digitális jólét nonprofit kft.

HUNEXPERT

Moór Gyula Jogtudományi Pályázat

2020

00110011 – MONDTA A BÍRÓ, ÉS FELMENTETTE A VÁDLOTTAT. AZ ALGORITMIKUS DÖNTÉSHOZATAL AKTUÁLIS KÉRDÉSEI

FERENCZ BÁLINT
ELTE-ÁJK
Phd. hallgató

00110011 – mondta a bíró, és felmentette a vádlottat. Az algoritmikus döntéshozatal aktuális kérdései

Moór Gyula Tudományos Pályázat

Ferencz Bálint

Absztrakt

A mesterséges intelligencia (MI) robbanásszerű fejlődése egyre csábítóbbá teszi az algoritmusok alapján történő döntéshozatalt, amely kiszámíthatóságot, megbízhatóságot, pontosságot és gyorsaságot ígér. Köztudott, hogy a bíróságok és a különböző hatóságok szinte fuldokolnak a beérkező ügyekben, amely egyaránt rossz a jogalkalmazóknak és az állami rendszerek igénybevevőinek. Ennek tükrében jelen tanulmány azt kívánja megvizsgálni, hogy ténylegesen milyen problémák vetődnek fel az emberi döntéshozatallal kapcsolatban, amelyek egyébként valóban megkérdőjelezik az emberi pontosságot és megbízhatóságot. Erre reflektálva azonban azt is megvizsgáljuk, hogy az MI valóban képes-e teljesíteni azokat a magas standardokat, amelyeket kitűzünk elé. Az emberi és gépi döntéseket a tanulmány során három paraméter alapján hasonlítjuk össze: kiszámíthatóság, objektivitás valamint körültekintés és megfontoltság. Arra keressük a választ, hogy ezen szempontok alapján melyik döntéshozatalt érdemes előnyben részesíteni.

Kulcsszavak: mesterséges intelligencia; szimbolikus MI; szubszimbolikus MI; átláthatóság; óvatosság; magyarázatadás

Tartalom

I. Bevezetés

A címben felvillantott kép reményeim szerint elég plasztikusan mutatja be azt a problémakört, amit jelen tanulmányban végig kívánunk járni. Adott egy rendszer (nevezhetjük robotnak, szoftvernek, ahogy szeretnénk), amely hoz egy jogilag kötelező döntést, miközben a döntés hátterét a perben állók vagy egyéb, technikai tudással nem rendelkezők nem értik (sőt, az informatikusok, szoftverfejlesztők sem feltétlenül). Probléma, ha nem értjük? Az autó működését értjük? Sőt, az internet működését értjük? A sor végtelenül folytatatható az atomenergiától elkezdve a repülésig, miközben már az MI is mindenhol jelen van a világunkban, és sok más problémát megold helyettünk – vannak-e akkor más ellenérvek az ellen, hogy ne bízzunk rá jogi döntéseket?

Valósággá válhat tehát a címben felvetett kép? Ha az MI és a jog kapcsolata vetődik fel, akkor az automatikus döntéshozatal témaköre megkerülhetetlen és bizony igen népszerű is. A robotbíró képe valamilyen oknál fogva elgondolhatóbbnak tűnik a robotügyvéd vagy a robot-vállalati jogászhoz képest – ezzel együtt ennek a képnek a "megvalósítása" legalábbis heves viták kereszttüzében áll.

Az algoritmikus döntéshozatal kapcsán korábban inkább az volt a kérdés, hogy lehetséges-e, hiszen ez kalkulusokon, algebrán és matematikán alapuló jogrendszer Leibniz óta jelenlévő vágyálom, amely hol erőteljesebben, hol marginálisabban képviselteti magát a jogi diskurzusban.[1] Ezzel szemben az algoritmusokon alapuló döntések most már a valóság részét képezik, amelynek – a tisztán technológiai lehetőségeken túlmenően – az ad különös fontosságot és izgalmat, hogy a bíróságok és különböző hatóságok ügyhátraléka, eljárásaik hossza ijesztő méreteket ölt[2], és az MI által felkínált lehetőség kifejezetten kézenfekvőnek, de mégis félelmetesnek és idegennek tűnik. A kérdés aktualitását mi sem mutatja jobban, hogy a

[1] Zsolt Ződi, „Hogyan változtatja meg a jog nyelvezetét a számítógép: A logika és a tekhné a jogban", *GLOSSA IURIDICA JOGI SZAKMAI FOLYÓIRAT* I., sz. 2 (2014): 115.

[2] Az Emberi Jogok Európai Bírósága előtt a 2018. évi statisztika alapján 56 350 ügy van (forrás: https://www.echr.coe.int/Documents/Stats_analysis_2018_ENG.pdf, elérés: 2020. február 15.), míg a magyar bíróság legfrissebb, bárki számára elérhető statisztikái alapján (2019. I. félév) csak a járásbíróságok, a közigazgatási-és munkaügyi bíróságok valamint a törvényszékek előtt mindösszesen 195 037 befejezetlen ügy fekszik, amelyből 85 917 peres ügy (forrás: https://birosag.hu/ugyforgalmi-adatok, 9. és 11. tábla, elérés: 2020. február 15.)

GDPR készítői is úgy gondolták, hogy valamit jogalkotói szinten mondani kell, így a 22. cikk – néhány kivételtől eltekintve – tiltja a joghatással járó automatikus döntéshozatalt.[3]

Jelen tanulmány célja hogyan olyan alapvető kérdéseket, problémákat mutasson be, amelyek az algoritmus döntéshozatallal kapcsolatban felmerülnek. Ezek a kérdések leginkább általános, az emberi / jogi és gépi döntéshozatallal kapcsolatos sajátosságokat igyekszenek feltárni, azzal, hogy olyan akár társadalmi, akár technológiai értelemben vett értékeket, eredményeket kívánok bemutatni, amelyek az algoritmikus döntéshozatallal kapcsolatos esetleges jövőbeli, kiterjedt jogalkotás során megfontolásra érdemesek. Tehát jelen írás a döntési módszerek és azok minősége közötti különbségekre kíván a hangsúlyt fektetni, de nem célja az algoritmus döntéshozatal jogszabályokban, normaszövegekben megjelenésének bemutatása. Fontos kiemelni, hogy témán szempontjából kizárólag az államnak vagy az állami szerveknek a joghatással járó döntéseit mérlegeljük, tehát a szintén algoritmikus döntéshozatalon alapuló célzott reklámok, egyéb magánszektorban felmerülő döntéshozatal kívül esik a tanulmány vizsgálati körén.

Jelen tanulmányban az algoritmikus döntéshozatallal kapcsolatos kérdéseket, problémákat ahhoz képest vetjük fel, hogy az emberi vagy pedig a gépi döntések-e a jobbak. Nyilvánvalóan egy döntés "milyenségét" sokféleképpen lehet megragadni, a tanulmány megírásakor azonban alapvetően három fokmérőt használunk:

- kiszámíthatóság, amely alatt azt értem, hogy azonos eseteket azonos módon, a különböző eseteket pedig különböző módon kerül elbírálásra

- objektivitás, amely alatt azt értem, hogy tehát a döntésben a társadalomi előítéletek nem jelentkeznek, társadalmi helyzetéből adódóan semelyik fél nem élvez előnyt vagy nem szenved hátrányt

[3] A GDPR. 22. cikk (1) bekezdése konkrétan így fogalmaz: „Az érintett jogosult arra, hogy ne terjedjen ki rá az olyan, kizárólag automatizált adatkezelésen – ideértve a profilalkotást is – alapuló döntés hatálya, amely rá nézve joghatással járna vagy őt hasonlóképpen jelentős mértékben érintené."

- körültekintés, megfontoltság, amelyek alatt azt értem, hogy a döntésben a rendelkezésre álló információk lehető legteljesebb köre figyelembe lesz véve és mérlegelésre kerül

1. Módszer és források

A tanulmány irodalomkutatáson alapszik, amely alapvetően az alábbi dokumentumok feldolgozását jelentették:

- az általánosabb, informatikai, pszichológiai kérdéseket taglaló művek esetében elsősorban a jelenlegi paradigma adó művek

- a kifejezetten az automatikus döntéshozatallal kapcsolatos publikációk esetében pedig lehetőség szerint a minél frissebb és újabb eredményeket felvonultató könyvek, szakirodalmi cikkek – utóbbiak esetében sajnos előfordult, hogy még folyóiratban nem megjelent, vagy anonim lektoráláson nem átesett publikáció kerül idézésre, amelynek oka, hogy szerencsére mind több kutató hozza nyilvánosságra előzetes részeredményeit online felületeken (kiváltképpen SSRN oldalán)

A szakirodalom tekintetében mindenképpen fontos hangsúlyozni, hogy kifejezetten a jog, a logika, az informatika és mesterséges intelligencia tudományközi kutatásában még nem született paradigma adónak tekinthető, összefoglaló könyv / kézikönyv. Ezt a tudományközi területet – mint az informatikával foglalkozókat általában – egyrészt a viszonylag gyorsnak mondható fejlődés valamint a módszerek sokasága jellemez, emiatt ezen témát is kifejezetten nehéz teljes aspektusában bemutatni. A tanulmány elkészítéséhez elsősorban angol nyelvű, nemzetközi szakirodalom került feldolgozásra.

2. A tanulmány felépítése

Tekintettel az MI-t körülvevő terminológiai “zavarra”, az II. fejezetben tisztázni kívánom, hogy mit értek MI alatt és ezzel kapcsolatosan bemutatásra kerül Mireille Hildebrandt tipológiája, aki a különböző MI módszereket vetítette rá lehetséges jogi szoftverekre. A III. fejezetben egyrészt kifejtésre kerül az emberi döntéshozatal kifejezetten problémás volta a Nobel-díjas Daniel Kahneman munkáján keresztül, majd erre reflektálva idézem az amerikai realizmus

egyik legfontosabb alakjának, Jerome Franknak a gondolatait. Ezen rész célja, hogy szemléltesse, melyek azok a főbb, az emberi gondolkodás sajátosságából adódó nehézségek, amelyek kívánatossá tehetnék az algoritmikus döntéshozatalt. A IV. fejezetben az ún. "Critical Code Studies" néhány jelentős megállapításán keresztül ismertetem általában a számítógépes szoftverekkel kapcsolatos fenntartásokat (kiemelve azok kiszámítható-kiszámíthatatlan jellegét), majd kifejezetten az MI korlátai kerülnek szemléltetésre. Ezen rész célja bemutatni azt, hogy a látszólagos, különösen médiában megjelenő népszerűségi hullám dacára milyen lényeges problémákat és kérdéseket vet fel az MI használata. Az V. fejezetben kifejezetten a jogi algoritmikus döntéshozatal mellett és ellen érvelő vélemények kerülnek bemutatásra azzal, hogy kifejtésre kerül egy olyan tényező, amely eddig az algoritmikus döntéshozatalról folytatott vita során explicit nem jelent meg, majd a Bevezetőben a „jó döntésként" megadott három paraméter (kiszámíthatóság, objektivitás, körültekintés / megfontoltság) szempontjai alapján összevetem az emberi és az algoritmikus döntéshozatalt. A Konklúzióban további kérdéseket vetek fel és összegzem a bemutatott eredményeket.

II. Mesterséges intelligencia

Az MI, mint gyűjtőfogalom, azon módszereket kívánja összefogni, amelyeknek a célja, hogy az emberi gondolkodást imitálva valódi, kognitív képességekkel rendelkező rendszereket / gépeket / szoftverek hozzon létre. Gyűjtőfogalomként használva azonban elmosódtak a különböző módszerek közötti olyan különbségek, amelyek alapvetően befolyásolják egy-egy ilyen szoftver működését.

Összefoglalóan elmondható, hogy az MI kutatásoknak kétféle iránya van: a szimbolikus és szubszimbolikus módszere.

Az MI elsősorban és először szimbolikus (vagy ahogy a szakirodalomban még nevezik: "good old fashion AI") rendszerek kifejlesztésével próbálkozott: ezek olyan logikai vagy egyéb programnyelven íródott szoftverek voltak, amelyek önálló szemantikával rendelkeztek és felülről-alulra ("top-down") megközelítéssel akartak egy-egy területet, tudásmennyiséget modellezni, az így létrejött szoftverek nevezték tudásrendszereknek ("expert systems"). Ezeknek a hátránya azonban az idő előrehaladtával egyre nyilvánvalóvá vált: egyrészt alapvető, az emberi tudás számára magától értetődő, köznapi információkat is modellezni kellett,

másrészt az egyes szakmai területek, professziók egyértelműnek tűnő szabályait sem lehetett egyértelműen, egyenesen lefordítani.[4]

Az MI jelenlegi felfutása a szubszimbolikus megközelítés eredménye. A módszer elnevezése onnan ered, hogy emögött nem találunk egy valamilyen logikát követő nyelvet, hanem "csupán" algoritmust és rengeteg adatot. A szubszimbolikus módszer arra épül, hogy a gép induktív módon, alulról-felülre irányuló logika alapján rengeteg adatból összefüggéseket tud kiolvasni, és ezen összefüggésekből tud döntést hozni egy új szituációra / helyzetre. A szubszimbolikus módszer foglalja magában a gépi tanulást (machine learning), amelynek egyik módszere a neuronhálókkal operáló deep learning. Ezek hatékonyságát az egyre jobb minőségű hardverek és az internet világából kinyert adatok segítették. A kezdeti felfutást követően azonban most már nagyon egyértelműen látszanak ennek a módszernek a hátrányai és korlátai, amelyek bővebben a III. fejezetben kerülnek kifejtésre.

A jogi szoftverek felosztása során Mireille Hildebrandt jogi algoritmusokkal foglalkozó cikkének felosztását alkalmazom.[5] Hildebrandt háromfajta szabályozási fajtát különböztet meg: a szövegalapút (text-driven), a kódalapút (code-driven) és az adatalapút (data-driven), tehát másfajta kontextusba helyezem ezeket a fogalmakat, de kifejező erejük miatt a továbbiakban ezeket használom az alábbi értelemben:

- kódalapú jogi szoftverek azok a szimbolikus mesterséges intelligenciát használó szoftverek, amelyek dedukciót, top-down módszert alkalmazva oldanak meg problémát

- adatalapú szoftverek azok a szubszimbolikus mesterséges intelligenciát használó szoftverek, amelyek induktív módszerrel, megfelelő adatmennyiség feldolgozását követően, a szoftver által felállított szabályszerűséget követően képesek minősíteni egy inputot

[4] Melanie Mitchell, *Artificial intelligence: a guide for thinking humans* (New York: Farrar, Straus and Giroux, 2019), 247. Idézi, hogy az egyik ilyen köznapi tudást "gyűjtő" Cyc vállalat szerint nagyjából 15 millió olyan összefüggést, megállapítást kell összegyűjteni, amellyel egy ember rendelkezik. Szerintük nagyjából az 5%-át sikerült modellezni.

[5] Mireille Hildebrandt, „Algorithmic Regulation and the Rule of Law", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, sz. 2128 (2018. szeptember 13.): 20170355, https://doi.org/10.1098/rsta.2017.0355.

Amint az a felsorolt módszerekből látszik, alapvetően másra és másra használható a kétféle megoldás. A kódalapú jogi szoftver elsősorban előre megadott és rögzített szabályokból történő következtetésre, az adatalapú szoftver pedig elsősorban jóslásra vagy valaminek a kategorizálására alkalmas.

A továbbiakban – a köznyelvhez hasonlóan – MI alatt az adatalapú szoftvereket és rendszereket fogom érteni azzal, hogy a rendszer, a szoftver és esetenként a robot szavakat felcserélve használom tekintettel arra, hogy jelen kérdések megválaszolásakor számunkra mindegy, hogy egy emberi felépítésű test adja meg az algoritmizált döntéseket, egy applikáció vagy egy számítógép. Szintén lényeges, hogy a továbbiakban a kérdést – kissé sarkosan – olyan szempontból vizsgáljuk meg, hogy feltételezzük, ezek emberi közbeavatkozás nélkül hoznának meg döntéseket (tehát nem az érdekel minket, hogy mennyivel lehetne hatékonyabb egy MI-vel kiegészülő emberi döntés hanem "pusztán" az, hogy egy tisztán emberi döntés és egy tisztán MI által hozott döntés mögött milyen előnyök és hátrányok jelentkeznek). Az emberi közbeavatkozás (ahogyan az angol nyelvű szakirodalom hívja: „human in the loop"[6]) és ehhez kapcsolódóan az ember és a gép együttműködése kézenfekvő megoldás, aminek szintén meg vannak a maga kérdései (pl.: Alarie-ék felvetik az ügyvédek számára az adatalapú szoftverek kötelező használatát a fölösleges perek elkerülése érdekében[7]), de ez nem képezi jelen tanulmány tárgyát.

A kódalapú rendszerekkel kapcsolatosan annyit mindenképpen szeretnék megjegyezni, hogy a jog területén sem dominánsak már ezek a megoldások. Ződi Zsolt sommás véleménye szerint "a szabályalapúak [szakértői rendszerek] pedig olyan szabályokat algoritmizáltak, amelyeket józan ésszel és egy nyolcadikos matematikai tudásával is bőven lehetett alkalmazni".[8] Kevin D. Ashley nemrég megjelent könyvében – mint az ilyen rendszerek egyik legismertebb fejlesztője, önkritikusan – a kódalapú szakértő rendszerekkel kapcsolatban felmerült problémákra a következő példákat hozza: a fejlesztőknek folyamatosan újra kellett fogalmazniuk a normaszövegeket, a logikai programozás nem képes tagadást bizonyítani (kizárólag abban az esetben, ha az összes szükséges feltevés hamis), és a nyílt-végű kifejezéseket a fejlesztők meg

[6] Harry Surden, „Artificial Intelligence and Law: An Overview", *Georgia State University Law Review* 35, sz. 4 (2019. június 1.): 1320–21.
[7] Benjamin Alarie, Anthony Niblett, és Albert H. Yoon, „How Artificial Intelligence Will Affect the Practice of Law", *University of Toronto Law Journal* 68, sz. 1 (2018. március 27.): 122.
[8] Zsolt Ződi, *Platformok, robotok és a jog új szabályozási kihívások az információs társadalomban* (Budapest: Gondolat, 2018), 46.

sem próbálták számítógépes nyelvre átfordítani.[9] A legmegdöbbentőbb információ, amelyet felhoz, az mégis egy formális törvényértelmezés, amelynek során két kutató egy jogszabályszöveg két mondatát elemezte, és a normalizálást követően arra jutottak, hogy annak 48 értelmezése lehetséges kizárólag szintaktikai alapon.[10] De a jogi szakértői rendszerek sikertelenségének okait már több más írás is vizsgálta, e helyütt annyit kívántam mindössze prezentálni, hogy a kódalapú jogi rendszerek jelenlegi tudásunk szerint nem fognak tudni áttörést hozni sem az algoritmikus döntéshozatal, sem általában véve a jogi gondolkodás és munka jelentős arányú kiváltása terén (lásd erről Ashley véleményét[11] valamint Leith írását[12] a kódalapú rendszerek hibáiról).

III. Emberi és gépi döntések sajátosságai

Jelen fejezetben kívánom bemutatni azokat a problémákat, amelyek az emberi valamint a gépi döntéssel szemben felmerülnek. Célom kifejezetten kritikai annyiban, hogy mindkét döntéstípusnak a gyöngeségét szeretném illusztrálni az ebben a fejezetben idézett szerzők példái, kutatásai alapján. Az ekképpen megismert hátrányokat fogom a későbbi fejezetekben egymáshoz hasonlítani.

1. Az emberi döntések

Annak az illúziójáról, hogy az ember egy racionális lény, amely objektív megfontolások alapján hozza meg a döntéseit, már réges-régen lemondott mind a tudomány, mind a közvélemény. Ennek fényében azonban az a kérdés még erőteljesebben fennáll, hogy mennyire tudjuk kivédeni a szubjektív elemeket, másik oldalról: mennyire tudjuk magunkra kényszeríteni az objektivitás és az észszerűség korlátját? Mennyire vagyunk képesek kizárólag a saját szakmánk szabályai alapján meghozni egy döntést?

Ezen alfejezet célja az emberi döntéseket kritizáló gondolatok bemutatása Daniel Kahneman és Jerome Frank munkái nyomán. Kahneman könyve alapján ismertetésre kerül, hogy sokkal jobban bízunk a saját döntéseink minőségében mint szabadna, és fel is vállalja azon véleményét, hogy előnyösebb lenne

[9] Kevin D. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (Cambridge New York Melbourne Delhi Singapore: Cambridge Univ Press, 2017), 49–52.
[10] Ashley, 45–46.
[11] Ashley, 11.
[12] Philip Leith, „The Rise and Fall of the Legal Expert System”, *International Review of Law, Computers & Technology* 30, sz. 3 (2016. szeptember): 94–106, https://doi.org/10.1080/13600869.2016.1232465.

bizonyos mennyiségben algoritmusra bízni őket. Ezt követően kifejezetten a bírói döntéseket kritizáló, az amerikai realizmus egyik prominens alakjának, Jerome Franknak a gondolatait idézem annak alátámasztására, hogy a bírói döntések is sokkal távolabb állnak az ideálistól, mint azt gondolnánk.

a) Az emberi döntések Daniel Kahneman kutatásai tükrében

Az, hogy az ember algoritmusként működik, sőt, bizonyos értelemben algoritmikusként hozza meg a döntéseit (vagy ha nem, érdemes lenne) egy bizarr de bizonyos értelemben plauzibilis kép. Erre kiváló példa Brian Christian és Tom Griffiths könyve, akik a különböző, emberi élethelyzetekre (pl.: miképpen szervezzük a feladatainkat) adnak algoritmusokon alapuló megoldásokat.[13] Azt persze ők is elismerik, hogy az élet egy komplexebb környezet annál, minthogy ezek a megoldások tökéletesen működjenek, ahhoz szükség van az eshetőség elfogadására, az idő és a pontosság közötti kompromisszumra és a becslések használatára.

Az emberi döntéshozatalról való gondolkodásra azonban az utóbbi évtizedben a legnagyobb hatást talán a Nobel-díjas pszichológus Daniel Kahneman gyakorolta. Ő azzal az újszerű és radikálisnak nevezhető gondolattal állt elő, hogy sokszor saját magunkat csapjuk meg, amikor azt gondoljuk, hogy egy-egy döntést racionálisan, mindenféle külső-belső körülménytől függetlenül hoztunk meg. Az általa lefolytatott kutatások éppen azt bizonyították be, hogy a külső hatások sokkal erőteljesebben befolyásolják a döntéseinket, mint azt korábban gondoltuk, és ezeket nem is tudjuk kizárni. Lényeges kiemelni, hogy Kahneman szerint ez alól senki nem tudja magát kivonni, még az egy-egy szakmát évtizede óta űző professzionális szakemberek sem, ideértve a bírókat is.

Mi ennek az oka? Röviden összefoglalva Kahneman a gondolkodásunkat kétfelé osztja: van az 1-es rendszer (ez a „gyors” gondolkodás) és van a 2-es rendszer („lassú gondolkodás”). Kahneman megállapította, hogy az 1-es rendszert sokkal gyakrabban használjuk egyrészt lustaságból (sokkal gyorsabban megadja számunkra a kért megoldást) másrészt túlzott önbizalomból (hiszen azt gondoljuk, hogy a felhalmozott tapasztalatunk és tudásunk alapján már nem kell annyit gondolkodnunk egy-egy kérdésen).

---

[13] Brian Christian és Tom Griffiths, *Algorithms to live by: The computer science of human decisions*, First U.S. Edition (New York: Henry Holt and Company, 2016).

Kahneman azonban kísérletek sokaságát hozza fel annak bizonyítására, hogy a döntéseink racionalitásához még mindig túl sok reményt fűzünk. Hogy rögtön egy jogi példát hozzunk fel: több, mint 15 éves tapasztalattal rendelkező bírókat arra kértek, hogy olvassák el egy bolti tolvaj aktáját, majd ezt követően "manipulált" dobogókockát dobattak velük, amely vagy 3-as, vagy 9-es eredményt adott. Ezután megkérdezték a bírókat, hogy a dobott számnál kevesebb vagy több hónapra büntetnék-e a tolvajt, majd megkérték őket, hogy pontosítsák, mennyi időre ítélnék el. Azok, akik 9-es dobtak, átlagosan 8 hónapot mondtak, akik 3-ast dobtak, átlagosan 5 hónapot.[14] „Miért teljesítenek rosszabbul a szakemberek az algoritmusokhoz képest? Az egyik ok, amelyet Meehl hibáztat, hogy a szakemberek megpróbálnak okosak lenni, sémákon kívül gondolkodni és amikor jósolni próbálnak, akkor tényezők összetett kombinációját veszik figyelembe. Az összetettség különös esetekben működhet, de gyakran inkább csökkenti az érvényességét. A tényezők egyszerű kombinációja jobb. Számos tanulmány bizonyította, hogy az emberi döntéshozók még akkor is gyöngébbek egy előrejelző formulánál amikor megkapják a formula által javasolt értékeket! Úgy érzik, hogy felülbírálhatják a formulát mert nekik további információjuk van az esetről, de gyakrabban tévednek, mint nem. Meehl szerint jó néhány eset van, amikor jó ötlet a döntést formulára bízni".[15] Kahneman felhoz egy másik nagyon fontos aspektust az emberi döntéshozatal minőségével kapcsolatban: "Egy másik oka a szakemberek döntéseinek rosszabb minőségének az, hogy az emberek javíthatatlanul következetlenek, amikor összefoglaló ítéletet kell alkotniuk összetett információkból. Amikor arra kérjük őket, hogy ugyanazokat az információkat értékeljék kétszer, rendszerint különböző választ adnak. Gyakran aggodalomra ad okot a kiszámíthatatlanság ilyen mértéke".[16]

Kahneman nem is rejti véka alá a véleményét azzal kapcsolatban, hogy mit a teendő, amennyiben kiszámítható, valódi, külső körülményektől mentes döntéseket szeretnénk: „Lehetséges mindenféle előzetes statisztikai tudás nélkül hasznos algoritmust fejleszteni. Egyszerű, jól kiegyensúlyozott, józan észen vagy meglévő statisztikán alapuló formulák általában a lényeges kimeneteleket elég jól megjósolják".[17] Viszont amint éppen a témánk szempontjából lényeges, egy algoritmus hibás döntésének az elfogadtatása nehéz a társadalommal szemben: „Ezzel szemben Meehl és az algoritmusok más pártolói határozottan állítják, hogy nem etikus intuitív ítéletekre bízni fontos döntéseket, ha egy algoritmus

[14] Daniel Kahneman, *Thinking, fast and slow*, 1st pbk. ed (New York: Farrar, Straus and Giroux, 2013), 148.
[15] Kahneman, 267.
[16] Kahneman, 266.
[17] Kahneman, p. 268.

rendelkezésre áll és kevesebb hibát követ el. Racionális érvelésük kényszerítő, de szembe megy a kijózanító pszichológiai valósággal: a legtöbb ember számára a hiba oka számít. Egy algoritmus hibás döntése miatt meghalt gyerek története megrendítőbb mintha ugyanez a tragédia egy emberi hiba miatt fordul elő, és az érzelmi intenzitásban megjelenő különbség rögtön átfordul erkölcsi előnyben részesítésbe".[18]

b) *Jogi döntéshozatal emberi jellege – Jerome Frank*

Ha élne, Jerome Frank minden bizonnyal mélyen egyetértene Kahneman nézeteivel, bár az kérdéses, hogy a végső konklúziójával (jobban járnánk-e az algoritmusokkal) kiegyezne-e.

Jerome Frank az amerikai realizmus egyik nagy hatású képviselője volt, amely mozgalom azt a célt tűzte a zászlajára, hogy a jogot, a jogrendszert a „valódi" működésében vizsgálja, tehát ne csak teoretikusan, „egyszerű" szabályrendszerként fogja fel.

Frank, maga is gyakorló bíróként a saját munkatársainak a kiszámíthatatlan döntéseire mutatott rá elég élesen. Alapvetően ő ezt két tényezőre vezette vissza: egyrészt a ténykérdések meglétére másrészt a bíróban meglévő előítéletekre és emberi mivoltukra.

A ténykérdések tekintetében Frank azt hangsúlyozta, vitát vállalva kortársaival, hogy a legegyszerűbb, legmechanikusabbnak látszó jogkérdésben is szükséges a ténymegállapítás, márpedig a tények nagyban befolyásolják egy per kimenetelét (ebben komoly vitája volt elődeivel és kortársaival pl.: Langdell-lel és Pound-dal, akik a szabályok elsőbbségében és kiszámíthatóságában hittek[19]). Ugyanazon cselekmények, események másképp és másképp lesznek bírók által megítélve, így a legegyszerűbben alkalmazandó jogszabály esetén is megannyi kimenetel lehetséges.[20]

Témánk szempontjából azonban Frank rámutatott egy még lényegesebb tényezőre: a bíró emberi mivoltára. Megannyi bírótípus létezik, mindegyik a maga előítéletességével, társadalmi

---

[18] Kahneman, p. 277.

[19] Neil Duxbury, „Jerome Frank and the Legacy of Legal Realism", *Journal of Law and Society* 18, sz. 2 (1991): 177–78, https://doi.org/10.2307/1410136.

[20] Jerome Frank, „Are Judges Human? Part 1: The Effect on Legal Thinking of the Assumption That Judges Behave Like Human Beings", *University of Pennsylvania Law Review* 80, sz. 1 (1931. november 1.): 31–33.

hátterével és meggyőződésével, és ezek keverednek azzal a jogi tudással, amelyet alkalmaznak. Kiemeli, hogy – bármennyire fáj kimondani – de korrupt, romlott bírók is vannak, akik, ha kell, ugyanúgy racionálisan és jól tudnak érvelni a kifejezetten rossz döntésük mellett, ezt mégis bajosan lehetne helyesnek nevezni.[21] De Frank azt sem hallgatja el előlünk, hogy a tisztességes, jó bírók is emberek: saját élményeiből merítve írja le, hogy hosszabb tanúkihallgatások során gyakornokként a feladata volt könyvek véletlenszerű elejtése annak érdekében, hogy a tisztelt bíró úr ne szunyókáljon el.[22]

Frankben mindezek ismeretében azonban nem az az álláspont alakult ki, hogy algoritmusokkal, robotokkal, egyenletekkel kell felváltani a bírókat (mi több: Frank ettől függetlenül nem veti el a megérzéseket és az abba vetett bizalmat[23]), hanem sokkal inkább az, hogy a bírókban tudatosítani kell saját előítéleteket,[24] a jogászoknak pedig tisztában kell lennie ezzel a helyzettel, ahogyan találóan megjegyzi: egy mérnökhallgatóknak sem csak a tökéletesen összerakott motor működését mutatják be, hanem törötteket is prezentálnak számukra, hiszen olyanokkal is találkozhatnak.[25]

2. Gépi döntések

Jelen alfejezet célja, hogy a gépi, számítógépes (szoftveres), és ezen belül pedig kifejezetten az adatalapú MI működési rendellenességeire felhívjam a figyelmet. Tekintettel arra, hogy az adatalapú MI is egyfajta szoftver, ezért először a Critical Code Studies elnevezésű tudományág néhány fontos gondolatát megidézve bemutatom, hogy a szoftverek futása, megbízhatósága közel sem olyan magától értődő, mint az a közvélekedésben jelen van. Ez tehát egyfajta ellenérv a szoftverek (és így az MI) használata ellen, de az adatalapú MI-vel kapcsolatosan felmerülő anomáliákra is külön kitérek mivel csak így érthető meg, hogy az MI látszólagos fejlődése közel sem annak az eredménye, hogy a gépek, robotok, szoftverek megtanultak pontosabban és jobban gondolkodni, mint az emberek, hanem kiterjedt alkalmazásuk inkább az erősebb hardvereinknek, a felgyülemlett adatmennyiségünknek és az egyre hatékonyabb algoritmusainknak köszönhetők.

---

[21] Frank, 34.
[22] Frank, 35.
[23] Charles L. Barzun, „Jerome Frank, Lon Fuller, and a Romantic Pragmatism”, *Yale Journal of Law & the Humanities* 29, sz. 2 (2018. szeptember 16.): 145, https://digitalcommons.law.yale.edu/yjlh/vol29/iss2/1.
[24] Frank, „Are Judges Human?”, 38–39.
[25] Frank, 34.

a) Általánosságban a szoftverek megbízhatatlanságáról

A python programozási nyelvvel ismerkedőket az egyik első meglepetés az alábbi, egyszerűnek mondható összeadás kimenetelekor éri: 4 + 2. A python ugyanis a következő választ adja: 42. Ha jobban belegondolunk, a válasz valahol meglehetősen észszerű, hiszen ugyanezt a logikát várnánk el a „szín + ház" bemenet esetén is. A python azért adja ezt a választ, mert külön kell „megmondani", hogy számokról beszélünk: ha ez megtörténik, természetesen kijön a megfelelő eredmény. Ebből is látszik, hogy mennyire furcsa képződmények a számítógépes nyelvek, és ebből adódóan a velük felépített rendszerek, számítógépek is milyen meghökkentő megoldásokra képesek.

A Critical Code Studies (CCS) egy viszonylag új tudományág, amelynek a fókuszában épp ilyen és ehhez hasonló, a számítógépes kódok furcsa viselkedését és ebből adódó társadalmi hatásokat produkáló kérdések állnak. Vizsgálati területe, hogy a szoftvereknek, leszűkítve pedig pontosabban a számítógépes kódnak milyen kölcsönhatása van a természetes nyelvvel, a társadalommal és a világunkkal, amelyben élünk. Például – kiragadva egy konkrétumot – változik-e a nyelvhasználatunk, a gondolkodásunk attól, hogy számítógépes nyelvek vesznek minket körül, és vica versa, a természetes nyelv hogyan jelenik meg a számítógépes kódokban, az emberi gondolkodás miképpen érhető tetten egy-egy szoftverben. Ahogy Frabetti írja: „Mégis, a nyelv és a kód közötti interakció átjárja a világunkat. Például az interneten keresztüli emberi kommunikáció ugyanúgy magában foglal természetes nyelvet (például amikor e-mailt írunk vagy chatvonalat használunk) mint a kódot (például a hálózatba kötött számítógépe különböző protokollját). A természetes nyelv és a kód interakciója történik minden egyes alkalommal amikor számítógépek mindennapi feladatokat hajtanak végre. „A nyelv egyedül", ahogy Hayles írja, „többé nem a megkülönböztethető vonása a technológiailag fejlett társadalomnak, az sokkal inkább a nyelv plusz a kód".[26] Vagy ahogyan Annette Vee fogalmazza: „A beszéd, az írás és a kód igénybevételének lehetősége komplex retorikai lehetőségekkel szolgál: egyik sem alapértelmezett, és mindegyik különböző, közvetített „szolgáltatást" ad. A komplex műveltség – amely magában foglalja a beszéd, az írás és a kódolás képességét – most már mind szükségessé válhat, hogy választani tudjunk ezekből a retorikai lehetőségekből".[27] A

[26] Federica Frabetti, *Software theory: a cultural and philosophical study*, Media philosophy (London ; New York: Rowman & Littlefield International, 2015), 39.
[27] Annette Vee, *Coding literacy: how computer programming is changing writing*, Software studies (Cambridge, MA: The MIT Press, 2017), 122.

nyelv és a kód egyik legizgalmasabb metszéspontja a performatív aktusok, hiszen a számítógépes kód is felfogható ilyenként, kérdés, hogy ezt hová tudjuk helyezni egy bírósági döntésben megjelenő performatív aktussal, ám ennek a kérdésével jelen tanulmányban nem kívánok foglalkozni.

A CCS kialakulásakor még az adatalapú MI nem volt annyira elterjedt, ezért különösen érdekes, hogy milyen gondolatok merültek fel a szoftverek kiszámíthatóságát illetően.

Frabetti könyvében visszatérő tárgy a szoftverek kiszámíthatóságából és kiszámíthatatlanságából eredő paradoxon. Egyik oldalról nyilván az szeretnénk, hogy egy szoftver minél kiszámíthatóbb és megbízhatóbb legyen, másik oldalról a társadalmunk olyan erős elvárásokkal van a technológiai fejlődésünkkel kapcsolatban, hogy azt a minél nagyobb ugrásokra, kockázatvállalásokra kényszeríti, ami értelemszerűen éppen kiszámíthatatlanságot, nem megbízható működést eredményez – ezt a paradoxont ráadásul már 1968-ban, a híres Garmisch-jelentésben megállapították.[28]

A másik ilyen ellentmondás bizonyos értelemben szintén a kiszámíthatatlanra szomjazás okozza: a kiszámíthatatlant ugyanis bizonyos értelemben kreatívnak gondoljuk, és mint ilyet, egy pozitív dolognak. „A rendszertípus túlságosan nagy kreativitása így éppen „rossz" kiszámíthatatlannak van feltüntetve – valami, ami túllépi és fenyegeti a projekt kivitelezését. Így aztán megint csak a technológia kiszámíthatatlan következményeivel, a [kód] átláthatatlanságával szembesülünk és ezeknek a kiűzhetetlenségével, amit a szoftverfejlesztők el akarnak érni; miközben ugyan a szoftverfejlesztés célja a kiszámíthatatlanság kiűzése a technológiából, a kiszámíthatatlan – mint a kreativitás jele – a technológia alapjaként van elismerve (és néha az emberi kreativitásnak tulajdonítják)".[29]

Cox és McLean a kiszámíthatóság tekintetében így írnak: „Valóban, a számítógépes programok a parancs és az ellenőrzés kettős regiszterén keresztül valósulnak meg – és ez a működés több szintjén történik így, és ekképpen jutnak el a végrehajtandó logikáig, ami eldöntöttnek és megváltoztatatlannak tűnik. Mégis a program először programozva volt. Tehát ugyan a felszínen olybá tűnik, mintha a számítógépes kód hasonlóan szuverén módon egyértelmű

[28] Frabetti, *Software theory*, 76.
[29] Frabetti, 84.

ágensként működne, konkrétan egy szuverén kódból eredő instrukciónak a parancsként történő végrehajtásaként, jelen könyvben mégis amellett fogunk érvelni, hogy lényeges módon ezen működések hibáknak és bugoknak van kitéve és lényeges kérdésekben úgy lehet tekinteni, mint ami kikerül az ellenőrzés alól pont úgy, mint a beszéd".[30]

A CCS-ben ekképpen megjelenő gondolatmenettel azt kívántam illusztrálni, hogy a szoftverek kiszámíthatóságába vetett hit legalábbis kérdőjeles: a klasszikus „ugyanolyan bemenetekre ugyanolyan kimenetet ad" logika bizonyosan sokszor megállja a helyét, de ha a saját életünket átgondoljuk, hányszor fordult már elő, hogy ugyanolyan helyzetben más eredményt produkált a számítógépünk vagy a szoftverünk? Hányszor fagyott már le ugyanolyan körülmények között, mint amivel korábban ezerszer használtuk (éppen erről szól az ismert, Turing által felfedezett megállási probléma, vagyis arról, hogy egy rendszer sosem tudja magáról előre megmondani, hogy egy adott inputra tud-e kimenetet adni[31])?

b) Mesterséges intelligencia hibaforrásai

Mint azt már a korábban is említettem, az MI óriási fejlődést mutatott be különösen az elmúlt évtizedben, de ezzel együtt a korlátjai is egyre világosabban látszanak. Meghökkentő például, hogy az MI-n alapuló algoritmusok mennyire tökéletesen képesek bemérni az ízlésünket és ezzel együtt különböző termékeket és szolgáltatásokat ajánlani nekünk – másik oldalról ugyanilyen meghökkentőnek érezzük, amikor egy automatikus robothang próbál „beszélgetni" velünk. De elég, ha az önvezető autók permanens ígéretére gondolunk: miközben visszatérő eleme az MI határtalan fejlődésének a GO játékban vagy a sakkban történő győzelem, a jóval egyszerűbbnek tűnő vezetési feladattal – a temérdek elégetett pénz dacára (már alsóhangon több, mint 100 milliárd dollár fölötti összegről beszélünk[32]) – még nem volt képes megbirkózni. De továbbra sem szolgálnak ki minket robotok a boltban, és a döntéseket is inkább még emberi

[30] Geoff Cox és Alex McLean, *Speaking code: coding as aesthetic and political expression*, Software studies (Cambridge, Mass: The MIT Press, 2013), 6.
[31] Jeffrey M Lipshaw, „Halting, Intuition, Heuristics, and Action: Alan Turing and the Theoretical Constraints on AI-Lawyering" 5 (é. n.): 44.
[32] „The Race To Driverless Technology | Leasing Options", elérés 2020. február 15., https://leasingoptions.co.uk/driverless-cars/index.html.

bírók hozzák. Hogyan lehet az MI egyrészt okosabb a világ legjobb sakkozójánál, Kaszparovnál ugyanakkor butább, mint egy kamionsofőr?

A válaszhoz először egy kicsit pontosabban meg kell értenünk, hogyan működik az adatalapú MI. Az egyik legtipikusabb példa egy automatikus spam szűrő készítése (a példát Harry Surden cikkéből[33] merítettem). Egy ilyen elkészítéséhez szükségünk van rengeteg megcímkézett adatra (tréning adatra), tehát egy olyan adatsorra, amiben egyrészt szerepelnek a spam levél adatai (a levél tárgya, szövege, feladója, származási helye stb.) másrészt szerepel, hogy az adott levél spamnek minősül vagy sem. Ha elegendő adatunk van, az algoritmusunk fel fogja ismerni a spam levél jellegzetességeit (pl.: tipikusan szerepel a tárgyukban a „nyerni" vagy a „hitelezés" szó), és egy annotálatlan (megcímkézetlen) adatot a jövőben képes lesz beazonosítani. Az MI tanulási folyamata közben a spam levél jellegzetességeit súlyozza akként, hogy minden egyes újabb adat alapján egy-egy meghatározott fontosságot tulajdonít az adott jellegzetességnek – ha a címkézés sikeres volt, akkor a súlyozáson nem változtat, amennyiben viszont a címkézés nem volt sikeres, helyesbíti, hogy melyik paraméter mekkora fontosságot kapjon. Az MI ennek köszönhetően tud dinamikus lenni és alkalmazkodni a külső körülményekhez. Az adatok különböző jellemzőit a továbbiakban paramétereknek, a paraméterek fontosságát pedig a továbbiakban súlyozásnak hívom. Éppen ezért kissé szerencsétlen az MI megnevezés, hiszen nem valódi intelligenciáról van szó. Ezért akarta ezt a módszert inkább „összetett információ feldolgozás" (complex information process) néven jelölni Herbert Simon és Allen Newell.[34] Erre a kérdésre utal Ződi, mikor a statisztikai adatokkal és statisztikai célú adathasználattal kapcsolatosan arra hívja fel a figyelmet, hogy annyiban komoly különbség van a statisztikai és a Big Data alapú eljárások között, hogy amíg a statisztikai számításokat előre meghatározott adatokkal és módszerekkel tesszük meg, addig a Big Data által felhasznált adatok spontán és folyamatosan termelődnek, olyan összefüggésekre rámutatva, amiknek mi sem vagyunk a tudatában.[35]

Mik tehát ennek a módszernek a korlátjai? A válasz nagyrészt az adatalapú szubszimbolikus voltában rejlik. Az MI-nek rengeteg adatra van szüksége, ahhoz, hogy hatékonyan tudjon

[33] Surden, „Artificial Intelligence and Law", 1312–14.
[34] David Leslie, „Raging Robots, Hapless Humans: The AI Dystopia", *Nature* 574 (2019. október 2.): 32–33, https://doi.org/10.1038/d41586-019-02939-0.
[35] Ződi, *Platformok, robotok ...*, 232–33.

működni,[36] és ez már sokat elárul róla: ha valami olyan helyzettel találkozik, amely számára ismeretlen, akkor biztos, hogy nem fog megfelelő kimenetet adni. Míg a Kaszparov elleni meccsen az ismeretlen helyzetek száma sok, de limitált, addig egy önvezető autó esetében éppen hogy végtelen mennyiségű új szituáció fordulhat elő. Ezt nevezik a „hosszú farok problémájának" (long tail problem) az az, hogy – különösen az életben előforduló minden helyzetre – nem tudjuk felkészíteni az MI-t.[37] Ez a probléma elvezet egy másikhoz is, amit „overfitting" vagy az általánosítás hiánynak nevezhetünk, nevesül: rendszeresen meg van a veszélye annak, hogy az MI-t egy nagyon specifikus helyzetre készítjük fel, és akár már csak egy kissé másabb problémára sem tud megfelelően reagálni.[38]

A másik sokat tárgyalt és kifejezett probléma az algoritmus összetettsége. A technikai részletek mély ismertetése helyett említés szintjén kell leszögezni, hogy a különböző neuronhálók és mélytanulási rendszerek egyszerűen annyira bonyolultak (annyiféle döntési csomópontot helyeznek a rendszerbe), hogy azok visszakövetése olyan mértékű nehézség, amit gyakorlatban nyugodtan hívhatunk lehetetlennek.[39] Ennek két lényeges következménye van: egyrészt egy-egy nem megfelelő kimenet esetén a hibafeltárás jelentősen megnehezül (hiszen eleve kérdés, hogy a tréning-adatban vagy magában az algoritmusban van-e a probléma), másrészt egy MI sokkal nehezebben vagy egyáltalán nem tudja visszaadni a felhasználónak, hogy miért hozott olyan döntést, amilyet.[40] Ez utóbb problémakör egyre nagyobb figyelmet kap, és az „önmagyarázó" MI lehetőségeit egyre többen kutatják. Ahogy Mitchell találóan megjegyzi: az iskolában is a legtorokszorítóbb pillanat, amikor oda kell adni a matek feladatunkat a tanárnak, hogy megnézze, miképpen jött ki az eredményünk, de egyben ebből is tudunk tanulni a legtöbbet.[41]

Az MI hatékonysága pedig szintén megkérdőjelezhető: a fenti overfitting problémával összefüggésben egy kutatócsoport megpróbált jóval egyszerűbb, szinte ökölszabálynak mondható algoritmust írni az Atari nevű játékra, amelyre korábban kifejezetten jó és hatékony MI algoritmust fejlesztettek viszonylag sok energiát és időt befektetve. Mint kiderült, az

[36] Gary Marcus és Ernest Davis, *Rebooting AI: building artificial intelligence we can trust*, First edition (New York: Pantheon Books, 2019), 37.
[37] Mitchell, *Artificial intelligence*, 101.
[38] Marcus és Davis, *Rebooting AI*, 23.
[39] Marcus és Davis, 59.
[40] Marcus és Davis, 41.
[41] Mitchell, *Artificial intelligence*, 107.

„ökölszabály”-szerű algoritmus alig ad vissza rosszabb eredményt, mint a speciálisan megalkotott algoritmus (ez meglepően rezonál Kahneman azon megállapításával, hogy az ember is éppen a túl sok tényező figyelembevétele miatt hoz rossz döntéseket).[42]

Ahogyan egyre több mindent MI-vel próbálunk megoldani, ezen problémák egyre erőteljesebben jelentkeznek. Gary Marcus és Ernest Davis friss könyvében éppen az ilyen kiszámíthatatlant és visszavezethetetlen döntések és hibák miatt hívják fel a figyelmet az MI veszélyeire, amikor arról írnak, hogy egészen addig, ameddig kizárólag Instagramra feltöltött képek taggelését rontja el az algoritmus, addig nincsen nagy baj, viszont ha magasabb a tét, és emberek meggyanúsítása múlik egy arcfelismerő szoftveren, akkor nem érhetjük be csupán annyival, hogy kicsit jobban működik, mint az ember.[43]

A téma mederben tartása érdekében nem kívánom az MI teljes történetét megidézni, azt azonban a teljes képhez hozzátartozik, hogy a fenti problémák már a ’80-as években ismertek voltak. Többek között például Dreyfus[44], valamint Winograd és Flores[45] kifejezetten erősen érveltek amellett, hogy az emberi gondolkodás és problémamegoldás – a látszólagos hasonlóságok ellenére – egészen másképp működik, mint az MI. Rengeteg érvvel támadták az MI módszereit, Dreyfus például hangsúlyozta, hogy minden adatnak, tevékenységnek a maga kontextusában van értelme és jelentősége, és a különböző kontextusok között az MI képtelen különbséget tenni. Például egy tipikus MI szoftvert lehetne készíteni arra, hogy különböző paraméterek alapján mondjuk meg egy adott személyről, hogy agglegény-e. Első ránézésre ez könnyűnek is tűnhet (van párja? férfi?), ám ekkor jönnek a zavarbaejtő példák, és kiderül, hogy az agglegényként jellemezhető személyek sora végtelen is lehet: agglegény-e Arthur, aki Alice-szal öt éve boldogan él, de nem házasodtak össze? Bruce, aki most válik, és folyamatosan hölgyekkel randizik? Charlie, gimnazista, aki 17 éves és a szüleivel lakik? David, 17 éves, aki 13 évesen elköltözött otthonról, most fiatal vállalkozó és playboy életet folytat? Eli és Edgar, akik homoszexuális kapcsolatban élnek? Faisal, akinek vallása szerint három felesége lehet, már kettővel rendelkezik és keres egy harmadikat? Gregory atya, aki katolikus pap? „Egy fogalom nem egyértelműen körülírható, egyszerű feltételek kombinációja hanem példaszerű

---

[42] Mitchell, 166.

[43] Marcus és Davis, *Rebooting AI*, 188.

[44] Hubert L. Dreyfus, *What Computers Still Can’t Do: A Critique of Artificial Reason*, 6. print (Cambridge, Mass.: MIT Press, 1999).

[45] Terry Winograd és Fernando Flores, *Understanding Computers and Cognition: A New Foundation for Design*, 24th printing (Boston: Addison-Wesley, 2008).

lista, amely sokféle tulajdonságot foglalhat magában" - írja Dreyfus. Ebből pedig egyértelműen következik, hogy „további példálózáshoz vezet, ha az agglegény jelentését kizárólag a „felnőtt" és feltételezhetően házasulandó" fogalmára szűkítjük le".[46] Ugyanígy Dreyfus kritizálja azt is, hogy tanulásnak nevezzük a fenti, paramétereket a hibák alapján dinamikusan módosító módszert is, hiszen az ember a hibáiból mélyebben tanul. Azt nem egyszerűen kijavítjuk, hanem strukturálisan átgondoljuk, hogy mi okozza a problémát, azonban az MI esetében az, hogy hol és melyik paraméternél keresse az MI a problémát már eleve előre, a programozó által el van döntve.[47] Jordan szintén egy nemrég megjelent cikkében reflektál az MI és az emberi gondolkodás között továbbra is meglévő szakadékra: „Az MI valami másra akart összpontosítani: az ember magasabb rendű vagy kognitív képességire, amelyek lehetővé teszik az értelemadást és a gondolkodást. Hatvan évvel később azonban a magasabb minőségű érvelés és gondolkodás megfoghatatlan maradt. A most MI-nek nevezett fejlődés leginkább mérnöki területen történik és alacsony szintű mintafelismeréssel és mozgásvezérléssel azonosítják ugyanúgy, ahogy a statisztikát, amelynek a lényege adatokban mintát találni és jól megalapozott jóslatokat tenni, hipotéziseket tesztelni és döntéseket hozni".[48]

Összességében kijelenthető, hogy az MI – dacára az elnevezésének – nem rendelkezik intelligenciával, legfeljebb az intelligencia imitálásának képességével.[49] Éppen ez a képessége teszi veszélyessé az MI-t: könnyen hihetjük azt, hogy az MI okosabb és jobb nálunk, holott valójában az MI meghatározott területeken, és az adatok pontosságának, elérhetőségének függvényében képes hatékonyan működni, és semmi garancia nincs arra, hogy ha egy területen sikeres (pl.: GO, sakk), akkor másikon is az lesz (pl..: orvoslás, jogi döntéshozatal). Mitchell ezt nevezi a transzformatív tudás hiányának: az ember képes arra, hogy valahol megszerzett tudását máshol kamatoztassa (pl.: aki pingpongozik az könnyebben tud átállni a teniszre, de saját témakörünkre vetítve egy sakkozó is biztosan rengeteg logikai összefüggést tudna meglátni egy-egy jogszabály értelmezésekor) ez viszont tökéletesen hiányzik az MI-ből.[50]

---

[46] Dreyfus, *What Computers Still Can't Do*, 49.
[47] Dreyfus, 114.
[48] Michael I. Jordan, „Artificial Intelligence—The Revolution Hasn't Happened Yet", *Harvard Data Science Review*, 2019. június 23., https://doi.org/10.1162/99608f92.f06c6e61. Elérés: 2020. február 21.
[49] Surden, „Artificial Intelligence and Law", 1308.
[50] Mitchell, *Artificial intelligence*, 164.

Látjuk tehát, hogy már eleve a kódalapú számítógépes szoftverek szintjén sem egyértelmű a kiszámíthatóság és a megbízhatóság teljes jelenléte, azonban ez az adatalapú MI esetében potenciálisan még nagyobb veszélyeket és problémákat tartogathat.

IV. Érvek az emberi és gépi döntések mellett

Bemutatásra került, hogy a szakirodalomban komoly érvek fogalmazódnak meg az ember döntések kiszámíthatóságával, minőségével szemben, de legalább ugyanilyen erős érvek szólnak az MI döntéshozatali mechanizmusaival szemben. Melyiket válasszuk? Melyiknek adjunk prioritást? Milyen olyan értékeket tudunk egyik vagy másik módszerhez kapcsolni, ami felülírhatja a velük járó hátrányt? Ezeknek a kérdéseknek próbálunk utánajárni először megvizsgálva az emberi döntés mellett szóló érveket majd megvizsgálva a gépi döntés mellett szóló érveket.

1. Emberi döntés mellett szóló érvek

a) ügyek egyedisége

Reuben Binns, az oxfordi egyetem kutatója plauzibilisen érvel amellett, hogy az embereket megilleti az egyedi elbírálás joga.[51] Elismeri, hogy lehetnek azonos vagy nagyon hasonló ügyek, azonban amellett foglal állást, hogy egy esetet megelőzően nem tudjuk, hogy milyen logika alapján kellene elbírálni, ezért ha általánosan, ugyanazt az algoritmust eresztjük rá az ügyek meghatározott körére, az könnyen figyelmen kívül hagyhat valamilyen olyan aspektust, ami lényeges.[52] Másképp megfogalmazva: „Lehetnek olyan tényezők, amiket rendszerszerűen nem lehet megfogni vagy olyanok, amelyek lényegtelenek valamennyi előző esetben, de hirtelenjében lényegessé válnak egy újban".[53] Másfelől ha tényleges emberi kontrollt szeretnénk az algoritmust fölött, akkor vagy valakinek mindig meg kellene vizsgálnia, hogy az adott ügyben esetlegesen milyen nóvum, újdonság van a korábbiakhoz képest, vagy már az ügy algoritmust általi eldöntését megelőzően el kellene dönteni, hogy ember vagy a gép hozzon-e

[51] Reuben Binns, „Human Judgement in Algorithmic Loops; Individual Justice and Automated Decision-Making", preprint (LawArXiv, 2019. szeptember 15.), https://doi.org/10.31228/osf.io/kz4s2. Elérés: 2020. február 21.
[52] Binns, 8.
[53] Binns, 9.

döntést – azonban meglátása szerint mindkét módszer továbbra is olyan energiát igényelne az ember részéről, ami tulajdonképpen fölöslegessé teszi az algoritmust.[54] Ahogy Binns fogalmaz: „Bár az algoritmikus rendszerek egyértelműen jól szolgálják az igazságosság egyik dimenzióját (konzisztencia), nem adják meg az egyedi igazságosságot és nem garantálják, hogy ne történjen diszkrimináció".[55]

b) felelősség, beszámoltathatóság, magyarázat

Mint korábban bemutatásra került, az adatalapú MI esetében rendkívül nehéz visszakövetni az egyes döntések hátterét és éppen ezért a felhasználó számára sem tud visszajelzést adni. Az MI ezen tulajdonsága sok szerző számára elfogadhatatlan. Davis a reciprok-elvet idézve amellett érvel, hogy az egész jogrendszerünk arra épül, hogy aki jogi döntést hozhat ránk, annak magának is a jog alatt kell állnia, márpedig jog alatt csak morális személyek állhatnak, az MI pedig egyértelműen nem az.[56] Wendel szintén nem tartja elhanyagolhatónak a magyarázatadást, ő elsősorban a társadalmunk demokratikus beállítódását hozza fel érvként: a jogviták lényege, hogy az emberek közötti nézeteltéréseket elsimítsa, méghozzá szemtől-szembe.[57] Markou és Deakin a döntés társadalmi vonására utalva rámutatnak: egyrészt egy jogi döntés nem leírni, rekonstruálni akar egy valóságot (pl.: hogy rákos-e az illető), hanem „ez egy normatív aktus azzal a céllal, hogy megváltoztassa a valóságot".[58] A magyarázatnélküliségre összefoglalóan Zódi amellett is érvel, hogy éppen emiatt nehezebb is megteremteni egy gépi döntés elfogadhatóságát: „Ugyanakkor az ember a saját „leegyszerűsítő" morális narratívái között él, és ezek alapján a megélt narratívák alapján akar magyarázatokat a világ dolgaira. Egyelőre elfogadhatatlannak látszik, hogy a gép „jobban tudjon" valamit, és elfogadjunk egy általa hozott és a hagyományos mércéinkkel abszurdnak tűnő döntést, arról nem is beszélve, hogy a gép nem

---

[54] Binns, 17.
[55] Binns, 13.
[56] Joshua P. Davis, „Artificial Wisdom? A Potential Limit on AI in Law(and Elsewhere)", *OKLAHOMA LAW REVIEW* 72, sz. 1 (2019): 62.
[57] W Bradley Wendel, „The Promise and Limitations of Artificial Intelligence in the Practice of Law", *OKLAHOMA LAW REVIEW* 72 (2019): 42–44.
[58] Christopher Markou és Simon Deakin, „Ex Machina Lex: The Limits of Legal Computability", SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 2019. június 21.), 30, https://papers.ssrn.com/abstract=3407856. Elérés: 2020. február 21.

fog tudni megnyugtató magyarázatot és indokolást adni, sőt (leegyszerűsítő) emberi mércéinkkel mérve még meggyőzőt sem".[59]

c) átláthatatlanság, értékek, előítéletek, adatminőség

Az MI esetében köztudott és tematizált probléma az adatokban rejlő előítéletesség, ezért ezt hosszabban nem is ragoznám. Ami azonban ehhez kapcsolódik és kevésbé említett, az – egy másik kontextusban ugyan, de ide is kapcsolódó releváns – a meglátás amelyet Primavera De Filippi hoz fel: eszerint az algoritmusokat gyakran privát cégek hozzák létre, többnyire hozzáférhetetlenek ráadásul olyan „értékeket" rejthetnek el bennük, amelyekről a társadalom nem is tudhat és ehhez kapcsolódóan az nem is ment át társadalmi vitán.[60] Ennél már csak egy rosszabb fordulhat elő: Harry Surden arra mutat rá, hogy bizonyos értékek megjelenése az algoritmusban nem is célzott szándék, hanem amolyan „mellékhatás", ráadásul a szándék hiánya miatt sokkal nehezebben kimutatható vagy explicit.[61] Ződi Zsolt a gépi működés átlátszatlanságát egy a római kori pontifexek működéséhez hasonlítja, „akik kiszámíthatatlanul mondhatták egy napra azt, hogy ez alkalmatlan a pereskedésre".[62] Mint az már az MI bemutatásakor megállapításra került, az adatok kulcsfontosságúak az MI működése szempontjából, ezért kifejezetten nem mindegy, hogy miképpen és hogyan, milyen minőségű adatokhoz jutunk hozzá. Az adatok minőségével kapcsolatban kevesebbet tárgyalt, de fontos nézőpont: Markou és Deakin reflektál arra a jelenségre, hogy az MI hatékony működéséhez annotált adatokra van szükség, amelyben jelenleg hiányt szenvedünk (bár hozzáteszik, hogy ez a fejlődést most nem akadályozza).[63] Ráadásul mind az adatok, mind az algoritmusok átlátszatlanságából következően közismert, hogy akár az adatok rosszindulatú manipulálásával, akár később, az algoritmust is át lehet verni.[64]

---

[59] Ződi, *Platformok, robotok ...*, 238.

[60] Primavera De Filippi és Samer Hassan, „Blockchain Technology as a Regulatory Technology: From Code Is Law to Law Is Code", *First Monday* 21, sz. 12 (2016. november 14.), https://doi.org/10.5210/fm.v21i12.7113. Elérés: 2020. február 21.

[61] Harry Surden, „Values Embedded in Legal Artificial Intelligence", SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 2017. március 13.), 3, https://papers.ssrn.com/abstract=2932333.

[62] Zsolt Ződi, „Gépek a jogban: Jogelméleti gondolatok a számítógépek jogalkalmazásáról", *JOGELMÉLETI SZEMLE* 2013, sz. 2 (é. n.): 212.

[63] Markou és Deakin, „Ex Machina Lex", 16.

[64] Claude Castelluccia és mtsai., *Understanding Algorithmic Decision-Making: Opportunities and Challenges.*, 2019, 33–37, http://publications.europa.eu/publication/manifestation_identifier/PUB_QA0618337ENN. Elérés: 2020. február 21.

2. Érvek a gépi döntések mellett

Az általam ismert legszélesebb, legmélyebb és legaktuálisabb véleményt a gépi döntések mellett (illetve az emberi döntésekkel szemben) Aziz Huq chicagói professzor írta, így elsősorban az ő tavaly megjelent cikkét idézem, reflektálva az eddig felhozott szempontokra.[65]

Huq véleményét a legtömörebben talán az alábbiak szerint lehetne összefoglalni: természetesen a gépi döntés sem tökéletes, vannak olyan, kifejezetten a gépi döntéseknél jelentő veszélyek (pl.: adattorzítás), amelyek hibákat generálnak. Ami azonban Huq állítása, hogy nincs racionális okunk azt gondolni, hogy a gépi döntés bármivel is rosszabb lenne, mint az emberi. A GDPR 22. cikkének, emberi közbeavatkozására rezonálva vitatja, hogy bármilyen komoly érvrendszer felmerült volna arra vonatkozóan, hogy egy gépnek miért kéne tudnia a saját döntését megindokolni. Akként fogalmaz, hogy szakadékot vél felfedezni a magyarázatadási kötelezettség nem megfelelő alátámasztása és a megalkotott jog között.[66]

Egyik legerősebb érve, hogy a saját agyunk és gondolkodásunk is egy fekete-doboz ("black-box"): tudományosan igazolható, hogy nem rendelkezünk olyan univerzális leírással, amely képes visszafejteni, hogy egy-egy ember miképpen jut egy-egy döntésre (ez annál is inkább így van, hiszen ha lenne ilyen módszer, akkor mesterséges intelligenciát is jóval könnyebb lenne építeni). Huq azokat, az általam is idézett szerzők által is felhozott problémát is visszautasítja, hogy az MI működése gyakorlatilag kijavíthatatlan: „Ebből adódóan a gépi tanulás rendszereiben a hibás leágazás visszakövethető. Gyakorlatilag az algoritmus felépítésének lényeges elemeinek következményei elkülöníthetők és megvizsgálhatók. Alternatívaként lehet többszörösen különböző alternatívákat adni egy algoritmusnak, tesztelve a kimenetet a bemenet folyamatos megváltozására akár a tényeket követően is".[67] Ehhez hasonló példát hoznak Castelluccia-ék, akik a „Local Interpretable Model-agnostic Explanations" modell kutatási elgondolását idézik, miszerint olyan rendszert még nem vagyunk képesek készíteni, amely globális módon megmagyarázza magát, de az megvalósítható, hogy egyes részeiben

[65] Aziz Z. Huq, „A Right to a Human Decision", SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 2019. május 3.), https://papers.ssrn.com/abstract=3382521. A szerző közlése szerint a tanulmány megjelenése 2020. májusában várható a Virgina Law Review-ban. Elérés: 2020. február 21.
[66] Huq, 14.
[67] Huq, 23.

magyarázatot adjon.[68] Idéznek olyan kutatásokat is, amelyek képesek az adattisztításra olyan módon, hogy az ne legyen torzító vagy előítéletes.[69]

Szintén meglepő és elgondolkodtató azon állítása, miszerint nem létezik teljes mértékben, kizárólag gépi döntéshozatal: "A GDPR. 22. cikke a „kizárólag automatikus meghozott" döntések lehetőségét látja előre. De valóban létezik ez a szörnyeteg? Talán nem – három okból. Először is, mindenféle gépi tanulásos módszer eredetét tekintve emberi tervezés és mérnöki választás gyümölcse. Nincs teljesen endogén algoritmus. És egy gépi tanulásos rendszer megtervezése nem gépies feladat".[70] Huq szerint az, hogy mikor jelenik meg az ember az algoritmus mögött (annak megalkotásakor, kijavításakor) igazából lényegtelen, mert biztosak lehetünk benne, hogy valamikor ember volt mögötte és – feltételezve, hogy hatékonya algoritmusban vagyunk érdekeltek – újra lesz mögötte ember.

Arra, hogy feltétlenül az eljárásban vagy az eljárás ellenőrzésére embert kellene beépíteni a gép ellenőrzésére, szintén elgondolkodtató ellenérveket hoz fel. Először is idézi Henry Friendly bírót, aki megállapította, hogy bármilyen további garanciát próbálunk beépíteni egy eljárásba, az az előny, amit kapna az ügyfél, elvész azáltal, hogy egy eljárásban alapvetően véges erőforrásunkat próbálunk elosztani.[71] Ezen logikán tovább haladva Huq állítása, hogy az emberi ellenőrzés egy ugyanolyan látszólag többletgaranciát adó, valójában az eljárást nehezítő, lassító és az eljárás hatékonyságát nem javító elem lenne. Sommás véleménye szerint: „Ahol az algoritmikus rendszer hibás, abból nem következik, hogy az ex post emberi ellenőrzés „igazságos". Valójában, amire minden okunk meg lenne, hogy igazságosnak gondoljuk, az egy jobb gépi döntés mint egy megbízhatóan megbízhatatlan emberi".[72]

Vitatkozik azzal az érvvel, hogy egy döntést feltétlenül meg kell-e indokolni (hiszen rengeteg olyan jog által hozott döntés van, amelyben ez nem elvárás az emberek részéről) valamint a döntések egyediségére vonatkozóan ő maga nem tud egyedibb igazságszolgáltatást elképzelni, minthogy van egy rendszer, amely az elérhető szinte összes lényeges paramétert figyelembe tudja venni. Véleményét azzal zárja, hogy szerinte az emberi döntéshozatalnak annyiban marad

[68] Castelluccia és mtsai., *Understanding Algorithmic Decision-Making*, 48.
[69] Castelluccia és mtsai., 46–47.
[70] Huq, „A Right to a Human Decision", 25.
[71] Huq, 39.
[72] Huq, 40.

relevanciája, hogy szülessenek olyan előremutató, újdonságokat felhozó döntések, amelyek a jogrendszer rugalmasságát biztosítják.[73]

3. A végső paraméter: az óvatosság

Bartosz Brozek tavaly megjelent könyvében a (nyugati) jogi gondolkodás jellemzőit igyekezett feltárni, olyan sajátosságok azonosításával, amelyek a jogi probléma megoldásainkat egyedivé teszik.[74]

Brozek – ellentmondva Kahnemanak– plauzibilisen érvel amellett, hogy az intuíció igenis fontos, kitörölhetetlen része a jogi problémamegoldásnak azzal együtt, hogy rengeteg tapasztalatra és óvatosságra van szükség ahhoz, hogy egy jogász a megérzéseit megfelelő módon kamatoztassa.[75] Teljesen ellentmond tehát Kahnemanak az intuíciók használata tekintetében, akinek a lényegi gondolata éppen az, hogy döntéseket nem bízhatunk a megérzéseinkre.

Sokkal lényegesebb azonban Brozeknek az a felsorolása, amely a jogi gondolkodás hat olyan elvét adja meg, amelyek egy-egy jó döntés megalkotásához szükséges.[76] Az alapelvek egyike pedig az óvatosság.[77]

Az MI köztudomásúlag képtelen az óvatosságra, hiszen a meglévő válaszaiból mindenképpen output-ot akar generálni. Tipikus példa, hogy amennyiben MI-vel szeretnénk képeket felismertetni két különböző emberről, feltételezve, hogy van elég tréning adat, az MI nagyon magas százalékos aránnyal fogja tudni megkülönböztetni őket. A probléma akkor keletkezik, ha újabb adatként egy teljesen új képet viszünk fel (például egy autóét), az MI kimenetként akkor is a két személy közül valakit fog azonosítani tekintettel arra, hogy csak ezeket a kimeneteleket ismeri, és egészen eltérő új input esetén sem képes azt mondani, hogy "elnézést, ez valami egészen más, nem tudom, hogy mi ez". "Természetesen, a transzparens és vonzó

---

[73] Huq, 52.
[74] Bartosz Brożek, *The Legal Mind: A New Introduction to Legal Epistemology*, 1. kiad. (Cambridge University Press, 2019), https://doi.org/10.1017/9781108695084.
[75] Brożek, 124.
[76] Brożek, 122.
[77] Brożek, 124.

módon becsomagolt megbízható algoritmus egyik problémája hogy "túlzott" bizalomhoz vezethet, amely során a kimenet megkérdőjelezhetetlennek és precíznek minősül. Egy valódi megbízható algoritmusnak képesnek kellene lennie kommunikálni a korlátait és ironikus módon, biztosítani, hogy nem bíznak meg benne eléggé"- írja Spiegelhalter.[78]

A gépi "gondolkodás" és döntéshozatali módszerek természetesen sok mindenben eltérnek az emberitől és a jogászitól, ami a gépi döntést – akár tudatalatt is – valóban kevésbé elfogadhatóvá teszik. Röviden utalok itt arra a tényezőre, amire Brozek például szintén rámutat, hogy egy jogász egyszerre képes induktívan (bottom-up) és deduktívan (top-down) gondolkodni[79], és ezen két módszer ötvözése jelenleg hiányzik az MI apparátusából (ld.: erre Mitchell[80] vagy Gary Marcusék[81] véleményét). De ugyanilyen lényeges Mitchell utal arra Lakoff és Johnson által bizonyított jelenségre, hogy az ember az alapvető fogalmait metafora-rendszereken keresztül érti meg (pl.: az időt pénznek fogjuk fel, ld.: "elpocsékoltam az időmet", "rosszul gazdálkodtam az időmmel" stb), amit nyilvánvalóan az MI nem tud magáévá tenni.[82]

Véleményem szerint azonban – kissé ironikusra véve a figurát – éppen az óvatosságra tekintettel kellene óvatosabbnak lenni az MI-vel. Ha valaki döntést hoz az ügyünkben, jogosan várjuk el, hogy képes legyen észrevenni az ügy azon nem szokványos jellegét / jellegeit, amelyeknek a döntés eredményében is meg kell jelenniük. Természetesen nincsen rá garancia, hogy egy emberi döntéshozó képes ezt minden esetben megtenni, de legalább a lehetőség meg van rá, amely a MI esetében teljes mértékben hiányzik. Markou és Deakin éppen a gépi kimenetelek minőségét kritizálja az új adatok kezelhetetlensége miatt: „Mivel a mélytanulásos rendszereknek a tréning adatokon túlmenően kell általánosítania a megoldásukat (pl.: kiejteni egy új szót vagy felismerni egy nem látott képet), az elérhető adatok úgy korlátozzák a teljesítményt, hogy nem képesek magas minőségű megoldásokat garantálni".[83]

---

[78] David Spiegelhalter, „Should We Trust Algorithms?", *Harvard Data Science Review*, 2020. január 31., https://doi.org/10.1162/99608f92.cb91a35a. Elérés: 2020. február 21.
[79] Brożek, *The Legal Mind*, 113–21.
[80] Mitchell, *Artificial intelligence*, 40.
[81] Marcus és Davis, *Rebooting AI*, 141.
[82] George Lakoff és Mark Johnson, *Metaphors we live by* (Chicago: University of Chicago Press, 2003).
[83] Markou és Deakin, „Ex Machina Lex", 17.

V. Nézetek összevetése

A tanulmány elején felsorolt három nézőpontot (kiszámíthatóság, objektivitás, körültekintés / megfontoltság) szempontjából értékeljük, hogy az irodalom alapján melyikben melyik típusú döntés előnyösebb.

1. Kiszámíthatóság

Mint láttuk, Kahnemann és Frank is erősen kritizálta az emberi döntéseket a kiszámíthatóságuk szempontjából. Véleményem szerint az emberi döntések mellett állásfoglalók (Binns, Wendel, Markou és Deakin) sem tudnak kifejezetten vitába állni azzal a kijelentéssel, hogy a gépi döntéshozatal jóval kiszámíthatóbb, így e tekintetben magasabb minőséget képvisel. Mint láttuk, Binns ezt megkerülve inkább az esetek egyediségével kívánta azt demonstrálni, hogy a kiszámíthatóság legalábbis fölösleges, ha olyan, újabb paramétereket kellene megfontolás tárgyává tenni, amelyeket nem építettek be az algoritmusba. A III. fejezet 2. alfejezetében magam idéztem a CCS néhány gondolatát a szoftverek kiszámíthatóságába vetett hit optimista voltáról. Nyilvánvalóan a CSS kritikája érvényes, és ennél még érvényesebb az a kijelentés, hogy az MI időnként kifejezetten meghökkentő eredményeket tud produkálni, amelyet az algoritmusok bonyolultsága miatt kifejezetten nehéz is visszakövetni. Az azonban bizonyos, hogy egy szoftver esetében folyamatos teszteléssel jóval nagyobb valószínűséggel tudjuk megtalálni az érthetetlennek tűnő hibát, míg az emberi kiszámíthatatlanság ennél jóval titokzatosabb és beláthatatlan, éppen ezért több eszközünk van a szoftverek kiszámíthatóságának szavatolására.

2. Objektivitás

Az emberi döntések objektivitása Kahneman és Frank részéről szintén komoly kritikában részesültek, Kahneman részéről inkább amiatt, mert véleménye szerint sokszor olyan tényezők is hatnak ránk, amelyeknek egyáltalán nem vagyunk tudatában, míg Frank arra is reflektált, hogy kifejezetten rosszindulatú (akár korrupt) bírók is vannak a rendszerben. Az objektivitás mércéjét azonban a gépi döntéshozatal is sokkal nehezebb állja ki: az adatok eleve nagyon könnyen tükrözik vissza a társadalomban meglévő előítéleteket, de nem nehéz az adatokat egyéb úton is torzítani. Ráadásul az algoritmus átláthatatlansága és bonyolultsága – mint azt de

Filippi és Surden érveléséből láttuk – egyáltalán nem garantálja, hogy esetleg magába az algoritmusba nincs beépítve valami olyan, egyes értékeket propagáló megoldás, amely esetleg diszkriminatívnak mondható. Huq az algoritmusok rossz döntését „védve" azzal válaszol erre, hogy az emberi döntések ugyanolyan visszakövethetetlenek, de ebből az érvelésből látszik, hogy e tekintetben nem feltétlenül van különbség ember és gép között. Összességében tehát az objektivitás szempontjából mind az emberi, mind a gépi döntéshozatal erősen megkérdőjelezhető és vélhetően az is marad.

3. Körültekintés / megfontoltság

A körültekintés és megfontoltság tekintetében plauzibilisnek látszana a gépi döntéshozatal behozhatatlan előnye tekintettel arra, hogy tulajdonképpen végtelen mennyiségű paramétert képes figyelembe venni. Azonban mint Mitchell és Spiegelhalter is utalt rá, az algoritmus kifejlesztésébe befektetett rengeteg idő és energia közel sem teszi sokkal hatékonyabbá egy jóval egyszerűbb algoritmusnál, tehát ismét igazzá válik a népi bölcsesség, miszerint a kevesebb néha több. Másrészt véleményem szerint idetartozik az óvatosság nézőpontja: a körültekintésbe és a megfontoltságba az is bele kell tartozzon, hogy a döntéshozó nemcsakhogy minden körülményt megvizsgált, de azokat – más döntésekhez, esetleg a társadalmi elvárásokhoz képest – képes radikálisan új nézőpontok szerint megvizsgálni. E tekintetben tehát azt gondolom, hogy határozottan az emberi döntés viseli a magasabb minőséget.

VI. Konklúzió

1. További kutatási kérdések

Bár vannak már jogi normák, amelyek tartalmaznak rendelkezést algoritmikus döntéshozatalra, ettől függetlenül de lege ferenda a leendő jogalkotók részére a fentiek szerint idézett irodalom alapján megfogalmazható néhány olyan aspektus, amelyet figyelembe kell venni az algoritmikus döntéshozatalra való szabályozási rezsim kialakításakor.

A legfőbb kérdés, és amely további kutatást igényel, hogy a társadalom mennyiben hajlandó az óvatosságra, önmagát elvileg kontrollálni képes emberi bíró lecserélni a gyorsabb, kiszámíthatóbb, de éppen ezért korlátozott döntési kimeneteket ismerő gép számára. Azt

gondolom, hogy egy ilyen szabályozás megalkotása előtt – ellentétben a GDPR általános és mindenre kiterjedő szabályozásával szemben – ügycsoportonként kellene szétválasztani azokat az eseteket, amelyeket hajlandóak vagyunk rábízni az algoritmusra, és azokat, amelyeket olyanként értékelünk, amelyekben szükségünk van az emberi óvatosságra és emberi körültekintésre. A kahnemani kutatásokat ismeretében ugyanis nehezen lehet amellett érvelni, hogy semmilyen ügykörre nem engedjük be az algoritmusokat, de ebben megkerülhetetlen a társadalom ingerküszöbének a kikutatása annak érdekében, hogy az így meghozott döntések társadalmi konszenzust élvezzenek. Amint Meehl-t idézve Kahneman írja: ha vannak jól bevált és működő rendszereink, akkor azokat kifejezetten nem etikus nem használni, mikor rendelkezésre állnak.

Ha elfogadjuk a fenti javaslatot, az viszont az alapjogias megközelítésű európai kultúrában okozhat nehezen feloldható és további kutatást igénylő problémát: ha bármilyen megfontolás mentén (pl.: ügycsoport, az eljárás szintje, pertárgy érték, felek személye) szétvágjuk az eljárásokat emberi bíró és algoritmus által eldöntött esetekre, akkor egy fontos aspektusban rendkívül lényeges különbség lesz az eljárások között. Éppen jelen tanulmány által követett érvelésekből következően mondhatja bármelyik csoport, hogy őt hátrány éri: az algoritmikus döntést kapók érvelhetnek azzal, hogy joguk van az emberi bírói döntéshez, hiszen az körültekintőbben, az eset esetleges nóvumait felderítő módon tud döntést hozni, miközben az emberi döntésben részesülők vitathatják, hogy miért a lassúbb és kiszámítatlanabb, megbízhatatlanabb igazságszolgáltatásban részesülnek. Filozófiai, erkölcsi és etikai szempontból szintén érdekes és megválaszolandó kérdés, hogy mit kezdünk azzal a szintén egyfajta európai értéknek tekinthető azzal Kanttól eredeztethető érveléssel, miszerint az embert nem tárgyiasíthatjuk, azonban egyértelműnek tűnhet, hogy ha egy ügyet adatokra redukálunk, akkor éppen ez történik (lásd erről egy kicsit más kontextusban Supiot könyvének idevágó részletét).[84]

2. Összefoglaló

Jelen tanulmány arra tett kísérletet, hogy bemutassa az algoritmikus döntéshozatal mellett és ellen szóló legfontosabb érveket, amelyhez elsősorban a bevezetőben lefektetett három alapértéket rendeltünk: a

[84] Alain Supiot, *Homo Juridicus: On the Anthropological Function of the Law*, ford. Saskia Brown, Paperback edition (London New York, NY: Verso, 2017), 44–67.

kiszámíthatóságot, az objektivitást és a körültekintést / megfontoltságot. Ehhez először áttekintésre kerültek az MI legfontosabb módszerei és típusai annak érdekében, hogy egyértelműsítsük: a tanulmányban az adatalapú, szubszimbolikus rendszerekről lesz szó. Ezt követően az emberi döntések jellegét Daniel Kahneman kutatásai alapján az emberi döntések sajátosságát elemeztük kiemelve azok kiszámíthatatlan jellegét, amely megfigyelést Jerome Frank jóval korábbi írásai is alátámasztottak a kifejezetten bírói döntésekre vonatkozóan. Annak érdekében, hogy az emberi és gépi döntések összemérhetőek legyenek, bemutatásra kerültek a köznyelvben MI-vel azonosított, adatalapú, szubszimbolikus rendszerekkel kapcsolatos fenntartások, problémák. Ezen áttekintések után, az aktuális szakirodalmat figyelembe véve ismertetésre kerültek az emberi döntést preferáló érvek majd a gépi döntést preferáló érvek, azzal, hogy Bartosz Brozek frissen megjelent, jogi gondolkodást elemző művének egyik lényeges elemét (óvatosság) kiemelve jelen tanulmány szerzője egy újabb érvet hozott fel az emberi döntések megtartása mellett. Végül a bevezetőben megadott három paraméter (kiszámíthatóság, az objektivitás és a körültekintés / megfontoltság) alapján megállapításra került, hogy a gép a kiszámíthatóság területén, az ember a körültekintés / megfontoltság területén az erősebb, és az objektivitás tekintetében mindkét döntésfajtának meg vannak a hátrányai.

Irodalomjegyzék

Alarie, Benjamin, Anthony Niblett, és Albert H. Yoon. „How Artificial Intelligence Will Affect the Practice of Law". *University of Toronto Law Journal* 68, sz. 1 (2018. március 27.): 106–24.

Ashley, Kevin D. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge New York Melbourne Delhi Singapore: Cambridge Univ Press, 2017.

Barzun, Charles L. „Jerome Frank, Lon Fuller, and a Romantic Pragmatism". *Yale Journal of Law & the Humanities* 29, sz. 2 (2018. szeptember 16.). https://digitalcommons.law.yale.edu/yjlh/vol29/iss2/1.

Binns, Reuben. „Human Judgement in Algorithmic Loops; Individual Justice and Automated Decision-Making". Preprint. LawArXiv, 2019. szeptember 15. https://doi.org/10.31228/osf.io/kz4s2.

Brożek, Bartosz. *The Legal Mind: A New Introduction to Legal Epistemology*. 1. kiad. Cambridge University Press, 2019. https://doi.org/10.1017/9781108695084.

Castelluccia, Claude, Daniel Le Métayer, European Parliament, és Directorate-General for Parliamentary Research Services. *Understanding Algorithmic Decision-Making: Opportunities and Challenges.*, 2019. http://publications.europa.eu/publication/manifestation_identifier/PUB_QA0618337ENN.

Christian, Brian, és Tom Griffiths. *Algorithms to live by: The computer science of human decisions*. First U.S. Edition. New York: Henry Holt and Company, 2016.

Cox, Geoff, és Alex McLean. *Speaking code: coding as aesthetic and political expression*. Software studies. Cambridge, Mass: The MIT Press, 2013.

Davis, Joshua P. „Artificial Wisdom? A Potential Limit on AI in Law(and Elsewhere)". *OKLAHOMA LAW REVIEW* 72, sz. 1 (2019): 51–89.

Dreyfus, Hubert L. *What Computers Still Can't Do: A Critique of Artificial Reason*. 6. print. Cambridge, Mass.: MIT Press, 1999.

Duxbury, Neil. „Jerome Frank and the Legacy of Legal Realism". *Journal of Law and Society* 18, sz. 2 (1991): 175. https://doi.org/10.2307/1410136.

Filippi, Primavera De, és Samer Hassan. „Blockchain Technology as a Regulatory Technology: From Code Is Law to Law Is Code". *First Monday* 21, sz. 12 (2016. november 14.). https://doi.org/10.5210/fm.v21i12.7113.

Frabetti, Federica. *Software theory: a cultural and philosophical study*. Media philosophy. London ; New York: Rowman & Littlefield International, 2015.

Frank, Jerome. „Are Judges Human? Part 1: The Effect on Legal Thinking of the Assumption That Judges Behave Like Human Beings". *University of Pennsylvania Law Review* 80, sz. 1 (1931. november 1.): 17.

Hildebrandt, Mireille. „Algorithmic Regulation and the Rule of Law". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, sz. 2128 (2018. szeptember 13.): 20170355. https://doi.org/10.1098/rsta.2017.0355.

Huq, Aziz Z. „A Right to a Human Decision". SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 2019. május 3. https://papers.ssrn.com/abstract=3382521.

Jordan, Michael I. „Artificial Intelligence—The Revolution Hasn't Happened Yet". *Harvard Data Science Review*, 2019. június 23. https://doi.org/10.1162/99608f92.f06c6e61.

Kahneman, Daniel. *Thinking, fast and slow*. 1st pbk. ed. New York: Farrar, Straus and Giroux, 2013.

Lakoff, George, és Mark Johnson. *Metaphors we live by*. Chicago: University of Chicago Press, 2003.

Leith, Philip. „The Rise and Fall of the Legal Expert System". *International Review of Law, Computers & Technology* 30, sz. 3 (2016. szeptember): 94–106. https://doi.org/10.1080/13600869.2016.1232465.

Leslie, David. „Raging Robots, Hapless Humans: The AI Dystopia". *Nature* 574 (2019. október 2.): 32–33. https://doi.org/10.1038/d41586-019-02939-0.

Lipshaw, Jeffrey M. „Halting, Intuition, Heuristics, and Action: Alan Turing and the Theoretical Constraints on AI-Lawyering" 5 (é. n.): 44.

Marcus, Gary, és Ernest Davis. *Rebooting AI: building artificial intelligence we can trust*. First edition. New York: Pantheon Books, 2019.

Markou, Christopher, és Simon Deakin. „Ex Machina Lex: The Limits of Legal Computability". SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 2019. június 21. https://papers.ssrn.com/abstract=3407856.

Mitchell, Melanie. *Artificial intelligence: a guide for thinking humans*. New York: Farrar, Straus and Giroux, 2019.

Spiegelhalter, David. „Should We Trust Algorithms?" *Harvard Data Science Review*, 2020. január 31. https://doi.org/10.1162/99608f92.cb91a35a.

Supiot, Alain. *Homo Juridicus: On the Anthropological Function of the Law*. Fordította Saskia Brown. Paperback edition. London New York, NY: Verso, 2017.

Surden, Harry. „Artificial Intelligence and Law: An Overview". *Georgia State University Law Review* 35, sz. 4 (2019. június 1.): 1305–37.

———. „Values Embedded in Legal Artificial Intelligence". SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 2017. március 13. https://papers.ssrn.com/abstract=2932333.

„The Race To Driverless Technology | Leasing Options". Elérés 2020. február 15. https://leasingoptions.co.uk/driverless-cars/index.html.

Vee, Annette. *Coding literacy: how computer programming is changing writing*. Software studies. Cambridge, MA: The MIT Press, 2017.

Wendel, W Bradley. „The Promise and Limitations of Artificial Intelligence in the Practice of Law". *OKLAHOMA LAW REVIEW* 72 (2019): 30.

Winograd, Terry, és Fernando Flores. *Understanding Computers and Cognition: A New Foundation for Design*. 24th printing. Boston: Addison-Wesley, 2008.

Ződi, Zsolt. „Gépek a jogban: Jogelméleti gondolatok a számítógépek jogalkalmazásáról". *JOGELMÉLETI SZEMLE* 2013, sz. 2 (é. n.): 196–212.

———. „Hogyan változtatja meg a jog nyelvezetét a számítógép: A logika és a tekhné a jogban". *GLOSSA IURIDICA JOGI SZAKMAI FOLYÓIRAT* I., sz. 2 (2014): 114–25.

———. *Platformok, robotok és a jog új szabályozási kihívások az információs társadalomban*. Budapest: Gondolat, 2018.