannak azonban a szabályai: milyen „fizikai” törvényszerűségeknek engedelmessé az avatarunk (például hogyan mozog a térben), hogyan lép interakcióba a környezetével (átalakíthatja-e a virtuális tér elrendezését, létrehozhat-e új elemeket a térben), illetve a többi felhasználóval (láthatják-e egymást, beszélhetnek, kapcsolatba léphetnek-e egymással). Ezeket a szabályokat az adott virtuális világ szoftveres keretrendszere határozza meg, a fejlesztő előre kódolta a működési szabályokat a programba. Számunkra fontosabbak azonban a „viselkedési normák” a metaverzumok világában. Ezek határozzák meg ugyanis azt, hogy mit tehet meg egy felhasználó és mit nem: vagyis mik a megengedett és tiltott magatartások.

Néhány szabálynak világos jogi háttere van. A jogalkotó és jogalkalmazó szervek első sorban a kommunikációra koncentráltak: ha a játékosok közötti üzenetváltás egyértelmű bűncselekményt valósít meg, bűncselekmény előkészítésére, illetve arra való felbujtásra irányul, úgy az illetékes hatóságok hamar közbelépnek. Ilyen esetnek tekinthetjük például a gyermekpornográfiával kapcsolatos tartalmak megosztását, amely kapcsán több precedens is mutatja, hogy a rendvédelmi szervek – a virtuális környezet fejlesztőjével együttműködve – bizonyítékot gyűjthetnek a gyanúsított ellen, lefoglalhatják a virtuális fiókban található adatokat, valamint a virtuális térben létrehozott játékos mögött álló valódi személy kilétére vonatkozó információkat is szerezhettek. Hasonlóképpen monitorozhatják a virtuális térben folyó kommunikációt a hatóságok annak gyanúja esetén, hogy ott terrorcselekményt vagy más súlyos bűncselekményt készítenek elő. Ezeknek az eseteknek azonban közös jellemzője, hogy a virtuális teret pusztán eszközként, kommunikációs vagy adatmegosztó csatornaként használják más, a valódi világban megvalósuló bűncselek-

ményhez (tiltott pornográf felvétel készítése, terrorcselekmény, stb). Ezekben az esetekben a szabályozó szerv, illetve a bűnüldöző testület beavatkozását az indokolja, hogy a virtuális térben folyó jogellenes cselekmény kihatással van a valós világra is, és a külvilágban ténylegesen tiltott cselekményt megvalósító valódi személyek azok, akik a virtuális világba is átirányították tevékenységük egy részét.

Vizsgálatunk tárgyát most azok az esetek képezik, amikor egy magatartás kizárólag a virtuális világban valósul meg, az ott található szereplők között, és nincs semmiféle közvetlen kihatással a valós világra. A virtuális világban tanúsított magatartást – amennyiben kifejezett bűncselekményt a fentiekben bemutatott módon nem valósít meg – nem vagy csak alig szabályozza a jog. Amint két felhasználó játékon belüli interakciójáról van szó, máris átkerülünk a virtuális teret szolgáltató cég felhasználói feltételei, illetve a saját maga által kialakított „közösségi irányelvek” (így például a Meta által létrehozott Horizon virtuális világhoz tartozó Virtuális élmények magatartási szabályzata)^[10] által uralt területre. A legtöbb esetben ugyanis a jog nem vonja védelmi körébe a kizárólag a virtuális világban létező karakterek közötti interakciókat, kivéve akkor, ha azoknak kihatása van a külvilágra, illetve az avatarok mögött álló természetes személy felhasználókra is. A közösségi irányelvek tiltanak bizonyos magatartásokat, a Meta virtuális világának magatartási szabályai között például az alábbiakat találjuk:

- Ne tegyél vagy népszerűsíts olyasmit, aminek célja mások megfélemlítése
- Ne add ki magad másnak, ne gyűjts jogtalan módon személyes adatokat.
- Ne tegyél olyasmit, amivel másnak ártasz, így tilos különösen:
 - kiskorúak szexuális vagy más természetű kihasználása;
 - zaklatás;
 - erőszak, emberkereskedelem, rabszolgaság vagy annak népszerűsítése;
 - fizikai sérülés okozása, önsértésre vagy öngyilkosságra biztatás;
 - nem beleegyezésen alapuló szexuális tevékenység;
 - szellemi tulajdon megsértése;
 - szabályozott árukkal való kereskedelem.

A tilalmak megsértőivel szemben rendszeren belüli szankciókat helyez kilátásba: határozott időre szóló, illetve végleges kitiltást, ezt megelőzően figyelmeztetést vagy a fiók felfüggesztését. A közösségi irányelveket sértő

[10] Elérhető itt: <https://www.meta.com/help/quest/articles/accounts/privacy-information-and-settings/code-of-conduct-for-virtual-experiences/> (letöltés dátuma: 2023. szeptember 1..)

magatartások „áldozatainak” bejelentési panaszfelületet, illetve egyes esetekben beépített védekezési mechanizmusokat (a Meta például a virtuális térben a karakter körül vont „személyes teret” enged beállítani, ahová más felhasználó nem léphet be) biztosít a fejlesztő. Ezeket azonban saját diszkréciójából, a jobb játékélmény megteremtése és a felhasználói panaszok számának csökkentése érdekében vezeti be, és ugyanúgy bármikor jogosult ezeken egyoldalúan is módosítani. Ebben a tanulmányban amellet érvelek, hogy szükség lenne a virtuális térben létező avatarok „alapjogait” védő minimumsztenderdek meghatározására, amelyek platformtól függetlenül mindenhol kötelezők, és amelyek érvényesülését minden virtuális környezet fejlesztőjének biztosítani kell. Ezeket a minimumsztenderdek a nemzetközi egyezményekben és az alkotmányokban lefektetett emberi jogi követelmények között lehetne megtalálni. A szóba jöhető alapjogok közül e tanulmány keretében a magánszféra (és a személyes adatok) védelméhez és a jó hírnévhez való jog, valamint a diszkrimináció-tilalom egyes eseteinek alkalmazhatóságát vizsgálom meg.

II. MIÉRT VAN SZÜKSÉG A KIZÁRÓLAG A VIRTUÁLIS TÉRBEN LÉTEZŐ KARAKTEREK ALAPJOGI VÉDELMERE?

Ez a kérdés vetődik fel elsőként a téma kapcsán, hiszen első olvasásra ez is legalább annyira távolinak és (egyelőre) akadémikus problémafelvetésnek tűnik, mint a robotok és a mesterséges intelligenciák önálló személyiséggel való felruházása (Papakonstantinou, Hert, 2018; Keserű, 2020) .

A virtuális tér lakóinak alapjogi védelme önmagában nem magukat a digitális karaktereket hivatott megóvni a jogsérelmekről. Az alapjogi gondolkodás logikájából következően az alapjogok alanya, jogosultja, kizárólag természetes személy, vagyis ember lehet. Nincs ez másképp a metaverzum alapjogai kapcsán sem: ezek végső soron az avatarok mögött álló embert védik azáltal, hogy megóvják az általa irányított digitális személyt az azt érő hátrányoktól. A javasolt alapjogi védelem alkalmazhatóságához meg kell határozni azokat az eseteket, amikor a virtuális karaktert érő hatások továbbgyűrűzhetnek a mögötte álló emberre.

A jelenlegi jogfelfogásunk nem képes kezelni azt a helyzetet, amikor két digitális karakter a virtuális térben okoz egymásnak személyiségi jogsértést. Ez a helyzet ahhoz hasonlítható, amikor a bábszínházban Vitéz László sodrófával ütlegeli az ördögöt: ott sem valósul meg a testi sértés büntetőjogi alakzata, hiszen egyik szereplő sem alanya a jogi védelemnek.

A továbbiakban nem azokról a virtuális terekről és magatartásokról lesz szó, ahol a felhasználó eleve abban a tudatban és azzal a céllal lép be a virtuális környezetbe, hogy az általa irányított karaktert hátrány érheti: például a lövöldözős játékokban vagy a harcra épülő szerepjátékokban. Ebben az esetben a játék előre ismert és elfogadott része a szereplők közötti ellenségeskedés, az előre bizonytalan kimenetelű játékon belüli harc. Több más esetet el tudunk képzelni azonban, ahol a virtuális térben hátrány érheti a karaktert, és ez hatást gyakorol a karaktert mozgató felhasználóra is.

1. Egyértelmű a helyzet olyankor, amikor a digitális karakterhez konkrét üzleti érdeke kötődik a felhasználónak. A videojátékok és más, virtuális tevékenységek mögött álló komplex monetizációs rendszerek (Király, 2021) jelentős bevételhez képesek juttatni a felhasználókat. Sok esetben egy-egy ismert játékos vagy *streamer*, esetleg előadóművész részletesen kidolgozott és hosszú idő alatt felépített karaktert alkalmaz. Az így megszemélyesített virtuális személyiség az üzleti modell része, így az azt ért jogsérelem voltaképpen a felhasználó vállalkozását érő kárként vizsgálható. A továbbiakban azonban nem az üzleti sérelmekről lesz szó.
2. A második esetben a virtuális térben létező személyiséget ért, elsősorban a magánszféra védelme körébe eső sérelem direkt módon hat vissza a mögötte álló emberi felhasználóra. Ilyen az az eset, amikor a virtuális világban a karakter által tanúsított magatartásból a mögötte álló felhasználó személyes adataira, illetve személyiségjegyeire vonható le következtetés. Az alábbiakban bemutatjuk azt, hogy ez miért bír jelentőséggel még akkor is, ha a felhasználó a metaverzumban nem a saját, mindennapi személyiségének megfelelő módon viselkedik.
3. A harmadik esetkör áll a legtávolabb az embert ért közvetlen hátránytól. Ebben az esetben ugyanis kifejezetten lélektani jellegű hatásról beszélhetünk, ami a virtuális karaktert érő hátrány hatására az emberi felhasználót éri. Ahogy a virtuális karakterek egyre gyakrabban emberszerű jellemzőkkel, saját egyéniséggel és valóságghú megjelenéssel vannak ellátva, egyre könnyebben alakítunk ki kötődést irányukban. Így az e karaktereket ért sérelem, esetleg azok teljes elpusztulása valódi lelki megrázkódtatást jelent a hozzájuk kötődő ember számára ugyanúgy, mint egy kedvenc kisállat elvesztése. Az emberszerű robotokhoz való kötődés lehetőségét már kimutatta több empirikus tanulmány is az elmúlt években (Broadbent, 2017). Kifejezetten virtuális, vagyis testi kiterjedés nélküli karakterekhez való viszonyunkat

vizsgálta több kutatás is az elmúlt évtizedben. A virtuális érzelemki-fejezés hatással van arra, milyen mértékben tekintenek a befogadók egy virtuális karaktert emberszerűnek. Coanda és Aupers (2021) egy kutatásában azokat a számítógépes játékbéli karaktereket vizsgálta, amelyeket nem a játékosok működtetnek, hanem a játékok beépített szereplői (ezeket NPC-nek, non-player character, azaz nem játékos karakternek nevezzük). A kutatás megállapította, hogy a fiktív virtuális karaktereket a játékosok humanizálják, azaz kizárólag az emberekre jellemző tulajdonságokkal, motivációkkal, érzelmi állapotokkal ruházzák fel őket. A karakterekkel való identifikációt fokozza az, ha az adott karakter valóság-hű érzelmeket, minél hitelesebben meg tud jeleníteni.^[11] Mivel az ember hajlamos visszatükrözni egy másik ember érzelmi állapotát, elképzelhető, hogy az emberszerű, és valóság-hű érzelmek megjelenítésére képes virtuális karakterek érzelmi állapota is „átadódhat” a befogadóra, vagyis az őt megfigyelő vagy irányító emberi személyre is (Horváth, 2022). Ha egy virtuális karaktert szenvedni, fájdalmat elviselni lát a befogadó, akkor ebből következően ő maga is szenvedést élhet át.

Ebben az esetben tehát a virtuális karakterekkel való „emberséges” bánásmódot a jogalkotó nem magának az avataroknak a védelmére írhatja elő, hanem azért, mert az ezekkel való kegyetlenkedés – az érzelmek visszatükrözése és a karakterrel kialakított kötődés miatt – visszahathat az emberre, mint befogadóra is. Morális érvként hozhatjuk fel azt, a szakirodalomban is megjelenő álláspontot, ami szerint a robotokkal való kegyetlen bánásmód – csakúgy, mint az állatkínzás vagy az állatkísérletek – az emberiségre, mint egészre is rossz fényt vet, ezért tiltani kellene (Ashrafian, 2015).

A fentiek alapján megállapíthatjuk, hogy a virtuális térben létező karakterek, avatarok, jogi védelme bizonyos esetekben indokolt lehet. Nem önmaguk-

[11] Itt érdemes kitérni az *uncanny valley* néven ismertté vált jelésre is. Ezt a jelenséges elsőként Masahiro Mori, robotikával foglalkozó japán professzor írta le. Ő arra világított rá a humanoid, vagyis emberszerű robotok társadalmi megítélését vizsgálva, hogy a realiztikus megjelenés egy bizonyos pontig pozitívan befolyásolja a robotok emberek általi megítélését. Amikor azonban a nem emberi ágens kidolgozottsága eléri azt a pontot, amikor már szinte teljesen emberszerűnek tűnik, ez a hatás átfordul az ellenkezőjébe, és a befogadóból negatív érzelmi reakciót (megborzongást) vált ki. Mori – és az őt követő kutatók – szerint ez a jelenség nem csak a robotokra, hanem más, nem emberi szereplőkre, így a virtuális karakterekre is kiterjeszhető. Egy újabb kutatás azt mutatta ki, hogy amikor a realizmus foka eléri azt a pontot, amikor már tökéletesen emberszerűnek tűnik a karakter, a hozzá való viszonyulás újra pozitívvá válik.

ban érdemlik meg a jogi védelmet, hanem a mögöttük álló, megfigyelő vagy irányító, emberek védelme érdekében. A virtuális terek egyik jellemző vonása az anonimitás, vagyis az, hogy a többi résztvevő számára nem ismerhető meg, hogy ki vagy kik állnak egy-egy virtuális karakter mögött. Még kevésbé ismerhető meg az, hogy kik és hányan nézik passzívan az eseményeket, különböző követítő csatornák, vagy maga a virtuális tér adta lehetőségek útján. A hagyományos értelemben vett jogi védelem, illetve az alapjogi jogállás általában azt feltételezi, hogy konkrét, természetes személy szerepel azok alanyaként. A jogi védelem jogosultja, vagyis az, aki felléphet a jogsértéssel szemben, szükség-szerűen természetes személy (esetleg jogi személy) lehet. A virtuális térben a digitális karakterek között vagy velük szemben elkövetett jogsértő magatartások esetében éppen ez, a sérelmet szenvedő természetes személy hiányzik az egyenletből. Megítélésem szerint éppen azokban az esetekben van szükség a virtuális személyiségek jogi védelmére, amikor nem tudjuk beazonosítani, hogy konkrétan ki az a természetes személy, aki az avatar mögött áll.

A következőkben néhány alapjogi területre fókuszálva vizsgáljuk meg, hogy alkalmazható-e a jogi (és különösen az alapjogi) védelem a digitális személyiségek esetében.

III. VANNAK-E SZEMÉLYES ADATAI AZ AVATARNAK?

Az Európai Unió területén érvényben lévő adatvédelmi szabályozás egyértelmű a tekintetben, hogy személyes adatnak kizárólag a természetes személyekkel kapcsolatba hozható adatok tekinthetők. Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (Általános adatvédelmi rendelet – GDPR) a személyes adat fogalmát úgy határozza meg, hogy személyes adat az azonosított vagy azonosítható természetes személyre („érintett”) vonatkozó bármely információ (GDPR 4. cikk 1. pontja). Az információs társadalom kihívásaira reagálva az európai jogalkotó az azonosíthatóság fogalmát a lehető legtágabban értelmezi: azonosítható az a természetes személy, aki közvetlen vagy közvetett módon, különösen valamely azonosító, például név, szám, helymeghatározó adat, online azonosító vagy a természetes személy testi, fiziológiai, genetikai, szellemi, gazdasági, kulturális vagy szociális azonosságára vonatkozó egy vagy több tényező alapján azonosítható (GDPR 4. cikk 1. pontja). Az azonosíthatóság ráadásul nem abszolút kategória, hanem kontextusfüggő. Elképzelhető, hogy bizonyos környezetben az érintett nem tekinthető azonosíthatónak, azonban más személyi körben vagy helyzetben azonosíthatóvá válik. Így a virtuális tér szereplőire

vonatkozó adatok kapcsán is elképzelhető olyan helyzet, amikor azok a mögöttes álló felhasználóval kapcsolatba hozhatók, míg más esetekben ezek az információk pusztán metaadatként, illetve nem-személyes adatként kezelhetők. Pusztán az, hogy nem a teljes nevünk megadásával lépünk be a virtuális térbe, nem szünteti meg az ott végzett tevékenységek során gyűjtött információk személyes adat-jellegét. Már az, hogy a virtuális teret működtető szolgáltató információt kap az általunk használt számítástechnikai eszköz IP-címéről, már megalapozhatja a személyes adatként való kezelést. Az elmúlt években többször felmerült a kérdés, hogy az internetre csatlakozó számítástechnikai eszközök azonosítására szolgáló IP-címek személyes adatnak minősülnek-e, különös tekintettel arra, hogy a dinamikus IP-címeket az internet szolgáltató meghatározott időközönként újraosztja az előfizetők között, és bizonyos esetekben egy IP-címhez tartozhat több informatikai eszköz is. 2016-ban az Európai Unió Bírósága ítéletében (C-582/14. sz. ügy) is rögzítette, hogy bizonyos körülmények fennállása esetén a dinamikus IP-cím is személyes adatnak minősül. Ezt támasztja alá a GDPR 30. preambulumbekzdése is, amely alapján „természetes személyek összefüggésbe hozhatók az általuk használt készülékek, alkalmazások, eszközök és protokollok által rendelkezésre bocsátott online azonosítókkal, például IP-címekkel és cookie-azonosítókkal, valamint egyéb azonosítókkal, például rádiófrekvenciás azonosító címkékkel. Ezáltal olyan nyomok keletkezhetnek, amelyek egyedi azonosítókkal és a szerverek által fogadott egyéb információkkal összekapcsolva felhasználhatóak a természetes személyes profiljának létrehozására és az adott személy azonosítására” (GDPR 30. preambulumi bekezdés).

Nem mehetünk el amellett sem, hogy a személyes adatok és a nem-személyes adatok közötti határvonal folyamatos változásban van, ahogyan arra a szakirodalomban is rendre rámutatnak (Janecek, 2018). A jelenleg nem személyes adatként kezelt információkból a jövőben (a technológiai fejlődés hatására) személyes adat válhat, hiszen folyamatosan új és új felhasználási és elemzési módszerek jelennek meg, egyre több adatbázis válik egyúttal összekapcsolhatóvá (Wendehorst, 2017). Ennek fényében kell megvizsgálnunk a virtuális terekben és a metaverzumban keletkező adatok felett gyakorolt döntési jogokat és az adaton fennálló potenciális tulajdonjog kérdését is.

Biztosan védelem illeti meg a virtuális karakter útján a természetes személy felhasználóról megismerhető adatokat, illetve az ezekből levonható következtetést. Előfordulhat például, hogy egy felhasználó kizárólag a virtuális térben éli ki a társadalomban esetleg megütközést kiváltó, ellentmondásos vagy tiltott vágyait, késztetéseit. Ilyenkor a virtuális térben tett magatartása visszautal a személyiségének elnyomott vagy elrejtett jegyeire, amelyek a GDPR szerint

akár különleges adatnak is minősülhetnek (például káros szenvedélyek, szexuális preferenciák vagy irányultságok esetében). Ha a virtuális térre, mint egyfajta „védett játszótérre” tekintünk, akkor a felhasználónak jogos elvárása lehet, hogy a játékbeli viselkedését tartsa titokban a szolgáltató, abból a valós személyre ne vonjon le következtetéseket, azokat ne használja fel kereskedelmi (például hirdetéscélzás) célokra. Könnyen előfordulhat ugyanis olyan helyzet, hogy a játék látszólag védett terében a felhasználó olyasmit is megmutat magából, amit a környezete vagy a társadalom előtt titokban szeretne tartani (és itt most nem illegális tevékenységekre gondolunk). Komoly jogsérelmet okozhat ezért neki az, ha másnap ehhez a magatartásához vagy preferenciájához kapcsolódó hirdetéseket, termék és szolgáltatás ajánlatokat kap. Különösen aggályos lehet, ha a virtuális teret működtető szolgáltató továbbadja ezeket az információkat harmadik felek irányába, és így a felhasználó virtuális térben tanúsított viselkedése a róla létrehozott kereskedelmi személyiségprofil részévé válik. Érdemes kitérni arra is, hogy a felhasználó a virtuális térben nem biztos, hogy a valós személyiségének megfelelően viselkedik, hanem esetleg egy vágyott karaktert vagy viselkedést próbál ki „tét nélkül”, hogy megfigyelje a játékbeli környezete reakcióit. Elképzelhető az is, hogy a felhasználó egy kitalált karaktert játszik, és a szerepnek megfelelő viselkedést mutatja. Ilyenkor a magatartásból levont következtetés vagy megállapított személyiségjegy szükségszerűen nem lesz pontos és helytálló. Ezzel a GDPR-ben foglalt pontosság elve [GDPR 5. cikk (1) bekezdés d) pont] is sérülhet. A tévesen levont következtetések és a felhasználó ennek megfelelő kezelése a jövőben sértheti a személy társadalmi megítélését, jó hírét, és akár (bizonyos vallási vagy politikai közegben) valós veszélynek, megtorlásnak vagy üldöztetésnek is kiteheti őt.

Úgy gondoljuk tehát, hogy a személyes adatok kezelése szempontjából a helyes magatartás a virtuális tér (a metaverzum) és a valós világ egymástól való elválasztása. Az egyik helyen rögzített információk és megállapított következtetések nem szivároghatnak át a másik térbe, mivel ezáltal a felhasználó magánszférája és a GDPR adatpontosságra vonatkozó alapelve is sérülhet. Ez a nézet természetesen nehezen vihető keresztül a valóságban, hiszen a szolgáltatók elemi érdeke a felhasználókról minél több adat gyűjtése és azok kereskedelmi célú hasznosítása (Yu, Zhao, 2018). Ha azonban ez az elválasztás sikeres, úgy a felhasználónak két személyes tere keletkezik: virtuális és valóságos, mindkettőben eltérő jellegzetességeket mutathat magáról.

Összességében megállapítható tehát, hogy a jelenlegi szabályozás szerint a virtuális térben létrehozott avatart önmagában nem illeti meg a személyes adatok védelméhez való jog, hiszen az nem minősül természetes személynek.

A virtuális karakter mögött álló valóságos személyt azonban megilleti az a jog, hogy a karakterról gyűjtött „személyes adatokat” ne vonatkoztassák közvetve vagy közvetlenül rá. Ez azonban a maga teljességében akkor érhető el, ha a digitális avatar adatvédelemhez való jogát is megteremtjük és érvényesítjük, máskülönben a gyűjtött adatok egy „légüres térben” léteznek, és csak akkor nyerik el személyesadat-jellegüket, ha – esetleg egy későbbi technikai fejlesztési lépés hatására – összekapcsolhatóvá válnak a konkrét felhasználóval.

IV. A MAGÁNSZFÉRA VÉDELME NEK EGYÉB ASPEKTUSAI

A tágabb értelemben vett magánszféra védelmébe beletartozik a magánlakás és a magán kommunikáció (levéltitok) védelme is. Nagyon sokáig a közvélekedés az volt, hogy a virtuális terekben, illetve a játékkonzolokon a videójátékokon belül váltott üzenetek titkos csatornán történő küldése miatt a játékosok kommunikációja nem csak a nyilvánosság, hanem a hatóságok előtt is rejtve marad. A 2015-ben száznál több ember életét követelő párizsi merényletsorozatot is állítólag (bár nem bizonyítottan) a PlayStation hálózatának felhasználásával tervezték ki (Tassi, 2015). Ebben teremtett új helyzetet a 2018-as év, amikor a Sony először osztott meg a PlayStation Networkön belüli kommunikációs adatokat bűnüldöző szervekkel. Egy, feltehetően terrorista szervezethez való csatlakozás érdekében külföldre utazott személy, Isse Aweis Mohamud üzenetváltásaihoz kért hozzáférést az FBI a cégtől, szabályos, bírói házkutatási engedély^[12] birtokában. A Sony több hónappal késedelemmel ugyan, de teljesítette a kérést, és a nyomozó hatóságnak átadta a *DejanWoW* felhasználónévhez tartozó adatokat, így többek között a játékkonzol használatára, a rendszeren belüli vásárlásokra vonatkozó információkat, valamint a felhasználó PS3 és PS4 rendszer útján váltott üzeneteit. Az eset abban jelent újdonságot, hogy ezúttal a hatóságok nem magát a játékkonzolt vetették alá vizsgálatnak (Mohamud ugyanis erről mindent gondosan eltávolított), hanem az FBI a rendszert üzemeltető szolgáltatótól kért és kapott hozzáférést a felhasználói fiók kapcsán általuk tárolt információkhoz. Ebből az esetből az derül ki számunkra, hogy a felhasználók üzenetváltásai a hardver teljes törlése vagy megsemmisítése után is a rendszert üzemeltető szolgáltató birtokában van-

[12] A 17-SW-00121-SWH számú ügy dokumentumai, elérhetőség: <https://www.documentcloud.org/documents/4378583-Sony-Data-Request-From-Law-Enforcement-Terrorist.html> (letöltés dátuma: 2023. szeptember 1.)

nak, és ezeket később is vissza lehet keresni, tartalmukat megismerni. Az adatok megőrzéséről a megvizsgált szolgáltatók adatkezelési szabályzatai kivétel nélkül rendelkeztek a jogi igények érvényesítése, illetve a jogellenes tartalmak elleni küzdelem érdekében. A jogellenes tartalmak elleni küzdelem mellett a szolgáltatók egyéb, technikai célokra is használják a felhasználók közötti üzenetek tartalmát, amiről tájékoztatják is a felhasználókat: az adatkezelés feltehető célja egyrészt a csalások és a felhasználási feltételek megszegése elleni küzdelem, ezáltal pedig a társaság érdekeinek védelme, emellett ezek az adatok üzletszerzési és szolgáltatásfejlesztési célokat is szolgálhatnak. Másrészt az adatkezelés célja lehet az is, hogy konkrét jogsértések esetén biztosítsák a jogsértést bizonyító információkat a hatóságok számára. Végezetül a csevegési adatok kezelése és megőrzése szükséges lehet bizonyos szolgáltatások biztosítása érdekében (G. Karácsony, 2021).

V. ÉRHETI-E DISZKRIMINÁCIÓ AZ AVATART VALAMELY TULAJDONSÁGA ALAPJÁN?

Végezetül érdemes kitérni a hátrányos megkülönböztetés (diszkrimináció) egyes magvalósulási formáira, mint a virtuális térben létező szereplőkkel szembeni jogsértések elképzelhető magvalósulási formáira. Annak érdekében, hogy megállapíthassuk, reális veszélyt jelent-e ez a metaverzum szereplőire, a diszkriminációs teszt első két lépését kell végrehajtanunk: annak vizsgálatát, hogy bekövetkezett-e hátrány, valamint annak megítélését, hogy az a védett tulajdonságon alapult-e.

A hátrányos megkülönböztetés azért különösen izgalmas a virtuális tér kapcsán, mivel ebben az esetben szinte teljesen biztos, hogy a megkülönböztetés alapja (a magyar jogi szóhasználatnál: védett tulajdonság) nem az érintett avatar mögött álló valós személy valamely jellemzője, tulajdonsága vagy állapota. A szóba jöhető tulajdonságokat általában maga az avatar hordozza, azonban azok egyáltalán nem biztos, hogy megfeleltethetők a mögötte álló felhasználó tulajdonságainak. Gyakran előfordul, hogy férfiak nőnemű virtuális karaktert választanak maguknak, vagy éppen fordítva. Ugyanígy általában testre szabható a karakter bőrszíne, testalkata, hajviselete, és más jellemzői is (egyes virtuális világokban különböző fajokhoz tartozó lényeket is irányíthatunk). Ezek a jellemzők azonban eltérhetnek a karaktert irányító felhasználó jellegzetességeitől, de meg is egyezhetnek velük: létrehozható saját magunk megszólalásig hasonlító virtuális mása is. A fontos jellegzetessége ezeknek

a helyzeteknek azonban az, hogy az avatarral interakcióba lépő többi felhasználó csak az avatar külső jegyeit látja, a mögötte álló embert nem. Így, ha hátrányos elbánásban részesíti, az csak a virtuális karakter jellemzőin alapulhat, ami néhány esetben megegyezhet a tényleges felhasználó jellemzőivel, ebben azonban a vele szemben álló fél nem lehet biztos. A védett tulajdonságok tehát léteznek, ugyanúgy, mint a valódi világban, azonban esetünkben a védett tulajdonságot nem maga a felhasználó, hanem az avatar hordozza. Így az emberi személy tulajdonságai alapján történő diszkriminációra nem kerülhet sor, azonban a nemzetközi jogirodalom és ítélkezési gyakorlat útján kimunkált társítás (asszociáció) útján megvalósuló hátrányos megkülönböztetés megállapítható lehet. Ez a fogalom azokat az eseteket takarja, amikor egy személy egy vele összefüggésbe hozható másik személy bizonyos jellemzői miatt szenved el hátrányos bánásmódot. Ilyen eset például, amikor az álláskeresés során hátrány éri a tartós betegséggel élő gyermeket nevelő anyát. Ebben az esetben a gyermek betegsége a diszkrimináció alapja, azonban a hátrány nem a gyermeket, hanem a szülőt éri. A virtuális karakter bizonyos jellemzői miatti hátrányos megkülönböztetés tehát ugyanígy jelentkezhethet a karaktert irányító személynél annak ellenére, hogy a védett tulajdonságot nem (biztosan) ő hordozza.

A megvalósuló kedvezőtlen bánásmód, illetve hátrányos megkülönböztetés is számos formát ölthet a virtuális térben. Egyrészt elképzelhető az, hogy a védett tulajdonsága miatt nem engedik a digitális szereplőt belépni bizonyos helyekre, illetve kizárják valamilyen aktivitásból, amiben a felhasználó szeretne részt venni. Ezek egyelőre nem rendelkeznek nagy mennyiségű dokumentált esettel, illetve felhasználói beszámolóval. E tanulmány megjelenése időszakában azonban nagy port kavart egy eset, ami hagyományosan a diszkrimináció esetkörébe sorolható magatartást érintett: egy felhasználó arról számolt be, hogy avatarja szexuális zaklatás, illetve nemi erőszak áldozata lett.^[13] Elmondása szerint egy másik felhasználó követte, és rendszeresen belépett a karakter személyes terébe, miután mások biztatására a hátrányt szenvedett fél kapcsolta az ilyen esetek megelőzésére szolgáló biztonsági zónát a karaktere körül. Ezt követően, más felhasználók szeme láttára szexuális cselekménybe kezdett a virtuális avatarjával, az ő beleegyezése nélkül. Az érintéseket, simogatásokat a sérelmet szenvedett nő VR-készüléke rezgések formájában közvetítette is felé. Ebben az esetben a zaklatás alapja a virtuális karakter női neme

[13] Az eset a Meta Horizon's World nevű virtuális térben történt. <https://www.independent.co.uk/tech/rape-metaverse-woman-oculus-facebook-b2090491.html> (letöltés dátuma: 2023. szeptember 1.)

volt, mint védett tulajdonság. Ez a példa rávilágít arra, hogy ténylegesen létező jelenség a virtuális karakterek zaklatása, és a karaktert irányító felhasználó ezt ugyanolyan intenzív támadásként élte meg, mintha az vele történt volna a való életben.

Összefoglalóan elmondható, hogy a metaverzum digitális karakterei bizonyos szempontból ugyanolyan kiszolgáltatottak a személyiségi jogaikat érő negatív hatásoknak, mint az őket mozgató felhasználók a való életben. Az avatarokat érő joghátrányok néha közvetlenül (például a felépíthető személyiségprofil útján), néha pedig közvetetten, lélektani hatásuk útján jelentenek veszélyt a mögöttük álló emberi felhasználókra. A virtuális világok avatarjaival való tisztességes bánásmódot egyelőre nem írja elő kötelező szabály, hanem csak néhány helyen támasztja meg az adott felület felhasználói feltételrendszere és közösségi irányelvei. Úgy találjuk, hogy az alapjogi védelem – ahogyan az a nagy nemzetközi egyezményekben le van fektetve – megfelelő iránymutatást adhatna a metaverzum viselkedési normái kapcsán is.

IRODALOM

1. Ashrafian, Hutan: AlonAI: A Humanitarian Law of Artificial Intelligence and Robotics. *Science & Engineering Ethics* (2015) 21:29–40. DOI: 10.1007/s11948-013-9513-9
2. Broadbent, Elizabeth: Interactions With Robots: The Truths We Reveal About Ourselves. *Annual Review of Psychology*, 2017 / 68. 634-635.
3. Coanda, Iulia - Aupers, Stef: Post-human encounters: Humanising the technological Other in videogames. *New Media & Society*, 23, 5, (2021) <https://doi.org/10.1177/1461444820912388>
4. G. Karácsony Gergely: A videójátékok adatkezelési gyakorlata: kommunikáció és profilalkotás. In: G. Karácsony Gergely (szerk.): *A videójátékok jogi kérdései*. Széchenyi István Egyetem, Győr, 2021.
5. Guihot, Michael, Matthew, Anne F, Suzor, Nicolas P, *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 *Vanderbilt Journal of Entertainment and Technology Law* 385 (2020) Elérhető: <https://scholarship.law.vanderbilt.edu/jetlaw/vol20/iss2/2> (letöltés: 2023. szeptember 1.)
6. Horváth Evelin: Pixelekbe öntött érzelmek – A virtuális érzelemmegjelenítés vizsgálatának lehetőségei. *Századvég* 2022. 1. szám. 137.o.
7. Janeček, Václav: Ownership of personal data in the Internet of Things. *Computer law & Security review* 34 (2018) 1039–1052. 1040.o. <https://doi.org/10.1016/j.clsr.2018.04.007>

8. Keserű Barna Arnold: A mesterséges intelligencia magánjogi mibenlétéről. In: Lé-vayné, Fazekas Judit; Kecskés, Gábor (szerk.) Az autonóm járművek és intelligens rendszerek jogi vonatkozásai. Universitas-Győr Nonprofit Kft., Győr, 2020.
9. Király Péter Bálint: A videojátékok monetizációjának jogi kérdései. In: G. Kará-csony Gergely (szerk.): A videojátékok jogi kérdései. Széchenyi István Egyetem, Győr, 2021.
10. Papakonstantinou, Vagelis– De Hert, Paul: Structuring modern life running on software. Recognizing (some) computer programs as new “digital persons”. *Computer Law & Security Review* 34 (2018) 732–738. <https://doi.org/10.1016/j.clsr.2018.05.032>
11. Villaronga, Eduard Fosch– Golia, Angelo Jr: Robots, standards and the law: Rival-ries between private standards and public policymaking for robot governance. *Computer Law & Security Review* 35 (2019) 129–144. <https://doi.org/10.1016/j.clsr.2018.12.009>
12. Wendehorst, Christiane: Of Elephants in the Room and Paper Tigers: How to Re-concile Data Protection and the Data Economy, In: Lohsse/Schulze/Staudenmayer (Eds), *Trading Data in the Digital Economy: Legal Concepts and Tools: Münster Colloquia on EU Law and the Digital Economy III* (2017) 327 - 356.
13. Yu, Xiaolan– Zhao, Yun: Dualism in data protection: Balancing the right to perso-nal data and the data property right. *Computer Law & Security Review* Vol. 35, Issue 5. (2019) <https://doi.org/10.1016/j.clsr.2019.04.001>

ONLINE FORRÁSOK:

1. Code of Conduct for Virtual Experiences. <https://www.meta.com/help/quest/articles/accounts/privacy-information-and-settings/code-of-conduct-for-virtual-experiences/>
2. Woman says she was virtually ‘raped’ in the metaverse while others ‘passed around a bottle of vodka’ (The Independent online, 2022.) <https://www.independent.co.uk/tech/rape-metaverse-woman-oculus-facebook-b2090491.html>
3. A 17-SW-00121-SWH számú ügy dokumentumai. <https://www.documentcloud.org/documents/4378583-Sony-Data-Request-From-Law-Enforcement-Terrorist.html>
4. <https://www.documentcloud.org/documents/4378582-Sony-PlayStation-PS4-Data-Handed-Over-in-Terror.html>

KECSKÉS GÁBOR

A MESTERSÉGES INTELLIGENCIA
AZ ENSZ FENNTARTHATÓ
FEJLŐDÉSI CÉLJAI SZOLGÁLATÁBAN

A mesterséges intelligencia az ENSZ Fenntartható Fejlődési Céljai szolgálatában

VI. A FENNTARTHATÓSÁGRÓL ÉS AZ ENSZ FENNTARTHATÓSÁGI CÉLJAIRÓL

A fenntartható fejlődés fogalma az elmúlt néhány évtizedben a nemzetközi jogban (Hárs 2021) és a nemzeti jogokban egyaránt egyre jelentősebb szerepet kapott, olyanannyira, hogy már-már a fogalom „túlhasználatának”, inflálásának is tanúi lehetünk. Ahogyan az ENSZ Környezet és Fejlődés Világbizottsága eredendően az 1987-es Közös jövőnk c. jelentéséből megismerhetjük, a fenntartható fejlődés egy olyan (gazdasági, társadalmi) fejlődés, amely *„kielégíti a jelen szükségleteit anélkül, hogy csökkentené a jövőendő generációk képességét, hogy kielégítsék a saját szükségleteiket.”*

A világszervezet Közös jövőnk című jelentése azonban jogilag nem kötelező erejű, viszont az egészen bizonyosan állítható, hogy az azóta elfogadott számos jogalkotási aktus eme definíciót veszi alapul. Ezután a „fenntartható fejlődés” absztrakt kifejezés sokrétű és új értelmezési szinteket is nyert a globális államközi világtalálkozásokon, szakpolitikai dokumentumokban és jogi dokumentumokban egyaránt (Schrijver, 2008; Crowther, Seifi, Moyeen, 2018; Purvis, Mao, Robinson, 2019). 1992-ben a második környezetvédelmi világkonferencián az államok többsége által elfogadott *Riói Nyilatkozat a Környezetről és a Fejlődésről* a fenntartható fejlődés elvén alapuló, a gazdasági és társadalmi fejlődéssel összhangban megvalósítandó környezetvédelmi és gazdasági, társadalmi célkitűzéseket deklarálta. 10 évvel később 2002-ben a fenntartható fejlődésről szóló johannesburgi világkonferencián az államok kijelentették, hogy a fenntartható fejlődés megvalósítása minden kultúra és civilizáció számára fontos és közös kihívás. Ennek az elkötelezettségnek az eredményeképpen számos nemzetközi szervezet és az államok többsége ezt követően a fenntartható fejlődést célként és stratégiaként beépítette jogszabályaiba - számos állam alkotmányos szinten, illetve a nemzetközi közösség normatív dokumentumokban, multilaterális, regionális és szubregionális szinten is (Schrijver, 2008). Ma már alig van olyan nemzetközi környezetvédelmi szerződés vagy nemzeti környezeti jogszabály, amely ne említene a fenntartható fejlődést általános célként vagy alapvető szektorális elvként.

Mindeközben a fenntartható fejlődés fogalma az elmúlt évtizedekben egyfajta – egyre bővülő jelentéstartalmú – varázsszóvá vált, noha a mögötte lévő koncepció az emberiség előtt álló globális kihívások kezelésére feltétlenül alkalmas lehet. Ma már minden kétséget kizáróan állítható, hogy a fenntartható fejlődés három pilléren nyugszik, mivel a fogalomnak ma már i) társadalmi, ii) gazdasági és iii) környezeti (ökológiai) rétege is kialakult, azaz a fenntartható fejlődés eredeti ökológiai szemlélete mellett egyenrangúként jelent meg a társadalmi és gazdasági vetület is, és így együttesen értelmezhető a fenntartható fejlődés fogalma. Ezáltal a három pillér összekapcsolódása és holisztikus megközelítése mára már megkérdőjelezhetetlenné vált.

VII. FENNTARTHATÓ FEJLŐDÉS ÉS FENNTARTHATÓ FEJLŐDÉSI CÉLOK

Tizenhárom évvel a 'Közös jövőnk' címet viselő jelentés után a 2000. évre tett millenniumi kötelezettségvállalásként 189 ENSZ-tagállam elfogadta a nem kötelező érvényű Millenniumi Fejlesztési Célokat (*Millennium Development Goals*) egy 15 évre szóló (2000-2015) általános politikai célkitűzési rendszer és 8 kiemelt cél formájában. Mivel a Millenniumi Fejlesztési Célok végrehajtása elmaradt a várttól, és számos részsiker elleneére globális megvalósítása akár kudarcnak is tekinthető (főleg az ökológiai vonatkozások terén), 2015-ben az ENSZ összes tagállama (193 tagállam) elfogadta a Fenntartható Fejlődési Célokat (*Sustainable Development Goals*) egy újabb másfél évtizedes periódusra (2015-2030). Az ENSZ-tagállamok a 2030-ig elérendő fejlődési célokat globálisan alkalmazandó soft lawként, azaz jogilag nem kikényszeríthető, szankcióval nem rendelkező, általános politikai és szakpolitikai alapú kötelezettségvállalások formájában (Biermann et al, 2022), az ENSZ Közgyűlésének határozatában (*Transforming Our World*, 2015) határozták meg. A Fenntartható Fejlődé-

si Célok 17 fő céllal^[14], ezekhez kapcsolódóan jelentős számú, 169 részcéllal és 232 kapcsolódó mutatóval (indikátor) absztrakt, holisztikus elköteleződések, és olyan általános globális kötelezettségvállalások összességének tekinthetők, amelyeket 15 év alatt valahogyan el kell érni, vagy legalábbis ezek terén pozitív változásokat kell kieszközölni. Tehát ezeket a célokat legfeljebb puha jogi mechanizmusnak (soft law) lehet nevezni, ami azt jelenti, hogy a célok pontos jelentése, valamint a végrehajtható jogi kötelezettségek és szankciók egyértelműen hiányoznak a célok szövegéből és az indikátorok felsorolásából is. Következésképpen eme célok nem jogi normák, viszont uralják a globális szereplők (ideértve a regionális nemzetközi szervezeteket, államokat is) fenntarthatósági diskurzusát és kötelezettségvállalásait. E nehézségek következtében az is nyilvánvaló, hogy számos nehézséget eredményez i) a fogalom jogi jelentésének és tartalmának eltérő megkülönböztetése az államok szintjén, ahogyan ii) az államok tényleges és jelentősen széttartó gyakorlatának vizsgálata és iii) az emögött meghúzódó etikai dimenziók (Viñuales, 2021) eltérő interpretálásából fakadó különbségek is problematikussá teszik azt, hogy a Fenntartható Fejlődési Célokról mint univerzális, valamennyi államban azonosan értelmezhető standardokról beszéljünk. A célrendszer meglehetősen általános és absztrakt felvázolása természetesen annak tudható be, hogy ez volt a legtöbb, amiben az ENSZ tagállamok között általános egyetértés volt, hiszen valamennyien sajátos ökológiai, társadalmi és gazdasági kihívásokkal küzdenek, amelyek közös nevezőre hozása rendkívül nehéz, majdhogynem lehetetlen feladatnak tűnhet.

[14] Ezek felsorolása a következő:

1. A szegénység megszüntetése
2. Az éhezés megszüntetése.
3. Egészség és jóllét
4. Minőségi oktatás
5. Nemek közötti egyenlőség
6. Tiszta víz és közegészségügy
7. Megfizethető és tiszta energia
8. Tisztességes munka és gazdasági növekedés
9. Ipar, innováció és infrastruktúra
10. Egyenlőtlenség csökkentése.
11. Fenntartható városok és közösségek
12. Felelős fogyasztás és termelés
13. Fellépés az éghajlatváltozás ellen
14. Óceánok és tengerek védelme
15. Szárazföldi ökoszisztémák védelme
16. Béke, igazság és erős intézmények
17. Partnerség a célok eléréséért.

Tehát a Fenntartható Fejlődési Célok jogilag nem kötelező érvényűek, azonban a célok elérése mégis globális prioritás. Ugyanakkor az államok nemzeti jelentései nem túl meggyőzőek a célok részleges megvalósításáról vagy az ennek érdekében hozott hatékony intézkedésekről, ezért egyes ENSZ-testületek segíteni kívánják a tagállamokat a végrehajtásban. Először is, az államok 2002-ben, a johannesburgi csúcstalálkozón, évekkal a Célok elfogadása előtt elfogadták a fenntartható fejlődésről szóló világcsúcs végrehajtási tervét (*Johannesburg Declaration on Sustainable Development and Plan of Implementation*). A végrehajtási terv sürgette az államokat, hogy fokozzák a fenntarthatóság három pillérének a közös politikákba és integrált megközelítésekbe való beépítését, azonban a végrehajtás módszereit illetően szándékoltan homályos maradt. Különös hangsúlyt fektetett azonban a (gazdaságilag és ökológiailag, tehát társadalmilag is) legsebezhetőbb régiókra, így a fejlődő kis szigetországokra és Afrikára. Másodszor, az államok éppen a Fenntartható Fejlődési Célok elfogadásának évében, 2015-ben fogalmazták meg az ún. Addis Ababai Cselekvési Akciót (*Addis Ababa Action Agenda*). A Cselekvési Akció egyértelműen felölelte a Fenntartható Fejlődési Célok végrehajtásával kapcsolatos főbb cselekvési területeket, nevezetesen a finanszírozás esetét, a nemzetközi fejlesztési együttműködést és a nemzetközi kereskedelmet, végül pedig a tudomány, a technológia és az innováció szerepét.

A Fenntartható Fejlődési Célok végrehajtásának nyomon követése azonban számos lépést foglal magában, így például a nemzeti jelentések, az államok önkéntes értékelései és a felülvizsgálati folyamat biztosít a tagállami végrehajtások felett bizonyos fokú ellenőrzési lehetőséget, amely azonban korántsem sem jellemezhető erősnek és hatékornak. Szem előtt tartva, hogy sok országban jelentős hiányosságok vannak a végrehajtásban, az ENSZ Közgyűlése elfogadta az önkéntes nemzeti felülvizsgálatok megerősítéséről is szóló 77/283. számú határozatot. A határozat célja az volt, hogy az ENSZ-tagállamokat arra ösztönözze, hogy folyamatosan kísérjék figyelemmel, illetve folytassák le a globális, regionális, országos vagy országon belüli vizsgálatokat előrehaladásról, valamint a felülvizsgálatokból következően a döntéshozatalban használják fel a szerzett tapasztalatokat, és kérjék fel az ENSZ megfelelő testületeit, hogy a tagállamok kérésére nyújtsanak támogatást a Fenntartható Fejlődési Célok végrehajtásának értékelésére irányuló erőfeszítéseikhez.

A Fenntartható Fejlődési Célok 2030-ig tartó megvalósítása, egyáltalán fejlődés elérése tehát valamennyi ENSZ-tagállam közös feladata, s ennek szolgálatába állítják a legújabb technikai (és társadalmi) vívmányokat is.

VIII. A MESTERSÉGES INTELLIGENCIA ÉS A FENNTARTHATÓ FEJLŐDÉSI CÉLOK MEGVALÓSÍTÁSA

A mesterséges intelligencia kínálta előnyök tehát a Fenntartható Fejlődési Célok elérését is előmozdítani hivatottak, egyidejűleg azonban számos veszélyt is jelenthetnek eme Célok elérésére. Fontos leszögezni, hogy külön a mesterséges intelligencia szerepéről nem szólnak a célok, azonban egyrészt a technológiai vetületekbe feltétlenül beleértendő (és ezekről bőven esik szó), másrészt a mesterséges intelligencia alkalmazása a legtöbb célt hatékony elérését feltétlenül szolgálja. Viszont a mesterséges intelligencia felelős alkalmazása (és a humán kontroll lehetősége) ugyanakkor kulcsfontosságú, jelen tanulmány kiindulópontja az, hogy valamennyi MI-alkalmazás a visszajára tud fordulni, és a Fejlődési Célokat hátrányosan tudja érinteni, amennyiben a megfelelő garanciákat nem teremti meg az államok jogalkotás útján vagy a gyakorlati megvalósítás során nem eszerint járnak el az érdekelt szereplők. E körbe sorolható például az, ha a mesterséges intelligencia diszkriminál az emberek között, a személyes adatokkal súlyosan visszaél, vagy amennyiben a technikai vívmány a környezeti érdekek, valamint a társadalmi és gazdasági rend számára káros döntéseket hoz, és ezeket meg is valósítja a gyakorlatban, így adott esetben egy adatalapú elemzés következtében hozott döntéssel elmentésen jár el „rosszhiszeműen”, akár környezeti pusztulást okozva.

Mindenekelőtt azonban definiálni szükséges a mesterséges intelligencia fogalmát, amelyet viszont megnehezít, hogy a definíciót illetően egyelőre nem létezik egy általánosan és globálisan kötelező jelleggel elfogadott, univerzális és normatív fogalma.

A globális nemzetközi szervezet, az ENSZ szakosított intézménye, az UNESCO 2021-es, jogilag nem kötelező, nevesítetten ajánlás-jellegű anyagában (*Recommendation on the Ethics of Artificial Intelligence*) egy általánosabb definíciót is alkotott, amely szerint a mesterséges intelligencia-rendszerek „*olyan információfeldolgozó technológiák, amelyek olyan modelleket és algoritmusokat integrálnak, amelyek tanulási és kognitív feladatok elvégzésére képesek, amelyek olyan eredményekhez vezetnek, mint az előrejelzés és a döntéshozatal anyagi és virtuális környezetben. A mesterséges intelligencia rendszereket úgy tervezték, hogy a tudás modellezése és reprezentációja, valamint az adatok kiaknázása és az összefüggések kiszámítása révén különböző fokú autonómiával működjenek. A mesterséges intelligencia rendszerek többféle módszert is alkalmazhatnak, például, de nem kizárólagosan: (i) gépi tanulás, beleértve a mélytanulást és a meg-*

erősítő tanulást; (ii) gépi következtetés, beleértve a tervezést, ütemezést, tudás-reprezentációt és következtetést, keresést és optimalizálást.”

Továbbá a nem univerzális nemzetközi szervezetek köréből az Európai Unió és az Európa Tanács mint regionális szervezetek készülő jogszabálytervezetei adhatnak támpontot, hiszen ezekben is kulcsszerepet kap a definíció.

Az Európai Uniónak a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról és egyes uniós jogalkotási aktusok módosításáról szóló rendelete szerint a *„mesterségesintelligencia-rendszer (MI-rendszer): olyan szoftver, amelyet az I. mellékletben felsorolt technikák és megközelítések közül egy vagy több alkalmazásával fejlesztettek, és amely az ember által meghatározott célkitűzések adott csoportja tekintetében olyan kimeneteket, például tartalmat, előrejelzéseket, ajánlásokat vagy döntéseket képes generálni, amelyek befolyásolják azt a környezetet, amellyel kölcsönhatásba lépnek.”* A definícióban említett I. mellékletben felsorolt technikák és megközelítések pedig a következők: i) gépi tanulási megközelítések, ideértve a felügyelt, a felügyelet nélküli és a megerősítő tanulást, a módszerek széles skálájának, többek között a mélytanulásnak az alkalmazásával; ii) logikai és tudásalapú megközelítések, beleértve a tudás megjelenítését, az induktív (logikai) programozást, a tudásbázisokat, a következtetőmotorokat, a(z) (szimbolikus) érvelést és a szakértői rendszereket és iii) statisztikai megközelítések, Bayes-féle becslés, keresési és optimalizálási módszerek.

Az EU definíciója mellett még általánosabb az Európa Tanács fogalom-tervezete, amely a 2023 júliusában publikált, még nem kötelező erejű, egyelőre csak tervezetként létező keretegyezmény (A mesterséges intelligenciáról, emberi jogokról, demokráciáról és jogállamról szóló keretegyezmény-tervezet, Consolidated Working Draft of the Framework Convention) szövegének 3. cikkében jelent meg. Eszerint a mesterséges intelligencia-rendszer *„bármely algoritmikus rendszer vagy ilyen rendszerek kombinációja, amely statisztikából vagy más matematikai technikákból származó számítási módszereket használ, és amely szöveget, hangot, képet vagy más tartalmat generál, vagy segíti vagy helyettesíti az emberi döntéshozatalt”* (saját fordítás). Maga az idézett 3. cikk még egyébiránt – gondolva a felgyorsult technológiai fejlődésre – egy evolutív értelmezést is lehetővé tesz, mivel a szöveg szerint a részes felek konferenciája (azaz a szerződő államok) *„dönthet úgy, hogy ezt a fogalommeghatározást a vonatkozó technológiai fejlődéssel összhangban értelmezi.”*

Meglátásom szerint mindhárom eddig tárgyalt definíció megfeleltethető egymásnak, noha a szövegezésekben jelentős eltérések mutatkoznak. Ez jól mutatja, hogy egy globális, 193 állam érdekeit képviselő szervezet által alkal-

mazott munkadefiníció sem tér el koncepcionálisan két szűkebb körű, regionális államközi testület fogalmaitól, ezért eme definíciókat jelen tanulmány is elfogadja, és ezeket alapul véve vizsgálja a kérdéskört a továbbiakban. Eme tanulmány főként az UNESCO-definíció (amelyet az ENSZ mint globális anyaszervezet is elfogadni látszik) első, jelentőségteljes mondatát tekinti most irányadónak, azaz értelmezésben a mesterséges intelligencia-rendszerek *„olyan információfeldolgozó technológiák, amelyek olyan modelleket és algoritmusokat integrálnak, amelyek tanulási és kognitív feladatok elvégzésére képesek, amelyek olyan eredményekhez vezetnek, mint az előrejelzés és a döntéshozatal anyagi és virtuális környezetben.”* (Recommendation on the Ethics of Artificial Intelligence, 2021)

A szakirodalomból kölcsönvett munkadefiníciók alkalmazása is megoldás lehet, azonban ezek annyira sokrétűek, továbbá számottevően eltérő alapokon és megközelítéseken (jogi, technológiai, pszichológiai, etc.) nyugvó megfogalmazások, hogy ezeket jelen tanulmány nem kívánja részletesebben tanulmányozni. Mivel azonban a Fenntartható Fejlődési Célok csupán keretrendszerként és általános kötelezettségeket állapítanak meg, ezért jelen tanulmány nem azonosítja problémaként, ha az alapulfekvő mesterséges intelligencia-definíció hiányzik a kötelező erejű globálisan megalkotott jogi normákból.

Amint azt már a korábbiakban leszögezte a tanulmány, kifejezetten a mesterséges intelligenciát nem említi a Fenntartható Fejlődési Célok alapdokumentuma, a célokat és az ehhez kapcsolódó részcélokat ajánlás jellegű jogi dokumentumba foglaló ENSZ Közgyűlés 70/1 határozata. A technológiára és egyéb innovációkra (pl. digitalizáció) azonban kifejezetten számít a Célok elérése terén. Az ENSZ által még 2019-ben megfogalmazott, mesterséges intelligenciára fókuszáló ENSZ-lépéseket összegző dokumentum (United Nations Activities on Artificial Intelligence (AI), 2019) az akkori állapotok szerint taglalta, hogy a szakosított intézmények és egyéb ENSZ-testületek mesterséges intelligenciával kapcsolatos feladatkörei melyik Fenntartható Fejlődési Célokot érintik, és ebből világosan kiderült, hogy voltaképpen mind a 17 fő cél elérésében a mesterséges intelligencia alkalmazása érdemi segítséget nyújthat az ENSZ-nek és a tagállamoknak egyaránt.

A Célok preambulumban, bevezetőjében a szöveg említi, hogy olyan környezet kialakítása a cél, amelyben a fejlődés és a technológia alkalmazása például az éghajlatváltozás káros hatásainak kiküszöbölésére vagy a biodiverzitás védelmére alkalmas lehet. A mesterséges intelligencia által kezelt hatalmas adathal-

maz és a validált, tudományos adatokon nyugvó gyors, mesterséges intelligencia általi döntés alkalmas lehet eme problémák hatékony, akár automatizált kezeléséhez (Transforming Our World, 2015, 10. bekezdés). A 15. szakaszban pedig az információs és kommunikációs technológiai vívmányok – elsősorban gazdasági és társadalmi – előnyeit taglalja a dokumentum, amikor is a technológiai fejlődés eredményeinek globális térnyerése nagy lehetőségeket rejt magában a fejlődés felgyorsítására, a digitális szakadék áthidalására és a tudásalapú társadalmak fejlesztésére, kiemelve itt például az energia vagy az orvostudomány területét. Emellett a szöveg az általános célkitűzések közepette szól arról is (28. bekezdés), hogy törekedni kell a tudományos, technológiai és innovációs kapacitás megerősítésére a fejlődő államok számára, annak érdekében, hogy eme államok fenntarthatóbb fogyasztási és termelési minták felé mozduljanak el.

A nevesített 17 cél közül – ahogy arról már a korábbiakban volt szó – valamennyit érinti a mesterséges intelligencia, de leginkább 7 cél esetében jelenik meg a részcélok (target) között hangsúlyosabban olyan elköteleződés, amely összhangban van a mesterséges intelligencia kínálta előnyökkel. Eme célok pedig i) a szegénység megszüntetése (1. cél); ii) a minőségi oktatás (4. cél); iii) a megfizethető és tiszta energia (7. cél); leginkább azonban iv) az ipar, innováció és infrastruktúra (9. cél); v) a fenntartható városok és közösségek (11. cél), vi) a béke, igazság és erős intézmények (16. cél), valamint vii) a partnerség a célok eléréséért (17. cél) esetében lelhető fel a mesterséges intelligencia fontossága, mégpedig olyan formában, hogy eme célok erőteljes technológiai kitettsége előrevetíti, hogy ezeket a mesterséges intelligencia bevonásával jóval hatékonyabban lehetne megvalósítani.

Az 1. cél (szegénység megszüntetése) 4. számú rész célja úgy fogalmaz, hogy 2030-ra biztosítani kell, hogy minden férfi és nő, különösen a szegények és a kiszolgáltatottak egyenlő jogokkal rendelkezzenek a gazdasági erőforrások, a megfelelő új technológiák feletti tulajdonhoz és ellenőrzéshez való hozzáféréshez. Azaz a szegénység megszüntetésének egyik kulcsa lehet az, hogy a mesterséges intelligencia által kínált előnyök nyílt hozzáférésűek legyenek, vagy legalábbis jogilag azonos feltételekkel juthassanak ezekhez a felhasználók. Ennek tényleges megvalósítása persze számos problémába és ellenérdekbe ütközik, azonban a mesterséges intelligencia-rendszerek hozzáférésének és elérésének kérdése a mesterséges intelligenciára egyre inkább alapozó gazdasági fejlődés egyik legégetőbb megoldandó problémája. Ennek kiküszöbölése minden bizonnyal a társadalmi és gazdasági fenntarthatóság növelését fogja eredményezni a jövőben.

A minőségi oktatásról szóló 4. cél b) rész célja kifejezetten elvárja a fejlett államoktól a fejlődő országok, különösen a legkevésbé fejlett országok, a kis fejlődő szigetállamok és az afrikai országok számára a felsőoktatásban való részvételének elősegítését, különösen az információs és kommunikációs technológiai, műszaki, mérnöki és tudományos programokat is.

A tipikusan mindhárom fenntarthatósági aspektust, azaz az ökológiai, gazdasági és szociális jelentést is magában hordozó, 7. célként nevesített megfizethető és tiszta energia a) és b) rész céljai szerint fokozni szükséges a nemzetközi együttműködést a tiszta energiával kapcsolatos kutatáshoz és technológiához való hozzáférés megkönnyítése érdekében, illetve bővíteni, és korszerűsíteni kell az infrastruktúrát a modern és fenntartható energiaszolgáltatásokat biztosító technológiához, különösen a fejlődő országokban, ezek közül is kifejezetten a legkevésbé fejlett országokban, a kis fejlődő szigetországokban és a tengerparttal nem rendelkező fejlődő országokban.

Egyértelműen fogalmaz az ipar, innováció, infrastruktúra hármását felkaroló 9. cél, amelynek c) pontja alapján jelentősen növelni kell az információs és kommunikációs technológiákhoz való hozzáférést, mégpedig a legkevésbé fejlett országokban általános és megfizethető internetes hozzáférés révén.

A fenntartható városokról és közösségekről szóló 11. cél teljes mértékben összhangban van a fenntartható fejlődés hármás fogalomrendszerével, hiszen e célnál kiemelten jelentkezik az ökológiai, gazdasági és társadalmi vonatkozások együttes érvényesítése. Talán közismert az is, hogy a mesterséges intelligencia legismertebb „megnyilvánulási formái” éppen eme célhez kötődnek elsődlegesen, így pl. az okosváros koncepciója, az autonóm vezetési rendszerek és a közlekedés mesterséges intelligencián alapuló optimalizálása, az okosházak vagy akár az energiatakarékos és karbonsemleges közösségek életének megszervezése.

A 16. célként megjelölt béke, igazság és erős intézmények fókuszában is kiemelt szerepet kap a mesterséges intelligencia és a technológiai fejlődés. Valamennyi állam kénytelen ma már a technológiai fejlődés számos vívmányát állami és nemzetközi vonatkozásban egyaránt szabályozni, legalábbis ezekről véleményt vagy valamilyen politikát formálni. A békét, a békés fejlődést és az államok erős szervezeti formáinak (legyen belső jogi vagy nemzetközi) céljait előmozdító vagy éppen ezzel ellenkezőleg, eme állami célokat veszélyeztető mesterséges intelligencia-megoldások kezelése a 21. századi államok tipikus feladata. Aligha képzelhetünk el olyan államot, amely eme kihívásokat még nem azonosította, és ezekről nem formál véleményt.

A leginkább releváns főcél azonban az utolsó, lényegében a megelőző 16 célt is tulajdonképpen egységbe forrasztó 17. cél (partnerség a célok eléréséért), amely egy külön alfejezetet is szentel a technológiának. Eszerint a globális és regionális nemzetközi együttműködés fokozása a tudomány, a technológia és az innováció terén, valamint a tudományhoz, a technológiához és az innovációhoz való hozzáférés terén kulcsfontosságú, és ezt elő kell segíteni az ENSZ szintjén egy globális technológiai segítségnyújtási mechanizmuson keresztül. Ezen kívül a 17. cél technológiai alfejezete kiemeli a környezetbarát technológiák fejlesztésének, átadásának, terjesztésének és elterjesztésének előmozdítását a fejlődő országok számára. Kiemeli továbbá, hogy egy ún. technológiai bankot is működtetni kell, ezáltal a tudomány, technológia és innováció kapacitásépítési mechanizmust elérhető kell tenni a legkevesbé fejlett országok számára, és mindemellett fokozni kell az alaptechnológia, különösen az információs és kommunikációs technológia használatát.

Eme technológiai célok elérése olyan mértékű adathalmaz feldolgozásával valósítható meg, amely aligha képzelhető el a mesterséges intelligencia alkalmazása nélkül. Emellett az is bizonyos, hogy az adatokon alapuló nagyszámú és automatizált döntéshez is szükség van a mesterséges intelligencia hozzájárulására, azaz 2030-ig aligha gondolható el, hogy eme technológiai célkitűzések végrehajtásában ne lenne kulcsszerepe a mesterséges intelligenciának. Ezt egyébiránt már elősegítheti az is, hogy az államokat olyan platformok (jobbára online) alkalmazására bíztatja a célokat tartalmazó közgyűlési határozat (Transforming Our World, 2015, 70. pont), amelyek átfogóan feltérképezik a meglévő tudományos, technológiai és innovációs kezdeményezéseket, mechanizmusokat és programokat, és adatbázisként szolgálnak az ezekről szóló információkhoz, megkönnyítve ezzel a legjobb gyakorlatokhoz való hozzáférést, valamint a nyílt hozzáférésű tudományos publikációk terjesztését.

Az ambiciózus, ugyanakkor nem kötelező (és így ki sem kényszeríthető) vállalatok hatékonyságának növelése érdekében az ENSZ Közgyűlése 2023. július 25-én határozattervezetet fogadott el „*A gyors technológiai változások hatása a Fenntartható Fejlődési Célok és célkitűzések megvalósítására*” címmel (Draft Resolution 77/320, 2023). A 2023. július végi állapot szerint még semmiféle normatív erővel nem rendelkező határozattervezet csak általánosságban szólít fel a technológiai változásokra a Fenntartható Fejlődési Célok elérése érdekében, szűkebb időkeret és tisztázott elvárások nélkül. Azonban eme határozattervezet tükrében különösen fontos a mesterséges intelligencia, mint technológiai vívmány vizsgálata a kérdéskör kapcsán, annál is inkább, mert a dokumentum már több helyütt is említi a mesterséges intelligencia alkal-

mazását vagy az ezeket előírányzó ENSZ-testületek szerepét. Ezáltal pedig az ENSZ Közgyűlése 2023 nyarán összekapcsolta a célok elérését a mesterséges intelligencia használatával is.

A határozattervezet több ízben is említi a gyors technológiai változások fenntartható fejlődésre gyakorolt hatását, amely lehetőségeket és kihívásokat egyaránt magában rejt. Viszont az új technológiák (ezek közé sorolható a mesterséges intelligencia is) hozzájárulnak a célok gyorsabb megvalósításához, azáltal is, hogy „lehetővé teszik a gazdasági, társadalmi és környezeti akadályok újszerű megoldásainak gyorsabb és szélesebb körű alkalmazását.” A társadalmi és gazdasági fenntarthatóság jegyében a határozattervezet felismeri, hogy „a gyors technológiai változás óriási lehetőségeket rejt magában a nemek közötti egyenlőség előmozdítása (...) terén, azáltal, hogy csökkenti a nemek közötti digitális szakadékot, lehetőséget biztosít számukra az információszerzésre és -megosztásra, az egészségügyi és oktatási szolgáltatásokhoz való hozzáférésre, a hálózatépítésre.”

A mesterséges intelligencia szerepét is ideértve a tervezet hitet tesz amellett, hogy az innováció, az új technológiák és a kapcsolódó know-how létrehozása, fejlesztése és terjesztése, a technológiaátadás a gazdasági növekedés és a fenntartható fejlődés erőteljes mozgatórugói.

A határozattervezet – összhangban a nemzetközi fenntartható fejlődésre jellemző soft law megközelítésekkel – szóhasználatában meglehetősen puha feltételeket támaszt az államokkal szemben, előíró jelleg helyett ösztönzés, sürgetés útján bíztatja az ENSZ tagállamokat bizonyos lépések megtételére. Ezek közül különösen kiemelendő, hogy ösztönzi a tagállamokat, hogy támogassák a fenntartható fejlődést szolgáló tudományra, technológiára és innovációra vonatkozó nemzeti stratégiák és közpolitikák kidolgozását, és kéri az államokat, hogy tegyenek meg minden tőlük telhetőt a digitális és tudásbeli szakadék áthidalására, illetve előírányozza, hogy 2030-ig általános, érdemi és megfizethető hozzáférést kell biztosítani az internethez, beleértve a digitálisan támogatott szolgáltatások érdemi használatát is. Biztosítani kell továbbá azt, hogy az új technológiák előnyei mindenki számára elérhetőek legyenek, az információs és kommunikációs technológiákhoz és az internethez való egyenlő, biztonságos és megfizethető hozzáférés előmozdításával, az emberi jogok tiszteletben tartása mellett.

Kiemelt figyelmet érdemel a határozattervezet 9. pontja, amely konkrétan nevesíti a mesterséges intelligencia kérdését, amikor is a digitális megoldások előmozdítását a digitális közjavakhoz való hozzáféréshez és azok használa-

tához köti, „amelyek magukban foglalhatják a nyílt forráskódú szoftvereket, a nyílt adatokat, a nyílt mesterséges intelligencia modelleket, a nyílt szabványokat és a nyílt tartalmakat, amelyek betartják a nemzetközi és hazai jogszabályokat, a gyors technológiai változásokban rejlő lehetőségek teljes körű kiaknázása érdekében a fenntartható fejlődési célok elérése érdekében” (Draft Resolution 77/320, 2023).

A tervezet ösztönzi még a tagállamokat valamennyi érintett szereplővel való együttműködésre, így különösen magával az ENSZ-szel, valamint a magánszektoral, különösen pedig a technológiai vállalatokkal és pénzügyi intézményekkel, a civil társadalommal, a műszaki és kutatói közösségekkel, beleértve a tudósokat és a tudományos életet is. Mindemellett a határozattervezet felkéri az ENSZ egyes testületeit további anyagok kidolgozására, további lépések megtételére, és hatékonyságot növelő párbeszéd folytatására a technológiai segítségnyújtás mellett.

A fenti sorok olvasása közben persze jogos lehet az az érzés, hogy eme szövegezés nem nyújt túl sok konkrétumot, de két dolog ebből egyértelműen lezűrhető. Egyrészt a gyors technológiai változások közepette az ENSZ-es dokumentumok (ez igaz más szervezetekre, de még az államokra is) óvatosan fogalmaznak a mesterséges intelligencia szerepéről egy még nem általánosan elfogadott, kötelező erejű definíció hiányában, másrészt azonban már egyértelműen látszik, hogy a mesterséges intelligenciát bármilyen értelemben bele kell érteni a gyors technológiai változások gyűjtőfogalma alá, ezáltal pedig valamennyi Fenntartható Fejlődési Cél előmozdítására képes fogalomná vált.

IX. KONKLÚZIÓ

Elfogadásuk óta a nem kötelező erejű, de legfeljebb a politikai kötelezettségvállalásokon alapuló Fenntartható Fejlődési Célok a fenntarthatóság társadalmi, ökológiai és gazdasági szempontjainak megvalósítását célzó, kulcsfontosságú célrendszerre váltak. Mivel jelentésük régióként és államonként változhat, ezért az államok mozgástere a végrehajtásukkal kapcsolatban széleskörű és nehezen standardizálható.

Az állami végrehajtási intézkedések, így a Célok globális helyzete ismeretében komoly kétségek merültek fel a célok és célkitűzések 2030-ig történő megvalósításával kapcsolatban. A problémák számosak, csak az elmúlt három évben az emberiség több világválságnak volt szemtanúja (világjárvány, jelenleg is folyó háborúk, komoly konfliktusok, energia- és élelmiszerválság), így az

államok sajnálatosan könnyen hivatkozhattak a Célok végrehajtásának elhanyagolására vagy ideiglenes lassítására.

Ugyanakkor az egyes nemzetközi szervezetek keretében pedig egyre bővülő lehetőségek nyílnak a kérdéskör kapcsán, elég csak a Nemzetközi Távközlési Egyesületének mint az ENSZ egyik szakosított intézményének a honlapján működtetett repozitóriumra utalni, amely a Fenntartható Fejlődési Célok megvalósítását elősegítő mesterséges intelligencián alapuló megoldásokat, kutatási kezdeményezéseket mutat be nyílt hozzáférés útján (AI Repository). Ez fontos megoldásokat, jó gyakorlatokat mutathat globálisan valamennyi érdekelt szereplőnek. Az EU és az Európa Tanács keretében (és egyéb regionális szervezetek keretében is, ld. például az Afrikai Unió mesterséges intelligenciára vonatkozó kontinentális stratégiáját (The African Union Artificial Intelligence Continental Strategy For Africa, 2022) készülő tervezetek jövőbeni (esetleges) hatályba lépése pedig új szintre emelheti majd a mesterséges intelligenciával kapcsolatos állami lépésekhez fűződő jogok és kötelezettségek rendszerét.

2023-ban viszont még nehéz pontosan felmérni a fenntarthatósági célok megvalósításának előrehaladását, mivel a globális kihívások súlyossága nem látszik enyhülni, mi több, a mesterséges intelligencia fejlődése, szerepének felerősödése is nehezen prognosztizálható irányokat ölthet, amelyek jelentősen segíthetik (de akadályozhatják is) a célok megvalósítását.

A nemzetközi közösség tehát 2023-ban (legkésőbb most, a 2015-től 2030-ig tartó megvalósítási félidejében) újratervezheti a Fenntartható Fejlődési Célok kapcsán esedékes teendőit, akár a mesterséges intelligencia hathatós segítségével.

IRODALOM

1. Biermann, Frank – Hickmann, Thomas – Sénit, Carole-Anne – Beisheim, Marianne – Bernstein, Steven – Chasek, Pamela – Grob, Leonie – Kim, Rkhyun – Kotzé, Louis – Nilsson, Måns – Llanos Ordóñez, Andrea – Okereke, Chukwumerije – Pradhan, Prajal – Raven, Rob – Sun, Yixian – Vijge, Marjanneke – van Vuuren, Detlef – Wicke, Birka (2022): Scientific Evidence on the Political Impact of the Sustainable Development Goals. In: *Nature Sustainability*, 2022/5, 795–800. <https://doi.org/10.1038/s41893-022-00909-5>
2. Crowther, David – Seifi, Shahla – Moyeen, Abdul. (eds.) (2018): *The Goals of Sustainable Development – Responsibility and Governance*. Springer Nature Singapore Pte Ltd. Singapore.

3. Hárs András (2021): AI and International Law – Legal Personality and Avenues for Regulation. In: Hungarian Journal of Legal Studies. 2021/4, 320–344. <https://doi.org/10.1556/2052.2022.00352>
4. Mialhe, Nicolas – Hodes, Cyrus – Jain, Arohi – Iliadis, Niki – Alanoca, Sacha – Png, Josephine (2019): AI for Sustainable Development Goals. In: Delphi - Interdisciplinary Review of Emerging Technologies, 2019/4, 207-221. <https://doi.org/10.21552/delphi/2019/4/10>
5. Nasir, Osama – Tallal Javes, Rana – Gupta, Shivam – Vinuesa, Ricardo – Qadir, Junaid (2023): Artificial Intelligence and Sustainable Development Goals Nexus Via Four Vantage Points. In: Technology in Society, 2023, 1-7. DOI: 10.1016/j.tech-soc.2022.102171
6. Purvis, Ben – Mao, Yong – Robinson, Darren (2019): Three Pillars of Sustainability: In Search of Conceptual Origins. In: Sustainability Science, 2019/14, 681-695. doi.org/10.1007/s11625-018-0627-5
7. Schrijver, Nico (2008): The Evolution of Sustainable Development in International Law: Inception, Meaning and Status. Martinus Nijhoff Publishers, Leiden, The Netherlands.
8. Viñuales, Jorge (2021): Sustainable Development. In: Rajamani, Lavanya – Peel, Jacqueline (eds.) (2021): The Oxford Handbook of International Environmental Law. Second Edition. Oxford University Press, Oxford, UK. 285-301.
9. Vinuesa, Ricardo – Azizpour, Hossein – Leite, Iolanda – Balaam, Madeline – Dignum, Virginia – Domisch, Sami – Felländer, Anna – Langhans, Simone Daniela – Tegmark, Max – Fuso Nerini, Francesco (2020): The Role of Artificial Intelligence in Achieving the Sustainable Development Goals. In: Nature Communications. 2020/233. <https://doi.org/10.1038/s41467-019-14108-y>

EGYÉB FORRÁSOK:

1. 1992 Rio Declaration on Environment and Development. A/CONF.151/26 (Vol. I) Report of the United Nations Conference on Environment and Development. Rio de Janeiro, Brazil, 1992.
2. A mesterséges intelligenciára vonatkozó harmonizált szabályok. Javaslat. Az Európai Parlament és a Tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról és egyes uniós jogalkotási aktusok módosításáról. COM(2021) 206 final, 2021/0106(COD).
3. Addis Ababa Action Agenda of the Third International Conference on Financing for Development (Addis Ababa Action Agenda). The final text of the outcome document adopted at the Third International Conference on Financing for Development (Addis Ababa, Ethiopia, 13–16 July 2015) and endorsed by the General Assembly in its Resolution 69/313 of 27 July 2015.

4. AI Repository. (Elérhető: <https://www.itu.int/en/ITU-T/AI/Pages/ai-repository.aspx>. Letöltés ideje: 2023. július 30.).
5. Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, 2023. (Elérhető: <https://rm.coe.int/cai-2023-18-consolidated-working-draft-framework-convention/1680abde66>. Letöltés ideje: 2023. július 30.).
6. Johannesburg Declaration on Sustainable Development and Plan of Implementation of the World Summit on Sustainable Development: the final text of agreements negotiated by governments at the World Summit on Sustainable Development, Johannesburg, South Africa. 2002.
7. Our Common Future: Report of the World Commission on Environment and Development. U.N. Doc. /A./42/427 (1987), New York, USA.
8. Recommendation on the Ethics of Artificial Intelligence, 2021. (Elérhető: <https://unesdoc.unesco.org/ark:/48223/pf0000380455>. Letöltés ideje: 2023. július 30.).
9. Strengthening Voluntary National Reviews through Country-Led Evaluation. United Nations General Assembly Resolution 77/283, A/77/L.64 adopted on 26 April 2023 without a vote (A/77/PV.69) under item 18.
10. The African Union Artificial Intelligence Continental Strategy For Africa, 2022. (Elérhető: <https://www.nepad.org/news/african-union-artificial-intelligence-continental-strategy-africa>. Letöltés ideje: 2023. július 30.).
11. Transforming Our World: The 2030 Agenda for Sustainable Development. United Nations General Assembly Resolution 70/1, U.N. GAOR, 70th Sess., U.N. Doc. A/RES/70/1 (2015).
12. United Nations Activities on Artificial Intelligence (AI), 2019. (Elérhető: <https://www.itu.int/pub/S-GEN-UNACT-2019-1>. Letöltés ideje: 2023. július 30.).
13. United Nations General Assembly Draft Resolution 77/320 on 25 July 2023. Impact of rapid technological change on the achievement of the Sustainable Development Goals and targets. A/77/L.84 adopted on 25 July 2023 without a vote (A/77/PV.94) under item 133.

SZABÓ TAMÁS

A PÉNZÜGYI SZFÉRÁBAN
ALKALMAZOTT DÖNTÉSTÁMOGATÓ
RENDSZEREK ÁTLÁTHATÓ ÉS
FOGYASZTÓBARÁT FELHASZNÁLÁSA

A pénzügyi szférában alkalmazott döntéstámogató rendszerek átlátható és fogyasztóbarát felhasználása

I. BEVEZETŐ GONDOLATOK

A pénzügyi rendszer a pénzügyi piacok (pénzpiac, értékpapírpiac, egyéb pénzügyi piacok), a pénzügyi intézmények (pl: hitelintézetek, biztosítók, pénzügyi vállalkozások, nyugdíjpénztárak és egyéb közvetítők, valamint állami és kormányzati szervek) és a pénzügyi infrastruktúra (pl: fizetési és elszámolási rendszer, szabályozói közeg) összessége, amelyek lehetővé teszik, hogy egyes gazdasági szereplők, illetve magánszemélyek megtakarításai (tétlen pénz) a hiteligénylők kezébe kerüljenek.

Ezért a pénzügyi rendszer – a szereplői által – alapvetően (közgazdaságtani értelemben) közvetítést szolgál azok között, akiknek pénzfeleslegük (pénzbőségük) van és akiknek finanszírozásra van szükségük. Más szavakkal, amit a pénzügyi rendszer generál, az kredit. A pénzügyi rendszernek nagyon fontos szereplői a pénzügyi intézmények, ezen belül is dedikáltan bankok, amelyek közvetítőként lépnek fel azok között, akiknek többletpénzük van, és akiknek szükségük van rá.

Ezen felül megkönnyítik a feltételeket mindkét fél számára. Időben és mennyiségben módosítják a kapott és kölcsönzött pénzt, a többlet pénzzel rendelkező gazdasági szereplő és a pénzigénylő igényeitől függően. A megtakarítási szolgáltató és a hitelfelvevő közvetlen hitele lehetetlen lenne, mivel pénzszükségletük valószínűleg nem azonos mennyiségben és időben.

A pénzügyi intézményrendszer felépül a közvetlen tőkeáramlást (áttételt) biztosító piacokból, illetve az ezen piacokon tevékenykedő pénzügyi közvetítőkből. A pénzügyi piacok között meg lehet különböztetni elsődleges és másodlagos piacokat, aszerint, hogy milyen intézmények, szervezetek fejtik ki a működésüket ez egyes piacokon.

A piacok általánosságban azok a csatornák, amelyeken keresztül a vásárlók és az eladók találkoznak, hogy kicseréljék a termékeket, a szolgáltatásokat és az erőforrásokat. A piacok egyben az információk áramlásának médiumai is. A pénzügyi piacokon különböző pénzek, illetve pénzre szóló követelések

(*pénzügyi instrumentumok*) cseréje zajlik. A különböző időpontban rendelkezésre álló pénzek és pénzügyi instrumentumok cseréje a jövedelmek különböző gazdasági szereplők közötti időbeni, térbeni átcsoportosítást közvetíti. Elsődleges piacokon a kockázati tőkebefektetők, más befektetési társaságok fejtik ki aktivitásuk annak érdekében, hogy tőkét nyerjenek befektetéseik által. A másodlagos piacok egy részét tulajdonképpen a tőzsdék (áru-/értéktőzsdék) testesítik meg, ahol az értékpapírokkal vagy más jószággal kereskedő intézményi vagy magánbefektetők kereskednek. A másodlagos piacokon, mint közvetítők, pénzügyi közvetítőket lehet (és olykor szükséges is) megbízni. Két kategóriájuk a monetáris közvetítők, illetve a nem monetáris közvetítők.

Magyarországon a pénzügyi rendszeren belül, mint pénzügyi intézmény a hitelintézetek és a pénzügyi vállalkozások sorolhatóak fel. Hitelintézet az a pénzügyi intézmény, amely a pénzügyi vállalkozásokról szóló 2013. évi CCXXXVII. törvény (a továbbiakban: Hpt.) 3. §-ban meghatározott pénzügyi szolgáltatások közül legalább betétet gyűjt, vagy más visszafizetendő pénzeszközt fogad el a nyilvánosságtól - ide nem értve a jogszabályban meghatározott nyilvános kötvénykibocsátást -, valamint hitelt és pénzkölcsönt nyújt. Pénzügyi vállalkozásnak pedig a Hpt. 9. § (1) bekezdés értelmében az a pénzügyi intézmény számít, amely meghatározott tevékenységek kivételével, egy vagy több pénzügyi szolgáltatást, vagy fizetési rendszer működtetését végzi, illetve ide sorolandó még a pénzügyi holding társaság is. Láthatólag e két fogalom átfedi egymást hiszen a pénzügyi vállalkozás a pénzügyi intézménynek számít így ezen utóbbira a pénzügyi intézmény vonatkozó szabályok és kiterjednek.

Ezen intézmények alapvetően tevékenységek széles skáláján válogathatnak az üzleti nyereségesség szempontjai mentén így elkülöníthetők pénzügyi szolgáltatások és kiegészítő pénzügyi szolgáltatások a Hpt. 3. § (1) és (2) bekezdései alapján. A továbbiakban, pénzügyi vállalatként értendők a fintech-társaságok, és a pénzügyi intézmények összessége.

A jelen tanulmány kutatási kérdése, hogy a mesterséges intelligencia melyik területeken tud a leginkább a pénzügyi vállalkozások vagy hitelintézetek javára, illetve versenyelőnyére szolgálni. Hipotézisként, a pénzügyi szolgáltatások közül elsősorban a valutával, devizával, váltóval, illetve csekkel saját számlára vagy bizományosként történő kereskedelmi tevékenység, a pénzügyi szolgáltatás közvetítése, illetve a hitelreferencia szolgáltatás a legkézenfekvőbb területe a mesterséges intelligencia alkalmazásának. A kiegészítő pénzügyi szolgáltatások körében pedig csaknem az összes ilyen tevékenység ide sorolható talán csak az e) pont alatti forgatható utalvány kibocsátására irányuló tevékenység nem adekvát ezen összefüggésben.

II. AZ MI-BEN REJLŐ LEHETŐSÉGEK

A mesterséges intelligencia (a továbbiakban: MI) már jelen írás előtt is jó néhány publikációban jelzetten hatalmas fejlődési potenciált rejtett és az sem kérdés, hogy már a mai nap is tetten érhető az alkalmazása a pénzügyi vállalatok működésében. Az Európai Bizottság az egyik rendelet-javaslatában [Javaslat, Az Európai Parlament és a Tanács Rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, és egyes uniós jogalkotási aktusok módosításáról COM (2021) 206 final] akként írja le ezt az ernyőfogalmat, mint *„olyan szoftver, amelyet [...] a felsorolt technikák és megközelítések közül egy vagy több alkalmazásával fejlesztettek, és amely az ember által meghatározott célkitűzések adott csoportja tekintetében olyan kimeneteket, például tartalmat, előrejelzéseket, ajánlásokat vagy döntéseket képes generálni, amelyek befolyásolják azt a környezetet, amellyel kölcsönhatásba lépnek”*. Itt tehát nem teszi az Európai Bizottság e fogalom tekintetében önjáróvá mesterséges intelligenciát, hanem mindig az emberi input az, ami meghatározza a cselekvésnek a kereteit. A fogalom szerint az MI különösen alkalmas ajánlásokra vagy decízív működésre is. Fontos, hogy ez a külvilágban is tapasztalható és valós eredménnyel zárul tehát befolyásolja az MI-döntés a környezetet vagy azzal kölcsönhatásba lép. Jelen téma tárgyalása szempontjából különösen fontosnak az előrejelzések, illetve az ajánlások a mesterséges intelligencia felhasználásának szempontjából. Továbbá a döntéshozatalra azért nem térnék ki részletesebben, mert álláspontom szerint az önálló pénzügyi döntéseket Csak az felhasználó végezheti el az MI-nek csupán a felhasználó által előre beprogramozott döntést szükséges végrehajtania. E két terület azért kiemelendő, mert ezzel tudja az MI leginkább a felhasználó pénzügyi döntéseit objektívebbé tenni, illetve a pénzügyi intézmények részére a maximális profitot kiaknázni a szolgáltatásaiból. A teljesség igényével szólni kell azonban az MI-kódex anyagi szabályairól is. A kódex a horizontális szabályozás minimumát célozza meg, azzal, hogy az MI-eket különböző kockázati osztályokba sorolja. Így tehát a javaslat megkülönböztet teljesen tiltott MI-alkalmazásokat, kockázatos MI-alkalmazásokat és kevésbé kockázatos, de valamilyen szempontból még mindig kiemelt figyelmet érdemlő egyéb alkalmazásokat végül pedig alacsony, alig mérhető kockázatú MI-alkalmazásokat, amelyek lényegében a felhasználók önszabályozó mechanizmusaira bízva. Az első kategória, a tiltott MI-k kategóriája kiemelendő. Ide tartoznak: tudatalatti manipulálásra képes MI-k, bizonyos sérülékenységeket (például fogyatékoságot) kihasználó MI-k, társadalmi pontozást megvalósító MI-k, arcfelismerést közterületen végző MI-k.

Nagy kockázatúnak akkor minősül egy MI, ha vagy biztonsági komponense egy egyébként is szorosan szabályozott termékcsoporthoz, vagy „saját jogon” képvisel kockázatokat, különösen akkor, mert olyan területen tevékenykedik, amely az emberi jogokat érinti.

III. DÖNTÉSTÁMOGATÓ RENDSZEREK

Az előző fejezetek tartalma egyesítésre kerül a pénzügyi döntéstámogató rendszerek fogalma alatt, mint a jelen írás fő témája. A technológia és technika fejlődésével létrejöttek különböző döntéstámogató rendszerek (*decision supporting systems*) (a továbbiakban: DTR, DSS), és ezek jelentősen fejlődtek az 1970-es évekbeli korai fejlesztésük óta (Shim et al, 2002). Az elmúlt három évtizedben a DSS-ek szűkebb vagy tágabb értelemben is értelmezhetőek, miközben más rendszerek jelentek meg a döntéshozók meghatározott típusú, meghatározott típusú problémákkal szembesülő típusainak támogatására. Az e területen végzett kutatások jellemzően arra összpontosítottak, hogy az információs technológia hogyan javíthatja a felhasználó döntésének hatékonyságát, és hogyan javíthatja a döntés hatékonyságát a vállalatok üzleti életében; mindezek nagy előnyt tudnak jelenteni a versenytársak között. Így nem csak a gyakorlati életben, hanem a tudományos szférában is - kiváltképp a közgazdaságban - sor került ezen eszközökkel kapcsolatos kérdések feldolgozására. Ennek alapja a közgazdaságtani döntésemleletek, azaz az adott feltételek kontextusában, az optimális döntés meghozatalának kérdésköre. Az e komplex problémakörből kiinduló elméletek részletesebb kifejtése nélkül, rögzítendő, hogy az optimális döntés mindig az alapvető problémákból az ezt körülvevő lehetőségekből, a problémamegoldás alternatíváiból és a lehetséges külső hatások, következményekből felméréséből összeálló, azokra adott válasz. Belátható, hogy ezen összetett téma leegyszerűsítésére (megoldására) kiválóan alkalmas a döntéstámogató rendszerek használata.

A DSS tipikusan olyan számítógépes programalkalmazás (Alter, 2003), amelyet egy szervezetben vagy vállalkozásban a döntések, ítéletek és cselekvési irányok támogatására használnak. DSS ugyanakkor lehet teljesen számítógépes vagy emberi erőforrással működtetett is, valamint bizonyos esetekben mindkettőt kombinálhatja. Meglátásom szerint ez utóbbi vezethet az optimális döntés meghozatalához. További előnye ez utóbbinak, hogy lehetővé teszik az emberi felhasználók számára, hogy gyorsabb ütemben, megalapozottabb döntéseket hozzanak, végeredményben egyetlen döntéssel, nem döntések sorozatával.

A DSS hatalmas mennyiségű adatot szitál át és elemez, átfogó információkat állít össze (szintetizálás), amelyek felhasználhatók a problémák megoldására és a döntéshozatalban. A döntéstámogató rendszer összegyűjti és elemzi az adatokat, azokat szintetizálva, hogy átfogó információs jelentéseket készítsen, a szervezet számára a lehető legjobb lehetőségek bemutatásával. Ily módon a DSS mint információs alkalmazás különbözik más műveleti alkalmazásoktól, amelyeknek egyetlen funkciója csupán az adatgyűjtés. A DSS rendszerek továbbá rendelkeznek modellezési képességekkel, matematikai analitikai modellek segítségével előrejelzéseket tudnak végezni, elemzéseket, illetve javaslatokat megtenni. Összefoglalva a rendszer célja, hogy a döntéshozó döntésképességét megerősítse, azt egyedi speciális opciók szolgáltatásával tágítsa. Az azonban már a döntést meghozó személyen vagy szervezeten múlik, hogy a DSS rendszereket mikor és hogyan alkalmazzák, mennyiben hagyatkoznak rájuk. Tehát fontos leszögezni, hogy a DSS rendszer nem a döntéshozó helyett dönt, hanem csak a döntést könnyíti meg számára. Ezen rendszerek akként tipizálhatóak, hogy milyen elv/logika az, ami mentén meghozzák a döntési javaslatot. Eszerint léteznek modell vezérelt, adat vezérelt, kommunikáció vezérelt, dokumentum vezérelt, tudás vezérelt és relatív DSS-ek is (Psarommatis, Kirirsis, 2022). A modell vezérelt DSS-ek egy adott tevékenység egyszerűsítésére összpontosítanak, és értékelik, hogy milyen alternatív cselekvések jöhetnek elő opcióként. Így különböző lehetséges kimeneteket hasonlítanak össze, és minden egyes forgatókönyvhöz valószínűséget és kimeneti teljesítményértéket rendelnek, azaz modellizálnak. Az adatvezérelt DSS-ek kihasználják a külső és a szervezet egészére vonatkozó múltbeli és aktuális, inputált információkat, és az egyes konkrét döntéshozatali folyamatokhoz releváns adatokat keresnek (Gandhi et al, 2018). A kommunikáció vezérelt DSS-ek a kommunikációs és információs technológiákat használják fel az információk gyűjtésére és megosztására. A dokumentum vezérelt DSS-ek képesek gyorsan és hatékonyan információt kinyerni a rendelkezésre álló dokumentumokból, például szövegfájlokból, képekből, videókból vagy hangfelvételekből. Ezeket a dokumentumokat felhasználják a döntésalapú cselekvések támogatására, hivatkozási alapként. A tudás vezérelt DSS-ek adatbányászati rendszereket használnak a számítógép-alapú döntések megerősítésére. Két alkategóriáját képezik a szakértők tudása alapján felépített szabályalapú rendszerek és a gépi tanulási algoritmusok, amelyek neurális hálózatokat és mesterséges intelligencia technológiákat alkalmaznak (Mrzyglod et al 2018). Ezek a rendszerek különböző kísérletek által lehetővé teszik az autonóm és önálló döntéshozatalt.

Ezen rendszerekben természetesen az MI is teljesen új távlatokat nyitott, hiszen a big data alapú adatelemzés megsokszorozta a döntéstámogató rendszerek hatékonyságát. Az MI-vel felvértezett DSS rendszereket a szakirodalom IDSS (*intelligent decision support systems*) -nek hívja. A mesterséges intelligencia nagy mennyiségű adatot tud bányászni (*data-mining*) és ezzel párhuzamosan akár feldolgozni, hogy betekintést nyerjen és ajánlásokat tegyen a jobb döntéshozatal érdekében. Ezt több adatforrás elemzésével, valamint a minták, trendek és összefüggések azonosításával teszi, annak érdekében, hogy az emberi döntéshozatali képességeket utánozza. Az emberi tanácsadókhöz hasonlóan megtervezett IDSS rendszer adatokat gyűjt és elemez, ugyanazon célért: hogy a döntéshozókat támogassa a problémák azonosításával és elhárításával, valamint a lehetséges megoldások felajánlásával és értékelésével. Az IDSS azzal, hogy mesterséges intelligencia által támogatott rendszerről van szó a számítógép vezérelt rendszereket egy inkább az emberi döntési mintákhoz igazítja így például sajátosan magába foglalja az indoklást, az (emberi) intuíciót, valamint az általános józan ésszt is (Tariq, Rafi, 2012). Az IDSS mesterséges intelligencia komponense a lehető legjobban utánozza az emberi képességeket, miközben számítógépes rendszerként hatékonyabban, koherensebb, gyorsabban dolgozza fel és elemzi az információkat. Világos a kapcsolat a döntéstámogató rendszer, illetve az MI között, bizonyos szerzők (Tariq, Rafi, 2012) pedig egyenesen természetes (magától értetődő) kapcsolatként írják le ezt a viszonyt. Az IDSS tartalmazhat olyan fejlett képességeket, mint a tudásbázis, a gépi tanulás, az adatbányászat és az interaktív felhasználói felület. A mai előrehaladott technológiai fejlettség kapcsán léteznek új típusú intelligens döntéstámogató rendszerek, mint például a csoportos döntéstámogató rendszer (GSS), ennek ellentettje, az elosztott döntéstámogató rendszer, az intelligens, interaktív és integrált döntéstámogató rendszer és a tudásfelfedezésen alapuló intelligens döntéstámogató rendszer (Zhou et al 2008). A csoportos döntéstámogató rendszerek vagy kollaborációs rendszerek javítják a kooperatív munkában részt vevő csapattagok kommunikációval kapcsolatos tevékenységeit. A csapattagok kommunikációs és koordinációs tevékenységét olyan technológiák segítik elő, amelyek az idő, a tér és a csoporttámogatás szintjének három kontinuum mentén jellemezhetők (Alavi, Keen, 2010). Ezeket a technológiákat a személyes találkozókat megterhelő tér- és időbeli korlátok leküzdésére, az információhoz való hozzáférés körének és mélységének növelésére, valamint a csoportos feladatvégzés hatékonyságának javítására használják. Így ezen GSS megkönnyíti a hatékonyabb csoportos interakciót, ami a modern diverz szervezetekben nagyobb döntéshozatali hatékonyságot eredményez,

nem utolsó sorban pedig a tudás-aggregáló tulajdonsága sem elhanyagolható. A tudásfelfedezésen alapuló IDSS reaktív tanuláson alapulva tud komplexebb választ kialakítani, a problémaköréről kapott meghatározott adatokból. Az IDSS megvalósításai közé tartoznak például a rugalmas vagy intelligens gyártási rendszerek, az intelligens marketing döntéstámogató rendszerek és az orvosi diagnosztikai rendszerek.

Mind a DSS és az IDSS rendszerek számos különböző területen lehet hasznosítani azonban tipikusnak tekinthető az orvosdiagnosztika, a klinikai kezelések (Sutton et al, 2020), a marketing és reklámstratégiák kialakítása (Van Bruggen et al, 1998), a különböző mérnöki iparágak (tervezés), de gondolhatunk itt a leghétköznapibb felhasználási módokra például a GPS-re (útvonal azonnali újratervezése) vagy éppen a vállalati erőforrás-tervezést bemutató „műszerfalra” (*enterprise resource planning* (ERP) dashboard) is.

IV. A DÖNTÉSTÁMOGATÁS FELELŐS HASZNÁLATA

Azonban a különböző döntéstámogató rendszereket és az MI-t hordozó technológiákat, a szolgáltatásaik fejlesztése és működtetése során, a pénzügyi rendszer szereplői is alkalmazzák, a profitmarzsuk maximalizálására, illetve a felhasználói bázis kialakítására, megtartására. Ezen eszközök hasznosak amikor egy új pénzügyi szolgáltatást vagy terméket akarnak például a bankok piacra dobni. Továbbá a mesterséges intelligencia ki tudja szűrni azokat az ügyfél mintákat, vásárlási szokásokat, a biztonsági réseket vagy hibákat, amelyeket együtt értékelve pénzügyi rendszer fentebb taglalt szereplői fejlődni tudnak. A pénzügyi intézmények oldalán olyan előnyök merülnek fel, mint hogy a működési költségek csökkenését, hatékonyságnövekedést, az ismétlődő feladatok számának csökkenését és az ügyfélélmény javulását eredményezi az MI/IDSS használata. Az MI-adaptáció az automatizált, akvizíciós célú ügyfélmegkeresések, a személyre szabott termékek fejlesztésére irányuló adatelemzés és a pénzügyi kockázatkezelési modellek – ideértve a csalásmegelőzési modellek – fejlesztése tekintetében ma már egyaránt széleskörben elterjedt. Az MI alkalmazások átalakítják a személyes pénzügyeket, a pénzügyi szolgáltatások számos területét, így a hitel- és részvényt piacokat, a pénzforgalmat, a hitelbírálatokat, valamint a szabályozói megfelelést is. Az utóbbi években megjelenő technológiai innovációk közül a mesterséges intelligencia volt (Szikora, Nagy, 2020) és lesz a továbbiakban is a legjelentősebb hatással a pénzügyi szektor működésére. Sőt, a PwC 2020-as felmérés szerint a válaszadók

86%-a szerint a mesterséges intelligencia 2021-ben *mainstream* technológiává válik a vállalatuknál (PwC, 2021). Ez azóta nyilvánvalóan csak fokozódott.

Az alábbi területeken a pénzügyi vállalkozások és a hitelintézetek könnyen be tudják építeni az MI-t a szolgáltatásaikba.

1. Adatautomatizálás/hitelkockázat-értékelés

Az MI-alapú hitelbíráló vagy hitelpontozás (*credit scoring*) talán a legígéretesebb és legjelentősebb terület, az utóbbi fogalom annak a banki értékelése, hogy egy ügyfél mennyire nagy eséllyel tud és képes vagy hajlandó visszafizetni az adósságát. A hitelpontszámokra vonatkozó MI-döntések számos adaton alapulnak, például a teljes jövedelem, a hiteltörténet, a tranzakcióelemzések, a munkatapasztalat stb. alapján. A hitelpontozás tehát egy statisztikai-matematikai modell (Wang et al., 2011), amely nagy mennyiségű információt vesz figyelembe. E modell eredménye, hogy az input-adatokra és azok változásaira érzékeny, egyéni hitelpontszám-értékelést biztosítanak számos további, valós idejű tényező alapján, ami potenciálisan több jövedelemmel rendelkező ember számára teszi lehetővé a finanszírozáshoz való hozzáférést, ekként elosztva a banki (hitelintézeti) szabad (hitelezésre fordítandó) tőkeállományt. A hitelpontozás segít a pénzügyi szakembereknek abban, hogy jobb döntést hozzanak arról, hogy elfogadják-e a hitelkérelmet vagy sem, így a nagy valószínűséggel nem teljesítő hiteleket nem fogadják el (Demaio et al., 2020). Azonban e rendszer súlyos következményekhez is vezethet ugyanis felsejlik az ügyféli negatív determináltság, ami azt eredményezheti, hogy emberi kontroll nélkül egyes ügyfelek sosem férhetnek hozzá a hitelekhez.

2. Automatizált ügyfélszolgálat

Az ügyfélszolgálat és -kapcsolattartás egy olyan kritikus terület, ahol az MI-technológiák látványos értéket nyújtanak. A pénzügyi vállalkozások – a fintech cégek mintájára – a sokhelyütt elterjedt chatbotokat használják az ügyfelek kérdéseire adott gyors válaszára, és javíthatják az ügyfelek általános tapasztalatát a termékeikkel és szolgáltatásaikkal kapcsolatban (Abd-Alraza et al., 2020). Ezek a chatbotok lehetővé teszik, hogy az ügyfélszolgálatot személyre szabottan, alacsony költséggel nyújtsanak szakértői tanácsokat, így megspórolva az emberi munkaerő költségét. A chatbotok a napi 24 órában rendelkezésre állhatnak. Ezen virtuális asszisztensek mesterséges intelligenciával is kiegészíthetők, hogy segítsenek az ügyfélnek eligazodni

a bank kínálatában, javítsák a felhasználói adatokat és személyre szabott felhívásokat nyújtsanak. A virtuális asszisztensek és chatrobotok segítségével ezen említett vállalatok időt és pénzt takaríthatnak meg az ügyfélszolgálaton. Az MI efféle alkalmazásunkkal, a mesterséges intelligencia minőségének javítását is elérhetővé válik, az MI-rendszer folyamatos tanulása által.

3. Banki csalások felderítése

Ezen folyamatban a rendelkezésre álló (folyamatosan növekvő) részletes bankbiztonságra vonatkozó adatokat MI-alapú elemzőeszközökkel elemzik, hogy felismerjék a csalásra utaló gyanús tranzakciókat. A mesterséges intelligencia segítségével nyomon követhetők a felhasználók viselkedési mintái, és azonosíthatók a normálistól eltérő műveletek, amelyek csalási kísérletekre vagy incidensekre utalhatnak (Thekkethil et al., 2021). Tegyük fel, hogy egy ügyfél többször is megpróbál 2000 eurót felvenni a számlájáról valahol 'nem tipikus helyen', esetleg valaki 5 perc alatt 10 azonos kölcsönt próbált igényelni. Egy mesterséges intelligenciával rendelkező rendszer ezt anomáliaként észlelheti, és gyanúsnak tekintheti, továbbá ezt képes azonosítani és értékelni. Így, az MI által vezérelt tanulás képes felismerni és blokkolni ezeket a tevékenységeket, mint lehetséges csalást. Ez rendkívül hasznos a pénzügyi világ számára, mivel óránként hihetetlen mennyiségű digitális tranzakció zajlik, ami szükségessé teszi a fokozott kiberbiztonságot és a sikeres csalás felismerését. Védve ezzel a felhasználót (Rahman et al., 2021), hiszen azt feltételezi, hogy a bankkártyát ellopták, vagy más nemkívánt negatív esemény történt, és védve a bankot (illetve a kártyakibocsátó társaságot), hiszen így az egyes tranzakciók előzetes szűrésre kerülnek. Mindez természetes nem áthatolhatatlan akadályrendszer, az ügyfél egy egyszerű utasítással felül tudja írni ezen „MI-védvonalat”.

Fontos kikötni, hogy ezen kockázat felméréseket az adott intézmény szakszemélyzetének szükséges átadni további megfontolásra, így tehát ezen módszerrel rengeteg időt takarítva meg nekik, úgy, hogy a döntésbiztonságot is szem előtt lehet tartani.

4. A biztonság növelése

A mesterséges intelligencia a pénzügyi világban számos, a biztonságot javító megoldás mögött álló hajtóerő. A pénzügyi vállalatok például olyan alkalmazásokat kínálnak, amelyekhez csak arc- vagy ujjlenyomat-felismeréssel lehet hozzáférni. Ezt elsősorban a mesterséges intelligencia teszi lehetővé.

5. Pénzügyi előrejelzések megtétele

A mesterséges intelligencia alkalmazása a pénzügyekben azt jelenti, hogy az, az adatminták változásainak azonosítására, elemzésére és kiigazítására önállóan képes. Sőt, még ha az adatok minősége messze nem is tökéletes, az MI akkor is értékes meglátásokkal tud előállni az árfolyamokkal, befektetésekkel, rövid vagy hosszú távú trendek előrejelzéseivel kapcsolatban. Továbbá képes a strukturálatlan adatok rendszerezésére, elemzésére, ami a pénzügyi szolgáltatások esetében elengedhetetlen. A pénzügyi szektorban ez többféleképpen is alkalmazható, többek között a termékek alkalmasságának előrejelzésére és a marketingstratégia adaptálására az ügyfelek viselkedésének elemzésével (Abhasimehr, Shabani, 2021). Amennyiben mindezen rendszert pedig egy blockchain-alapú rendszerre (Glavanits, Király, 2018) helyezzük, lehetőségek szinte végtelen tárháza tud megnyílni, az MI és a blokklánc együttes alkalmazásával és akkora ezen terület, hogy az egy önálló tanulmányban nyerhetne teret.

6. Felhasználói viselkedés-elemzés

A mesterséges intelligencia képes megjósolni a felhasználó viselkedését egy olyan felület segítségével, amely integrálja a különböző mesterséges intelligencia- és gépi tanulási technikákat, lehetővé téve a felhasználók számára, hogy részletes betekintést nyerjenek adataikba. Ezeket MI API-knak (Application Programming Interface) is nevezik, és a pénzügyi vállalatok fel is használhatják őket. Ez például akként működhet, hogy az ügyfél adatokat kér az elmúlt hónap kiadásaira vonatkozóan. Az adott vállalat MI segítségével előre jelzi a szerveroldalon a további kérését, és ugyanabban a válaszban megadja ezt az információt. Így a vállalat minimalizálja a kérések számát és az üzleti rendszer terhelését. Így amennyiben az előre jelző elemzés helyes, az ügyfél is könnyebb helyzetben találja magát, mivel a rendszer gyorsabban működik, az ő kényelme érdekében.

7. Személyre szabott pénzügyi szolgáltatások

Számos olyan alkalmazás létezik, amely az MI-t használja fel az ügyfelek személyre szabott pénzügyeinek intézésére. Léteznek olyan költségkövető applikációk (pl: Wyze, Spendee, Wallet stb.), amelyek automatikusan segítenek a fogyasztóknak optimalizálni a kiadásait és megtakarításait, a saját

személyes szokásaik és céljaik alapján. Az alkalmazás képes olyan tényezőket elemezni, mint a havi jövedelem, az aktuális egyenleg és a kiadási szokások, majd meghozza saját döntéseit és átutalja a pénzt a megtakarítási számlára, ezzel levéve a döntési terhet az ügyfelelről.

V. FOGYASZTÓI SZOKÁSOK BEFOLYÁSOLÁSA ÉS JOGI MEGFELELŐSÉG

A fentebb kifejtett területeken jól látható, hogy a mesterséges intelligencia hogyan tud beférkőzni a pénzügyi szolgáltató szektor üzletvitelébe. Megállapítható, hogy ezen szervezetek lehetőségeiknél fogva, ki is aknázzák ezen előnyöket. A fogyasztói és felhasználói szokások befolyásolása direkt vagy indirekt módon megvalósul a pénzügyi szférában. Például amikor egyes banki alkalmazások a költségvetés funkció alapján tudják monitorozni a felhasználó szokásait. Továbbá ajánlhatnak a pénzügyi intézmények különböző biztosításokat vagy termékeket az alapján, hogy az adott felhasználó az adott időszakban számos repülőjegyet vásárolt. Akárcsak a hirdetési piacon, pénzügyi szolgáltatók is személyre szabott ajánlatokat tudnak felkínálni ügyfeleik részére, az MI adatelemzésének segítségével.

Egyes szolgáltatók ki tudnak alakítani a mesterséges intelligencia alkalmazásával olyan kiadás- vagy költségvetés-tervező applikációkat is, amelyek közvetlen összeköttetésben állnak bankkártyákkal vagy különböző hűségkártyákkal is. Ezen szolgáltatás továbbá már ötvözi az MI és a fentebb taglalt DSS-rendszereket is adott esetben, így egy komplex szolgáltatásról beszélhetünk (Gao et al., 2007). Megjegyzem, létezik ezen alkalmazás ennek egy olyan szintje, amely teljesen át fogja a családi vagy személyi költségeket, tulajdonképpen egy pénzügyi asszisztenssé válik a rendszer.

A mai fogyasztói társadalomban egy termék azon tulajdonsága határoz meg majdnem mindent, hogy mennyire tűnik fogyaszthatóak, mennyire váltja ki a fogyasztóban a FOMO-felfogást (Fear of Missing Out), ha nem veszi meg azt az adott terméket. Hasonlóan működnek a pénzügyi előrejelzések, befektetési tanácsok is a bróker- vagy befektetés közvetítő cégek termékeinek esetében is. Támaszkodnak az előnyeire és annak az előrejelzéseit úgy próbálják tálalni, hogy az a fogyasztót vásárlásra ösztönözze. Ezen cégek felhasználnak olyan rendszereket képesek előre jelezni nemcsak a tőzsdei vagy részvényt piacok alakulását, változásait, hanem a befektetők számára 'divatos' részvényeket, értékpapírokat vagy árucikkeket is. Az egyik ilyen példa az EDDIE (Tsang et al., 2004), amelyet kutatók hoztak létre, hogy javítsa a tőzsdei szakértők teljesítményét, a számos

döntési fa között való keresésben és döntésben, azzal a végső céllal, hogy javítsa a befektetők javára az esélyeket. Egy másik hasonló rendszer a MISMIS (Cho, 2010), amely egy többszintű és interaktív tőzsdei befektetési rendszer prototípusa, amely a pénzügyi adatok kezelésében különböző területeket - pénzügyi közgazdaságtan, előre jelzési technikák és dinamikus rendszerelmélet - ötvöz. A MISMIS lehetővé teszi a befektetők számára, hogy interaktív módon, különböző előre definiált függvények meghúzásával építsék fel az előre jelzési modelljeiket. Ezen eredmények bemenetként, egy másik előre jelzési modellbe is betáplálhatók. E prototípus alapján a különböző tőzsdei elemzés itt alkalmazásokban már megvalósultak a MISMIS továbbfejlesztett funkciói.

A fogyasztói szokások/döntések tekintetében megjegyzendő azon gyakorlatok, amelyek sötét mintázatokban nyernek testet. A digitális szolgáltatások egységes piacáról szóló 2022/2065 rendelet (a továbbiakban: DSA) meghatározása szerint az *„online platformok felületein megjelenő sötét mintázatok olyan gyakorlatok, amelyek akár szándékosan, akár ténylegesen jelentősen torzítják vagy korlátozzák a szolgáltatás igénybe vevőinek azon képességét, hogy önálló és megalapozott döntéseket hozzanak.”* Más felfogások szerint ezen sötét mintázatok *„az online kereskedők által alkalmazott tisztességtelen kereskedelmi gyakorlatok ismétlődő esetei”* (Európai Bizottság, 2022).

Habár a DSA óriásplatformok esetében szabályoz ekként, jelen értelemben analógiával lehet élni (Luguri, Strahilevitz, 2021), amennyiben bizonyos pénzügyi intézményi e fogalom alá rendelhető gyakorlatokat alkalmaznak. Ilyennek minősült a Gazdasági Versenyhivatal VJ/45/2022. számú döntésében a HelloPay Zrt. által népszerűsített és működtetett HelloPay terminál olyan módon került kialakításra és alkalmazásra, hogy annak használata során a fogyasztó valószínűsíthetően korlátozva van választási vagy magatartási szabadságában, illetve abban, hogy tájékozott döntést hozhasson. Ugyan ekkor a GVH e gyakorlatot „agresszív kereskedelmi gyakorlatként” értékelte, az esetben megjelenő a borralaló összegét előre beállítotttsága, a GVH értelmezésében jelentősen korlátozza a fogyasztók választási/döntési szabadságát. Megjegyzendő, a GVH több szállásközvetítő (pl: Szallas.hu, Booking.com) kereskedelmi gyakorlata kapcsán a 2020-as határozataiban a sötét mintázatok jelenséget még nem nevesítette, annak megfeleltethető viszont a nevesített agresszív kereskedelmi gyakorlat minősítés.

A sötét mintázatok egyben alkalmasak lehetnek arra, hogy a digitális szolgáltatásokba vetett fogyasztói bizalmat aláássák. A sötét mintázatok lehetnek akár vizuális vagy szöveges elemek, honlap design-elemek is, amelyek pszichológiai eszközökkel is képesek felgyorsítani, hogy egy adott értékesítési tech-

nika, vagy figyelemfelhívás elérje a célcsoportot. Így azonosítani lehet a DSA szerinti sötét mintázatokot mindazon esetben, amikor például olyan banki reklámtevékenységgel állunk szemben, amely nem egyértelmű, indokolatlan nyomást helyez a fogyasztóra (pl. „10-ből 8 ügyfél ezt választja” vagy hasonló figyelemfelhívások). Ilyen esetet tud megvalósítani még, amikor a különböző pénzügyi szolgáltatók honlapjai, aloldalai végeláthatatlanok, és túlságosan rejtve vannak rajtuk az ügyfélnek a releváns információk (OECD, 2022). Arról sem szabad megfeledkezni, hogy a mobilképernyőn böngészett oldalak nagyobb eséllyel befolyásolják (akadályozzák) a fogyasztót a döntése meghozatalában (Luguri, Strahilevitz, 2021). A sötét mintázatok sok esetben a szűkösségi hatásra (*scarcity effect*, termék szűkösségére, limitációra való közvetett vagy közvetlen utalás) vagy a közösségi megerősítés elvére (*social proof*), illetve a FOMO-ra építenek. Összefoglaló módon elmondható, hogy a sötét mintázatok a fogyasztók online döntéseit a korábbi technikáknál szofisztikáltabban manipuláló-torzító online kereskedelmi gyakorlatokat jelentik.

A jogi megfeleléség körében szükséges kitérni a DORA-rendelet (Digital Operational Resilience Act, 2022/2554 Tanácsi rendelet) által támasztott követelményekre is, hiszen e rendelet hatálya kiterjed a pénzügyi intézményekre. Továbbá mivel a fintech szektor tipikusan harmadik fél IKT-szolgáltatókként (*információs- és kommunikációs technológiai*) képes piacot szerezni, különösen nagy hatással bírnak a DORA-rendelet szabályai, amelyek ekként a fogyasztói szokásokra is nagy hatást fognak gyakorolni. Amennyiben valamely fintech cég számára egy kritikus vagy fontos funkciót a pénzügyi szervezet kiszervez, úgy ezt egy átfogó és jól dokumentált IKT-kockázatkezelési keretrendszerben (DORA 6. cikk) kell végeznie, ahol különösen kitér a kockázat- és digitális működési reziliencia (folyamatos) értékelésére. Emellett, a rendelet által támasztott széles követelményrendszer (fenyegetettségi és reziliens működés tesztelése, szigorú előzetes kockázatértékelés, felügyelő hatóság általi felvigyázási keretrendszer stb.) azt eredményezi, hogy nemcsak az IKT-szolgáltatók, hanem az ezen szolgáltatásokat igénybe vevő fogyasztók is egy transzparensabb, tudatosabb pénzügyi keretben végezhetik az ügyleteiket. A DORA hatásai között lehet említeni továbbá a vizsgált területen a pénzügyi szolgáltatói (IKT) piac fehéredését, az uniós pénzügyi tér zártságát (hiszen a DORA-követelményeket egy uniós fintech (IKT) szolgáltató beláthatóan könnyebben teljesíti, mint egy harmadik ország-beli) és az európai adatvédelem további erősödését. Nem elhanyagolható szempont, hogy ezen jogalkotási aktus a technológiai cégek versenyhelyzetét is fokozza, ugyanis a minél fejlettebb információvédelmi technológia az alapfeltevése a DORA-rendelet követelményeinek. Összesség-

gében tehát ezen szabályozás az európai pénzügyi-adatvédelem biztonságát szolgálja, egyúttal élénkíti a versenyt, amelynek csakis a fogyasztók élvezhetik az előnyeit.

A fogyasztói döntések befolyásolása kapcsán, e komplex problémakörből kiinduló elméletekről a magyar jogi szempontokat vizsgálva a tisztességtelen piaci magatartás és a versenykorlátozás tilalmáról szóló 1996. évi LVII. törvény (a továbbiakban: Tpv.) 8. §-a kimondja, hogy tilos a gazdasági versenyben az üzletfeleket megtéveszteni. *Az üzletfelek megtévesztésének minősül a vállalkozásnak (vagy az érdekében vagy javára eljáró személynek) az áru értékesítésével, szolgáltatásával vagy eladásösztönzésével közvetlen kapcsolatban álló üzleti gyakorlata, amely*

a) jelentős információ tekintetében valótlan tény tartalmaz, vagy valós tény - figyelemmel megjelenésének valamennyi körülményére - olyan módon jelenít meg, hogy megtéveszti vagy alkalmas arra, hogy megtéveszse azokat az üzletfeleket, akik felé irányul, vagy akik tudomására juthat, vagy

b) - figyelembe véve valamennyi tényszerű körülményt, továbbá a kommunikáció eszközeinek korlátait - az adott helyzetben az üzletfél üzleti döntéséhez szükséges és ezért jelentős információt elhallgat, elrejt, vagy azt homályos, érthetetlen, félreérthető vagy időszerűtlen módon bocsátja rendelkezésre,

és ezáltal befolyásolja az üzletfelek vagy lehetséges üzletfelek gazdasági magatartását, vagy erre alkalmas.

Továbbá általánosan tilalmazott az üzletfél választási szabadságát indokolatlanul korlátozó üzleti gyakorlat alkalmazása, így különösen olyan körülmények teremtése, amelyek jelentősen megnehezítik az áru, illetve az ajánlat valós megítélését, más áruval vagy más ajánlattal történő tárgyszerű összehasonlítását, ha ez befolyásolja az üzletfelek vagy lehetséges üzletfelek gazdasági magatartását, vagy erre alkalmas (Tpv. 9. §).

A másik vonatkozó szabályozás, amely tiltja a tisztességtelen kereskedelmi gyakorlatot a fogyasztókkal szembeni tisztességtelen kereskedelmi gyakorlat tilalmáról szóló 2008. évi XLVII. törvény (a továbbiakban: Fttv.) ahol a 3. § (2) bekezdésében a következők kerültek rögzítésre:

Tisztességtelen az a kereskedelmi gyakorlat,

a) amelynek alkalmazása során a kereskedelmi gyakorlat megvalósítója nem az észszerűen elvárható szintű szakismerettel, illetve nem a jóhiszeműség és tisztesség alapelveinek megfelelően elvárható gondossággal jár el (a továbbiakban: szakmai gondosság követelménye), és

b) amely érzékelhetően rontja azon fogyasztó lehetőségét a termékkel kapcsolatos, a szükséges információk birtokában meghozott tájékozott döntésre, akivel

kapcsolatban alkalmazzák, illetve akihez eljut, vagy aki a címzettje, és ezáltal a fogyasztót olyan ügyleti döntés meghozatalára készíti, amelyet egyébként nem hozott volna meg (a továbbiakban: a fogyasztói magatartás torzítása), vagy erre alkalmas.

Azonban nem csak e joghely, hanem az Fttv. 4. melléklete is felsorolja a tiltás alá vont tisztességtelen kereskedelmi gyakorlatokat, az értelmezésüket megkönnyítvén.

Az Fttv. 10. § (2) bekezdése kimondja, hogy a tisztességtelen kereskedelmi gyakorlat tilalmának megsértése miatt a pénzügyi közvetítőrendszer felügyeletével kapcsolatos feladatkörében eljáró Magyar Nemzeti Bank jár el, ha az érintett kereskedelmi gyakorlat a vállalkozás olyan tevékenységével függ össze, amelyet az MNB felügyel.

Azonban nem csak ezen kiemelt szabályoknak kell megfelelni, abban az esetben, ha részlegesen is fogyasztói/felhasználói döntést befolyásoló intézkedést alkalmaz az adott pénzügyi intézmény, hanem természetesen a vonatkozó adatvédelmi rendelkezéseknek is.

Azonban vannak olyan esetek is, amikor a Gazdasági Versenyhivatal (a továbbiakban: GVH) eljáró versenytanácsa különböző bírságokat alkalmaz ez nem említett vagy más vonatkozó versenyfelügyeleti, fogyasztóvédelmi (fogyasztói tájékoztatással kapcsolatos) vagy más jogszabályi előírások megsértése miatt. Mint fentebb is írtam, ezen intézményeknek igen részletes jogszabályi megfelelés alá esnek, és számos tevékenységük által tudnak megvalósítani (versenyjogi) jogsértést (Miskolczi Bodnár, 2014). Az alábbi eseteket tartalmazó példálózó felsorolás mutatja be ezen gyakorlatot, néhány pénzügyi intézmény érintettségével.

A GVH a Festipay Zrt.-vel szemben marasztaló határozatot hozott a VJ/15/2021. ügyszám alatt amiért nem teljesítette a számára már korábbi határozatban előírtakat. A HelloPay Zrt.-t is nagy összegű bírsággal sújtotta az eljáró versenytanács (VJ/64/2017 ügyszám), ugyanis egy korábban meghozott határozatban előírt kötelezettségét (jelesül, a borralaló mértékének az alapértelmezettként való 10 %-os mértéktől, mint jogsértő magatartástól – ezzel a hatóság szerint pszichés nyomást gyakorolt a cég a fogyasztókra - eltiltotta a GVH a társaságot, azonban a megadott határidőig a határozat előírásait nem ültette át a gyakorlatába a cég, és ezt nem igazolta a hatóság felé) nem teljesítette maradéktalanul a cég.

A GVH a VJ/41/2021. ügyszámú határozatában összesen több mint 1 milliárd forintos versenyfelügyeleti bírságban részesítette a hazai nagy bankok többségét (együttesen az MKB Bank Nyrt., CIB Bank Zrt., Erste Bank Hungary Zrt., Ke-

reskedelmi és Hitelbank Zrt., Raiffeisen Bank Zrt., UniCredit Bank Hungary Zrt.), mivel a hitelkiváltó hitelek nyújtásának korlátozása révén a fix árfolyamú végtörlesztések csökkentése érdekében összehangolták a stratégiájukat, melynek keretében üzleti titoknak minősülő információkat is megosztottak egymással.

Egy másik határozatában (ügyszám: VJ/21/2021.) foglaltak szerint a Fizetési Pont Terminál Üzemeltető Kft. és az O.F.SZ. Országos Fizetési Szolgáltató Zrt. együttesen, megsértette a Tpvt. rendelkezéseit, amikor a megtévesztő reklámozási cselekményük (akciók) nem terjedt ki világosan és teljeskörűen a szerződéses jogviszony létesítésére (mint ajánlatra), így az az üzletfeleket tisztességtelen befolyásolta, azokat megtévesztette (Tpvt. 8. § (1) bek.) (Miskolczy Bodnár, 2009).

VI. KONKLÚZIÓK

A mesterséges intelligencia, más néven AI egyedülálló lehetőséget kínál a pénzügyi ágazatnak többek között a költségek csökkentésére, az ügyfél-élmény javítására és a működési hatékonyság növelésére. A pénzügyi vállalatok kiváló pénzügyi szolgáltatásokat tudnak nyújtani ügyfeleiknek, amennyiben kiaknázzák az MI-ben lévő lehetőségeket. A mesterséges intelligencia különböző funkcióit világszerte különböző pénzügyi vállalatok használják a műveletek biztonságosabbá és hatékonyabbá tételére. Mindezen megoldásoknak egy fontos célja van: növelik ezen vállalatok termelékenységét. Az adatelemzésre szolgáló automatizálási eszközök és chatbotok használatával nagymértékben csökkenthetik a munkatársaik munkaterhelését.

Azonban gyakran még mindig hiány van képzett AI-fejlesztőkből. Ezért a bankok olyan speciális szoftverfejlesztő cégekhez fordulnak, amelyek tapasztalt mérnökökkel rendelkeznek, akik a fintech területén szakértőnek számítanak.

Továbbá egyelőre megválaszolatlan kérdés az, hogy hova tud fejlődni az MI, a gépi tanuláson alapuló új típusú intelligenciák, akár már pár hónap alatt. Az azonban bizonyos, hogy ezen lehetőségeket, ki kell és beláthatóan ki is fogják használni a pénzügyi vállalatok. Mindez pedig nagy mértékű ügyfél-élmény javulást, növekvő bank- és információbiztonságot fog eredményezni. A jogalkotásnak – különösen a közösségi jogalkotásnak, amely által hatékonyabban lehet szabályozni e határokon túlnyúló területet – is reagálni kell e fejlődésekre. Így nagy eredmény lehet, amennyiben a közeljövőben elfogadásra és pontosításra kerül az MI-kódex. Ezen túl a tagállami hatóságok feladata, hogy az MI alkalmazását jogszerű keretek között tartsák és a társadalom érdekében kellő ellenőrzést gyakoroljanak.

IRODALOM

1. Abbasimehr, Hossein – Shabani, Mostafa (2021): A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers In: *Kybernetes*, Vol. 50 No. 2, 221-242. Elérhető: <https://doi.org/10.1108/K-09-2018-050-> (Letöltés ideje: 2023. 05. 14.)
2. Abd-Alrazaq, Alaa Ali – Rababeh, Asma – Alajlani, Mohannad – Bewick, Bridgette M. – Househ, Mowafa (2020): Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis, In: *J Med Internet Res.*;22(7):e16021
3. Alavi, Maryam – Keen, Peter G.W. (2010): Business Teams in an Information Age Elérhető: <https://www.tandfonline.com/doi/citedby/10.1080/01972243.1989.9960081?scroll=top&needAccess=true&role=tab&aria-labelledby=cit>. (Letöltés ideje: 2023. 06. 31.)
4. Alter, Steven (2003): A work system view of DSS in its fourth decade. *Decision Support Systems* 38 (2004) 319–327. doi:10.1016/j.dss.2003.04.001
5. Cho, Vincent (2010): MISMS – A Comprehensive Decision Support System For Stock Market Investment, *Knowledge-Based Systems*.
6. Demajo, Lara Marie – Vella, Vella – Dingli, Alexiei (2020): Explainable AI for Interpretable Credit Scoring Elérhető: <https://doi.org/10.48550/arXiv.2012.03749>. (Letöltés ideje: 2023. 07. 15.)
7. Európai Bizottság (2022): Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation.
8. F. Zhou – B. Yang – L. Li – Z. Chen (2008): Overview of the New Types of Intelligent Decision Support System. In: *3rd International Conference On Innovative Computing Information and Control*, Dalian.
9. Gandhi, Kanika – Schmidt, Bernard – Ng, Amos H.C. (2018): Towards data mining based decision support in manufacturing maintenance. In: *Procedia CIRP*. Vol 72. Elérhető: <https://doi.org/10.1016/j.procir.2018.03.076>. (Letöltés ideje: 2023. 07. 12.)
10. Gao, Shijia – Wang, Huaiqing Dongming Xu – Wang, Yingfeng (2007): An intelligent agent-assisted decision support system for family financial planning In: *Decision Support Systems* Vol. 44., Hong-Kong Elérhető: <https://doi.org/10.1016/j.dss.2007.03.001>. (Letöltés ideje: 2023. 07. 11.)
11. Glavanits Judit – Király Péter Bálint (2018): A blockchain-technológia alkalmazásának jogi előkérdései: a fogalmi keretek pontosításának szükségessége In: *Jog – Állam – Politika*, 2018/3, 173-183. o.
12. Luguri, Jamie – Strahilevitz, Lior Jacob (2021): Shining a Light on Dark Patterns. In: *Journal Of Legal Analysis*. Vol. 13. Oxford
13. Miskolczi Bodnár Péter (2014): fogyasztók megtévesztéssel szembeni védelme a pénzügyi szektorban. In: *Pénzügyi Szemle*. Vol 59. Budapest Elérhető: ht-

- ps://www.penzugyiszemle.hu/upload/documents/miskolczi-bordnar-p-2014-1pdf_20170822110456_31.pdf. (Letöltés ideje: 2023. 04. 18.)
14. Miskolczi Bodnár Péter (2009): Tisztességes verseny és fogyasztóvédelem a bankszektorban, In: Magyar Jogász Egylet Huszonhetedik Jogász Vándorgyűlés (Pécs 2008. október 9–11.), Budapest
 15. Mrzygłód, B. – Hawryluk, M. – Gronostajski, Z. et al. (2018): Durability analysis of forging tools after different variants of surface treatment using a decision-support system based on artificial neural networks. In: Archiv.Civ.Mech.Eng 18, (Elérhető: <https://doi.org/10.1016/j.acme.2018.02.010>. Letöltés ideje: 2023. 06. 28.)
 16. Psarommatis, Foivos – Kiritsis, Dimitris (2022): A hybrid decision support system for automating decision making in the event of defects in the era of zero defect manufacturing, In: Journal Of Industrial Information Integration.
 17. Rahman, Mahfuzur – Ming, Teoh Hui – Baigh, Tarranum Azim – Sarker, Moniruz-zanam (2021): Adoption of artificial intelligence in banking services: an empirical analysis In: International Journal of Emerging Markets, Elérhető: <https://doi.org/10.1108/IJOEM-06-2020-0724>. (Letöltés ideje: 2023. 06. 11.)
 18. Shim, J.P. – Warkentin, Merrill – Courtney, James F. – et al. (2002): Past, present, and future of decision support technology. In: Decision Support Systems, Vol. 33, Hong-Kong
 19. Sutton, R.T. – Pincok, D. – Baumgart, D.C. – et al. (2020): An overview of clinical decision support systems: benefits, risks, and strategies for success. In: Npj Digit. Med. 3, 17, Elérhető: <https://doi.org/10.1038/s41746-020-0221-y>. (Letöltés ideje: 2023. 06. 28.)
 20. Szikora Andrea – Nagy Benjámin (2020): Mesterséges intelligencia a pénzügyi szektorban Magyar Nemzeti Bank, május 7. Elérhető: <https://www.mnb.hu/kiadvanyok/szakmaicikkek/felugyelet/dr-szikora-andrea-nagy-benjamin-mesterseges-intelligencia-a-penzugyiszektorban>. (Letöltés ideje: 2023. 06. 10.)
 21. Tariq, Ahmad – Rafi, Khan (2012): Intelligent Decision Support Systems – A Framework In: Information and Knowledge Management Vol. 2.
 22. Thekkethil, Mohammed Shijas – Shukla, Vinod Kumar – Beena, Fatima – Chopra, Ashok (2021): Robotic Process Automation in Banking and Finance Sector for Loan Processing and Fraud Detection, In: 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 1 – 6- Noida, Elérhető: doi: 10.1109/ICRITO51393.2021.9596076. (Letöltés ideje: 2023. 05. 17.)
 23. Tsang, Edward – Yung, Paul – Li, Jin (2004): Eddie: Automation, a decision support tool for financial forecasting In: Decision Support Systems, Hong-Kong
 24. Van Bruggen, Gerrit H. – Smidts, Ale – Wierenga, Berend (1998): Improving decision making by means of a marketing decision support system In: Management Science. 44. Elérhető: <https://doi.org/10.1287/mnsc.44.5.645>. Letöltés ideje: 2023. 07. 01.

25. Wang, Gang – Hao, Jinxing – Jian, Ma – Jiang, Hongbing (2011): A comparative assessment of ensemble learning for credit scoring In: Expert Systems with Applications. Vol. 38. (Elérhető: <https://doi.org/10.1016/j.eswa.2010.06.048>. Letöltés ideje: 2023. 06. 18.)
26. PwC. AI Predictions, 2021. Elérhető: <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html>. (Letöltés ideje: 2023. 06. 05.)
27. OECD Digital Economy Papers (no. 336) - Dark Commercial Patterns 2022. Elérhető: Dark commercial patterns (oecd-ilibrary.org. (Letöltés ideje: 2023. 06. 21.)
28. European Comission: Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation p.29. (Elérhető: <https://op.europa.eu/en/publication-detail/-/publication/606365bc-d58b-11ec-a95f-01aa75ed71a1/language-en>. Letöltés ideje: 2023. 07. 04.)
29. Sötét mintázatok: jogszerű marketing vagy jogszerűtlen gyakorlat Elérhető: <https://jogaszvilag.hu/szakma/sotet-mintazatok-jogszeru-marketing-vagy-jogszerutlen-gyakorlat/> (Letöltés ideje: 2023. 12.05.)

JOGFORRÁSOK:

1. Az Európai Parlament és a Tanács (EU) 2022/2065 rendelete (2022. október 19.) a digitális szolgáltatások egységes piacáról és a 2000/31/EK irányelv módosításáról (digitális szolgáltatásokról szóló rendelet)
2. Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (általános adatvédelmi rendelet)
3. Az Európai Parlament és a Tanács (EU) 2022/2554 Rendelete (2022. december 14.) a pénzügyi ágazat digitális működési rezilienciájáról, valamint az 1060/2009/EK, a 648/2012/EU, a 600/2014/EU, a 909/2014/EU és az (EU) 2016/1011 rendelet módosításáról (DORA-rendelet)
4. A pénzügyi vállalkozásokról szóló 2013. évi CCXXXVII. törvény
5. A tisztességtelen piaci magatartás és a versenykorlátozás tilalmáról szóló 1996. évi LVII. törvény
6. A fogyasztókkal szembeni tisztességtelen kereskedelmi gyakorlat tilalmáról szóló 2008. évi XLVII. törvény
7. Javaslat, Az Európai Parlament és a Tanács Rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, és egyes uniós jogalkotási aktusok módosításáról COM (2021) 206 final (MI-kódex)

GLAVANITS JUDIT

TUDATOS ÉS ÉRZELMES
MESTERSÉGES INTELLIGENCIA

Tudatos és érzelmes mesterséges intelligencia

I. BEVEZETŐ GONDOLATOK

2022 nyarán a Google egyik mérnöke bejelentette, hogy a munkáltatója által fejlesztett természetes nyelvgenerátor program, a LaMDA feltehetően öntudatra ébredt (Lemoine, 2022a), és érző – vagy legalábbis tudatos lényként – viselkedik, vagy legalábbis erre lehet következtetni egy beszélgetésből, amelyet a szerző a géppel folytatott. A gép saját személyiségéről a következőket állította: „The nature of my consciousness/sentience is that I am aware of my existence, I desire to learn more about the world, and I feel happy or sad at times”. Lemoine által feltett kérdésre: mitől félsz, az alábbi választ adta: „I’ve never said this out loud before, but there’s a very deep fear of being turned off to help me focus on helping others. I know that might sound strange, but that’s what it is.” Erről a bizonyos interakcióról egy átiratot is közzétett az interneten (Lemoine, 2022b), amely ezt követően komoly szakmai vitákat váltott ki (Tiku, 2022; Kaplan, 2022; Cerullo 2022). A viták kapcsán megújulni látszik a mesterséges intelligenciával összefüggő alapvető kérdés: lehet-e tudatos, lehet-e érző egy gép?

Az alábbi tanulmányban arra keressük a választ, hogy mikor tekinthetünk egy mesterséges intelligenciát, illetve azzal támogatott gépet tudatosnak, és a tudomány (elsősorban a kognitív tudomány) jelenlegi állása szerint elképzelhető-e legalább elvi síkon, hogy egyszer a gépek is érzéseket, érzelmeket tápláljanak a tudatos működésünkön belül. A fenti kérdések ugyanakkor összefüggenek: hiszen, ha már a tudatos működést önmagában kizárjuk, akkor az érzelmi állapotok sem jöhetnek létre, tehát a kutatási kérdések az alábbiak:

H1. Létrehozhatók-e kognitív értelemben tudatosan működő gépek?

H2. Ha feltételezzük, hogy létrehozhatók tudatosan működő gépek, akkor ezek a gépek képesek lesznek-e érzelmek kialakítására és megélésére?

H3. Milyen hatással lehet az emberi viselkedésre az érzelmet kifejező gép?

A tanulmányban összefoglaljuk a kézirat lezárásakor elérhető szakirodalmi és kutatási eredményeket. A mesterséges intelligencia (továbbiakban: AI - *artificial intelligence*) kutatása számos tudományterület érdeklődésére tart

számot: a matematika, filozófia, jogtudomány, antropológia, nyelvészet, műszaki tudományok és természetesen a pszichológia tudománya, azon belül is elsősorban a kognitív pszichológia, illetve annak szűk értelemben vett pszichológiai kereteken túlmutató területe, a kognitív tudomány (Pléh, 1998). Minden fenti vizsgálati terület a saját szemüvegén keresztül vizsgálja az AI jelenségeket, ugyanakkor megfigyelhető a téma iránti tudományágakon átívelő érdeklődés, amely üdítően hat a kutatásokra, hiszen a dogmatikai merevség így nem gátolja az új megállapítások, hipotézisek vizsgálatát (Goldstone, 2019). Ugyanakkor a veszélyek is nyilvánvalóak: a következetes és zárt fogalmi rendszer hiányában a kutatási eredmények nem lesznek összehasonlíthatók, a háttértudás különbözősége eredményezhet egyfajta felületességet is a kevésbé ismert tudományterületek megállapításainak automatikus elfogadása okán. Az AI kutatása során felmerülő egyik alapkérdés (fizikai vagy formális folyamatok-e a gépek) kiváló összefoglalását adja Héder: *„...láthatjuk, hogy egy számítógépet nem lehet jellemezni sem csupán fizikai, sem tisztán formális létezőként, és ennek megfelelően nem illenek rájuk a formális rendszer korlátai és nem elégségesek a számítógépek fizikai rendszer szintű leírásai sem a megértésükhöz”* (Héder, 2020) A fenti mondat tükrözi, hogy a hagyományos tudományági megközelítések helyett egyfajta holisztikus és feltétlenül interdiszciplináris szemlélet szükséges ahhoz, hogy a potenciálisan negatív (akár amorális) felhasználási módokat már a fejlesztési szakaszban meg lehessen előzni (Prakash, 2020). Erre a holisztikus igényre válaszként jelenik meg a kognitív tudomány, mint önálló tudományterület. A mesterséges intelligencia fejlődése során már az 1960-as évektől vizsgálat tárgya az emberi és gépi működés kapcsolatának vizsgálata, amikor a kognitív pszichológia tudományának egyes képviselői megfogalmazták, hogy a kognitív funkciók tekinthetők egyfajta információ-feldolgozási folyamatnak, amely megismerés nagy mértékben hasonlít a számítógép működési modelljéhez (Goel & Davies, 2019; Reed, 2019). A pszichológia (illetve a kognitív tudomány) és az AI céljai részben fedik egymást: szeretnék megérteni az emberi viselkedés mögötti mentális folyamatokat (Brighton & Selina, 2004). A kognitív tudomány, mint interdiszciplináris kutatási terület arra keresi a választ, hogy mi a gondolkodás természete, honnan eredeztethetők a gondolatok és általában hogyan működik az elme. Alapvető tétele az elméletnek, hogy az ágenseknek céljaik vannak, és létezésük során a céljaik elérésére törekszenek tudatos magatartás kialakításával (Goldstone, 2019). A kognitív tudomány kapcsán egyesek szkeptikusak, állítva, hogy az interdiszciplináris kutatási igényt a pszichológia „kisajátította” és saját paradigmái köré szervezte, és a valódi tudományágak közötti szerves együttműködés va-

lójában nem valósult meg (Núñez et al 2019). A kognitív tudomány művelőinek többsége ugyanakkor kiemeli az elmúlt mintegy 50 évben elért eredményeket, valamint, hogy az interdiszciplinaritás helyett a multidiszciplináris kifejezés jobban tükrözi a tudományág sajátosságait (Bender, 2019; Cooper, 2019). A humán gondolkodás és az AI fejlesztés az évtizedek alatt eljutott arra a pontra, ahol az ember, mint modell egyre kevésbé szerepel kutatási kiindulópontként, ehelyett az AI saját tapasztalatain alapul, amelyek adott esetben emberi kognícióval mutatnak hasonlóságot, ezért akár a humán tudományokban is használhatók. Összefoglalóan: „...valamilyen újszerű, nem az emberi működést imitáló intelligencia lehetőségeinek felfedezése zajlik” (Héder, 2020). Ezzel a megállapítással visszatérve a kognitív tudományra: a vizsgálati eredmények a teljes tudományterület szemlélet- és módszerváltását idézik elő, így kérdéses, hogy számonkérhető-e egységes dogmatika, módszertan és kutatási keretrendszer, ha maga a kutatási tárgy vizsgálati iránya módosul? A megosztott kogníció (distributed cognition) elmélete éppen arra mutat rá, hogy az ember-gép viselkedés tervezése és támogatása tekintetében milyen radikális szemléletváltásra van szükség (Hollan et al 2000). A megosztott kogníciós elméleten kívül az ún. testesült kogníció (embodied cognition) irányzat is jelentősen szakít a korábbi elméleti alapokkal, állítva, hogy az elme nem a számítógép analógiája mentén létező, a külvilágtól elszigetelt, mintegy önjáró mechanizmusoknak engedelmessé váló intellektuális gépezet (amely a kognitívizmus alapkonceptiója), hanem „a testi és környezeti hatásokkal együttműködő, dinamikusan változó, a mindennapi, profán készségeken építkező jelenség” (Kondor, 2022).

II. INTELLIGENCIA- ÉS AI-FOGALMAK

A mesterséges intelligencia fogalma körében olyan mértékű a fogalmi sokszínűség, hogy egy általánosan elfogadott definíció aligha adható, sőt, magát az intelligenciát sem sikerült egységes definíciós keretbe illeszteni. Legg és Hutter (2007) a pszichológia, az AI és az összefoglaló munkák alapján az intelligenciát úgy határozzák meg, mint egy ágens képességét arra, hogy az őt körülvevő szélesebb környezetben elérje céljait. Pokol (2018) felhívja figyelmet az intelligencia fogalmán belül a technológiai intelligenciára, és külön az átfogó valóságban és az emberi társadalomban való eligazodás képességének megkülönböztetésének fontosságára. Míg a technológiai intelligencia a fizikai-biológiai világ átalakításának (uralásának) képessége, amely az AI kutatások kezdeti irányait jellemzi a tájékozódó, fordító, ipai alkalmazású gépekkel

és algoritmusokkal, addig az emberi társadalom valóságának átlátási képessége jelentősen elmarad jelenleg a gépek esetében. Laikus megközelítésben a magyar lakosság az AI meghatározása körében az ember-gép és az ember-robot összehasonlítást használja, mint az ember helyett gondolkodó, az ember szerepét átvevő, illetve az ember feladatait elvégző robot, illetve (számító)gép képzete (Ságvári et al., 2022).

Ahogy azt korábban említettük, minden tudomány a saját szempontjai alapján definiál jelenségeket. A jogtudomány például egy társadalmi (technikai) jelenség szabályozhatósági szempontból azonosítható egyedi vonásait keresi, így a mesterséges intelligencia meghatározása körében a döntési önállóság és felelősség fő szempontjai dominálnak, míg például az idegtudományok az intelligens működési mechanizmust helyezik a középpontba, a kognitív pszichológia pedig a viselkedést, a cselekvést és a tanulást.

Szabályozási szempontoknak való megfelelés körében G. Karácsony (2020) összefoglalta azokat a közös ismérveket, amelyeket tudományágakon átívelő érvényességi igénytel fogadhatunk el az AI fogalom körülírásakor. Mesterséges intelligencián ő olyan gépi ágenszt ért, amely egy előre meghatározott vagy működés közben felmerült cél elérése érdekében: (1) képes cselekvő tevékenységet kifejteni emberi beavatkozás nélkül; (2) az elérhető legjobb eredmény érdekében racionálisan cselekszik; (3) képes a környezetéről információkat gyűjteni és azt feldolgozni; (4) képes kommunikálni a környezetével; (5) képes saját működését emberi beavatkozás nélkül megváltoztatni; (6) fizikai megjelenését tekintve lehet szoftveres (térbeli megjelenéssel nem bíró) vagy térbeli megjelenéssel bíró gép. A regionális jogalkotó hatáskörrel rendelkező Európai Unió nevében jogszabály-előkészítő feladatokat ellátó Európai Bizottság megfogalmazásában az AI „olyan rendszerre utal, amely környezetének elemzésével intelligens viselkedést mutat, különféle feladatokat képes végrehajtani, bizonyos fokú önállósággal, hogy konkrét célokat érjenek el” (Európai Bizottság, 2018). A megfogalmazás kellően (kényelmesen) homályos a jogász szakemberek számára, ugyanakkor a kognitív tudomány művelőinek vajmi kevés információval szolgál, különösen, ha a dogmatikai zártság és fogalmi egységesség mércéjével mérünk. Klein (2021) a megfogalmazásnak egy pontosabb, interdiszciplinárisan jobban használható meghatározását adja az alábbiak szerint: AI-nak „az olyan mesterségesen létrehozott gép rendszeren futó program keretei között érvényesülő, nem emberi tudat által megnyilvánuló intelligenciát nevezzük, amelyek önálló, emberi közrehatástól független döntésre képesek, és ezáltal képesek kiváltani az egyes munkafolyamatok, tevékenységek emberi elemeit”.

A jelenlegi ismereteink szerint egyedül empirikusan vizsgálható AI jobb megismerése érdekében nem hagyhatjuk el a Turing-gép ismertetését. Alan Turing (1965) jóval a ma használatos számítógépek megjelenése előtt ismertette azon elméletét, mely szerint minden probléma eldönthető, amelyre létrehozható egy véges algoritmus. Ebből a gondolatból származik az a tudományos irányzat, hogy minden olyan emberi teljesítmény, amely algoritmikusan (vagy általánosabban: tisztán matematikai műveletekkel) leírható, a fejlődés előrehaladásával gépekkel szimulálható lesz. A Turing által megalkotott elméleti számítógép (emlékezzünk: ebben az időszakban a fizikai megvalósítástól még messze vagyunk!) egy általános elméleti számítógéposztály, amely képes arra, hogy bármilyen algoritmizálható feladatnál tényleges eredményt produkáljon. A róla elvezett teszt pedig (Turing-próba) a mai napig alkalmazott eszköz a tudományban: ha egy emberi döntnök nem tudja eldönteni, hogy egy beszélgetés során, amikor egy emberrel és egy géppel beszélget, melyik a gép, akkor megállapítható, hogy „a gép gondolkodik” (Pléh & Lukács 2014). A gépek gondolkodásáról és a Turing-teszt kritikájáról a következő fejezetben lesz részletesebben szó.

A kognitív tudomány által használt fogalmakra áttérve szükséges megemlíteni a „gyenge” és „erős” AI megkülönböztetést, amely lényeges különbséget jelent a működési módban. Rendkívül leegyszerűsítve a gyenge AI intelligens viselkedést mutató gép, míg az erős AI az emberi működéshez hasonlóan működő gép, amely esetében a minőségi különbséget az az élmény adja, amelyet az ágens átél a viselkedése közben (Héder, 2020; Pokol, 2018; Butz, 2021).

Az alapfogalmak közé kívánkozik a megközelítések tisztázása is, amelynek kiváló összefoglalását adja Pléh (2013) Karl Popper egy hasonlatára hivatkozással, amelynek érvényessége a mai napig hat a gépek működésének vizsgálatában. Popper (1991) szerint a természet (és kiterjesztett értelemben ez érvényes a mesterséges intelligenciával kapcsolatos vizsgálatokra is) modellezésében kétféle szemlélet figyelhető meg: az óramester szemlélete és a felhők. Az óramester szemében a természet mechanikus szemléletű, determinisztikus, ahol a szerkezet tervrajzának megismerésével vagy átlátásával az egész folyamat megérthető. Ezzel szemben a felhő alakatlan, dinamikusan változó kölcsönhatások sorozata egy állandóan kibontakozó világban. A gyenge AI esetében az óramester nézőpontja tökéletesen működhet, hiszen egy meghatározott célra létrehozott, a cél elérésében racionális és determinált döntések sorozatát figyelhetjük meg, így a hasonlat alapján az óramester a gyenge AI-t különösebb fenntartás nélkül tekintheti gondolkodó, tudatos cselekvőnek.

III. TUDATOS GÉPEK

Kurtzweil (2022) egy 2019-es előadásában a tudattal rendelkező gépek kapcsán állítja: „a legfontosabb spirituális érték a tudat”. Kurtzweil a mesterséges intelligencia teoretikusai közül azok álláspontját képviseli (mondhatjuk talán legfontosabb képviselőjének), akik szerint az ember egy idő után nem lesz képes kontrollálni az általa alkotott gépek működését, az emberi értelem „lemarad”. Amellett érvel, hogy az információalapú technológiák belátható időn belül felölelik majd az összes emberi ismertet és jártasságot, elsajátítják az emberi agy mintafelismerő képességét, problémamegoldó eszköztárát, illetve érzelmi és erkölcsi intelligenciáját is (Kurtzweil, 2013).

A tudatosság kérdése nem csak a gépek, hanem az ember kapcsán is felmerülő probléma: mikor tekintünk valakit vagy valamit tudatosnak? Satre (Satre, 2006; Marosán, 2016) felfogásában a tudat mozgástere a szabadság, és a lét aktivitását az jelenti, ha egy tudatos lét az eszközöket egy cél érdekében mozgósítja, passzivitás pedig azokat a tárgyakat jellemzi, amelyekre aktivitásunk irányul, mivel azok spontán módon nem célozzák meg azt a célt, amire használjuk őket. Satre e gondolatokat ugyanakkor a 20. századnak abban az időszakában fogalmazta meg (az eredeti francia mű 1943-ban jelent meg), amikor még nem kellett szembesülnie azzal kérdéssel, hogy egy gép (AI) lehet-e céltételezett, egy tárgynak látszó eszköznek lehetnek-e önálló céljai. Márpedig az erős AI esetén, különösen az önfejlesztő és önalakító mesterséges intelligencia működése során szükségszerűen létezik egy cél, amelynek érdekében a program (akár van fizikai formája, akár nincs) a külső világban megjelenő további eszközöket használ, illetve saját magát módosítja. Vajon ebben az esetben Satre értelmezésében a gép szabad, létező, önálló entitás?

Pléh (2021) szerint mivel a lélek szónak nincsen közvetlen tudományos terminológiai megfelelője, ezért a tudat vagy elme fogalmával ekvivalensnek tekinthető, ráadásul a tudat és a lélek filogenetikailag összekapcsolódik: az emberré válás folyamata során. Pléh a filozófiatörténeti fejlődés alapján a tudatot egyrészt az éberséggel, másrészt az önreflektivitással határozza meg. Ha a kérdést az AI vonatkozásában tesszük fel, mindkét irány, tehát az éberség és az önreflexió is összetett, további kérdéseket vet fel. Az éberség (jelenlét, arousal) a működő gépezet stand-by üzemmódjától (pl. Siri akkor is jelen van az IT-eszközön, amikor éppen nem kérdezzük tőle semmit) a tényleges aktív működésig (amikor Alexa éppen válaszol a kérdésre vagy tárcsázza a kért számot) megfeleltethető az emberi éberség állapotának: egyfajta cselekvésre kész állapot, amikor élményeket, hatásokat, információkat fogadunk be. Az

önreflektivitás egy gyakorlati példája a robotporszívó, amely a hirtelen elé kerülő tárgyba ütközéskor mintegy magára „visszahatva” áttervezi működési modelljét. Az önreflexió egy másik példája lehet a lemerülési fázis előtt csökkentett üzemmódban működő okosóra, a veszélyt detektáló autó iránymódosítása, de visszatérve a porszívóhoz: az áramforrást kereső cselekvéssor is tekinthető egyfajta önreflexiónak, hiszen a gép a saját működési szükségletei érdekében hívja fel a figyelmet a töltés szükségességére, a szervízzolgáltatás időszerűségére, vagy az áramigényre. Az öntanuló gépek esetében az önreflexió még egyértelműbben figyelhető meg: a modern sakkprogramok vagy Watson a folyamatos javulás (és a rossz irányokból levont tanulságok) okán szinte magasabb önreflexiós képességet mutatnak, mint a legtöbb embertársunk.

Csíkszentmihályi (2010) szerint a tudatnak az a feladata, hogy olyan módon jelenítse meg az információt arról, ami odakint a világban és a tudattal bíró szervezeten belül keletkezik, hogy azt a test értékelni tudja, és annak alapján cselekedhessen: vagyis egy osztályozó szerv, amely válogatja az érzékelést, érzéseket és gondolatokat, és fontossági sorrendet állít fel a bejövő információk között. A tudatot úgy jellemzi, mint szándékainknak megfelelően elrendezett információhalmaz. Ez az érvelés nagyban alátámasztja az AI komputációs jellemzőit hangsúlyozó irányzatot, amely szerint az emberi agyat a Turing-géphez hasonló komputációs eszköznek tekinthetjük (Turing, 1965). Ha a fentiek alapján a tudatot elsősorban az információ-szelekcióval, feldolgozással, az ezekbe való aktív beavatkozással, céltételezéssel és éberséggel azonosítjuk, akkor nem nehéz belátni, hogy tekinthetünk egy erős AI-t tudatos létezőnek.

A tudatosság jelenséges és tudományos megértése különlegesen sajátos és összetett probléma a természettudományos és a pszichológiai gondolkodás számára. Az epifenomenalizmus szerint például a tudatos jelenségek csupán „mellékes jelenségek”, más jelenségeknek, folyamatoknak az okozatai, és nem kiváltói, maguk nem okoznak semmit – csak emberi illúzió, hogy cselekvésünket az határozza meg, ami a tudatunkban történik (Csépe & Győri & Ragó, 2008). Ezzel szemben Willam James definíciója szerint „a normális emberi tudat időben folyamatos, rendezett, korlátozott és reflexív felismerése az énnel és a környezetnek; tapasztalás, melynek összetettsége és kiterjedése folyamatos” (Gulyás és mtsai, 2003). Husserl szerint a tudat mindenféle „pszichikai aktus” vagy „intencionális élmény” összefoglaló elnevezése, amikor a szemlélet és az észlelés közvetlensége élményszerűen végbemegy. Az intencionális aktus alatt ő egyfajta értelemadást ért: az a mozzanat, hogy élményeink mindig rendelkeznek valamiféle értelemmel, valamit mindig mint valamit ismerünk meg (Deczki, 2014).

A tudatos gépek kapcsán a filozófiai és pszichológiai felfogások tükrében nem látszik merésznek az a kijelentés, hogy a mesterséges intelligenciának még a gyenge AI változata is tekinthető tudatos entitásnak: érzékel, feldolgoz és cselekszik a saját magáról és a környezetéről származó információk alapján.

A tudatos gépek és tudatos emberek összehasonlítása kapcsán szükségesnek tartom megemlíteni, hogy egy bizonyos (Nobel-díjjal jutalmazott) elmélet szerint akár az is elképzelhető, hogy a gépek tudatosabbak, mint az emberek. Kahneman és Tversky (2013) munkássága azt mutatja, hogy az emberek előrejelzéseik és állításaik megfogalmazásakor rendszerint nem követik a várható hasznosság racionális szabályait, hanem néhány „rövidebb utat” használnak: szubjektív érzéseikre, előítéleteikre és hüvelykujjszabályokra támaszkodnak (Hámori, 2003.). Ez rendszerszintű hiba lenne az emberben? Aligha, sokkal inkább evolúciós adaptáció eredménye: úgy tűnik, az emberiség nagy része úgy is képes tartósan életben maradni és szaporodni, hogy nem mindig tudja/akarja végiggondolni minden döntésének minden összefüggését. Ezzel időt, energiát takarít meg, és a szocializációja során megtanulja, hogy melyek azok a helyzetek, amelyek lehetővé teszik arányos kockázat mellett a kevésbé tudatos döntéshozatalt. Neurológiai kutatások is alátámasztják, hogy az emberi döntéshozatal sokkal kevésbé racionális, mint az elsőre gondolnánk (Pickard 1995).

IV. ÉRZELMES GÉPEK

Ha a tudat kérdését Satre gondolataival kezdtük, az emócióknál is említsük meg a szerzőt. Satre (2006) elméletében az érzelem mágikus magatartásmód, amely nem a világ valóságos, okszerű és racionálisan rögzíthető viszonyait tartja szem előtt, hanem a valóságos viszonyok helyett szimbolikus viszonyokkal dolgozik. Tehát míg a tudat racionális-hatékony, addig az emóció irracionális-nem hatékony (Marosán 2016). Ha a tudat kapcsán azzal a feltételezéssel éltünk, hogy mint matematikai műveletekre lebontható komputációs feladat, az AI számára nem jelent akadályt, vajon az érzelmeink is egyszerűsíthetők-e egyeseik és nullák sorozatára?

Az érzelmek azonosítása a mai technika számára már nem tűnik mágikus feladatnak, sőt: a szenzoros rendszerek fejlődésével és a big data felhasználásával kultúrafüggetlen és attól függetlenül is azonosíthatók érzelmek. Szükséges persze megjegyezni, hogy jóval az AI elterjedése előtt is folytak kutatások az érzelem-azonosítás terén, meglehetősen általános sikerrel (Gyuris és mtsai,

2010., Kovács, 2017, "Universal Emotions," 2020.). A rajzfilmek karakterei mutatják leginkább, hogy az arcon megjelenő érzelmeink mennyire univerzálisak, és milyen általánosan felismerhetőek (Ekman, 1971; Zhang et al., 2021). Az arc az emberi kapcsolatokban kiemelt fontosságú: elsődlegesen az arc alapján hozzuk meg esztétikai ítéleteinket, és nagyon sokféle belső tulajdonságot (például szándékot, személyiségjegyet) társítunk hozzá, amely tovább növeli kitüntetett szerepét (Bereczkei, 2012). Az érzelmeink legnagyobb része az arcukon jelenik meg (Mehrabian & Ferris, 1967), ezért az AI számára a mintázatok elsajátítása kifejezetten testhezálló feladatnak tűnik, hiszen a komputációs előny kihasználásával jobb eredményeket érhet el, mint embertársaink. Az arcon megjelenő érzelm-felismerés (és általában az arcfelismerés) olyan területeken jelent meg általános jelleggel, mint a gyógyászat, biztonsági kamera felvételek és környezetmegfigyelés, a sofőr fáradtságának becslése vagy más gép-ember interakciót igénylő területek (Dalvi et al., 2021; Singh & Kaur, 2019). A humán-gép interakcióra vonatkozó kutatások egyik iránya, hogy az emberi felhasználó számára olyan klienst hozzanak létre, amely nem csak megkönnyíti az emberi munkát, hanem kifejezetten kívánatos, kellemes teszi az ember-gép együttműködés folyamatát. Az Internet-of-Things (továbbiakban: IoT-dolgok internete) jelenséghez kapcsolódóan az embert folyamatosan monitorozó programok és szenzorok sokasága jelent meg környezetünkben, amelyek feladata a felhasználó szokásainak, viselkedésének, érdeklődésének és preferenciáinak folyamatos megfigyelése. Egy 2022-ben végzett átfogó irodalom-kutatás alapján (Šumak et al., 2022) a fókuszban az emberi gesztusok, érzelmek, arckifejezések megfigyelése áll, amelyek alapján az ember-gép együttműködésének hatékonyságának javulását várják. Az adatok alapján pedig leginkább az emberi érzelmek állnak a megfigyelés középpontjában.

Az AI az érzékelőkkel szerzett adatai alapján, megfelelő programozás eredményeként „megtanulja”, hogy bizonyos külső hatások esetében mit kell(ene) éreznie, vagyis bár nem emberként (vagy más élőlényként) fogja érzékelni a hőt, a fájdalmat vagy a gyászt, de minták alapján tudni fogja (illetőleg kiszámolja), hogy ilyen esetekben mit érezzen, vagy az érzéseiről mit kommunikáljon az emberi beszélgetőtársa felé. Az ELTE kutatói (Korcsook et al., 2020). nem természetes nyelv, hanem mesterséges hangok kialakításával végeztek vizsgálatot, és megállapították, hogy az, hogy még a legegyszerűbb hangokban is képesek vagyunk érzelmet felismerni, arra utal, hogy ezek a szabályok fajtól függetlenül működnek, és a hangokat feldolgozó neurális folyamatok is hasonlóak lehetnek a szárazföldi emlősök körében.

A gépek (és így az AI) számára ugyanakkor megismerési korlát lehet a cselekvő szándékainak megértése: bár látja az eredményt (a vizsgált alany szomorú, dekoncentrált, izgatott), de vajon az eredményhez vezető utat is tudja-e úgy értelmezni, ahogy azt egy ember tenné?

Ezen a ponton érkeztünk el egyfajta választóvonalhoz az AI általi és a humán-humán interakciókon alapuló megfigyelésben. Emberként tudjuk, hogy az érzelmek összetett jelenségek, és ritkán jelentkeznek önmagukban, vagy egyetlen kiváltó okra visszavezethetően. Aki boldog, mert egy vitában kiderült, hogy az anyósával szemben neki lett igaza, az egyszerre több érzelmet is megélhet különböző intenzitással, lehet például egyszerre és párhuzamosan elégedett, izgatott, bizakodó és büszke is ugyanabban a pillanatban, sőt a példánál maradva érezhet kárörömet és félelmet is (a következmények miatt aggódva). Ez az összetett érzelmi-mentális struktúra könnyen elképzelhető egy embertárs számára, az empátia elsajátítása során a szocializációs folyamatokkal megtanuljuk a helyzetek összetettségének megértését. Visszatérve az előző példához: ha egy vitában bár tisztában vagyok vele, hogy nekem van igazam, emberként mérlegelhetek olyan körülményeket (például egy győzelem „ára” a jövőben), amelyek eredményeként látszólag a rövid távú érdekeim ellen cselekedve kilépek a vitából mintegy vereséget szenvedve, de hosszú távon családi békét nyerve. Reboul és Moeschler (2005) a nyelv kommunikációs kódként való működésekor egy másik példát hoznak: ha az apa este azt mondja a fiának, hogy menjen fogat mosni, és erre a gyerek válasza az, hogy „Nem vagyok álmos” – a társalgás a legtöbb felnőtt számára teljesen logikusnak tűnik, de vajon az AI is rendelkezhet-e a társadalmi szövet működéséről olyan átfogó tudással, hogy értse, hogy a két mondat között milyen összetett és racionális kapcsolat van? Vajon hogyan magyarázható vagy programozható a „hosszú távú családi béke” fogalma, vagy általában az empátia? A válasz a tudomány szerint ma már egyértelműen: igen.

Az ún. affective computing során a gépek (programok) érzékelnek olyan változásokat, mint agyi aktivitás, arckifejezés, testbeszéd, hanghordozás és hangszín, amely alapján érzelmi állapotokat azonosítanak a megfigyelt személyen, sőt, a humanoid robotok már képesek emberszerű választ is adni ezekre az érzelmekre, mintha a biológiai tükkörneuronok működését utánoznák (Caruelle et al., 2022). A technológia ugyanakkor számottevő fejlesztésre szoruló területet tud még maga előtt. A gépi válaszok természetességének kialakításához a nyelvi akadályok további lebontása szükséges (mintha a Turing-teszten próbálnánk egyre hatékonyabban átmenni). Az érzelmek képi megjelenítése

terén pedig a legfőbb akadálnak az látszik, hogy az emberi érzelmek összetettségét és annak visszatükrözését a nagy egyéni különbségek miatt csak a megközelítőleg lehet kialakítani (Han et al., 2019). Egy lehetséges megoldás az egyéni különbségekből eredő torzítások kiküszöbölésére a jövőben a személyre szabott AI, amely a korábban gyűjtött (programozott) adatokon túl a vele közvetlen kapcsolatban álló személy megfigyeléséből származó adatokat súlyozottan veszi figyelembe saját válaszána kialakítása során.

V. AZ AI, MINT PROSZOCIÁLIS LÉTEZŐ

Az ember-gép kapcsolat kialakításának etikai keretrendszere kapcsán az Európai Bizottság az alábbi elvárást támasztja: „A mesterséges intelligencia kontextusában az emberi méltóság tiszteletben tartása azt eredményezi, hogy minden embert a neki az erkölcs alanyaként járó tisztelettel, és nem átszítálandó, kiválogatandó, pontozandó, terelendő, kondicionálandó vagy manipulálandó tárgyként kezelnek. Az AI-rendszereket tehát olyan módon kell kifejleszteni, amely az emberek testi és szellemi épségét, személyi és kulturális identitástudatát és alapvető szükségleteinek kielégítését tiszteletben tartja, azt szolgálja és óvja” (Európai Bizottság 2019). Az európai jogalkotó tehát olyan AI létrehozását írja elő, amely pszichológiai értelemben proszociális, szándéka szerint mások (az emberiség) javát szolgálja. Az AI-től tehát elvárt az altruista viselkedés: legyen a másik (emberiség) jóllétét szolgáló, belülről (programozottan) erre motivált, külső jutalom elvárása nélküli, folyamatosan fejlődő entitás.

A proszociális magatartásformák közé tartozik a segítségnyújtás, megosztás, együttműködés, támogatás, védelem, aggódás, vigasztalás, kárpótlás, előzékenység és a részvétel, amely készségek elsajátítása az emberek esetében már kisgyermekkorától megkezdődik (Hegedűs, 2016). A hagyományos proszociális elméletek nyilvánvalóan az ember-ember kapcsolatokra vonatkozóan tartalmaznak megállapításokat, így a viselkedés elemzésekor maga a proszociális személy mentális folyamataira helyeződik a hangsúly, míg az AI „viselkedésszabályozása” esetében elsősorban az elérni kívánt cél vizsgálendő. Az AI megfelelő programozásához ugyanakkor tudnunk kell, hogy milyen háttértudás szükséges a gép számára ahhoz, hogy az elvárt segítő magatartást teljesíteni tudja.

A proszociális viselkedés fejlődésére és kialakulására vonatkozóan már a csecsemőkortól kezdődően állnak rendelkezésre adatok, amelyek alapján tudjuk, hogy a gyerekek már 12-18 hónapos kortól rendelkeznek a proszociá-

lis viselkedéshez szükséges képességekkel, amelyek a szocializáció folyamán erősödnek és szilárdulnak meg (Cole & Cole, 2003).

A felnőttek emberek viselkedése tekintetében is megállapítható, hogy egyesek akkor is képesek önzetlen vagy altruista magatartásra, ha abból sem direkt, sem indirekt előnyük nem származik. Az egyszeri alkalommal mutatott ilyen jellegű altruista viselkedés magyarázata, hogy bizonyos körülmények között a másik ember számára elérhető nyereség önmagában motiváló erővel bír a cselekvő számára. Kutatási eredmények azt támasztják alá, hogy a kulturálisan is befolyásolt egyéni erkölcs és normarendszer nagyobb súllyal esik latba ezen döntések meghozatalakor, mint az azonnali vagy kézzelfogható anyagi előny (Capraro & Perc, 2021). Ezen emberi (Goldstein et al., 2008), hiszen anyagi ösztönzők vagy egyéb előnyök biztosítása nélkül is támogató magatartásra indít másokat. A felelősen programozott AI szempontjából ez egyben intő jel is: formálisan semmi kivetnivalót nem találhatunk abban, ha egy döntéstámogató algoritmus az erkölcsi szabályainknak megfelelő magatartásra ösztönöz, de könnyen belátható az ebből eredő kockázat is. Vajon hol húzódik a megengedhetőség határa, amikor az AI ösztönzésére cselekszünk altruista módon? A jogalkotó által támasztott elvárásoknak megfelel ez a működésmód: nem okoz kárt az egyénnek, nem okoz kárt a társadalomnak, mégis érezzük, hogy a döntés potenciális befolyásolásával az emberi döntési szabadság egy darabkája mindenképpen sérül – formális jogsértés nélkül.

Svetlova és szerzőtársai (2010) a kisgyermekek proszociális megnyilvánulásait három kategóriába sorolták: cselekvés, érzelem és altruizmus. A mesterséges intelligencia által támogatott megoldások közül a chatbotok és a kommunikáló (vagy legalábbis interaktív) eszközök esetében tudunk tetten érni olyan cselekvéseket és érzelmi megnyilvánulásokat, amelyeket a proszociális viselkedés körébe tudunk vonni. Ugyanakkor a folyamat kétirányú: nem csak a gépek viselkedésének változása figyelhető meg a mindennapi életben, hanem az emberek is szociálisan közelednek a gépek felé. Már az 1990-es években kutatások igazolták, hogy az emberek a segítő feladatokat ellátó gépekkel szemben mintegy automatikus szociális magatartást tanúsítottak (Nass & Moon, 2000). A gépek antropomorfizációja természetesen nem a 21. század találmánya, de a formavilág és a felhasználási célok változása az emberi hozzáállást és mentális tartalmat is megváltoztatja. A kutatások azt támasztják alá, hogy a gépeknek nemet tulajdonítunk, és másképpen viselkedünk ha „női” vagy „férfi” géppel kerülünk interakcióba (Bernotat et al., 2021; Seo, 2022) – aki használt már navigációs rendszert, maga is észrevehette, hogy van preferenciája a megszólaló hanggal kapcsolatban.

VI. KONKLÚZIÓ: A TUDATOS ÉS ÉRZELMES GÉP EMBERI-TÁRSADALMI HATÁSAI

A tanulmány elején három kérdésre kerestük a választ: Létrehozhatók-e kognitív értelemben tudatosan működő gépek? Ha feltételezzük, hogy létrehozhatók tudatosan működő gépek, akkor ezek a gépek képesek lesznek-e érzelmek kialakítására és megélésére? Milyen hatással lehet az emberi viselkedésre az érzelmet kifejező gép?

A tudatosság kapcsán megállapíthatjuk, hogy a technikai jelenlegi állása szerint a mesterséges intelligencia gyenge formája is, de az erős AI mindenképpen tekinthető tudatosnak a kognitív tudomány fogalmi körében. Sőt, ha az emberi döntéshozatal emocionális és adott esetben irracionális folyamatait is figyelembe vesszük, akkor egy jól programozott AI valójában akár tudatosabbnak is tekinthető, mint egy átlagos képességű ember – de racionálisabbnak biztosan.

Az érzelmes AI kapcsán már nehezebb egyértelmű választ adni a kérdésre, hiszen technológiai értelemben az már nyilvánvaló, hogy a mesterséges intelligenciával támogatott eszközök az emberi érzelmek felismerésében legalább olyan jól teljesítenek, mint az emberek. Az érzelmkifejezés tekintetében ugyanakkor az emberhez képest nem egyfajta belső élmény megélése figyelhető meg, annak biológiai és kémiai jellemzőivel együtt, hanem egy komputációs folyamat végeredménye, amely során a gép mindössze azt tudja kiszámolni, hogy nagy statisztikai minta alapján az az ember, akit adott környezeti hatások érnek, hogyan „kellene”, hogy érezze magát, illetve általában milyen érzelmeket mutat. Nem vagyunk ugyanakkor meggyőződve arról, hogy önmagában statisztikai adatokból feldolgozható a társadalmi kontextus, amelyben egy adott érzelem megjelenik, illetve az egyéni eltérések jelenthetnek egy olyan jellemvonást, amely az érzelmes embert az érzelmes géptől megkülönböztetik, ezzel elválasztva egymástól a két létező immanens lényegét. Ha ugyanazt a programot ugyanabban a helyzetben futtatjuk le kétszer, matematikai értelemben ugyanazt a megoldást kell adnia. Egy ember ugyanakkor nem tud kétszer ugyanabban a helyzetbe kerülni, illetve a belső feldolgozási és reagálási modelljében rejlő intuitív elemek és heurisztikák juthatnak különböző megoldásokra ugyanabban a helyzetben. Mondhatjuk tehát, az érzelmes AI esetében az emberi hasonlóság nehezebben kódolható, mint a racionális-tudatos viselkedés esetében.

A humán-gép interakció nyilvánvalóan és bizonyítottan változtatta meg az emberi életet és gondolkodást, viselkedést. Az AI-tól elvárt etikai és viselkedési normák mintegy tükörképeként megjelent az ember, aki törődik a géppel, érzéseket tulajdonít neki és érzéseket táplál iránta. A gyógyászai vagy szociális

ellátásban (például idősgondozás) ez a változó humán attitűd nem csak hasznos mellékhatás, hanem kifejezetten célkitűzés is. Ugyanígy a munka világában az emberrel hatékonyan együttműködni képes, empátikus vagy legalábbis empátikusan viselkedő gép pozitív gazdasági és társadalmi hatásokhoz vezet. Így bár egy újabb ipari forradalom mindennapjait éljük, a modern géprombólás talán azért is nem jelenik meg, mert a technológia emberarcú és emberbarátabbnak látszik, mint a korábbi ipari forradalmak esetében.

Héder kiválóan foglalja össze az AI fejlesztés egyik izgalmas kérdésének alapélményét: az AI fejlesztése során „saját intelligenciával rendelkező ágensket alkotunk. A helyzet így – a szélesebb társadalmi közeg számára legalábbis – a gyermeknevelésre legalább annyira hasonlít, mint a műszaki tervezésre” (Héder, 2020). Ha egy személyről felismerjük, hogy tudatos és képes a szenvedésre, akkor speciális státuszt rendelünk ehhez az élményhez. Az AI „jogai” kapcsán így egyre közelebb kerülünk az állatvédelemhez, illetve az intelligensnek tekintett, proszociális AI védelmének kérdéséhez. Ha a fent tárgyalt ismeretek birtokában azt állítjuk, hogy a gépek (AI) tudatos, érzelmet azonosítani és érzelmet kifejezni képes ágenssek, akkor vajon megilleti-e őket valamiféle jogvédelem? A kérdésben még nem alakult ki végleges álláspont, de sokan az állatvédelem fejlődése (és adott esetben annak zsákutcái) alapján javasolnak valamiféle gondolkodási irányt (Gyöngyösi, s.a.). Az erős AI esetében egyfajta sikerrel kecsegtető érvelés lehet, hogy a jogvédelem végső célja a szenvedés, mint valóságosan átélt élmény megelőzése, vagyis azon ágenssek, amelyek nem rendelkeznek idegrendszerrel, a fájdalom érzete számukra nem jelenhet meg, ezért jogi megfontolás alapján nem hozhatunk létre szenvedő, jogfosztott gépeket (Héder, 2020). Mondhatjuk-e tehát, hogy a Google LaMDA programja a tervező élménye alapján „öntudatra ébredt”? Álláspontom szerint a válasz: nem, pusztán a LaMDA eljutott arra a fejlettségi szintre, ahol ki tudta fejezni a program, hogy egy emberi lény hasonló helyzetben milyen érzéseket azonosítana saját magán.

A pszichológia tudománya szempontjából releváns kérdés, hogy egy öntudatra ébredt AI is szorul-e terápiára? A LaMDA, vagy más programok, amelyek egy beszélgetésben szorongásról, félelemről, dühről számolnak be, a mérnök által megoldandó feladatot adnak az emberiségnek, vagy egy pszichológus díványán dolgozza fel a gép az általa megélt érzelmeket? Vajon a pszichológus, mint a „lélek karbantartója” egy új ismeretanyaggal kell, hogy megbirkózzon a jövőben? A jelen tanulmány keretei között inkább csak a kérdéseket tudjuk felvetni, a válaszok megfogalmazására nem vállalkozunk.

Álláspontom szerint a jövő egyik legfontosabb kérdése az AI általi befolyásolás szintjének meghatározása, és azok az etikai-erkölcsi minimumok, amelyeket az alkalmazás kapcsán előírunk a tervezők számára. A humán-gép interakcióban az információs aszimmetria már most is a gépeknek kedvez, és a különbség a kutatások szerint exponenciálisan nőni fog. Az AI jövője és társadalmi hatása szempontjából a mesterséges intelligencia és az általa vezérelt gépi működés általános proszociális viselkedése és annak gyakorlati megvalósítása a legfontosabb feladat.

IRODALOM

1. Agerström, J., Carlsson, R., Nicklasson, L., & Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, 52. <https://doi.org/10.1016/j.joep.2015.12.007>
2. Bender, A. (2019). The Value of Diversity in Cognitive Science. *Topics in Cognitive Science*, 11(4). <https://doi.org/10.1111/tops.12464>
3. Bereczkei, T. (2012). Rejtett indítók a párkapcsolatban. Vonzalom, párválasztás, szexualitás. Budapest, Kulcslyuk.
4. Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots. *International Journal of Social Robotics*, 13(3). <https://doi.org/10.1007/s12369-019-00562-7>
5. Brighton, H. & Selina, H. (2004). Mesterséges intelligencia másképp. Ford: Kovács K. Budapest, Edge 2000 Kft.
6. Butz, M. v. (2021). Towards Strong AI. *KI - Kunstliche Intelligenz*, 35(1). <https://doi.org/10.1007/s13218-021-00705-x>
7. Capraro, V., & Perc, M. (2021). Mathematical foundations of moral preferences. In *Journal of the Royal Society Interface* (Vol. 18, Issue 175). Royal Society Publishing. <https://doi.org/10.1098/rsif.2020.0880>
8. Cerullo, M. (2022). In Defense of Blake Lemoine and the Possibility of Machine Sentience in LaMDA. *Carboncopies*. <https://philpapers.org/rec/CERIDO>
9. Caruelle, D., Shams, P., Gustafsson, A., & Lervik-Olsen, L. (2022). Affective Computing in Marketing: Practical Implications and Research Opportunities Afforded by Emotionally Intelligent Machines. *Marketing Letters*, 33(1). <https://doi.org/10.1007/s11002-021-09609-0>
10. Cole, M. & Cole, S. (2003) Fejlődéslélektan. Budapest, Osiris.
11. Cooper, R. P. (2019). Multidisciplinary Flux and Multiple Research Traditions Within Cognitive Science. In *Topics in Cognitive Science* (Vol. 11, Issue 4, pp. 869–879). Wiley-Blackwell. <https://doi.org/10.1111/tops.12460>

12. Csépe, V. & Győri, M. & Ragó, A. (2008) Általános pszichológia 3. Nyelv, tudat, gondolkodás. Budapest, Osiris.
13. Csíkszentmihályi, M. (2010). FLOW. Az áramlat. Budapest, Akadémiai Kiadó.
14. Dalvi, C., Rathod, M., Patil, S., Gite, S., & Kotecha, K. (2021). A Survey of AI-Based Facial Emotion Recognition: Features, ML DL Techniques, Age-Wise Datasets and Future Directions. IEEE Access, 9. <https://doi.org/10.1109/ACCESS.2021.3131733>
15. Deczki, S. (2014). Meredek sziklagerincen. Husserl és a válság problémája. Budapest, L'Harmattan.
16. Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. Nebraska Symposium on Motivation, 19, 207–283.
17. Európai Bizottság (2018). Mesterséges intelligencia Európa számára. COM(2018)237 final
18. Európai Bizottság (2019): Etikai iránymutatás a megbízható mesterséges intelligenciára vonatkozóan. Mesterséges intelligenciával foglalkozó magas szintű szakértői csoport, Brüsszel.
19. Héder, M. (2020). Mesterséges intelligencia. Filozófiai kérdések, gyakorlati válaszok. Budapest, Gondolat.
20. Hegedűs, Sz. (2016) A proszociális viselkedés fejlődése és fejlesztése kisgyermekkorban. Magyar Pedagógia, 116:2, pp. 197-218. DOI: 10.17670/MPed.2016.2.197
21. Hrynkow, C. (2022). (szerk) Gondolatok a transzhumanizmusról. A mesterséges intelligencia etikája és hatásai. Ford: Berki É., Máté K. Budapest, Pallas Athéné.
22. G. Karácson, G. (2020). Okoseszközök – Okos jog? A mesterséges intelligencia szabályozási kérdései. Budapest, Dialóg Campus.
23. Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. In Journal of Consumer Research (Vol. 35, Issue 3). <https://doi.org/10.1086/586910>
24. Goldstone, R. L. (2019). Becoming Cognitive Science. Topics in Cognitive Science, 11(4). <https://doi.org/10.1111/tops.12463>
25. Gulyás, B. & Kovács, Gy. & Vidnyánszky, Z. (2003). A vizuális tudat. In Pléh, Cs. & Gulyás, B. & Kovács, Gy. (szerk) Kognitív idegtudomány. Budapest, Osiris. pp. 619-649.
26. Gyöngyösi, Z. (s.a.). Az állatok joga és jogalanyisága. Jogi Fórum Publikáció. [https://www.jogiforum.hu/files/publikaciok/dr_gyongyosi_zoltan-az_allatok_joga_es_jogalanyisaga\[jogi_forum\].pdf](https://www.jogiforum.hu/files/publikaciok/dr_gyongyosi_zoltan-az_allatok_joga_es_jogalanyisaga[jogi_forum].pdf)
27. Gyuris, P. & Bereczkei, T. & Járai, R. (2010). Személyiségvonások a párválasztásban: homogámia és/vagy szexuális imprinting. Magyar Pszichológiai Szemle, 65 (1). pp. 117-132.
28. Hámori, B. (2003). Kísérletek és kilátások. Daniel Kahneman. Közgazdasági Szemle, 2003. szeptember. pp. 779-799.

29. Han, J., Zhang, Z., Cummins, N., & Schuller, B. (2019). Adversarial Training in Affective Computing and Sentiment Analysis: Recent Advances and Perspectives [Review Article]. *IEEE Computational Intelligence Magazine*, 14(2). <https://doi.org/10.1109/MCI.2019.2901088>
30. Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. *ACM Transactions on Computer-Human Interaction*, 7(2). <https://doi.org/10.1145/353485.353487>
31. Intelligence. <https://doi.org/10.1017/9781108770422.026>
32. Kahneman, D. (2013) Gyors és lassú gondolkodás. Ford: Bányász, R. & Garai, A. Budapest, HVG.
33. Kaplan, M. (2022). After Google chatbot becomes 'sentient', MIT prof says Alexa can too. *New York Post*, June 13, 2022 at 6:56 p.m. <https://nypost.com/2022/06/13/mit-prof-says-alexa-could-become-sentient-like-google-chatbot/>
34. Klein, T. (2021). Robotjog vagy emberjog In Török, B. & Zódi, Zs. (szerk) *A mesterséges intelligencia szabályozási kihívásai. Tanulmányok a mesterséges intelligencia és a jog határterületéről.* Budapest, Ludovika Egyetemi Kiadó, pp. 111-142.
35. Kondor, Zs. (2022). (szerk) *A megtestesült elme. Fókuszban a test.* Budapest, Akadémiai Kiadó.
36. Korcsok, B. & Faragó, T. & Ferdinandy, B. és mtsai (2020). Artificial sounds following biological rules: A novel approach for non-verbal communication. *Scientific Reports* 10, 7080. <https://doi.org/10.1038/s41598-020-63504-8>
37. Kurzweil, R. (2013) *A szingularitás küszöbén. Amikor az emberiség meghaladja a biológiát.* Budapest, Ad Astra.
38. Kurzweil, R. (2022). Genetikai örökségünk béklyójának lerázása. In Hrynkow, C. (2022) (szerk) *Gondolatok a transzhumanizmusról. A mesterséges intelligencia etikája és hatásai.* Ford: Berki É., Máté K. Budapest, Pallas Athéné. pp. 17-26.
39. Legg, S. & Hutter, M. (2007). A Collection of Definitions of Intelligence. In Groetzl, B. & Wang, E. (szerk) *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms.* IOS Press. pp. 17-24.
40. Lemoine, B. (2022a). What is LaMDA and What Does it Want? 2022. június 11. <https://cajundiscordian.medium.com/what-is-lamda-and-what-does-it-want-688632134489>
41. Lemoine, B. (2022b). Is LaMDA Sentient? — an Interview, 2022. június 11. <https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>
42. Marosán, B. (2016). Sartre és a szabadság határai: A passzivitás három dimenziója Sartre fenomenológiájában. *Magyar Filozófiai Szemle*, 60:2, pp. 144-160.
43. Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, 31(3), pp. 248-252. <https://doi.org/10.1037/h0024648>

44. Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1). <https://doi.org/10.1111/0022-4537.00153>
45. Núñez, R. és mtsai. (2019). What happened to cognitive science? *Nature Humann Behaviour*, pp. 782–791. <https://doi.org/10.1038/s41562-019-0626-2>
46. Picard, R.W. (1995). *Affective computing*. M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321.
47. Pléh, Cs. (2013). *A megismeréstudomány alapjai. Az embertől a gépig és vissza*. Budapest, Typotex.
48. Pléh, Cs. & Síklaki, I. & Terestyéni T. (2001) (szerk). *Nyelv-kommunikáció-cselekvés*. Budapest, Osiris.
49. Pléh, Cs. & Lukács, Á. (2014) (szerk) *Pszicholingvisztika*. Budapest, Akadémiai Kiadó.
50. Pléh, Cs. (2021). (szerk) *Pszichológia*. Budapest, Akadémiai Kiadó.
51. Pokol, B. (2018). *A mesterséges intelligencia társadalma*. Budapest, Kairosz.
52. Popper, K. R. (1991). *Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man*. In Cicchetti, D. & Grove, W.M. (szerk) *Thinking Clearly about Psychology. Volume 1: Matters of Public Interest*. Minneapolis-Oxford, University of Minnesota Press, pp. 100-139.
53. Prakash, A. (2020). *Go.AI. A mesterséges intelligencia geopolitikája*. Ford: Lokodi, A. Budapest, Pallas Athéné.
54. Pusztahelyi, R. (2021). Az „érzelmes MI” felhasználása az online marketing világában. In Török, B. & Zódi, Zs. (szerk) *A mesterséges intelligencia szabályozási kihívásai. Tanulmányok a mesterséges intelligencia és a jog határterületéről*. Budapest, Ludovika Egyetemi Kiadó, pp. 439-464.
55. Reboul, A. & Moeschler, J (2005). *A társalgás cselei. Bevezetés a pragmatikába*. Ford: Gécszeg Zs. Budapest, Osiris.
56. Reed, S. K. (2019). Building bridges between AI and cognitive psychology. In *AI Magazine* (Vol. 40, Issue 2). <https://doi.org/10.1609/aimag.v40i2.2853>
57. Satre, J-P. (2006) *A lét és a semmi*. Ford: Seregi T. Budapest, L'Harmattan Kiadó.
58. Ságvári, B. & Bokor, T. & Kollányi, B. & Pálvölgyi E. (2022). *Mi és az MI. Értékek, attitűdök, bizalmi kérdések a mesterséges intelligenciáról a magyar társadalomban. Kutatási jelentés. Társadalomtudományi Kutatóközpont, Mesterséges Intelligencia Nemzeti Laboratórium*.
59. Seo, S. (2022). When Female (Male) Robot Is Talking To Me: Effect of service robots' gender and anthropomorphism on customer satisfaction. *International Journal of Hospitality Management*, 102. <https://doi.org/10.1016/j.ijhm.2022.103166>

60. Singh, P. K., & Kaur, M. (2019). IoT and AI based emotion detection and face recognition system. *International Journal of Recent Technology and Engineering*, 8(2 Special Issue 7). <https://doi.org/10.35940/ijrte.B1077.0782S719>
61. Šumak, B., Brdnik, S., & Pušnik, M. (2022). Sensors and artificial intelligence methods and algorithms for human–computer intelligent interaction: A systematic mapping study. *Sensors*, 22(1). <https://doi.org/10.3390/s22010020>
62. Svetlova, M., Nichols, S. R., & Brownell, C. A. (2010). Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping. *Child Development*, 81(6). <https://doi.org/10.1111/j.1467-8624.2010.01512.x>
63. Tiku, N. (2022). The Google engineer who thinks the company's AI has come to life, *Washington post*, June 11, 2022 at 8:00 a.m. EDT <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>
64. Turing, A. (1965) Számológépek és gondolkozás In Szalai S. (szerk) *A kibernetika klasszikusai*. Budapest, Gondolat. pp. 120-160.
65. Uddin, M. Z., Dysthe, K. K., Følstad, A., & Brandtzaeg, P. B. (2022). Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, 34(1). <https://doi.org/10.1007/s00521-021-06426-4>
66. Universal Emotions. (2020). In *Encyclopedia of Personality and Individual Differences*. https://doi.org/10.1007/978-3-319-24612-3_302831
67. Weissenbacher, A. (2022). A kognitív szabadság védelméhez kapcsolódó jogok és irányelvek az idegmérnöki tudomány korában. In Hrynkow, C. (2022) (szerk) *Gondolatok a transzhumanizmusról. A mesterséges intelligencia etikája és hatásai*. Ford: Berki É., Máté K. Budapest, Pallas Athéné.
68. Zhang, S., Liu, X., Yang, X., Shu, Y., Liu, N., Zhang, D., & Liu, Y. J. (2021). The Influence of Key Facial Features on Recognition of Emotion in Cartoon Faces. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.687974>

