THEMIS

THEMIS

THEM

ELTE
Állam- és Jogtudományi Kar
doktori iskoláinak
elektronikus folyóirata

2023. december

Kiadó: ELTE Állam- és Jogtudományi Kar

Kiadó székhelye: 1053 Budapest, Egyetem tér 1–3.; www.ajk.elte.hu

Felelős kiadó: prof. dr. Sonnevend Pál, az ELTE Állam- és Jogtudományi Kar dékánja

Megjelenik minden évben kétszer.

HU ISSN 2064 0900

DOI 10.55052

Tartalom

Contents

I. RÉSZ
JOGI TANULMÁNYOK

Dr. Badinszky Áron
ELTE ÁJK Polgári Jogi Tanszék
Témavezető: dr. Darázs Lénárd egyetemi tanár
DOI: https://10.55052/themis.2023.2.6

A beavatkozási pontok stratégiája - észrevételek és javaslatok a mesterséges intelligencia szabályozásával kapcsolatban[1]

Meggyőződésem, hogy a XXI. század jogászai számára a legnagyobb próbatételt az úgynevezett „diszruptív technológiák", közöttük különösen a mesterséges intelligencia (MI) és az azon alapuló megoldások megjelenése és széleskörű alkalmazása jelenti.

Az MI térhódítása a negyedik ipari forradalom legfőbb jellemzője. Ezt a napjainkban is elképesztő elánnal zajló revolúciót a számítási kapacitások – korábban technikai okokból kivitelezhetetlen – exponenciális növekedése indukálta, amit gyorsított az ilyen technológiák árának rohamos csökkenése. Ennek következtében nemcsak egyre szélesebb kör számára lettek elérhetőek az MI-t valamilyen formában működésükbe integráló eszközök, alkalmazások, de sokkal komplikáltabb feladatok megoldására is képesekké váltak. A teljesség igénye nélkül ilyen, MI alapú technológia az önvezető autók önvezetésének záloga, a gépi látás, vagy az ún. nagy nyelvi modellek elterjedése, melyek fejlettségére ékes példa az utóbbi időben a köztudatba berobbant ChatGPT. Adattudománnyal és MI-vel foglalkozó szakértői körökben gyakorta úgy fogalmaznak: az MI területén mostanra kaptuk meg a kulcsot azokhoz az ajtókhoz, amelyek létezése ismert volt, azonban a technológiai akadályok miatt nem tudtuk kinyitni őket. Gondoljunk csak bele, hogy korunk emberének „tartozéka", az okostelefon hány olyan funkciót képes ellátni, amihez 20 évvel ezelőtt még egy szobányi különböző eszközre lett volna szükségünk. Az MI alapú technológiák pedig nemcsak az ilyen eszközökben találhatók meg, de olyan területen is elterjedtek, amelyekhez korábban nélkülözhetetlennek véltük az emberi intelligencia jelenlétét: ilyen az orvostudomány, a mezőgazdaság vagy akár maga a jog.

[1] A jelen tanulmány a K-142232 OTKA_22 alapkutatási pályázat keretében született a Kulturális és Innovációs Minisztérium és az NKFIH támogatásával.

Az alábbiakban bemutatom az egyik módját annak, miként volna lehetséges megközelíteni e korszakalkotó technológiát a szabályozás szempontjából. Ennek érdekében az olvasó elé tárom, hogyan lehet praktikus jogászként a mesterséges intelligenciához közelítenünk és mit kell figyelembe vennie a jogalkotásra felkent szerveknek, ha ezt a terület normatív keretek közé kívánják szorítani. Már ha erre egyáltalán van valódi esély…

1. Helyzetkép és distinkciók

Ugyan a mesterséges intelligencia már az 1950-es évek óta ismert fogalom[2], igen sokat kellett várni arra, hogy a számítástechnikai kapacitások elérjék azt a szintet, hogy az addig csak papíron létező megoldások a valóságban is működjenek, majd mindennapi életünk elengedhetetlen részeivé váljanak. Az Intel vállalat társalapítója Gordon E. Moore 1965-ben fogalmazta meg állítását[3], miszerint a mikrochipek teljesítménye körülbelül kétévente a duplájára nő, miközben áruk körülbelül a felére csökken. A Moore-törvényként fennmaradt ökölszabály kiállta az idő próbáját, azonban az utóbbi két évtizedben a fejlődés sebessége sokkal inkább nevezhető exponenciálisnak. Az évezred elején a csúcsteljesítményű processzorok ~10.000.000 tranzisztor segítségével végezték a számításokat, ez 2020-ra mintegy ötvenmilliárdra nőtt[4], de a Moore-törvény időtállóságát bizonyítja, hogy az Apple tavaly mutatta be legújabb, „polgári” használatra szánt processzorát, amelyben 114 milliárd tranzisztor biztosítja az elképesztő teljesítményt.[5]

Fontos már most tisztáznunk, hogy amit a köznyelv mesterséges intelligenciának hív az alapvetően magas szintű matematikai, statisztikai, fizikai módszereken és számítástechnikai megoldásokon alapuló algoritmusok működésének eredménye, amelyhez az „üzemanyagot” a hatalmas mennyiségű adattömegek biztosítják. A mesterséges intelligencia tehát voltaképp nem egy „megfogható” entitás, hanem – jellemzően[6] – ember által létrehozott, az emberi gondolkodás egyes részterületeit, mint például az érzékelést, kategorizálást, mintázatok felismerését szimuláló algoritmusok

[2] Moor 2006. 87.
[3] Tardi 2022
[4] Roser és Ritchie, 2022
[5] Apple Inc. 2022
[6] Az ún. generatív MI technológia megjelenésével – pl.: ChatGPT, Dall-E, Google Bard – a társadalom számára is világossá vált, hogy léteznek olyan algoritmusok is, amelyek működésük során új algoritmusokat képesek létrehozni egy-egy probléma vagy feladat megoldására.

kimenete. Az output lehet egy döntés, statisztikai valószínűség, vagy bármilyen „emberi” interakció a külvilággal, például „hallás”, „látás”, illetve szövegalkotás.

Az MI öncélú létrehozásának nincs sok értelme. Pusztán azért, hogy az emberi gondolkodás egy szeletét[7] szimuláljuk nem éri meg kutatásba, fejlesztésbe fektetni. Természetesen a tudománynak szüksége van ezekre az eredményekre is, de magától értetődő, hogy az MI-t elsősorban azért alkottuk meg, hogy az élet különböző területén használjuk az arra épülő alkalmazásokat. Egy olyan MI algoritmus elkészítése, ami hatalmas rendelkezésre álló mintákat elemezve trendeket tud megfigyelni és ezek alapján döntési alternatívákat javasol, megfelelő kontextus nélkül nem sok gyakorlati – értsd.: monetizálható, piaci – haszonnal rendelkezik. De „engedjünk be” egy ilyen alkalmazást a tőzsdére és addig soha nem látott eredményeket érünk el. Az MIT Techology Review-ban megjelent tanulmány szerint a 2000-es évi csúcsidőszakban a Goldman Sachs New York-i központja 600 részvénykereskedőt alkalmazott. Napjainkra mindössze két kereskedő maradt 200 számítástechnikussal, a háttérben automatizált kereskedő-programok vették át a munka nagy részét”.[8]

Egyáltalán nem mindegy azonban, hogy ezek a nagy hatású alkalmazások milyen jogi keretek között működhetnek, hiszen számos olyan életviszonyunkat befolyásolhatják, ahol a legkevésbé sem szeretnénk, hogy működésük valamilyen okból hibás vagy akár kártékony legyen.

Jogászként közelítve a kérdéshez azt látjuk, hogy az MI szabályozás szűk értelmezési tartományát voltaképp mind az algoritmusok, mind az ezek kondicionálásához használt és működésükhöz nélkülözhetetlen adatok képezhetik, de távolabbról nézve ide tartozik az MI-t alkalmazó területek, ágazatok regulációja is. Mielőtt azonban elmélyülnénk a témában, elengedhetetlen éles és világos határt húznunk az MI *alkalmazási formáinak, felhasználási módjainak* szabályozása, és e tanulmány fókuszának*, magának a mesterséges intelligenciának a szabályozása* között. Amikor az utóbbiról, tehát az MI szabályozásáról beszélünk, akkor úgy vélem, magára az algoritmusok megtervezésének, létrehozásának konkrét folyamatára, illetve a kész algoritmus működésére és annak felügyeletére vonatkozó keretrendszer

[7] Jelenleg a „szűk” MI korát éljük, amikor az algoritmusok még „csak” az emberi intelligencia egy-egy részterületét képesek reprodukálni. A fejlődés azonban hamarosan – egyesek szerint már a következő években, az óvatosabb szakértők szerint csak a 2030-as években – elhozhatja az általános mesterséges intelligenciát, ami már képes lesz a teljes humán intellektust reprodukálni.

[8] Brynes 2017.

kialakítására kell gondolnunk a tárgyi hatály kijelölésekor. Ebbe a körbe tartozhatnak azok a normák, amelyek meghatározzák az MI technológiai kritériumait, tehát definiálják, mit nevezhetünk egyáltalán MI-nek, és lefektetik az MI biztonságos és megbízható működését szavatoló szakmai és adminisztratív követelményeket. Ahogyan az alábbiakban is látni fogjuk, ezek a keretek egyaránt vonatkozhatnak az algoritmusokra és az MI alapjául szolgáló adatkészletekre is. A tanulmány során a regulációnak erre az irányára „*sui generis MI szabályozásaként*" fogok hivatkozni.

A mesterséges intelligencia felhasználási módjainak szabályozását éppen ezért más szemléletmóddal kell megközelítenünk. Ebben az esetben abból volna kívánatos kiindulnunk, hogy az adott területen, akár szoftveres alkalmazásban használt, akár eszközbe – pl.: drónba – épített, kvázi „manifesztálódott" MI-nek már eleve immanens részét képezik azok a követelmények, amelyekre a fentiekben már utaltunk. Kedvező esetben ezeknek való megfelelés hiányában nélkül nem kerülhetett volna piacra, „nyilvános" felhasználásra az adott algoritmus. Ilyenkor tehát véleményem szerint a szabályozásnak már nem kell részletesen kiterjednie a technikai, szakmai és etikai alapokra, hanem elég az adott felhasználási módra – akár ágazat-, akár termékspecifikusan – jellemző körülményeket átfognia. Az is előfordulhat, hogy azokat az életviszonyokat, amelyek között az MI alapú eszközt, szoftvert alkalmazzák, már kimerítően és részletesen szabályozták és körülvették a jogi garanciákkal, így mindössze arra van szükség, hogy néhány MI specifikus értelmező vagy kiegészítő normát illesszenek a rezsimbe. Például az önvezető autók esetében úgy vélem, igen kikristályosodott szabályok vonatkoznak a veszélyesüzemi felelősségre, a közúti közlekedésre vagy a személyszállításra, azonban eddig még semmi nem tette indokolttá annak a kérdésnek a rendezését, milyen etikai elvek alapján döntsön egy magát irányító autó, ha életek – praktikusan a sofőr és mások – között kell választania. Ennek megfelelően ezeket a szabályrendszereket hozzá kell igazítani az MI megjelenésével felmerülő sajátosságokhoz, és finomhangolásra van szükség, hogy ismét hézagmentesen zárjon a jogrendszer. Ilyen, az MI által jelentősen érintett rezsimek a teljesség igénye nélkül a versenyjogi, a fogyasztóvédelmi szabályozások, az adatvédelem – különösen a személyes adatok, az egészségügyi adatok védelmének – joga, de a jogrendszerek gerincét képező polgári jog és büntetőjog normarendszere és az alapjogok köre sem marad érintetlenül.

Világosan kell látnunk azt is, hogy számtalan olyan jelenség van, amelynek a létezését az MI tette lehetővé, ezekre kiemelt figyelmet kell fordítani, és

viszonylag gyors megoldásokat kell(ene) adni a jogrendszereknek, mivel működésük rengeteg kockázattal jár. Korábban például sokan nem gondolták volna, hogy egy nap olyan, tökéletesen személyre szabott hirdetésekkel fog minket „bombázni" az internet, amelyeket MI alapú algoritmus állít össze a böngészési tevékenységünk elemzése és az okostelefonon használt alkalmazásokon keresztül gyűjtött adatok segítségével. Az említett adatgyűjtési tevékenység sokszor kiegészül valamilyen, szintén az MI által lehetővé tett „lehallgatással", az ún. természetes nyelvfeldolgozás[9] módszereivel. Ennek ellenére az MI-vel támogatott, *hypertargetingnek* nevezett marketingtechnológia napjaink egyik legelterjedtebb és leghatékonyabb – bár fogyasztóvédelmi szempontból olykor aggályos módszerei ellenére is egyelőre meglehetősen alulszabályozott – értékesítési módszere. Hiszen használatával mindenkinek azt a terméket tudják reklámozni, amit a legnagyobb valószínűséggel meg is vásárol. Az EU Parlamentje két éve döntött az akkor még formálódó *Digital Services Act* (DSA) rendelet olyan kiegészítéséről, ami legalább egyes veszélyeztetett csoportok – gyermekek – tekintetében megtiltja ezeknek a technológiáknak az alkalmazását.[10] Véleményem szerint tehát a MI *sui generis* szabályozásáról folytatott diskurzus mellett elmaradhatatlan a *MI-n alapuló technológiák felhasználására vonatkozó* intézkedésrendszerek kialakításáról is beszélni, azonban a továbbiakban figyelmünket fordítsuk az előbbire.

2. A sui generis MI szabályozás tárgya

Annak érdekében, hogy egyes életviszonyokra hatékonyan működő jogi keretrendszert tudjunk kialakítani, elmaradhatatlan a szabályozás tárgyának pontos definiálása, körülhatárolása, hiszen a jogalkotó ez alapján tudja orientálni tevékenységét. A tárgyi hatály kijelölésekor a mesterséges intelligencia mint önálló entitás nem értelmezhető, mivel egyelőre a tudományos világban is megoszlanak a vélemények arról, mit tekinthetünk egyáltalán MI-nek. Úgy gondolom, ha ki is alakulna egy általánosan elfogadott MI definíció, akkor is nehéz dolga volna a szabályalkotónak, hiszen az MI nem

[9] A természetes nyelvfeldolgozó algoritmusok képesek az emberi beszéd felismerésére, és annak a számítógépek számára érthető formátummá alakítására. legelterjedtebb és leginkább közismert megjelenési formái az okoseszközeinkben található személyi asszisztensek, mint a Siri vagy az Alexa.

[10] Goujard 2021.

egy konkrét, állandó, világosan körülhatárolható létező, hanem jelenségek, eszközök, technológiai megoldások folyamatosan bővülő halmaza.

Kétségtelen, hogy az említett, MI mögött álló, annak létezését lehetővé tévő komplex informatikai és matematikai megoldások rendszere sokkal inkább megfogható, definiálható elemekből áll, ennek folytán pedig nagyságrendekkel alkalmasabb arra, hogy azokra nézve normákat alkossunk. Észszerű tehát, ha az MI szabályozás célkeresztjében nem az emberi absztrakciós képességekkel is alig megragadható, sokszor még fizikai megjelenéssel sem rendelkező létező MI, hanem azok a „építőkockák" állnak, amelyekből az MI-ként ismert jelenség felépül. Az Európai Unió Bizottsága kristálytisztán határozza meg ezeket az elemeket: *„bármely lehetséges jövőbeli politikai kezdeményezésről szóló megbeszélések céljából fontosnak tűnik tisztázni, hogy a mesterséges intelligencia fő alkotóelemei az **„adatok"**,* valamint az ***„algoritmusok"**.*[11]

Amennyiben ezeket az alkotóelemeket közelebbről vizsgáljuk, észrevehetjük, hogy rendelkeznek olyan paraméterekkel – pl.: technikai kialakítás részletei –, és olyan attribútumokkal, tulajdonságokkal – pl.: a működés átláthatósága, nyomon követhetősége, megmagyarázhatósága –, amelyek szabályok alkotásával befolyásolhatók, illetve amelyekre nézve lehetséges bizonyos követelményeket megfogalmazni. E követelmények megjelenhetnek jogszabályokban, etikai normákban, vagy bármilyen más, egy adott szférában kötelező erővel bíró regulációban. Feladatunk tehát ezen alkotóelemeket feltérképezni, s olyan „beavatkozási pontokat" meghatározni, ahol eredményesen lehet alakítani az MI körébe tartozó technológiák létrehozását és működését. A kutatásaim során arra a megállapításra jutottam, hogy e beavatkozási pontok közül a legfontosabbak az alábbikban tárgyalandók.

3. Adatok

Az adatot szokás a XXI. század kőolajának is nevezni, hiszen amióta rendelkezésre állnak azok a technológiák, amelyek segítségével nagy mennyiségben lehet „nyers" adatokat kinyerni, tárolni, rendszerezni és elemezi, ez az erőforrás a gazdaság és a mindennapi élet számos egyéb területén hozott előrelépést. A már említett tökéletesen személyre szabott

[11] Fehér könyv a mesterséges intelligenciáról: a kiválóság és a bizalom európai megközelítése. COM(2020) 65 final (2020)

hirdetéseken alapuló marketingtől kezdve az adatalapú kormányzati és üzleti stratégiai döntések meghozásán át egészen a Facebook hírfolyamunk összeállításáig ma már szinte nincs olyan terület, mely működésének zálogát ne a hatalmas mennyiségű strukturált adat jelentené. A digitalizáció az elmúlt néhány év legfontosabb fejlesztési iránya volt az üzleti világban, a szolgáltatásoktól kezdve a gyártásig mindent automatizálni, „okosítani" szerettek volna a vállalkozások, s e szemlélet a közigazgatásba is egyre inkább begyűrűzött. Az adat az, ami a folyamatokat „okossá" teszi, minél több adattal rendelkezik egy szervezet, annál okosabb.[12]

Az MI létezése sem képzelhető el adat, mégpedig megfelelő mennyiségű és minőségű adat nélkül. Az MI-t létrehozó algoritmusok működése, pontossága, megbízhatósága alapvetően azoktól az adatoktól függ, amelyeket a tanításukhoz és tesztelésükhöz felhasználnak. Minél nagyobb és jobb minőségű az adatállomány, annál jobban „tud az MI tanulni", azaz felismerni az adatokban rejlő finom összefüggéseket, ember számára rejtett mintákat is. Fontos leszögeznem, hogy az „éles" működés során minden MI rendszer annyira biztonságos, pontos és megbízható, amennyire a tanításához és teszteléséhez felhasznált adatok azok. *„A gépi tanulás során a számítógépek olyan előrejelző modelleket futtatnak, amelyek már létező adatokból tanulva előre jelzik a jövőbeni viselkedéseket, kimeneteket és trendeket. Bármely gépi tanuló algoritmus és az általa hozott kisebb döntések attól az adatkészlettől függenek, amin betanították. Minél több adathoz hozzáfér, annál jobb előrejelző modellt tud alkotni*".[13] Amennyiben tehát a tanítási adatkészletben hiba vagy torzítás van, az kedvezőtlen irányba befolyásolhatja az algoritmus működését. Pontos és körültekintően összeállított adatok hiányában téves, részrehajló eredmények születhetnek, ez pedig végső soron kárt, jogsérelmet vagy egyéb negatív következményeket okozhat. Ennek megfelelően az MI szabályozásának első lépését a bemeneti adatokra vonatkozó következetes, átlátható és minden érintett szereplő számára teljesíthető rezsim kialakítása jelenti. Jelen fejezet tehát az MI *„előállításához" szükséges* adatokra fókuszál

Mielőtt a konkrét szabályozási lehetőségek bemutatásába kezdek, szeretném, hogy az olvasó is világosan lássa: szinte lehetetlen *minden* MI-vel kapcsolatba hozható adatkészletre nézve kötelező szabályokat alkotni, és véleményem szerint nem is indokolt. Nem feltétlenül szükséges jogi vagy egyéb kötelező erővel bíró normákat alkotni például az olyan algoritmusok

[12] Global Industry Analysts Inc. 2021.
[13] Buiten 2019. 51.

alapjául szolgáló adatkészletekkel kapcsolatban, amelyek kisebb jelentőségű ipari folyamatot tesznek hatékonyabbá. Például olyan beépített MI-t alkalmazó eszközt illetően, ami autók gyártása során hibákat keres a fényezésben, nem elengedhetetlen, hogy az algoritmus tanításához használt adatkészlet összeállítása során – ami pl.: rengeteg képet tartalmaz hibás festésekről – szigorú, jogi erővel rendelkező szabályokat kövessenek, és lássuk be nem is életszerű. Ha azonban a példában szereplő algoritmus nem festési, hanem hegesztési hibákat keres a karosszérián, mindjárt más a helyzet, hiszen hibás működése esetén akár emberéletekben eshetne kár. Hasonló a helyzet az olyan MI rendszernek a működése során, amely demográfiai és kriminálstatisztikai adatokat elemezve alkot predikciót egy bűnelkövető visszaesési esélyeiről[14], vagy azoknak az MI alapú rendszereknek az esetében, amelyek emberi felügyelet nélküli hivatali ügyintézést, netán HR feladatokat végeznek.[15] Nem mehetünk el említés nélkül az egészségügyben egyre szélesebb körben alkalmazott MI mellett sem, ahol az adatkészletek minőségén emberéletek múlhatnak.

Véleményem szerint napjaink egyik legfontosabb feladata, hogy azonosítsuk azokat az ágazatokat, felhasználási környezeteket, ahol az MI alkalmazása emberek életét számottevő mértékben befolyásolhatja. Különösen körültekintően kell kezelnünk azokat a helyzeteket, amikor az MI „saját” – tehát emberi kontroll nélküli – döntéseket hozhat, sorsokat formáló „véleményt” alkothat.

Szintén fontos feltérképezni azokat az eseteket, ahol az MI az emberi döntést befolyásoló tényező lehet, például egy onkológiai diagnosztikai alkalmazás részeként. Első lépésként ebben a két körben is azokra az adatkészletekre kell koncentrálni, amelynek elemei személyes adatok, illetve különösen ezek speciális kategóriái, pl.: bűnügyi- vagy egészségügyi adatok vagy egyéb érzékeny információk, pl.: pénzügyi információk. Ilyenkor ugyanis a rendszer elégtelen, rosszul összeállított, esetleg hiányos adatkészletekből fakadó hibás működése szinte kivétel nélkül alapjogsérelmet eredményez vagy jelentős kárt okoz. Ismét hangsúlyozom: olyan garanciális szabályok, minőségi követelmények megalkotása szükséges, amelyek biztosítják, hogy az algoritmusok működése során torzítás- és részrehajlásmentes eredmények szülessenek.

[14] Fry 2020. 75.
[15] C. Lennox 2021. 65.

Úgy vélem az MI létrehozásához szükséges adatkészletekkel[16] összefüggő szabályozásnak két lehetséges fókuszpontja van. Az egyik az algoritmusok tanításához és teszteléséhez szükséges adatbázisok összeállítására, kialakítására, illetve ezen adatkészletek „minőségi követelményeire" fókuszál, ugyanis a megbízható MI alapja az elfogulatlan, torzításmentes és kiegyensúlyozott input-adat. A másik a kiberbiztonság, aminek középpontjában az említett adatkészletek védelme, minőségi állapotuk megőrzése áll.

3.1. Adatminőségi követelmények

Nem szabad elfelejtenünk, hogy az algoritmusok „hozott anyagból" dolgoznak, s azon adatkészlet alapján kondicionálják működésüket, amit a fejlesztőik „eléjük tesznek". Nem kérdeznek vissza, hogy biztosan jók-e az adatok, hanem fenntartások és kritika nélkül teszik, amire tervezték azokat. Ennélfogva, ha egy képzeletbeli hitelminősítő algoritmus olyan adatbázisra épül, amelyben a fizetésképtelen adósok 75%-a a vidéki lakossághoz tartozik, akkor is kisebb eséllyel fog hitelt adni egy nyíregyházi lakosnak, ha helyzete egyébként minden más bírálati szempont alapján megegyezik a fővárosi igénylőkével.

> *„Ha a kulcsfontosságú adatokat szándékosan vagy véletlenül nem biztosítják az algoritmus számára, annak teljesítménye nagyon gyenge lehet. Nem igazolható az a gyakran emlegetett feltevés sem, hogy ha elegendő adatot gyűjtünk össze, az algoritmusok nem lesznek elfogultak. Az algoritmusok torzítása többféleképpen is bekövetkezhet. Először is, előfordulhat, hogy az általunk gyűjtött adatok részrehajló mintavételezésből származnak, ezért maga az adatminta torz. Másodszor, elfogultság merülhet fel, mert az összegyűjtött adatok a meglévő társadalmi egyenlőtlenségeket tükrözik. Amilyen mértékben megtalálhatók a társadalomban a kirekesztés vagy a diszkrimináció nyomai, az adatok is ezt fogják tükrözni. Például a faji csoportok letartóztatási arányában mutatkozó különbségeket reprodukálhatja a bűnismétlés kockázatát kiszámító*

[16] Adatkészletek alatt értem azokat az adatsorokat vagy nagyméretű adatbázisokat, amelyek elemei már tisztításon, rendszerezésen átesve, kvázi „konyhakész" állapotban vannak a további műveletekhez, pl.: algoritmusok kondicionálásához.

algoritmus. […] Röviden, a gépi tanulás megerősítheti a diszkrimináció meglévő mintáit."[17]

A PredPol esete[18]

Az előbbiekből is kitűnik, hogy a torz, nem megfelelően összeállított adatkészletek veszélye, hogy korrekció hiányában képesek konzerválni bizonyos szociális problémákat, negatív társadalmi szituációkat, és elvágni a változás lehetőségétől a hátrányos helyzetű csoportokat. Az USA számos államában használják a PredPol – Predictive Policing – nevű, MI alapú alkalmazást, amelynek segítségével a bűnüldöző szervek optimalizálhatják kapacitásaikat a bűnmegelőzés érdekében. A rendszer működésének lényege, hogy a korábbi évek bűnözési adatai alapján előre jelzi azokat a gócpontokat a településen belül, ahol adott napon nagyobb esélye van egy adott bűncselekmény elkövetésének. Így a rendőrség minden nap optimalizálni tudja járőrtevékenységét, méghozzá nem kis sikerrel. A „klasszikus" módszer hátulütője, hogy a véletlenszerűen szétosztott járőrök átlagosan nyolc évente kerülnek egy betörés 100 méteres körzetébe. Ezzel szemben a PredPol lehetővé tette, hogy a rendőröket célzottan az algoritmus által kijelölt bűnözési gócpontokra küldjék, bízva abban, hogy puszta jelenlétük elrettenti a bűnelkövetőket a cselekvéstől, illetve, ha mégis történik valami, gyorsabban tudnak arra reagálni arra. Az eredmény drámai volt. Az algoritmus bevezetése után néhány hónappal több Los Angeles-i körzetben számottevően visszaesett a bűnözés, volt olyan körzet, ahol 13%-kal, de a kaliforniai Alhambra városában pl.: 32%-os csökkenést mértek a betörések számában. Ennek a megközelítésnek azonban van egy igazán komoly hátulütője. Azzal, hogy a rendőröket a magas kockázatú helyekre küldte, az algoritmus nem kívánt öngerjesztő folyamatot is indukált. Ha ugyanis a magasabb bűnözési mutatókkal rendelkező, szegényebb környékekre több egységet irányít, azok több bűncselekményt észlelnek, ezeknek az adatai pedig bekerülnek az algoritmus alapjául szolgáló adatbázisba. Így, az adatkészlet torzulása következtében jövőben még inkább kockázatosnak fogja tartani az adott környéket, s még több rendőri egységet vezényel oda, s így tovább. A helyi lakosságot egy idő után kifejezetten zavarni fogja a megnövekedett rendőri jelenlét, sőt nem kizárt, hogy elnyomásként, diszkriminációként tekint arra.[19]

[17] Buiten 2019. 52.

[18] Fry 2020. 168-174.

[19] Hasonló eset történt New Yersey államban is. Az esetről kiváló és részletes tanulmány számol be: Sankin és mtsai.: *„Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them – The Markup* (2021)".

Úgy gondolom, a bemutatott esetből az olvasó számára is egyértelművé vált: ha a PredPol működése során a fejlesztők figyelmet szenteltek volna a rendszer működéséhez szükséges adatkészletben – azaz a bűnügyi nyilvántartásban – már az első pillanattól rejlő, illetve később kialakuló anomáliák kezelésére, elkerülhető lett volna a cég körüli botrány kirobbanása.

Természetesen, ha egy MI rendszer fejlesztője nem rendelkezik ezzel az alapos attitűddel, a jogalkotó kezében széles eszköztár áll rendelkezésre ahhoz, hogy legalább minimális szinten beavatkozzon a helyzetbe. Véleményem szerint olyan szabályokat volna érdemes kidolgozni, amelyek minőségi követelményeket határoznak meg az adatkészletekre nézve. Különböző szigorúságú standardokat, kritériumoknak kell előírni annak fényében, hogy az adatokra épülő algoritmust milyen területen használják fel, és az általa létrehozott eredményeknek milyen lehet a hatása, kockázata. Világos, hogy ha az MI alkalmazás emberek életét is befolyásolhatja, a legmagasabb szintű garanciákat kell meghatározni annak adatkészleteivel szemben, és figyelemmel kell lenni arra, hogy az adatok az alkalmazási terület jellegzetességei, sajátosságai szerint legyenek összeállítva.

Az adatkészletek minőségét és állapotát nemcsak a tervezéskor kell vizsgálni, hanem a MI teljes életciklusa során monitorozni kell azokat, és szükség esetén az eredeti célnak megfelelően korrigálni is szükséges tartalmukat. Ha az algoritmus dinamikus adattömegből dolgozik, tehát működése során változik vagy bővül az eredeti adatok köre, ki kell dolgozni azokat az eljárásokat, amelyekkel az új adatokat a régiek torzítása nélkül beépítik a rendszerbe, hogy az algoritmus továbbra is megbízható eredményeket adhasson.

Külön említést érdemelnek azok az esetek, amikor egy-egy MI alkalmazás outputja később a rendszer inputjává válik, ahogyan ez a PredPol esetében is történt, hiszen ilyenkor még gyorsabban alakulhat ki „torzítási spirál". Az adatminőségre vonatkozó kritériumokat jelenleg jobbára csak íratlan ágazati jó gyakorlatok rögzítik, de egyre több kellemetlen[20] eset támasztja alá, hogy milyen messze vezető következményei lehetnek a hanyagul, felelőtlenül vagy nem kellő körültekintéssel összeállított MI kondicionáló adatkészleteknek.

[20] A tanulmányban említett példákon túl érdemes tanulmányozni a holland kormány lemondásához vezető botrány esetét, amit egy szociális juttatásokkal kapcsolatos csalást felderítő, rosszul kialakított kormányzati MI rendszer hibás, ezért diszkriminatív működése generált. Lásd pl.: Henley: *Dutch Government Resigns over Child Benefits Scandal*. (2021); Heikkilä: *Dutch Scandal Serves as a Warning for Europe over Risks of Using Algorithms* (2022)

3.2. Kiberbiztonság

A számítógépek megjelenésével és különösen az internet elterjedését követően egyre nagyobb számban voltak jelen a digitális világban azok, akik nem feltétlenül jó szándékkal, tiszta motivációval rendelkeztek. Amikor nyilvánvaló vált, hogy a Föld különböző pontjain elhelyezett adattároló szerverek a világhálón keresztül bárhonnan hozzáférhetőek, valóságos aranybánya tárult a kiberbűnözők elé. Az illegális üzletágban banki és személyes adatok, titkos kormányzati nyilvántartások és minden információ, amit hálózatra kötött számítógépen tároltak kiszolgáltatottá vált a „szakemberek" számára. Lelki szemeink előtt ilyenkor megjelenik egy kép a családi ház pincéjében, házikabátban gubbasztó hackerről, ám természetesen nemcsak magánszemélyek, hanem egész vállalkozások, bűnszervezetek és államok is jelentős energiát öltek és a mai napig ölnek bele abba, hogy adatokat szerezzenek másokról. Természetesen nem kellett sok idő ahhoz, hogy meginduljon a támadások elleni védelmi szolgáltatásokat kínáló ágazat működése, és ma már éppen az ide tartozó cégek a legnagyobb munkaadói az úgynevezett etikus hackereknek. Amennyiben a kiberbiztonságra úgy tekintünk, hogy az az MI betanításához, működéséhez felhasznált adatok védelmének letéteményese, rögtön egyértelművé válik, hogy a jól működő MI szabályozási rezsim nem képzelhető el átfogó kiberbiztonsági szabályok nélkül. Példaként tekintsük meg az Európai Bizottság álláspontját:

> *„Fontos, hogy a szabályozás megfelelő keretet biztosítson az MI-vezérelt innováció és az MI-megoldások elterjesztése számára, kezelve ugyanakkor a technológia használata, vagy az azzal való kapcsolatba kerülés által eredményezett lehetséges kockázatokat, a kiberbiztonsági aggályokat is beleértve. Ez azt jelenti, hogy gondoskodni kell a „kiberbiztonságról", a visszaélések megakadályozásának értelmében (például az MI algoritmusok meghackelése vagy az MI algoritmus által feldolgozott adatok manipulációja), valamint az olyan mechanizmusok beépítéséről, amelyek biztosítják a fogyasztók biztonságát és a hatékony jogorvoslatot az áldozatok részére kár esetén, továbbá megkönnyítik a vizsgálatokat az MI rendszer sérülése esetén."*[21]

[21] A mesterséges intelligenciáról szóló összehangolt terv. COM(2018) 795 final ANNEX (2018), 19.

Ahogy a Bizottság is utalt rá, a lehetséges MI adatreguláció másik területe maguknak a „kész” adatbázisoknak a védelme, ezzel pedig a bennük szereplő elemi adatok integritását szavatolása, így meghatározva mind a külső biztonságra vonatkozó szabályokat, mind a belső stabilitásra, robosztusságra vonatkozó kritériumokat. Külső biztonság alatt azokat a szabványokat, szakmai standardokat, eljárásokat értem, amelyek célja az adatok védelme a kívülről jövő befolyásolási kísérletekkel – pl.: hackertámadásokkal – szemben, és szavatolják azt, hogy a rendszerek alapja ne legyen kiszolgáltatva a rosszindulatú személyeknek.

Nem szabad elfelejtenünk, hogy már egy kis módosítás is hatalmas veszélyeket hordoz magában, gondoljunk csak arra, ha a bűnügyi nyilvántartásba olyan személyről készült fényképet csempésznek, aki valójában nem tett semmi jogelleneset. Azok az MI arcfelismerő algoritmusok, amelyek pl. a reptereken a tömegek ellenőrzésére, szűrésére szolgálnak, a felismert arcokat a bűnügyi nyilvántartással vetik össze. Ha olyan input adatokat kapnak, mely szerint az érintett személy bűnelkövető, szinte azonnal jelezni fogják a hatóságoknak. Ennek az esetnek természetesen a fordítottja sem kevésbé veszélyes, hiszen gondoljunk csak bele abba, mi történhet, ha egy körözött bűnöző arcképét illetéktelenek törlik a bűnüldöző arcfelismerő rendszerek adatbázisából. Szintén veszélyes, ha az algoritmusok tanítására használt adatbázist módosítják kívülről, hiszen így eleve rossz, valótlan, torzított adatokkal kondicionálják az algoritmust, ami ezáltal később az „éles” alkalmazás során hibás eredményeket produkál annak ellenére, hogy működési mechanizmusa kifogástalan. Úgy vélem, ennek érdekében olyan szabályokat kell alkotni, amelyek technikai, elsősorban informatikai és kiberbiztonsági követelményeket állítanak azokkal az adatbázisokkal szemben, amelyeket MI tanításához, teszteléséhez és működtetéséhez használnak.

Nem feledkezhetünk meg a belső biztonságot szolgáló keretek létrehozásáról sem. A belső biztonságot úgy tudjuk megteremteni, ha gondoskodunk róla, hogy az adatbázisok létrehozói értsenek ahhoz, amit csinálnak. Messze vezető kérdés volna e sorok között tárgyalni, de bizonyos MI alkalmazások hibás működése lehet olyan kockázatos az emberekre nézve, és képes olyan károkat okozni, hogy nem engedhetjük meg, hogy kizárólag bootcampekben végzett programozók vagy hobbi-adatelemzők vegyenek részt a felhasznált adatkészletek kialakításában. Természetesen magánvéleményemet osztom meg az olvasóval, de meggyőződésem, hogy pl.: banki szoftverek, bűnüldöző algoritmusok vagy egészségügyi felhasználásra szánt MI rendszerek

tervezésekor legalább soft-law formában szükséges volna a munkában résztvevők minimális szakmai, képzettségi szintjének előírása. E szabályok – amelyekkel kapcsolatban kívánatos volna, hogy nemzetközileg elfogadott szabványok formájában épüljenek be az „MI-univerzumba" – célja olyan stabil környezet megteremtése, amelyben biztosak lehetünk abban, hogy az algoritmusok működése nem az alapok rogyadozása miatt omlik össze.

4. Algoritmusok

A mesterséges intelligenciával foglalkozó irodalomban kiemelt szerepet foglal el annak az 1997-es sakkjátszmának a visszatérő felidézése, amelyben az IBM Deep Blue számítógépe MI alapú „sakktudásával" felülkerekedett az addig veretlen sakknagymesteren, Garry Kasparovon. Az eseményt a mai napig olyan mérföldkőnek tekintik, ami jelzi: az ember által alkotott gép igenis képes emberként gondolkodni. Sajnálatos módon a számítógép, ami a szimbolikus jelentőségű diadalt[22] aratta, semmit sem érzett tette jelentőségéből. A helyzet az, hogy nem is tudott volna érezni, hiszen nem volt más, mint a mérnökök által precízen megtervezett algoritmusok összehangolt működése, amit egy célra terveztek: kiszámolni és megtalálni a sakktáblán végrehajtható valamennyi összes lépés közül azt, amire ellenfele már nem tud reagálni. A sakknagymester 20 évvel később így emlékezett vissza az eseményre, ami sok szempontból az algoritmusok korának hajnalát jelentette: *„Annyira lenyűgözött a Deep Blue játéka, annyira azzal voltam elfoglalva, vajon mire lehet még képes, hogy észre sem vettem: problémáim nagy része saját csapnivaló játékomból ered, nem pedig a gép zseniális lépéseiből."*[23]

4.1. Tipológia

A Merriam-Webster[24] szótár szerint: *„az algoritmus (fn.) egymást követő lépésekből álló folyamat valamely probléma megoldására, vagy eredmény elérésére*" Ezt jól egészíti ki a Cambridge Dictionary[25] definíciója: *„[az algoritmus] leginkább számítógépek számára adott matematikai utasítások és*

[22] Fry 2020. 17.
[23] Kasparov 2017. 191.
[24] https://www.merriam-webster.com/dictionary/algorithm#other-words (2023.02.25.)
[25] https://dictionary.cambridge.org/dictionary/english/algorithm (2023.02.25.)

szabályok készlete, ami segít kiszámolni egy adott kérdésre a választ." Az algoritmusok, mint konkrét probléma megoldására tervezett „eszközök" nem újkeletű dolgok, azonban a technológia fejlődése grandiózus új horizontokat nyitott felhasználásuk előtt. „*Matematikai műveletek sorozatát vesszük – akár az aritmetika, az algebra, az analízis, a logika vagy a valószínűségszámítás területéről – és alakítjuk őket számítógépes programmá. A való világból vett adatokat táplálunk be, és bonyolult számításokra ösztönözzük a cél elérése érdekében. Ezáltal lesz a számítástudomány valódi tudománnyá, és gépeink általuk valósítják meg a modern kor csodálatos vívmányait*."[26] Ahogyan az előbbi idézetet jegyző matematikus, Hanna Fry is rávilágít: megszámlálhatatlanul sok algoritmus létezik, mindegyik más és más feladatra optimalizálható, különböző sajátosságokkal, előnyökkel és hátrányokkal rendelkeznek[27]. Annak érdekében, hogy az MI szabályozásával kapcsolatban beavatkozási pontokat találhassunk, azaz az algoritmusok olyan attribútumait, amelyek befolyásolhatók normatív eszközökkel, alapszinten ismernünk kell az MI-hez használt algoritmusok legjellemzőbb csoportjait.[28]

Klasszifikációs algoritmusok

Ezek az algoritmusok képesek valamilyen jellemzők alapján kategóriákba sorolni adatokat. Ilyen megoldások segítségével működik a *hypertargeting*, de a kézírás szöveggé alakításában vagy a telefonunkon található képek felismerésében és címkézésében is tetten érhetjük azokat, nem beszélve a hangfelismerő, „beszédértő" szoftverekről, amiket széles körben alkalmaznak pl.: autókban vagy digitális asszisztenseknél.

Sorba rendező algoritmusok

A prioritási sorrendet felállító algoritmusok rengeteg lehetőséget rendeznek legoptimálisabb formába az adott igényhez. A Google találatokat, vagy a Netflix javaslatokat is ilyen algoritmusok priorizálják nekünk, de a Waze vagy más útvonaltervezők is ezek segítségével javasolják a legrövidebb utat, matematikai módszerekkel kiszámolva és rendezve az összes lehetséges eljutási mód közül a legoptimálisabbakat.

[26] Fry 2020. 20.
[27] Uo.
[28] Uo. 20-23.

Asszociációs algoritmusok

E kategóriába tartozó algoritmusok olyan területeken is segítenek megtalálni az összefüggést adatok között, amelyek sokszor az emberi intelligenciának sem egyszerű. Például a webáruházak személyre szabott ajánlásaikat általában úgy készítik, hogy kosarunk tartalmát elemezve más, hasonló kosárral rendelkező vevőkhöz kapcsolnak, majd később azokat a termékeket ajánlgatják nekünk pl.: hírlevelek formájában, amit a velünk kapcsolatba hozott személyek is megvásároltak. Az asszociációs algoritmusok felelősek társkereső alkalmazások felhasználóinak összekötéséért is a közös érdeklődési kör vagy egyéb, a szoftverbe integrált kapcsolatrendező elvek szerint.

Szűrő algoritmusok

Aki már viselt aktív zajcsökkentésre képes fejhallgatót, tudja, milyen nagyszerű érzés a külvilág zörejeinek beszűrődése nélkül élvezni kedvenc zenénket a délutáni csúcsban, tömött metrón. Ezt azok az algoritmusok teszik lehetővé, amelyek elkülönítik a lényeges jelet a zajtól. Az okosasszisztensek, mint Siri vagy az Amazon Alexa szintén ilyen megoldást használ, hogy felismerje tulajdonosa hangját egy társaság zajában. Természetesen nemcsak fizikai zajokat lehet szűrni, hanem többek között a Facebook és Twitter hírfolyamunkat is ilyen algoritmusok szondázzák és távolítják el belőlük azokat az információkat, amiket szokásaink alapján érdektelennek vélnek.

A felsorolt megoldások szinte egyszer sem önmagukban álló, „típusalgoritmusként” alkotják a MI-t, azt számos különböző funkcióval rendelkező algoritmus összehangoltan, egymást kiegészítve hozza létre. E komplex működés eredményeképp tudjuk ujjlenyomatunkkal vagy arcunkkal feloldani telefonunk kijelzőjét. Fontos különbséget kell tennünk a tervezők által megszabott logikai-számítási műveletek végrehajtására tervezett „klasszikus” és a gépi tanuló algoritmusok között is. Előbbiek emberek alkotta utasítások sorozatát követve jutnak el a kívánt eredményig, míg utóbbiak az élő organizmusok tanulási sémáit követik.

> *„Ha analógiát keresünk, gondolhatunk arra, ahogy a kutyánkat tanítjuk pacsit adni. Nincs szükség precíz utasítások összeállítására és átadására. A tanításhoz elegendő pontosan tudni az elérendő célt, és jutalmazni a helyes viselkedést. Egyszerűen arról van szó, hogy megerősítsük, amikor jól csinálja, ne foglalkozzunk a rosszul sikerült kísérletekkel, és gyakoroljunk minél többször, hogy végül maga*

jöjjön rá, mit is várunk el tőle. Az algoritmusokat hívjuk gépi tanuló algoritmusnak (…) A gépnek elég adatokat adni, megjelölni a kívánt célt, megerősíteni, amikor jól csinálja – és ráhagyni, milyen úton módon ér célba."[29]

4.2. Beavatkozási lehetőségek

A gépi tanuló algoritmusok előnye a „klasszikus" algoritmusokkal szemben, hogy olyan problémák megoldására is alkalmasak, amelyekre az emberi utasítássorok által programozott társaik nem, hiszen minden, az utóbbi kategóriába tartozó algoritmus olyan „okos", mint a készítői. Ehhez képest a gépi tanuló algoritmusok előre nem specifikált helyzetekkel is megbirkóznak, például tárgyakat ismernek fel képeken vagy szöveget fordítanak. Van azonban egy jelentős hátulütőjük, ami egyben a velük kapcsolatos szabályozás kiindulópontja lehet: az esetek nagy többségében még készítőik sem tudják, hogyan csinálják.

Az úgynevezett „feketedoboz-hatásból", azaz a döntési modell átláthatatlanságából fakadó kihívások mellett a gépi tanuló algoritmusok további jellemzői, mint a kiszámíthatatlanság, az összetett, részben autonóm működés mind veszélyt jelenthetnek az alapjogok érvényesítésére, az esetleges MI működéssel szembeni hatékony jogérvényesítésre, és megnehezíthetik a vonatkozó jogszabályok betartásának ellenőrizhetőségét.[30] Véleményem szerint ezekre a kérdésekre az algoritmusok életciklusának különböző szakaszaiban kell reflektálnia a szabályalkotónak, amennyiben a megbízható mesterséges intelligencia kereteit kívánja megteremteni. Itt szeretném megemlíteni, hogy az adatoknál már felvázolt beavatkozási pontok, megoldási javaslatok az algoritmusok esetében is megfontolandók, terjedelmi megfontolásokból azonban nem ismétlem meg a fentieket. Most pedig nézzük, hol lehetséges beavatkozni az algoritmusok világába, s milyen típusú, tartalmú szabályozási konstrukció képzelhető el.

[29] Fry 2020. 27.

[30] Fehér könyv a mesterséges intelligenciáról: a kiválóság és a bizalom európai megközelítése. COM(2020) 65 final (2020).

4.2.1. Tervezés (működési elv)

Napjainkra meghaladottá vált az a szemlélet, miszerint a technológia semleges, nem rendelkezik erkölcsi-etikai beállítottsággal, értékrenddel, hanem az embertől függ, milyen célra használja azt. Egy traktort például lehet mezőgazdasági munkákra is használni, s élelmiszert termelni, de gazdája ugyanúgy kiszánthatja vele a szomszéd termését, ha esetleg elmérgesedig köztük a viszony. Épp a tavaly kirobbant orosz-ukrán konfliktus kapcsán arról lehetett olvasni: az ukrán földművesek magukra hagyott orosz tankokat vontatnak el traktorjaikkal. A mesterséges intelligencia megjelenésével ez a helyzet gyökeresen megváltozott. Az algoritmusok esetében ugyanis fennáll a lehetősége annak, hogy tervezőik – véletlenül, vagy szándékosan – beépítsék saját „DNS-üket" a számítási modellbe, ami így későbbi működése során világnézetüket, értékrendjüket is tükrözheti. Ha például egy MI alapú HR asszisztens programozója szerint a férfiak alkalmatlanabbak egy bizonyos pozícióra, taníthat olyan viselkedést – pl. súlyozási szempontokat – az algoritmusnak, hogy az később az ő ebbéli meggyőződésének megfelelő eredményeket produkáljon. Az előbbi példa igen sarkos, ám ez nem azt jelenti, hogy nem áll fenn az ilyen okból fakadó torzítás veszélye.

Éppen ezért meggyőződésem, hogy nagy szükség van olyan szilárd, univerzális MI etikai keret kialakítására, ami alapvető értékeket, kvázi etikai kódexet határoz meg minden algoritmus tervezésére, s akár speciális követelményeket támaszt egyes nagy kockázatú rendszerekkel[31], vagy magukkal a tervezőkkel szemben. A tervezési folyamat során különösen figyelni kell, hogy etikus módon történjenek a dolgok, ennek hiányában ugyanis az életciklus további szakaszaiban sem lesz kivédhető az esetlegesen részrehajló, az alapjogokra veszélyt jelentő vagy más módon káros működés.

Arra az esetre, ha a tervezés során figyelmen kívül hagyják vagy akár szándékosan megszegik az etikai követelményeket, ki kell dolgozni azokat a mechanizmusokat, amelyekkel érvényt lehet szerezni a szabályos, etikus magatartásnak, illetve szükség esetén szankciók alkalmazásának lehetőségét is meg kell teremteni. Tekintve, hogy az MI alkalmazása nem ismer országhatárokat, a szakemberek, köztük Tilesch György és Omar Hatamleh egyetlen mindent lefedő, globálisan végrehajtható, a jogalkotás során

[31]Alapvetően ezt a logikát követi az Európai Unióban készülő átfogó MI rendelettervezet is, lásd: European Commission., Proposal of the Artificial Intelligence Act. COM(2021) 206 final (2021)

alkalmazható MI-etikai kódot tartalmazó etikai keretrendszer létrehozását javasolják.[32] Mivel már most számos szakma rendelkezik szigorú etikai kódexekkel és részletesen kidolgozott szankciórendszerrel, amelyeket független szakmai szervezetek tartatnak be, ezeket viszonylag könnyen és gyorsan lehet átszabni a teljes MI fejlesztő- és üzemeltető értéklánc számára. Így a makroszintű szabályozás mellett megvalósulhat a hatékonyabb, egyéni szintű reguláció is, teszik hozzá a szerzők.[33] Az etikai követelmények előírása természetesen az algoritmusok teljes életciklusa során indokolt, azonban véleményem szerint a tervezés szakaszában legfontosabb, mivel a nem etikusan kialakított algoritmusok veszélye a végfelhasználáshoz közeledve egyre kevésebé kezelhető.

4.2.2 Tanítás és tesztelés

A tanítás során a gépi tanuló algoritmusnak a feljebb már bemutatott kutyás metodikához hasonlóan kondicionáljuk működését. Ebben a szakaszban kulcsszerephez jutnak a korábban tárgyalt tanítási adatkészletek is, azonban egyáltalán nem mindegy, hogyan viselkedik az algoritmus, amikor találkozik ezekkel az adatokkal. Az egyik legérdekesebb kérdés ebben a szakaszban annak vizsgálata lehet, miként dolgozza fel az algoritmus a rendelkezésre álló tesztadatokat, még pontosabban: milyen mértékben befolyásolják azok a kimeneti eredményeket:

> *„Amennyiben a tanító adatkészlet elemei nem reprezentálják megfelelően annak az alkalmazási környezetnek az adatait, amiben a rendszernek később működnie kell, könnyedén torzított eredmények születhetnek. Az algoritmusok ugyanis váratlan és nem kívánt eredményeket produkálhatnak, ha olyan szituációval találkoznak, ami jelentősen eltér a tanítás során megismert helyzetektől. Egy friss tanulmány szerint egy MI algoritmus felülmúlta az emberi bőrgyógyászokat a bőrrák képekről való felismerésében, azonban felmerült egy lehetséges veszély is: mivel az algoritmust főként olyan képeken tanították, amin túlnyomóan fehér bőrű páciensek mintái voltak láthatók, a sötétebb árnyalatú bőrök esetében valószínűleg rosszabbul teljesíthetett volna. (…)*

[32] Tilesch és Hatamaleh 2021. 184.
[33] Uo.

> *Felmerül tehát a kérdés: hogyan nézhet ki a megfelelő tesztkörnyezet? Az első megfontolandó tényező az algoritmus típusának kiválasztása. Egy megfelelő rezsim több algoritmus adott probléma megoldásával kapcsolatos teljesítményét tesztelné, s vetné össze a specifikus teljesítménymutatókkal. (…) Ezt követően az algoritmus tanítási adatkészleten produkált eredményeit először egy tesztelési adatkészlet elemeivel, majd reális forgatókönyvek adataival kellene összevetni, minden esetben figyelve arra, hogy az adatok megfeleljenek a majdani alkalmazási környezet tulajdonságainak."*[34]

A tanítás során kiemelten fontos tehát, hogy olyan adatok felhasználásával és olyan körülmények között vizsgáljuk az algoritmusok teljesítményét, ami hűen reprezentálja a jövendőbeli való alkalmazási terüket. Ennek érdekében véleményem szerint szükséges olyan szabályokat alkotni, amelyek pontosan előírják az MI készítők számára, hogy a tesztelési folyamatok során a tervezett felhasználási környezettől függően milyen elemeket, sajátosságokat kell figyelembe venniük a nem kívánatos – pl.: diszkriminatív, vagy nem a valóságot tükröző – működés megelőzése érdekében. Különös figyelmet kell fordítani azokra az algoritmusokra, amelyek viselkedése működésük során változik, ilyen a gépi tanuló algoritmusok jelentős része. Ki kell dolgozni továbbá azokat a felügyeleti eszközöket, amelyek segítségével folyamatosan monitorozni lehet, hogy az adott MI továbbra is az eredeti célnak megfelelő módon funkcionál-e, nem tanult-e meg helytelen vagy működését torzító mintákat.

4.2.3 Eredmények – a feketedoboz probléma

Annak érdekében, hogy a mesterséges intelligenciában valóban megbízzanak az emberek, szükségszerű, hogy legalább felületesen ismerjék azokat a körülményeket, működési elveket, folyamatokat, amelyek az MI által létrehozott eredmények, döntések hátterében állnak. Röviden: átlátható, megmagyarázható és indokolható MI-re van szükség.

Sajnálatos módon ez nem minden esetben lehetséges, hiszen például egy mélytanuló neurális háló esetében sokszor maguk a programozók sem tudják, pontosan hogyan állította elő a rendszer az adott kimenetet. Ez nem csoda,

[34] Buiten 2019. 52.

hiszen az emberi agy nem képes befogadni és elemezni az input és az output közötti összetett, akár több milliárdnyi köztes lépést.[35] Nem vitás, hogy még neurális hálóknál jóval egyszerűbb algoritmusok megértése sem várható el az „utca emberétől", akár végfelhasználó, akár más módon találkozik az MI-vel, például érintett egy gép által hozott döntésben.

Ezért fontos az olyan szabályozási ökoszisztéma kialakítása, ami megköveteli az MI rendszerek bizonyos szintű átláthatóságát, előírja a laikusok számára is „megmagyarázható", megérthető döntési folyamatok követelményét, és rögzíti: az MI alkalmazása során a felhasználók számára is egyértelművé kell tenni, hogy nem szenvednek hátrányt, megbízhatnak a MI-ben és nem éri őket semmilyen jogi vagy erkölcsi sérelem. A felsorolt aggályok jelentős része az úgynevezett „*feketedoboz-hatás*" megszüntetésével, vagy – mivel e jelenség teljes orvoslása a technológia bonyolultsága miatt aligha képzelhető el a közeli jövőben –, a probléma szisztematikus kezelésével tompítható. Ugyan nem írhatjuk elő – és ez nem is volna indokolt –, hogy minden egyes MI rendszer működését részletesen megismerhessék a felhasználók és más érintettek, azonban a megmagyarázhatóság létfontosságú követelmény az olyan nagy kockázatú és „nagy téttel" bíró rendszerek esetében, amelyek előfeltétele a felhasználói bizalom (pl.: katonai, klinikai, igazságszolgáltatási- vagy bűnüldözési előrejelző, illetve banki rendszerek stb.).[36]

A feketedoboz-hatás legnagyobb veszélyét azonban nem az jelenti, hogy a felhasználók nem látnak a folyamatok mögé, hanem az, hogy az MI már-már okkult működése számottevően megnehezíti, számos esetben teljesen kizárja az esetleges jogsérelmet követő jogérvényesítés lehetőségét. „*Nem nehéz belátni, hogy a feketedoboz jelenti a legfőbb gátat az elszámoltathatóság és a megfelelő szabályozási keretrendszer létrehozása előtt az MI által okozott károk felelősségének megállapítására.*"[37] Az önvezető autók létezését jó részben MI alapú technológiának köszönhetjük, így a balesetek elkerülésére szolgáló komplex rendszer is ezen nyugszik. Abban az esetben, ha az autó balesetet okoz, kihívást jelenthet a károkozó magatartás és a bekövetkezett kár közötti okozati összefüggés bizonyítása. Már önmagában azt is nehéz bizonyítani, hogy a kárhoz vezető döntések sorozatát az algoritmus hibája okozta-e, hiszen szinte lehetetlen pontosan megállapítani, hol volt az a defektus, ami megindította az események láncolatát. Szintén nem volna

[35] Tilesch és Hatamaleh 2021. 139.
[36] Tilesch és Hatamaleh 2021. 140.
[37] Uo.

könnyű az algoritmus részrehajló működésének bizonyítása akkor sem, amikor egy rosszul kalibrált MI döntésének következtében nem vesznek föl valakit egy olyan pozícióba, amire önéletrajza és tudása alapján tökéletesen megfelel, hiszen van olyan rendszer, ami 25.000 jellemzőt vizsgálva választja ki[38] az alkalmas jelöltet az adott munkakörre.

Az átláthatatlanság csökkentésének esélyeit tovább rontja, hogy az MI rendszerek tervezői, gyártói jellemzően üzleti titokra hivatkozva nem osztják meg a nyilvánossággal az algoritmus működési dokumentációját, nem tárják fel annak belső logikáját, ezzel még a korlátozott tanulmányozási lehetőségtől is megfosztva a hozzáértőket.

Az átláthatóság problémaköre és az abból fakadó kihívások megnyugtató rendezése az MI szabályozása előtt álló legnagyobb kihívás. Vannak, akik az egész MI-vel kapcsolatos regulációt valójában a transzparencia szabályozásának tartják.[39] Mások úgy vélekednek, hogy az új MI szabályozási és támogatási rendszerekben a politikának tudatosan ki kell választania azokat az irányokat, amelyek a feketedoboz-jelenség csökkentésére irányulnak.[40] Az egészen bizonyos, hogy a jövőbeli regulációs törekvéseknek – akármilyen irányból érkeznek is – hangsúlyt kell helyezniük azoknak a technikai, jogi szabályoknak és etikai elveknek a kidolgozására, amelyek átláthatóbbá teszik az MI rendszerek működését. Úgy vélem fontos, hogy az átláthatóságot, megmagyarázhatóságot, s ezzel végső soron az elszámoltathatóságot megalapozó normák ne egy-egy terület már kialakult szabályozási rezsimjébe, ágazati jogszabályaiba épüljenek be, hanem egy átfogó MI kódex részeként minden algoritmussal szemben megfogalmazott, horizontális követelménnyé váljanak.

4.2.4. Érthető vagy pontos?

A feketedoboz-hatás további problémák megjelenését indukálja, melyek az algoritmusok potenciális teljesítményével, illetve az általuk generált eredményekkel kapcsolatban merülnek föl. Általánosan elfogadott tény, hogy az MI rendszerek átláthatóságának növelése csökkenti azok hatékonyságát, hiszen nem egyszer az algoritmusok emberek számára érthetetlen,

[38] C. Lennox 2021. 65.
[39] Buiten, 2019.
[40] Tilesch és Hatamaleh 2021. 185.

követhetetlen működésének köszönhetők a lenyűgöző – vagy csak „szimplán" midennapi életünket megkönnyítő – eredmények. *„Olyan rendszer tervezése, ami kézzelfogható magyarázatokat is generál a megoldások mellé, értékes időt és mérnöki teljesítményt igényel. Éppen ezért fel kell tennünk a kérdést, mikor kifizetődő az átláthatóság megkövetelése, egyensúlyozva az átláthatóság hasznát és egy ilyen rendszer létrehozásának költségeit."*[41] Az MI algoritmusok képesek arra, hogy rövid idő alatt váratlan, az emberek által számításba sem vett megoldásokat nyújtsanak egy adott problémára, s éppen ez teszi azokat pótolhatatlanná a különböző alkalmazási területeken. Legtöbb esetben ráadásul olyan műveletek, számítások tömegét végzik el, amelyekhez az emberi munkaerőnek nagyágrendekkel több időre lenne szükség. További vitathatatlan előnyük, hogy képesek olyan mintázatokat is észrevenni adattömegekben, amit az emberi agy képtelen, hiszen memóriánk, befogadóképességünk véges és nem tudunk több milliárd elemet egyben vizsgálni. *„Az átláthatóság tekintetében lehetséges kompromisszumot kötni az MI számítási kapacitása és megmagyarázhatósága között. Az emberek számára is könnyen értelmezhető lineáris számítási modellek csak egyszerű kapcsolatok megmagyarázására képesek, míg a komplex megoldásokat tartalmazó, többfunkciós módszerek értelmezése nehéz lehet. Amennyiben nagyobb átláthatóságot kívánunk, úgy el kell fogadnunk, hogy olyan MI rendszereket használunk, amelyek pontossága alul múlja azt a szintet, amire technológiai szempontból képesek lennének."*[42]

Az átláthatósággal járó haszon a döntéssel járó kockázattól függhet, ezért a szabályozásnak kontextusfüggőnek kell lennie, és a biztonságot, a méltányosságot és a magánéletet fenyegető kockázatokon kell alapulnia. Megkövetelhetjük az átláthatóságot olyan döntéseknél, amelyek a döntéshozón kívül másra is nagy hatással vannak. Kevésbé fontos döntéseknél a jobb rendszerteljesítmény érdekében kisebb átláthatóságot is választhatunk.[43] Ha tehát azt szeretnénk, hogy az algoritmusok működése átlátható legyen, bizonyos mértékben fel kell áldoznunk a hatékonyságot, ráadásul annak kockázata mellett, hogy az átláthatóvá tett működés az emberek számára továbbra is értelmezhetetlen számítások, kódok tömegét fogja jelenteni.

41 Buiten 2019. 57.
42 Uo. 58.
43 Uo.

5. Szabad-e vakon bízni az MI-ben?

Az MI által létrehozott eredményekkel kapcsolatos szabályok kialakításának másik jelentős irányaként azt kell meghatároznunk, hogyan viszonyulunk a döntésekhez, amelyeket az algoritmusok generáltak. Egyre több és több élethelyzetben kerül sor MI támogatott döntéshozatalra, nem egy ezek közül olyan szituáció, ami jelentősen befolyásolhatja a döntésben érintett személy, csoport vagy entitás életét, boldogulását. Bűnüldözés, igazságszolgáltatás, orvosi diagnosztika, banki ügyintézés – mind olyan területek, ahol biztosan kell tudnunk: a döntés valóban a legjobb szakemberek (gépek?) kezében van.

Paul Zilly esete[44] kiváló példája azoknak a helyzeteknek, amelyek elkerülése érdekében komoly regulációs célokat kell megfogalmaznunk. 2013 februárjában az amerikai Wisconsin államában Paul Zilly ellopott egy fűnyírót. A férfi ügyvédei kedvező vádalkut kötöttek, s a férfi teljes nyugalommal sétált be a tárgyalásra, hiszen komoly büntetésre nem lehetett számítani, letöltendő börtönbüntetés pedig szóba sem jöhetett. Zilly pechjére azonban a bírák egy COMPAS elnevezésű kockázatértékelő rendszert használtak, aminek működési mechanizmusát üzleti titok övezte. A COMPAS a számításait olyan kérdőív válaszai alapján végezte, amiben olyan kérdések szerepeltek, mint: „Egyetért-e ön azzal, hogy ha valaki éhes, annak joga van lopni?” Az algoritmusnak, ami feladatát körülbelül 70%-os pontossággal látta el, egy célja volt: megállapítani a visszaesés valószínűségét. Annak ellenére, hogy korántsem nevezhető pontosnak, a wisconsin-i bíráknak kötelező volt figyelmbe venni a rendszer eredményeit az ítélet kiszabásakor. Miután a bíró megismerte Zilly pontszámát – ami nem volt túl jó –, jobban hitt az algoritmusnak, mint az ügyvédeknek, s a vádalku elutasítása mellett megduplázta a büntetést: egy év fogház helyett két év fegyházra ítélte a szerencsétlen tolvajt.

Az MI-vel való felelősségteljes együttműködés kereteinek kialakítása a szabályozás egyik központi kérdése. Vitán felül áll, hogy a felelősen használt technológia hatalmas segítség a mindennapokban, hiszen közreműködik döntéseink meghozatalában, számos folyamatot automatizál, egyszerűsít. Sőt, az okosasszisztensek, a személyre szabott tartalmak vagy az önvezető autók olyan, korábban csak a filmvásznon létező álmokat váltanak valóra, amelyek láttán hajlamosak vagyunk teljesen félretenni az MI-vel szembeni fenntartásainkat. A kifejezetten emberközpontú MI alkalmazások korában

[44] Fry 2020. 75-76.

azonban nem szabad elfelejtenünk a technológia felelőtlen, hozzá nem értő, kritika nélküli alkalmazásának veszélyeit sem. Nem szabad megfeledkeznünk a Zilly-hez hasonló esetekről sem, amikor a technológiában való vak bizalom az emberi gondolkodás és értékek gondolkodás nélküli feladását eredményezi.

Felmerül tehát a kérdés: érdemes-e szabályozni ember és MI együttműködésének kereteit, szükséges-e olyan etikai, jogi, magatartási szabályokat kodifikálni, amelyek a fejlődés és az innováció lehetőségének fenntartása mellett megvédenek minket attól, hogy túlzott mértékben az MI által létrehozott eredményekre hagyatkozzunk, illetve kontrollálják együttműködésünket a technológiával? A válasz egyértelműen igen, s úgy gondolom, a kívánatos reguláció súlypontjai a következők lehetnek:

1.) A legfontosabb összegyűjteni és kategóriákba sorolni azokat az alkalmazási területeket, ahol az MI befolyásolni képes az emberi döntések eredményét, az élethelyzetek alakulását, s ezek közül kiemelten kezelni azokat az eseteket, amikor az algoritmus hibás, torz eredményeinek figyelembevétele közvetlen kockázatot jelent az alapjogokra. Az egyes kategóriákat különböző kockázati besorolással volna érdemes ellátni, és az egyes szintekhez különböző részletességű, szigorúságú szabályokat társítani, bizonyos szint alatt pedig mellőzni az emberi kontroll szükségességét. Egy hitelbíráló rendszerben hozott elutasító döntés esetében a fenntartások nélküli végrehajtás nagyobb veszélyeket hordoz, mint egy olyan fogyókúrás alkalmazás felülvizsgálata, ami sportolási tevékenységünk monitorozását követően tippeket nyújt az étrendükkel kapcsolatban.

2.) Fel kell térképezni, hogy az adott, nagyobb kockázati kategóriába eső döntési szituációkat milyen mértékben befolyásolják az MI által létrehozott eredmények, mennyiben segítik vagy helyettesítik az emberi mérlegelést, s ezek alapján kell kialakítani azokat a protokollokat, amelyek az MI döntés humán döntésbe való beépítését határozzák meg. E körbe tartozik annak a kérdésnek a rendezése, hogy az MI kimeneteit kötelező vagy csak ajánlott figyelembe venni a döntés meghozatala során, esetleg csak egyfajta támogatásként, megerősítésként kell tekinteni arra.

3.) Az MI rendszerek kimenetei a legtöbb esetben nem konkrét válaszok, hanem számokban kifejezett kockázati, valószínűségi esélyek és skálák, de nem ritka az egyszerű igen/nem válasz sem. Az algoritmus és rendszer fejlettségétől függően tehát az igen komplex eredményektől a „kétbites” válaszokig minden elképzelhető. Magától értetődő, hogy a különböző outputok nem egyenlő súllyal, hatékonysággal használhatók föl az emberi döntéshozatal

során, így érdemes olyan rezsimet alkotni, amely az MI rendszer fejlettsége, válaszainak szofisztikáltsága alapján határozza meg azok beépíthetőségét az emberi döntéshozatalba. Ha például egy pénzügyi trendeket elemző, brókereket segítő szoftver egyes gazdasági események valószínűségét egy skálán határozza meg, sokkal egyszerűbb döntést hoznia a felhasználóknak, mintha csak annyit tudnának meg, hogy növekedés vagy csökkenés várható.

E tanulmány kereteit ugyan bőven meghaladná az emberi tényező szerepének részletes elemzése, azonban nagyon fontos szót ejteni erről is a humán-MI kooperációk során. Lehet akármennyire tökéletes egy rendszer, ha felhasználói nincsenek felkészülve a használatára, semmit sem ér a technológiai fejlettség. A társadalmi szintű MI tudatosság kiépítése, az MI-vel való együttélésre való céltudatos felkészülés elengedhetetlen ahhoz, hogy ne váljunk az általunk teremtett technológia szolgáivá. Az emberek digitális életterének, a közösségi médiának a lelke a mesterséges intelligencia, és rengetegen válnak az ajánló algoritmusok által végtelenül felkínált tartalmak rabjaivá.

Az MI uralta univerzum természetesen nem veszélytelen. Széles körben tárgyalt jelenség pl. az ún. visszhangkamrák megjelenése, ami az MI algoritmikus tartalomválogatás eredménye. Ennek oka, hogy az algoritmus felhasználói profilok alapján rangsorolja a javaslatokat, keresési eredményeket és hírcsatornákat, és ezzel olyan információs buborékokat hoz létre felhasználóik körül, amelyben folyamatosan a személyes igényeiknek, érdeklődésüknek, ideológiájuknak megfelelő tartalommal találkoznak, de mással nem. Annak ellenére, hogy tömegek élnek ilyen visszhangkamrákban, nem igazán hallani olyan érzékenyítő, edukációs programokról, amelyek társadalmi szintű szemnyitogatást végeznek. Gondoljunk csak bele: ha ezzel a mindennapi jelenséggel is alig vannak tisztában a felhasználók, mi garantálja, hogy az ennél sokkal fontosabb területeken – mint az egészségügy, az igazságszolgáltatás vagy a fogyasztói jogok – alkalmazott MI ismeretével, veszélyeivel, tudatos és felelős használatával kapcsolatban nagyobb a világosság?

6. Záró gondolatok

Az MI keretek közé szorítására napjainkban a világ vezető politikai és gazdasági hatalmai kiemelt feladatukként tekintenek, ami nem csoda, hiszen

a technológia megjelenése olyan mértékben rúgja fel gazdasági és szociális konvenciókat, amelynek hatását még korántsem tudjuk felmérni. Az MI fejlődése egyelőre csak kezdeti szakaszában jár, azonban hétről-hétre találkozunk új alkalmazási területekkel, áttörésekkel és olyan tudományos jóslatokkal, amelyek egyre közelebbi időpontra teszik az általános – azaz az emberével azonos szintű – MI megjelenését. Nem kérdés tehát, hogy a technológiát még azelőtt kell szabályozni, amíg az egyáltalán véghezvihető és a feladat komplexitása nem haladja meg a létrehozóinak kapacitásait.

Tanulmányom – mely voltaképp gondolatkísérletként is értelmezhető – célja egy olyan megközelítés bemutatása volt, amely a mesterséges intelligencia szabályozásának egyfajta, némiképp idealizált módját tárta az olvasó elé. Ennek érdekében igyekeztem különbséget tenni az MI, mint technológia és az azon alapuló megoldások szabályozása között, kiemelve, hogy előbbi csak akkor képzelhető el és valósítható meg, ha nem önálló entitásként, hanem adatok és algoritmusok eredményeként tekintünk a mesterséges intelligenciára. E két alkotóelem, „beavatkozási pontként" tud funkcionálni a jogalkotó számára, hiszen rendelkeznek olyan jellemzőkkel, tulajdonságokkal, amelyekre nézve normatív tartalommal rendelkező szabály alkotható. Így a jogalkotó vagy más, akár soft-law alkotói potenciállal rendelkező entitás közvetett módon tudja befolyásolni az MI létrehozásának, működésének folyamatát és körülményeit, ezzel elősegítve, hogy biztonságos, megbízható és az emberi értékeket tiszteletben tartó technológia jöjjön létre.

Az említett, beavatkozási pontként kezelhető alkotóelemek részletesebb elemzése során felhívtam a figyelmet azokra a gócpontokra, amelyek mentén érdemes kialakítani a szabályozási struktúrát, mert ha e területek, kérdések rendezettek, az MI-vel járó kockázatok is jelentős mértékben mérséklődnek. Ilyen kulcsfontosságú az adatok körében az adatminőségre vonatkozó garanciák kialakítása és érvényre juttatása, valamint a minél robusztusabb kiberbiztonság megteremtése. Az algoritmusok esetében pedig követelményeket lehet megfogalmazni a tervezés folyamatával kapcsolatban, valamint a tanítási és tesztelési fázisra vonatkozóan. Külön figyelmet kell fordítani az MI működésének következtében létrejött eredményekre. Ezek körében kiemelt szerepük van átláthatóságnak és a megérthetőségnek, mivel ezek a feketedoboz-hatás ellenére is kívánatos célok.

Felhasznált irodalom

Apple unveils M1 Ultra, the world's most powerful chip for a personal computer – Apple Inc. (2022), https://www.apple.com/newsroom/2022/03/apple-unveils-m1-ultra-the-worlds-most-powerful-chip-for-a-personal-computer/ (2022.09.13.)

Buiten, Miriam C.: Towards Intelligent Regulation of Artificial Intelligence. *European Journal of Risk Regulation* 10. sz. 1 (2019): 41–59. DOI https://doi.org/10.1017/err.2019.8.

Brynes, Nanette: As Goldman Embraces Automation, Even the Masters of the Universe Are Threatened. *MIT Technology Review.* (2017) https://www.technologyreview.com/2017/02/07/154141/as-goldman-embraces-automation-even-the-masters-of-the-universe-are-threatened/ (2023. 10. 12.)

C. Lennox, John: A mesterséges intelligencia és az emberiség jövője. Budapest, Harmat Kiadó, (2021)

Fry, Hanna: Emberek és gépek – Hogyan tartsuk a kezünkben az irányítást a mesterséges intelligencia korában? Budapest, HVG Könyvek, (2020)

Goujard, Clothilde: EU lawmakers vote to ban online ads targeting children amid broader tech crackdown – *Politico* (2021). https://www.politico.eu/article/eu-lawmaker-rule-out-online-ads-target-children/ (2023.02.25.)

Heikkilä, Melissa: Dutch Scandal Serves as a Warning for Europe over Risks of Using Algorithms. *Politico* (2022). https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/ (2023.10.21.).

Henley, Jon: Dutch Government Resigns over Child Benefits Scandal. *The Guardian*, 2021. január 15., szak. World news. https://www.theguardian.com/world/2021/jan/15/dutch-government-resigns-over-child-benefits-scandal (2023.10.21.)

Inc, Global Industry Analysts. Global Big Data Market to Reach $234.6 Billion by 2026. (2021.) https://www.prnewswire.com/news-releases/global-big-data-market-to-reach-234-6-billion-by-2026--301322252.html. (2023.10.12.)

Kasparov, Garri: Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins. London, Hodder & Soughton, (2017)

MIT Technology Review. As Goldman Embraces Automation, Even the Masters of the Universe Are Threatened. (2017) https://www.technologyreview.com/2017/02/07/154141/as-goldman-embraces-automation-even-the-masters-of-the-universe-are-threatened/ (2023.10.23.)

Moor, James: The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine* 27, sz. 4. (2006.) https://doi.org/10.1609/aimag.v27i4.1911.

Heikkalä, Melissa: Dutch Scandal Serves as a Warning for Europe over Risks of Using Algorithms. *Politico* (2022), https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/ (2023.10.11.)

Roser, Max, és Hannah Ritchie: Technological Change. Our World in Data, (2021) https://ourworldindata.org/technological-change (2023.02.25.)

Sankin, Aaron, Dhruv Mehrotra, Surya Mattu, és Annie Gilbertson: Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them – The Markup. (2021). https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them (2022.09.13.)

Tardi, Carla: What Is Moore's Law and Is It Still True? https://www.investopedia.com/terms/m/mooreslaw.asp (2022) (2022.09.13.)

Tilesch, György és Omar Hatamaleh: Mesterség és Intelligencia. Budapest, Libri, (2021).

***

Strategy of „intervention points" – comments and suggestions on the regulation of artificial intelligence

Summary

In this contribution, I would like to elaborate on a specific approach of the regulation of the defining technology of our world, artificial intelligence. Classic legal regulation is unable to work with entities that lack legally graspable attributes. AI is one of these, so-called disruptive entities. It has yet to have

internationally approved definition, and lacks legally adjustable features as well. However, if we look behind the curtains, we realize, that AI practically consists of two very concrete components from the legal and regulatory perspective: data and algorythms. Focusing on these two parts, scholars and regulators can set a viable AI regulatoy environment, that can keep the pace of exponential developement and stand the test of time. That is why I briefly mapped the possibilities of regulating AI through data and algorythms, and also depicted some risks that come with it.