

Az emberi döntéshez való jog

Aziz Z. Huq*

[105 Virginia Law Review - (megjelenés: 2018)]

Absztrakt

A számítástechnikai technológiák közelmúltbeli fejlődése aggodalmat keltett az emberi hatalomnak a gépi döntéshozókra való áttelődése miatt. A börtönbüntetésektől kezdve a hitelek jóváhagyásán át a főiskolai jelentkezésekig a vállalati és állami szereplők egyre inkább a gépi tanulási eszközökre (a mesterséges intelligencia egy alcsoportjára) támaszkodnak a javak elosztásában és a kényszerítésben. A gépi tanulási eszközökről úgy vélik, hogy háttérbe szorítják, sőt kioltják az emberi cselekvőképességet olyan módon, hogy fontos egyéni érdekeket áldoznak fel. Az ilyen aggodalmakra kialakulóban lévő jogi válasz az emberi döntéshez való jog. Az európai jog az általános adatvédelmi rendeletben már elfogadta ezt a gondolatot. Az amerikai jog, különösen a büntető igazságszolgáltatás területén, már ugyanebben az irányban halad. Egyetlen joghatóság sem határozta meg azonban pontosan, hogy mit jelent ez a jog, és nem indokolta meg egyértelműen a létrehozását.

Ez a cikk az emberi döntéshez való jog jogi lehetőségeit vizsgálja. Először is meghatározom e jog technológiai plauzibilitásának feltételeit az állami cselekvéssel szemben alkalmazva. A technológiai predikátumok megértéséhez meghatározom azokat a határokat, amelyek mentén a gépi döntések különböznek az emberi döntésektől. Az ilyen technológiai kontextualizálás lehetővé teszi annak árnyalt feltárását, hogy az emberi és a gépi döntéseket elválasztó réseknek miért lehet normatív jelentőségük, vagy egyáltalán van-e. E technológiai elszámolás alapján elemzem az emberi döntéshez való jog normatív tétjét. Három lehetséges normatív indoklást veszek figyelembe: (a) az egyéni érdekekre való hivatkozás a részvételhez és az észérvek megadásához; (b) az állami cselekvés nem kellően átgondolt vagy egyénre szabott minőségével kapcsolatos aggodalmak; és (c) negatív externáliákon alapuló érvek. E három indok gondos elemzése azt sugallja, hogy nincs általános indok arra, hogy az állam emberi döntéshez való jogot fogadjon el. A nem kellően indokolt vagy pontos döntésekkel kapcsolatos normatív aggályokat, amelyek különösen erősen hatnak a jogi képzeletre, más módon lehet a legjobban kezelni. Hasonlóképpen, az algoritmikus eszközök által a társadalmi hatalom aszimmetriáit előidéző módokkal kapcsolatos aggályokat nem lehet az emberi döntéshez való joggal kivédeni. Sőt, ahelyett, hogy határozottan támogatnák az emberi döntéshez való jogot, a rendelkezésre álló bizonyítékok inkább arra utalnak, hogy a "jól kalibrált gépi döntéshez való jog" normatív szempontból megalapozottabb.

* Frank és Bernice J. Greenberg Chicagói Egyetem jogi kara. Köszönet Faith Laken-nek a nagyszerű kutatási segítségért. Köszönet Tony Caseynek, David Driesennek, Lauryn Gouldinnak, Daniel Hemelnek, Darryl Linek, Anup Malaninak, Richard McAdamsnak, Eric Posnernek, Julie Roinnak és Lior Strahilevitznek a figyelmes beszélgetésekért. A

Chicagói Egyetem, a Houstoni Egyetem és a Syracuse University School of Law workshop résztvevői szintén átgondolt visszajelzéseket adtak. Minden hiba az enyém.

Tartalomjegyzék

| | | |
|----------------|---|----|
| | Bevezetés | 1 |
| I. | Az hdöntéshozatal jog jogi megfogalmazása | 8 |
| A. | Az emberi h aló európai jog | 8 |
| 1. | Előzmények..... | 8 |
| 2. | A22..... GDPR cikke | 9 |
| B. | Az amerikai jog és az h aló jog | 11 |
| C. | Egy újszerű h álószerű formája | 13 |
| II. | A gépi h aló döntések közötti különbség | 15 |
| A. | A gépi tanulás mint az h aló | 15 |
| B. | A gépi és az emberi h aló megkülönböztetése | 19 |
| 1. | Hogyan h aló és az emberi döntések a működés | 19 |
| 2. | Más (emberi és h aló)..... átláthatatlansága | 21 |
| C. | Az emberi és a h aló összefonódása | 25 |
| D. | Következtetés..... | 28 |
| III. | Az emberi döntéshozatal jog jogi megfogalmazása | 29 |
| A. | A döntéshozatal pontossága | 31 |
| B. | Tantárgyi szempontú indokok..... | 33 |
| 1. | Részvétel | 33 |
| 2. | Indoklás Adás | 37 |
| C. | Osztályozóval szembe forduló területek..... | 44 |
| 1. | Indokolt állami fellépés | 44 |
| 2. | Az egyéni h aló jog | 46 |
| D. | Rendszerszintű aggályok..... | 60 |
| E. | Az emberi döntéshozatal való "jog" h aló | 63 |
| Következtetés: | Joga van a h aló döntéshozatal jog jogi megfogalmazása | 53 |

Az emberi döntéshez való jog

Bevezetés

Minden tektonikus technológiai lökés háziasított gabonától az első Az okostelefonok felbuzdulása: egy új fordít el. Az új társadalom születési fájdalmai egyszerű aggodalmak is vannak a hatalom elosztásával kapcsolatban - hogyan lehet megszerezni, és hogyan lehet elveszíteni. A gépi tanulás néven ismert számítástechnikai technológia hirtelen fejlődése a közvélemény legújabb, tintás aggodalmait váltotta ki. Ezek az új technológiák komplex algoritmusokat, zúgynevezett gépi tanulási eszközöket alkalmaznak a köz- és kormányzati adatok hatalmas halmazaira, hogy olyan feladatokat hajtsanak végre, amelyek korábban meghaladták az egyszerű emberek képességeit.³ A vállalati és állami szereplők egyre inkább ezekre az eszközökre támaszkodnak, hogy "az emberek életét és megélhetését befolyásoló döntéseket hozzanak - a hitelek jóváhagyásától kezdve a toborzáson, a bírósági ítélethezatalon át a főiskolai felvételig".⁴

Ennek eredményeképpen sokan úgy érzik, hogy elveszítik kontrollt az életükkel kapcsolatos kulcsfontosságú döntések felett.⁵ Attól tartanak, hogy a gépek olyan kritikus fontosságú kérdéseket oldanak meg, amelyek az egyén számára nem ismerhetők vagy nem ellenőrizhetők.⁶ Ennek eredményeképpen az egyének elveszítik az elemi emberi cselekvőképességüket, és ennek megfelelően kiszolgáltatottá válnak egy embertelen és embertelen gépi logikának. Egyesek szerint "maga a gondolat", hogy egy algoritmikus rendszer

¹ A társadalmi átalakulások e témáján: JAMES C. SCOTT. AGAINST THE GRAIN: A DEEP HISTORY OF THE EARLIEST STATES (2018), és RAVI AGRAWAL, INDIA CONNECTED: HOW THE AZ OKOSTELEFON ÁTALAKÍTTJA A VILÁG LEGNAGYOBB DEMOKRÁCIÁJÁT

² Az algoritmus egyszerűen "egy bizonyos cél eléréséhez szükséges jól meghatározott" Joshua A. Kroll et al, *Felelős algoritmusok*, U165. PA. L. REV. n.633,64014 *Lásd még* THOMAS H. CORMEN ET AL „INTRODUCTION TO ALGORITHMS (102d ed. 2001) (az algoritmus definíciója: "bármely jól definiált számítási folyamat, amely bemenetként valamilyen értéket vagy értékek halmazát veszi, és kimenetként valamilyen értéket vagy értékek halmazát állítja elő" (a hangsúlyok kihagyva)). A számítástechnikai feladat, annak atomi szintjén, a soros algoritmusok végrehajtását foglalja magában. MARTIN ERWIG, ONCE UPON AN ALGORITHM: HOW STORIES EXPLAIN COMPUTING 2-4 (2017).

³ A gépi tanulás egy általános célú nagyobb értelemben "olyan algoritmusokat és rendszereket foglal magában, amelyek a tapasztalattal javítják tudásukat vagy teljesítményüket". PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA (32012); *lásd még* ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING (23d ed. 2014) (a gépi hasonló fogalmakkal történő meghatározása). A gépi tanulás felhasználási módjairól lásd Susan Athey, *Beyond Prediction: Using Big Data for Policy Problems*, SCIENCE, Febr3. ,2017 , at483 , <http://science.sciencemag.org/content/355/6324/483>. full (megjegyezve a strukturált gépi tanulás használatát előrejelzési problémák megoldására). A projekt technológiai hatókörét és a releváns fogalmak meghatározását az *alábbiakban*, a kísérő megjegyzés81 szövegében tárgyalom. Az "algoritmikus eszközök" és a "gépi tanulás" kifejezéseket felváltva fogom használni, annak ellenére, hogy az algoritmusok osztálya technikailag sokkal nagyobb.

⁴ Kartik Hosanagar és Vivian Jair, *We Need Transparency in Algorithms, But Too Much Can Backfire*, HARV. BUS. REV., 201823. július; Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, GEO105. L.J. (1147,11492017) ("[T]oday, advanced machine-learning algorithms are creating a hatalmas mértékben automatizált társadalom, transforming many facets of life. Számos termék és szolgáltatás, köztük az e-mail spamszűrők, az orvosi diagnózisok, a termékmarketing és az önvezető autók ma már a gépi tanuló algoritmusoktól és azok elképesztő előrejelzői képességtől és gyorsaságtól függ.")

⁵ Shoshana Zuboff, *Nagy Más: J* 20. INFO. TECH. 75,75 (2015) (az információs kapitalizmus "új formáját írja le, [amely] az emberi viselkedés előrejelzésére és módosítására törekszik, mint a bevétel és a piaci ellenőrzés eszköze").

⁶ *Lásd pl.* Rachel Courtland, *The bias detectives*, SCIENCE, június (a nyilvánosság körében a gyermekbántalmazás felderítésére szolgáló algoritmikus kockázati pontszámokkal kapcsolatos aggodalmak dokumentálása, mivel azok nem veszik figyelembe "az [egy] élet megváltoztatására irányuló erőfeszítést").

egy fontos döntést a ~~h~~ adatok alapján, eleve "igazságtalannak" tűnik.⁷ A gépek, mondják, végzetesen hiányolják az "empátiát".⁸ Mások számára a gépi döntések veszélyesen kifürkészhetetlenek, átláthatatlannak, és ezért veszélyesen kiszámíthatatlannak tűnnek.⁹ Ami még rosszabb, hogy a kormányok és a vállalatok szabadon használják ezeket az eszközöket, hogy népességüket taxonómiaiilag osztályozzák, az egyéni viselkedést megjósolják, sőt, a viselkedést és a preferenciákat olyan módon manipulálják, amely új előnyhöz juttatja őket az algoritmikus osztályozás emberi alanyaival szemben. Még a politikai választás alapvető feltételei is veszélyeztetettnek tűnnek.¹⁰ Ugyanakkor, amikor a gépi tanulás a polgárok és a kormányzat (vagy a nagyvállalatok) közötti interakció alapvető feltételeit készül megváltoztatni, úgy tűnik, hogy a robotika és a gépi tanulás fejlődése a munkaerőpiac nagy részét fogja kiszorítani mind a munkás-, mind a fehérgalléros munkaerőpiacról.¹¹ Félelmetes jövő rajzolódik ki, amelyet hatalmas gazdasági zavarok jellemeznek, ahol az emberek elveszítik az irányítást számos központi életválasztás felett, és az alapvető fogyasztói és politikai preferenciák már nem igazán a sajátjuk.

Ez a cikk egy születőben lévő ~~sz~~ analízisről szól ezekre a félelmekre: annak lehetőségéről, hogy bizonyos feltételek mellett az embernek joga van emberi döntéshez, nem pedig egy tisztán automatizált folyamat által hozott döntéshez (amit én "gépi döntésnek" nevezek). Az európai jog felkarolta ezt az elképzelést. Az amerikai jog, különösen a büntető igazságszolgáltatás területén, kacérkodik vele.¹² Ebben a cikkben az a célom, hogy teszteljem ezt az újjászülető javaslatot, megvizsgáljam a technológiai lehetőségekkel való kapcsolatát, és megállapítsam, hogy ez a javaslat meggyőző válasz-e a növekvő elosztási, politikai és ismeretelméleti aggodalmakra. Komolyan véve az algoritmikus jövő által felvetett, a méltányossággal, a magánélet védelmével és a hatalommal kapcsolatos aggodalmakat, itt azt kívánom elemezni, hogy az állami szereplők emberi döntéshez való joga meggyőző választ jelent-e erre.

A vizsgálat motiválására tekintsünk ~~h~~ nyilvános vitában már kibontakozó aggodalmak ~~E~~gy ápolói szakszervezet például kampányt indít, amelyben arra szólítja fel a betegeket, hogy a technológiai értékelés helyett emberi orvosi ítéleteket követeljenek.¹³ Az Accenture 2018 felmérésében megkérdezett betegek többsége pedig határozottan elutasította, hogy ember helyett gép kezelje őket.¹⁴ Amikor Kaliforniában a pénzzel fizetett óvadékot felváltotta

⁷ Reuben Binns et al., "Az emberi lényt százalékra redukálja": *Perceptions of ~~h~~ Algorithmic Decisions*, PROC. OF THE CHI2018 CONFERENCE ON HUMAN FACTORS IN COMPUTINGSYSTEMS (3852018).

⁸ VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW ~~h~~ PROFILE, POLICE, AND PUNISH THE POOR (302018).

⁹ Will Knight, *The Dark Secret at the Heart of AI*, ~~M~~EV. ,2017 , április11 , 2017, <https://www.technologyreview.com/s//the-dark-secret-at-the-heart-of-ai/604087> ("Az ezeket a szolgáltatásokat működtető számítógépek önmagukat programozták, és ezt olyan módon tették, amit nem értünk. Még az alkalmazásokat fejlesztő mérnökök sem tudják teljesen megmagyarázni a viselkedésüket. ").

¹⁰ *Lásd pl. Daniel Kreiss és Shannon C. McGregor, ~~W~~irms shape political communication: The work of Microsoft, Facebook, Twitter, and Google with campaigns during the US2016 presidential cycle*, POL35. COMM. 155, 156-57 (2018) (a közösségi médiában dolgozók szerepének leírása a kampányok alakításában).

¹¹ A mára általánossá vált nézetet *lásd ~~h~~ Robots will take our jobs. We'd better plan now, before it's too late*, GUARDIAN, Feb. 1, 2018, <https://www.theguardian.com/commentisfree//feb//robots-take-our-jobs-201801> amazon-go-seattle. Egy árnyaltabb beszámolóért lásd MARTIN FORD, RISE OF THE ROBOTS: TECHNOLOGY AND THE THREAT OF A JOBLESS FUTURE 282-83 (2017) (árnyaltabb és feltételes elemzés).

¹² *Lásd az alábbi szöveg kiséző megjegyzését* 81.

¹³ National Nurses United, *When it ~~n~~ insist on a registered nurse*, May14 ,2013 , <https://www.nationalnursesunited.org/insist-registered-nurse>.

¹⁴ Accenture, *Accenture Consumer2018 Survey on Digital Health*, https://www.accenture.com/hu-en/insights/consumer-survey-digital-health?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosvitals&stream=top-stories.

"kockázatalapú tárgyalás előtti állami bíróságok bírái figyelmeztettek, hogy "[a] technológia nem helyettesítheti az ügyészek és a vádlottak ügyvédei által birtokolt mélyreható igazságügyi ismereteket, tapasztalatot és szakértelmet a bűnüldözés terén".¹⁵ Flint 2018, városában (MI) megszüntették egy rendkívül hatékony gépi tanulási eszköz használatát a hibás vízvezeték azonosítására, és a közösség nyomására visszatértek az emberi döntéshozatalhoz, amely sokkal alacsonyabb találati aránnyal rendelkezik a hibás csövek felderítésében.¹⁶ Vagy gondoljunk csak a Cathy O'Neill adattudós által elmondott anekdotában összefoglalt aggodalomra: Egy arkansasi nőtől, Catherine Taylortól megtagadják a szövetségi lakhatási támogatást, mert megbukik egy automatizált, "webes, adatgyűjtő" háttérelőnézésen.¹⁷ Csak amikor "egy lelkiismeretes ember" veszi a fáradságot, hogy megvizsgálja a gépi eredmény minőségét, derül ki, hogy Taylor tévesen került piros zászlóra.¹⁸ O'Neill anekdotája szépen érzékelteti azt a félelmet, hogy a gépek igazságtalanok, érthetetlenek vagy összegeyztetetetlenek lesznek az elemi emberi cselekvőképesség hajlításával.

Az **his** nem az automatizált döntéshozatal értékének eltérő megítélése a nagy repülőgépgyártók - és szabályozó hatóságok - eltérő megközelítéseiben is megmutatkozik a kereskedelmi légitársaságok "mesterséges intelligenciával" működő robotpilótáival kapcsolatban.¹⁹ Míg az Airbus a nagyfokú automatizálást részesíti előnyben, a Boeing az emberi pilótákra bízta az irányítást ²⁰(bár talán nem mindig eléggé?²¹). Az amerikai jog az embert részesíti előnyben a géppel szemben. Az olyan robotpilóta-funkciók, amelyeket a pilóták nem tudnak "gyorsan és határozottan kikapcsolni, hogy megakadályozzák, hogy a robotpilóta beavatkozzon az irányításukba", tilosak a szövetségi szabályozás alá tartozó repülőgépeken.²² A rendelet valójában egy szűken vett jogot formál az emberi pilótához, legalábbis az utolsó esetben.

Az emberi döntéshez **h**addigi legfontosabb megfogalmazása az európai jogban található. Áprilisban az 2016, Európai Parlament új adatvédelmi rendszert léptetett életbe a

¹⁵ Quentin L. Kopp, *Replacing Judges with Computers is Risky*, **N** REV. BLOG, Feb 18, 2018, <https://blog.harvardlawreview.org/replacing-judges-with-computers-is-risky/>.

¹⁶ Alexis C. Madrigal, *How a Feel-Good AI Story Went Wrong* **Flint**, THE ATLANTIC, Jan. 3, 2019, <https://www.theatlantic.com/technology/archive//how-machine-learning-found-flints-lead-pipes///201901578692>.

¹⁷ CATHY O'NEILL, A MATEMATIKAI PUSZTÍTÁS FEGYVEREI: **N**ÖVELI A NAGY ADATMENNYISÉG AZ EGYENLŐTLENSÉGET ÉS VESZÉLYEZTETI A DEMOKRÁCIÁT 152-53 (2016).

¹⁸ *Id.* 153.

¹⁹ Jack Stewart, *Don't Freak Out Over Boeing's Self-Flying Plane-Robots Already Run the* **WIRED**, June 9, 2017, <https://www.wired.com/story/boeing-autonomous-plane-autopilot/>.

²⁰ *Id.*

²¹ James Glanz et al., *Why Regulators Fear the Two Boeing 737 Crashes Happened for Same Reason*, N.Y. TIMES, Mar. 13, 2019, **N**.

²² 14 C. F. R. §21.1329(a)(1); *lásd még* 14 F. R. §A légi járművet irányító pilóta közvetlenül felelős a légi jármű üzemeltetéséért, és ő az utolsó hatóság a légi jármű üzemeltetése tekintetében. "). Halálos balesetek előfordulnak az automatizált rendszerek hibája miatt. Júniusban 2008, az Air France járata az Atlanti-óceánba zuhant 447, a 228 fedélzeten tartózkodó összes ember meghalt. A francia légügyi hatóságok által ezt követően lefolytatott vizsgálat megállapította, hogy a balesethez vezető eseménysorozatban a robotpilóta váratlan kikapcsolása is szerepelt. Bureau d'Enquêtes et

d'Analyses, *Final Report on the accident on st June 2009 to the 1st June 9 to the Airbus 330A-203 registered F-GZCP operated by the Air France flight AF Rio447 de Janeiro*, (17 July 2012), <https://www.bea.aero/docspa/fcp2009090601.en/pdf/fcp090601.en.pdf>.

általános adatvédelmi rendelet ²⁴Az általa felváltott jogi rendszerrel ellentétben ^{24a} májusban bevezetett GDPR 2018jogilag kötelező érvényű, még akkor is, ha az Európai Unió ("EU") tagállamai nem hoznak végrehajtási jogszabályokat. Ennélfogva a bíróságon közvetlenül érvényesíthető, jelentős pénzbírságok révén. ²⁵22A GDPR cikke felruhazza a természetes személyeket "azzal a joggal, hogy ne lehessen olyan, kizárólag automatizált adatkezelésen alapuló döntés tárgya, ideértve a profilalkotást is, amely őt érintő joghatásokat eredményez vagy őt hasonlóképpen jelentős mértékben érinti".²⁶ Ez a jog kiterjed a magán- és egyes (de nem minden) állami szervekre.²⁷ Első pillantásra ez egy meglehetősen általános hatályú opt-outot jelent az automatizált döntéshozataltól az emberi döntéshozatalig.²⁸

A GDPR extraterritoriális ²⁹is rendelkezik. Olyan platformokra is vonatkozik, mint Google és a Facebook, amelyek az EU-n belül nyújtanak szolgáltatásokat.²⁹ És az amerikai jog is tesz óvatos lépéseket a hasonló emberi döntéshez való jog felé. A Wisconsin Legfelsőbb Bíróság például 2016, úgy ítélte meg, hogy az algoritmus által generált kockázati pontszám "nem tekinthető meghatározó tényezőnek annak eldöntésében, hogy az elkövető biztonságosan és hatékonyan felügyelhető-e a közösségben", a megfelelő eljárás értelmében.³⁰ Ez a határozat kizárja az óvadék megállapításának teljes automatizálását. Az emberi bírónak is részt kell vennie a folyamatban. A wisconsini bíróság döntése valószínűleg nem bizonyul egyedülállóan. A gépi tanulás állami alkalmazása általánosságban véve éles panaszokat váltott ki az eljárási igazságosság és a méltányosság szempontjából.³¹ Továbbá, a hatodik kiegészítés esküdtszéki tárgyaláshoz való jogát mindeddig alapvetően

²³ Bizottság ~~2016~~rendelete HL (L 119), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679> [a továbbiakban "általános adatvédelmi rendelet" vagy "GDPR"]; lásd még Christina Tikkinen-Piri, Anna Rohunen és Jouni Markkula, *EU General Data Protection Regulation*: COMP34. L. & SEC. REV. 134, 134-35 (2018) (a GDPR hatályba léptetési folyamatának dokumentálása).

²⁴ Lásd a személyes adatok feldolgozása vonatkozásában az egyének védelméről és az ilyen adatok szabad áramlásáról szóló 1995. októberi 24 95/46/EK európai parlamenti irányelvet, 95/46/EK, cikk. 1,1995 HL (L 281.) (EK) [a továbbiakban: 95/46/EK irányelv].

²⁵ Bryce Goodman & Seth Flaxman, *European Union regulations on algorithmic ~~data~~ and a "right to explanation"*, arXiv preprint at 1606.08813, (2016) (a nem kötelező érvényű irányelv és a jogilag kötelező érvényű rendelet közötti különbség magyarázata az európai jog szerint).

²⁶ GDPR 22. § (1) BEKEZDÉS.

²⁷ GDPR 4. § (7) és (8) bekezdés (az "adatkezelő" ~~mint~~ kulcsfontosságú fogalmak meghatározása). A rendelet azonban nem vonatkozik a bűnügyi és biztonsági vizsgálatokra. Lásd a GDPR 2. § d) pontját.

²⁸ Amint azt alább kifejtem, nem ez a GDPR egyetlen olyan ~~amely~~ ^{amely} úgy értelmezhető, hogy emberi döntéshez való jogot teremt. Lásd az ~~alábbi~~⁶³ szöveg kísérő megjegyzéseit 68.

²⁹ A GDPR extraterritoriális hatályával ~~is~~ ^{is} eltérések vannak a ~~valamely~~ ^{valamely} a mértéktől kezdve, lásd Griffin Drake, *Navigating the Atlantic: A bizonytalanság tengere közepette az uniós adatvédelmi megfelelés megértése*, 91 S. CAL. L. REV. 163,166 (2017) (a GDPR értelmében az amerikai vállalatokra leselkedő ~~új~~ ^{új} kockázatok dokumentálása), a riogatóság lásd Mira Burri, *The Governance of Data and Data Flows in Trade Agreements: The Pitfalls of Legal Adaptation*, U 51.C. Davis L. Rev. (65,922017) ("A GDPR sok szempontból túlzottan terhes és jelentős extraterritoriális hatásokkal jár.").

³⁰ State v. Loomis, N.W881.2d 749,760

³¹ Lásd pl. Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And' Its Biased Against Blacks*, PROPUBLICA (May23 ,2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [[https://perma. cc/LT-67-ELPG](https://perma.cc/LT-67-ELPG)] (a gépi

tanulással kapcsolatos eszközök büntetőjogi kontextusban történő kritizálása).

a bírósági tényfeltárásnak való ellenállás érdekében.³² De olyan koncepcionális ok, amiért a hatodik módosításra ne lehetne hivatkozni, hogy kizárja az algoritmusok által generált eredmények legalább bizonyos formáit a büntetőjogi ítélethozatal kontextusában. Sőt, úgy tűnik, hogy a fortiori következik, hogy egy olyan jog, amely kizárja az esküdtszék bíró általi helyettesítését, szintén megakadályozná, hogy azt egy egyszerű gép helyettesítse.

Ezek csak a jéghegy csúcsai. Gépi tanulás elterjedését tekintve még az elején járunk. Az első igazán működőképes, komplex vizuális és szöveges mintafelismerésre képes "mélytanuló" eszközöket alig egy évtizede tesztelték sikeresen.³³ Idővel még sok emberi képesség fog feleslegesnek bizonyulni.³⁴ A lehetséges kizorolások hatása valószínűleg "mélyreható" lesz.³⁵ A londoni University College kutatócsoportja 2016, például olyan gépi tanulással működő eszközt fejlesztett ki, amely százalékos⁷⁹ sikerességgel képes volt megjósolni az Európai Bíróság döntéseit.³⁶ A gépi bírák ideje még nem érkezett el. De a horizonton várokozik.³⁷ Tehát még ha az emberi döntéshez való jog ma még fantáziadúsnak is tűnik, hamarosan eljön az a nap, amikor ez tűnik az egyetlen válaszlépésnek egy helyrehozhatatlan technológiai tengerváltozásra. Tehát megérett az idő arra, hogy megkérdezzük, mit is jelent ez pontosan.

Ebben a cikkben először az emberi döntéshez helyezem el a kortárs technológiai kontextusban. Ezáltal pontosan meg tudom határozni azokat a határokat, amelyek mentén a gépi döntések különböznek az emberi döntésektől. Fókuszomban a kapcsolódó technológiák egy csoportja áll, amelyet gépi tanulásnak nevezünk. Ez a

³² Lásd pl. *Apprendi v. New Jersey*, U530.S. (40007) (kifejtve, hogy az ötödik és hatodik módosítás "vitathatatlanul feljogosítja a vádlottat arra, hogy az esküdtszék minden kétséget kizáróan megállapítsa, hogy a vád tárgyát képező bűncselekmény minden elemében bűnös" (változtatás az eredetiben) (belső idézőjelek kihagyva) (idézte *United States v. Gaudin*, U 515.S. (506,1995510))).

³³ Jürgen Schmidhuber, *Mélytanulás neurális hálózatokban*: (2015); Yoshua Bengio, *Machines Who Learn*, SCIENTIFIC AMERICAN, June at2016, (46,48 megjegyezve, hogy a mélytanulás 2005-ben "jött be"). A mélytanulásról lásd: Yann LeCun et al., *Deep Learning*, SCIENCE 521 (436,4372015) (a mélytanulás leírása a sztochasztikus gradiens felemelkedés használatáról az előrejelzés minőségének folyamatos javítása érdekében); lásd még a kísérő megjegyzés alatti szöveget. 116.

³⁴ Mégis könnyű túlbecsülni valószínű rövid távú változások nagyságrendjét, és figyelmen kívül hagyni a következetesebb hosszú távú dinamikát, lásd Rodney Brooks, *The Seven Deadly Sins of AI Prediction*, MIT TECH. REV., Oct6, 2017, <https://www.technologyreview.com/s//the-seven-deadly-sins-of-ai-predictions/609048>.

³⁵ Lásd Erik Brynjolfsson és Tom Mitchell, *What can AI learning do? Workforce implications*, SCIENCE 35815 30, 15-3031 (2017) (a gépi tanulás legújabb fejlődésének köszönhetően "még nagyobb és gyorsabb átalakulást [a munkaerőpiacokon] jósolnak"). A különböző típusú gépi tanulást alkalmazó vállalatok száma rendkívül gyorsan nőtt az elmúlt három-négy évben. Andrew Nusca, *The Current State of Artificial Intelligence, According to Nvidia's CEO*, FORTUNE, Mar. 22, 2016, 9: 00AM, <http://fortune.com///artificial-intelligence-nvidia/20160322>.

³⁶ Nikolaos Aletras, Dimitrios Tsarpatanis, Dreouciuc-Pietro & Vasileios Lamos, *Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective*, PEER2 J. COMP. SCI. e93 (2016); lásd még Chris Johnston, *Mesterséges intelligencia bíró kifejlesztett a UCL tudósok*, A GUARDIAN, Okt. 24, 2016, <https://www.theguardian.com/technology//oct//artificial-intelligence-judge-university-college-london-201624-computer-scientists#https://www.theguardian.com/technology//oct//artificial-intelligence-judge-university-201624-college-london-computer-sci> ("Nem gondoljuk, hogy az AI helyettesítené a bírákat vagy ügyvédeket, de úgy gondoljuk, hogy hasznosnak találnák az ügyekben a bizonyos kimenetelhez vezető minták gyors azonosításában." (idézi az algoritmus alkotóját)).

³⁷ Sean Braswell, *Mindenki álljon fel a főbíró robotért!*, OZY (június 11, <http://www.ozy.com/multiplatform-rise-for-chief-justice-robot/41131>); Eugene Volokh, *Chief Justice Robots*, DUKE 68 L.J. (1135,11422019) (a "mesterséges intelligenciával rendelkező bíró létrehozásának ötletét dolgozza ki, akit jogi döntések meghozatalára használhatunk"). A "gépi

tanuláson alapuló döntések" közigazgatási bíróként való alkalmazásának lehetőségéről szóló árnyalt vitát lásd Coglianese & Lehr, *Supra* note at 4,1186-91. Az ilyen elképzelések elleni meggyőző érvert lásd Emily Berman, *A Government of Laws and Not of Machines*, B98.U. L. Rev. (12772018).

a mesterséges intelligencia ma leggyorsabban formája.³⁸ Az emberi döntéshez való jogot nem lehet meghatározni vagy értékelni az emberi döntéshozatal és az ilyen gépi tanulási technológiák által hozott döntések közötti technikai különbségek megértése nélkül. A gépi tanulás tervezésének és végrehajtásának gondos elemzése ugyanis azt mutatja, hogy az emberi döntéshez való jog gondolata a gyakorlatban nem egyértelmű. A jelenlegi technológiai korlátok miatt sem igazolható bizonyos okokból. A gépi döntések átláthatatlanságára és autonómiájára vonatkozó néhány gyakori panaszt, mint mutatom, technikai tények veszélyeztetik vagy meghiúsítják. Az emberi közreműködéshez való joggal kapcsolatos állítások - javaslom - jobban érthetőek, ha az ilyen közreműködés időzítésén, nem pedig pusztán tényén alapulnak.

E technikai alapokkal a ~~h~~elemzem az emberi döntéshez való jogot a normatív célokkal összefüggésben, amelyek előmozdítása érdekében az jog hihető módon értelmezhető. Három, egymástól némileg eltérő indoklást elemzek. Az első elemzési irányvonalam a gépi döntésnek kitett egyén érdekeit veszi sorra. Ezek közül a leglényegesebbek az egyén döntéshozatalban való részvételén és az indoklás lehetőségén alapulnak. A második elemzési szempont azt követi nyomon, hogy egy gépi eszköz milyen módon lehet önmagában véve kifogásolható (anélkül, hogy figyelembe vennénk az egyéni jogokat). Itt az egyéni mérlegelés hiányával kapcsolatos aggodalmakra összpontosítok, valamint arra, hogy a gép nem képes indoklással ellátott ítéleteket hozni. Végezetül a dinamikus, rendszerszintű hatásokat vizsgálom a negatív tovagyrűzés formájában, különösen a társadalmi hatalommal kapcsolatban. Az ilyen típusú érvelések közül végül egyik sem nyújt biztos alapot az emberi döntésekhez való joghoz való joghoz. Inkább *a jól kalibrált gépi döntéshez való jog* (eddig látszólag ellentmondásos) lehetőségét kínálom fel. Azaz a gépi döntések gyakran képesek az embernél kisebb számú hamis pozitív és hamis negatív eredményt tartalmazó osztályozásra, és kevesebb torzító torzítással képesek cselekedni. De ahogy a kritikusok joggal megjegyzik, az elmélet nem gyakorlat. A kormányok által alkalmazott eszközök közül sokan hajlamosak arra, hogy erősen hibásak legyenek.³⁹ A jól kalibrált gépi döntéshez való jog, és nem a hasonlóan hibás emberi döntéshozatalhoz való visszatérés, normatív értelemben ugyanolyan hihető és erőteljes, mint az ismertebb ellenpélda. Ez a következtetés, hogy világos legyen, hangsúlyozottan nem a gépi tanulás elterjedése ellen felhozott elosztási, politikai és normatív kifogások elutasítása. Inkább azt remélem, hogy megmutatom, miért nem megfelelő válasz az emberi döntéshez való jog. Azáltal, hogy tisztázom a szűk normatív határokat, remélem, hogy megelőzhetjük a jogi befektetéseket egy alkalmatlan ötletbe, mielőtt az visszavonhatatlanul beágyazódna a jogba.

Elemzésem itt arra összpontosít, hogy ilyen jog hogyan vonatkozhatna az állami fellépésre, és nem a magánjogi fellépésre, három okból kifolyólag. Először is, az Egyesült Államok kiemelkedő jogi keretei, a GDPR hatályával ellentétben, nagyrészt (bár nem kizárólag) az állami fellépésre irányulnak. Ennek megfelelően az állami fellépésre való összpontosításnak van értelme a jelenlegi amerikai szabályozási környezet magyarázata és értékelése szempontjából. Másodszor, míg az algoritmikus eszközök magánfelhasználásának köre széles és heterogén, az algoritmikus eszközök állami felhasználása mindeddig homogénebb. Ez teszi őket az elemzés egyszerűbb tárgyává. Az algoritmusokat ma már a magán-tevékenységek sokkal szélesebb körében alkalmazzák. Ezek a Google PageRank-től kezdve

³⁸ *Lásd a kísérő megjegyzés alatti szöveget (a ~~h~~5 meghatározása)* Nem vagyok egyedül ezzel a fókusszal. A jogtudósok egyre nagyobb figyelmet szentelnek az új algoritmikus technológiáknak. Vezető példákért lásd Kroll et al., 2. lábjegyzet; Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, U164. REV. 871,929 (2016) ("keretrendszer" kidolgozása a gépi tanulási technológiáknak a negyedik módosítás elemzésébe való integrálásához);

Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, U163. PA. L. REV. 327, 383-84 (2015) (hasonló); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward A Framework to Redress Predictive Privacy Harms*, B.C55. L. REV. 93,109 (2014) (érvelve az "eljárási adatjogi tisztességes eljárás mellett [a Big Data elemzési folyamatainak méltányosságának szabályozása érdekében] a személyes adatok (vagy metaadatok ...) felhasználásának módja tekintetében").

³⁹ AI Now, *AI Now Report2018*, 18-22. o. (2018), [link](#)

algoritmus,⁴⁰ az bevételi források használt "fintech",⁴¹ a stroke kockázatának kiszámítására használt orvosi műszerek,⁴² valamint a mérnökök által új stabil szeretlen vegyületek azonosítására használt "fintech".⁴³ Az algoritmikus eszközök olyan új alkalmazásokba is beágyazódnak, mint a hangfelismerő szoftverek, a fordítószoftverek és a vizuális felismerő rendszerek.⁴⁴ Ezzel szemben az állam a mai napig a gépi tanulás fantáziátlan felhasználója. A büntető igazságszolgáltatás vezető adaptálónak tűnik.⁴⁵ Más szabályozási területek egyelőre lemaradásban vannak. Néhány államban a jóléti bürokráciák elkezdtek algoritmikus eszközöket alkalmazni a kedvezményezettek közötti válogatáshoz.⁴⁶ Ugyanez igaz szövetségi szinten is. A Környezetvédelmi Ügynökség például gépi tanulást használ egyes mérgek hatásainak értékelésére.⁴⁷ Az Internal Revenue Service gépi tanulási eszközt használ a csalások és visszaélések előrejelzésére.⁴⁸ Mivel ezek az átvételek inkább kivételnek, mint szabálynak tűnnek, a gépi tanulás állami felhasználása élesebb vizsgálati célpontot jelent. Harmadszor, ahol az állam algoritmikus eszközöket használ, az gyakran a szabadság, a mozgásszabadság, a testi épség vagy az alapjövedelem megvonásával jár. Ennek megfelelően ezek a gépi döntések vitathatatlanul a legnyomósabb körülményeket jelentik az emberi döntéshez való jog elfogadásához. Azzal, hogy a vizsgálatot arra a területre összpontosítom, ahol az emberi döntéshez való jog prima facie a legnyomósabb, a leghatékonyabban tudom tesztelni az e jog mellett és ellen szóló érveket.

A cikk három lépésben halad. Az I. felsorolja azokat a módokat, amelyekkel a jog az emberi döntéshez való jogot megalkotta vagy megalkothatta. Ez a rendszertani vállalkozás azt mutatja, hogy egy ilyen jog messze nem hipotetikus vagy fantáziadús. A II. rész meghatározza a figyelembe veendő számítástechnikai eszközök osztályát, és megvizsgálja, hogy az ilyen eszközök mennyiben különböznek az emberi döntéstől. Míg ez a rész

⁴⁰ Lásd pl. David Segal, *The Dirty Secrets of Search*, N.Y. TIMES, **Id.** 2011BU1.

⁴¹ Lásd Falguni Desai, *The Age of Artificial Intelligence in Fintech*, **June** 30, 2016, 10: 42PM, <http://www.forbes.com/sites/falgunidesai///the-age-of-artificial-intelligence-in-fintech20160630> [<https://perma.cc/EK-DDY893>] (leírja, hogyan használják a fintech cégek a mesterséges intelligenciát a befektetési stratégiák javítására és a fogyasztói pénzügyi tevékenységek elemzésére).

⁴² Lásd pl. Benjamin Letham, Cynthia Rudin, ~~Tim~~ McCormick, & David Madigan, *Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model*, ANNALS9 APP. STAT. 1350 (2015).

⁴³ Lásd pl. Paul Raccuglia et al., *Machine-learning-assisted materials discovery ~~guided~~ experiments*, NATURE (533732016) (új vanádiumvegyületek azonosítása).

⁴⁴ LeCun et al., *Supra* note at 33,438.

⁴⁵ Lásd Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, DUKE68 L.J10432019) (példák gyűjtése); lásd még Andrew Guthrie Ferguson, *Policing Predictive Policing*, WASH94. U. L. REV. 1109, 1122-44 (2017) (a prediktív rendőri eszközök gondos katalógusát nyújtja). Az ilyen felhasználások mostanában a bevándorlás területére is kiterjednek. Spencer Woodman, *Palantir Provides the Engine for Donald Trump's Deportation Machine*, INTERCEPT (Mar. 2, 2017, 1: 18PM), <https://theintercept.com///palantir-provides-the-engine-for-donald-trumps-deportation-machine/20170302> (beszámoló arról, hogy a Belbiztonsági Minisztérium ("DHS") egy magánvállalkozónak 41millió dolláros szerződést ítelt oda egy "Investigative Case Management" rendszer létrehozására, amely lehetővé teszi a DHS számára, hogy "hozzáférjen egy hatalmas "ökoszisztéma" adataihoz, hogy megkönnyítse a bevándorlási tisztviselők számára mind a célpontok felfedezését, mind pedig az ellenük indított ügyek létrehozását és adminisztrációját").

⁴⁶ Lásd EUBANKS, *fenti* jegyzet, 8,14-38. o.

⁴⁷ Lásd U.S. Env'tl. Prot. Agency, *ToxCast Fact Sheet* <http://www.epa.gov/pdfs/2012/documents/toxcast-fact-sheet.pdf>.

⁴⁸ David DeBarr & Maury Harwood, *Relational Mining for Compliance*, presented at the Internal Revenue Service Research Conference (2004), <http://www.irs.gov/pub/irs-soi/debarr04.pdf>. Az élelmiszerbiztonság területén tett hasonló erőfeszítésekről lásd: *Commissioners' Fellowship Program: Final Report Abstracts*, U.S. Food & Drug Admin., <https://www.>

fda. [gov/AboutFDA/WorkingatFDA/FellowshipInternshipGraduateFacultyProgr/CommissionersFellowshipProgram/ucm413253. htm](http://www.fda.gov/AboutFDA/WorkingatFDA/FellowshipInternshipGraduateFacultyProgr/CommissionersFellowshipProgram/ucm413253.htm).

a leírás alapján azt is hogy a technológiai kontextus hogyan alakítja vagy zárja ki a jog bizonyos megfogalmazásait. III. rész ezután egy ilyen jog lehetséges normatív alapjaival foglalkozik. Ezen indokok gondos taxonómiáját nyújtja. Ezután megmutatja, hogy miért nem elégségesek mind. Végül egy rövid következtetés megfordítja a dolgot elemzési szempontból, hogy megvizsgálja, elképzelhető-e a jól kalibrált gépi döntéshozó jog, sőt, meggyőzőbb feltételekkel védhető-e, mint az emberi döntéshozó jog.

I. Az emberi döntéshozó jog jogi megfogalmazása

Ez a rész azt dokumentálja, hogy az emberi döntéshozó jogot az emberi döntéshozó jog. A "jog" kifejezést itt tágan használom, hogy az amerikai jogon túl az európai jogra is kiterjedjen, és mind a magánjog, mind a közjog területére kiterjedjen, megragadva az állami és nem állami cselekvés szabályozását. Ezt a széles látószögű szemléletet alkalmazom annak érdekében, hogy megérthessük a feltételezett jog számos aspektusát: az okokat, amelyek miatt megfogalmazódik; a kontextusokat, amelyekben alkalmazzák; és a korlátokat, amelyekkel védik. Ez a vizsgálat nagyrészt leíró jellegű. A jelenlegi jogi környezet áttekintésével "konceptuális bizonyítékot" nyújtok arra vonatkozóan, hogy az emberi döntéshozó jog nem annyira idegen gondolat, hogy azt eleve el kelljen vetni. Ugyanakkor a jelenlegi jog homályosságai és korlátai további bizonyítékot szolgáltatnak arra, hogy milyen nehézségek jelennek az egyének ilyen jogának megteremtésére irányuló erőfeszítésekben.

A. Az emberi döntéshozó európai jog

Az európai jog valamilyen formában az emberi döntéshozó joghoz való jogot, mivel már 1978. ennek a jognak eddig nem volt sok gyakorlati jogi hatása, a GDPR hatályba lépése konkrét hatást válthat ki. A 22 GDPR cikkének történelmi előzményei némi fényt vetnek az ilyen jog kialakításának és igazolásának nehézségeire is.

1. Előzmények

Az emberi döntéshozó jog egyik korai előzménye 1978a francia "számítástechnikáról, az aktákról és a szabadságjogokról" szóló törvény.⁴⁹ Ennek 2a törvénynek az eredetileg elfogadott cikke tiltotta az automatizált profilalkotás vagy személyiségvizsgálat hatósági alkalmazását.⁵⁰ Amennyire meg tudom állapítani, ezt az intézkedést soha nem alkalmazták a gépi tanulásra, amely egyébként akkoriban még nem volt általános használatban. Tizenhét évvel később az Európai Parlament és a Tanács kihirdette az adatvédelmi irányelvet.⁵¹ Ez utóbbi nem rendelkezett önálló joghatással az egyénekre nézve, de kötelezte az Európai Unió tagállamait, hogy hozzanak megfelelő jogszabályokat. 15Az irányelv cikke kötelezte a tagállamokat, hogy teremtsenek jogot arra, hogy "ne lehessen olyan döntésnek alávetni, amely joghatással jár ... vagy jelentős mértékben érinti őt, és amely kizárólag bizonyos személyes szempontok értékelésére szolgáló automatizált adatfeldolgozáson alapul".⁵² Az országok végrehajtották ezt

⁴⁹ Loi n°78 -17 du 17 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT&000006068624 dateTexte=19781031>.

⁵⁰ "Aucune décision de justice impliquant pour un comportement humain ne peut avoir pour fondement un traitement automatisé d'informations donnant une définition du profil ou de la personnalité de l'intéressé.". *Id.* 2. Egy

második záradék ugyanezt a szabályt kiterjesztette a közigazgatási határozatokra. *Id.*

⁵¹ 95/16 irányelv.

⁵² *Id.* Cikk. 15(1).

különböző módon, időnként magán- és az állami szereplők is. A francia 1978. törvényhez hasonlóan azonban úgy tűnik, hogy ezeknek az intézkedéseknek a gyakorlatban kevés hatása volt.⁵⁴ Szintén kevés bizonyíték van arra, hogy a 15. cikk adatvédelmi irányelvbe való beillesztését motiváló aggodalmakról van szó.⁵⁵ Ennek megfelelően az irányelv története kevés útmutatót vagy megvilágítást nyújt arra vonatkozóan, hogy miként kell az emberi döntéshozatalhoz való jogot konceptualizálni vagy igazolni.

2. A 22. GDPR cikke

Az 2017. Európai Bizottság határozata az ideje "egy alapvető lépést tenni a polgárok alapvető jogainak megerősítése érdekében a digitális korban".⁵⁶ Ennek eredménye az általános adatvédelmi rendelet ("GDPR") lett. A május óta hatályos 2018. GDPR az európai adatvédelmi és adatbiztonsági szabályok átfogó átalakítása. A korábbi irányelvvel ellentétben közvetlenül a hatálya alá tartozó adatokat kezelő vállalatokra és állami intézményekre hat. Több millió eurós bírságokat is kilátásba helyező büntető rendelkezéseket tartalmaz.⁵⁷ Kilencvenkilenc cikke egyéb témák széles körét is lefedi, és az én értekezésem inkább a vonatkozó nyelvezetre összpontosít, mintsem szinoptikus áttekintést nyújtana.

A GDPR 22. cikkének (1) bekezdése biztosítja a magánjog számára azt a jogot, hogy "ne lehessen olyan, kizárólag automatizált adatkezelésen alapuló döntésnek alávetni őket, ideértve a profilalkotást is, amely rájuk nézve joghatással jár, vagy hasonlóan jelentős mértékben érinti őket".⁵⁸ Az Európai Unió által létrehozott Európai Bizottság adatvédelmi munkacsoportja szerint a 22. cikk (1) bekezdése csak akkor alkalmazandó, ha "a döntéshozatali folyamatban nem vesz részt ember".⁵⁹ Ezzel a megfogalmazással némi feszültséget okozva a munkacsoport iránymutató dokumentuma azt is javasolja, hogy csak az automatizált

⁵³ *Lásd pl.*, Decreto legislativo giugno 30 1996: Codice in Materia di Protezione dei Dati Personali, <http://www.camera.it/parlam/leggi/deleghe/dl03196.htm> (megkülönbözteti a közszereplőket, és abszolút tilalmat ír elő számukra, de jogot teremt a magánszereplők automatizált döntései elleni tiltakozásra).

⁵⁴ *Lásd Lee Bygrave, Minding the Machine: Az EK adatvédelmi irányelv 15. cikke és az automatizált döntéshozatal* COMP17. L. & SEC. REV. 17,21 (2001) (a jogot "kártyavárként" jellemzi). Úgy tűnik, hogy csak egy német határozat érintette ezt a jogot. *Lásd Isak Mendoz & Lee A. Bygrave, The Right to not be Subject to Automated Decisions based on Profiling, in EU INTERNET LAW* n11.36 (Tatiana-Eleni Synodinou et al. eds. 2017) (egy német határozat ismertetése, amely megállapította, hogy a hitelponozási rendszerek kívül esnek a cikk és a nemzeti 15. végrehajtási jogszabályok hatályán).

⁵⁵ Bygrave azt állítja, hogy a cikket "az a lehetőség, hogy ... az automatizálás csökkentheti a személyek szerepét a fontos döntéshozatali folyamatok alakításában". Bygrave, *Supra* note at 54, De 18. ez egy tautológia, nem pedig annak magyarázata, hogy miért kell a gépi döntéseket az emberi döntésekkel szemben előnytelenül kezelni.

⁵⁶ Európai Bizottság, *Questions and Answers Data Reform Package* (2017), [eu/rapid/press-release_MEMO--en171441.pdf](http://europa.eu/rapid/press-release_MEMO--en171441.pdf).

⁵⁷ Jan Philipp Albrecht, *Hogyan fogja a GDPR megváltoztatni a világot*, EUR2. ROT. L. REV. 287,287 (2016) (a szankciórendszer leírása).

⁵⁸ GDPR Art. 22(1).

⁵⁹ cikk 29. Adatvédelmi munkacsoport, *On Automated individual decision-making and Profiling for the purposes of Regulation /2016/679*, at (Feb 20. 6,)2018, <https://iapp.org/resources/article/wp-draft-guidelines-on-29-automated-individual-decision-making-and-profiling/> [a továbbiakban: a munkacsoport iránymutatásai].

határozat kivonná azt a 22. cikk hatálya alól.⁶⁰ A GDPR azonban nem magyarázza meg, hogyan kell értékelni az ilyen felülvizsgálat minőségét.⁶¹ A tilalom által érintett automatizált gépi tanulási eszközök pontos köre továbbra is kérdéses. A 22. cikk (2) bekezdése ebből kiindulva kizárja a "szerződés megkötéséhez vagy teljesítéséhez szükséges"; az "uniós vagy tagállami jog által egyébként engedélyezett"; vagy "az érintett kifejezett hozzájárulásán alapuló" adatkezelést.⁶² Úgy tűnik, hogy a mai napig nincs olyan európai állam, amely a 22. cikk (2) bekezdése alapján kivételt fogadott volna el. E kivétel hatálya tehát korlátozottnak tűnik, ⁶³míg a többi kivétel hatálya értelmezésre vár. Attól függően, hogy a "kifejezett hozzájárulás" fogalmát hogyan határozzák meg, a rendelkezés hatálya tágabb vagy szűkebb lehet. Könnyen elképzelhető, hogy a magánszervezetek magától értetődően szereznek ilyen hozzájárulást; az azonban már más kérdés, hogy az állami munkáltatók vagy a jóléti szervek megtehetik-e ugyanezt. Ennek megfelelően nehéz megjósolni a 22. cikk (1) bekezdésében foglalt jog pontos hatályát vagy gyakorlati hatását.⁶⁴

A cikk 22 nem az egyetlen olyan GDPR-nak, amely az adatkezeléssel szembeni jogként értelmezhető. A cikk lehetővé teszi a természetes személyek számára, hogy "korlátozásokat érjenek el" a jogellenes vagy pontatlan adatfeldolgozással szemben, ⁶⁵míg a 21. cikk "tiltakozási joga" előírja, hogy a jogalany "ne dolgozza fel többé" egy személy adatait, ha "kényszerítő erejű jogos okokra" hivatkozik.⁶⁶ Tekintettel az "adatfeldolgozás" és a "profilalkotás" tág fogalom meghatározására, ⁶⁸ezek az egyéb rendelkezések valóban széleskörűek lehetnek. A nemzeti kormányok általi végrehajtás, illetve a magán-, nemzeti és magánszféra által benyújtott végrehajtási keresetek hiánya miatt,

⁶⁰ *Id.* (kizárva a 22. cikk (1) bekezdéséből azokat az eseteket, ha ember "a végső döntés meghozatalakor más tényezőket vizsgál és vesz figyelembe").

⁶¹ Nem világos, hogy mit értelmessé. Lásd Michael Veale & Lilian Edwards, *Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling*, COMP34. L. & SEC. REV. 398, 401(2018) ("Az, hogy a "kizárólagosság" e kibővített fogalmát hogyan lehetne gyakorlatilag értékelni az adatkezelő vagy az érintett szempontjából, az egyik jelentős szürke terület, amelyet az iránymutatás maga után hagy."). Ezen túlmenően, ha egy automatizált folyamat nem változtatja meg a "jogokat", vagy nem jár "egyenértékű vagy hasonlóan jelentős" hatással, a munkacsoport szerint az nem tartozik a 22. cikk (1) bekezdésének hatálya alá. A munkacsoport iránymutatásai, fenti megjegyzés, 59,21-22. o. (megjegyezve, hogy bár a célzott reklám jellemzően nem tartozik a 22. cikk (1) bekezdésének hatálya alá, a célzott reklámozás "tolakodó jellege", az egyén elvárásai és az üzemeltetőnek a személy "sebezhető pontjairól" való tudása miatt a 22. cikk (1) bekezdésének hatálya alá tartozhat).

⁶² GDPR 22. cikk (2) bekezdés.

⁶³ A GDPR a hozzájárulást is nagyon szigorúan meghatározott módon határozza meg.

⁶⁴ A GDPR előzményét, az első irányelvet az európai közjogban alkalmazott arányosság elvének fényében értelmezték. Lásd Charlotte Bagger Tranberg, *Proportionality and data protection in the case law of the European Court of Justice* INT'L DATA PRIVACY L. 239-40239, (2011) (az esetjog összefoglalása). Az arányossági felülvizsgálat GDPR-ra való alkalmazásának bizonytalansága tovább nehezíti a törvény útjának előreljelését.

⁶⁵ GDPR Art. 18.

⁶⁶ *Id.* Cikk. 21(1).

⁶⁷ A vonatkozó rész meghatározása szerint ez nem olyan műveletet vagy műveletsort, amelyet személyes adatokon vagy személyes adatok csoportjain végeznek, akár automatizált eszközökkel, akár nem". *Id.* Art. 4(2).

⁶⁸ Ez a meghatározás szerint magában foglalja "a személyes adatok bármely formáját, amely a személyes adatok felhasználásából áll egy természetes személyre vonatkozó bizonyos személyes szempontok értékelésére, különösen az adott természetes személy munkahelyi teljesítményére, gazdasági helyzetére, egészségére, személyes

preferenciáira, érdeklődésére, megbízhatóságára, viselkedésére, tartózkodási helyére vagy mozgására vonatkozó szempontok elemzésére vagy előrejelzésére". *Id. Art. 4(4).*

vagy szupranacionális hatóságoknak köszönhetően jelentős bizonytalanság van azzal kapcsolatban, hogy mit jelentenek ezek a rendelkezések, nem is beszélve arról, hogy megfelelően értelmezik-e őket úgy, hogy tartalmazzák az emberi döntéshez való jogot.

A GDPR legfőbb eleme, ameddig a figyelmet az algoritmikus döntések magyarázatához való potenciális jog, amely a dokumentum máshol található.⁶⁹ Az emberi döntéshez való jog, akár a 22. cikk (1) bekezdésében, akár máshol van rögzítve, eddig ellentétesen kevés tudományos figyelmet kapott.⁷⁰

B. Az amerikai jog és az emberi döntéshez való jog

Az amerikai bíró pontos analógja a GDPR-nak. Ez a szakasz három olyan jogi területet vizsgál fel, amelyeken utalások észlelhetők. Ezek összevonása feltárja az emberi döntésekhez való jog bizonytalan árnyékát, amely a szövetségi és az állami jogszabályok között húzódik meg.

Először is, az alkotmányjog mind a szövetségi szinten már most is egyéni jogokat teremt a döntéshozatal olyan formáihoz, amelyek nem egyeztethetők össze a gépi tanulás legalábbis bizonyos formáival. Ezek közül talán a legjelentősebb az, hogy az Alkotmány különböző jogokat teremt az esküdtszék döntéséhez, amelyek közül a legjelentősebb az esküdtszéki tárgyaláshoz való hatodik módosítás szerinti jog büntetőügyekben.⁷¹ Ha most figyelmen kívül hagyjuk a vádalkuk mindenütt jelenlévő elterjedtségét, az Alkotmányban az esküdtszék döntésére való jogosultság az emberi döntéshez való jog egyik fajtája. Az eredeti alkotmány rendszerében az esküdtszék beágyazása az állami hatalom gyakorlásának kritikus helyein a demokrácia iránti elkötelezettséget fejezte ki.⁷² Az állampolgár esküdként való kötelezettsége szükségszerűen "aktív kötelezettség" volt, amely feltételezte az emberi ítélőképesség autonóm gyakorlását a potenciálisan zsarnoki államhatalommal szemben.⁷³ Ha a Hatodik Alkotmánymódosítást sérti az, hogy az emberi polgárok testülete helyett bíró jár el, akkor nehéz elképzelni, hogy a büntetőeljárásban a döntő döntő szerepet hogyan játszhatná el egy gép. Tehát az esküdtszéki tárgyaláshoz való jog a Hatodik Alkotmánymódosítás és a párhuzamos alkotmányos rendelkezések értelmében, amelyek állítólagosan biztosítják az esküdtszékhez való hozzáférést, valószínűleg magában foglalja az emberi döntéshez való jogot. És egy gépi tanulási eszköz eredményének felhasználása egy vádlott büntetésének a törvényes maximumon túli megemelésére sérti a vádlottnak a Hatodik Alkotmánymódosítás azon jogát, hogy esküdtszék döntsön a tényekről, ha a tények, amelyek alapján

⁶⁹ Vita folyik arról, hogy a GDPR-t úgy kell-e értelmezni, hogy az ilyen jogot teremtse. Vö. Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*, INT'L DATA PRIVACY L. (762017) (érvel a magyarázathoz való jog következtetése ellen) és Andrew Selbst & Julia Powles, *Meaningful information and the right to explanation*, INT'L DATA PRIVACY L. (233,2342017) (a jog "pozitív felfogását" ajánlja).

⁷⁰ Lásd Veale & Edwards, *Supra* note at 61,400-01 (jelenlegi megfogalmazás); lásd még Meg Leta Jones, *The right to a human in the loop*: SOC47. STUD. SCIENCE (216,2242017) (a 22. cikk létezését megjelölve).

⁷¹ U.S. Const. art. III § cl2., 3 ("The Bill Crimes ... shall be by Jury"); VI. módosítás (a "pártatlan esküdtszékhez" való jog); VII. módosítás (az "esküdtszéki tárgyaláshoz" való jog bizonyos polgári ügyekben); lásd még *Apprendi v. New Jersey*, U.S.530. (466,4772000) (megerősítve a Hatodik Alkotmánymódosítás esküdtszéki tárgyaláshoz való jogát a bírósági ténymegállapítással szemben).

⁷² Lásd Akhil Reed Amar, *The Bill of Rights as a Constitution*, YALE100 L.J.(1831991); lásd még Mark DeWolfe Howe, *Juries as Judges of Criminal Law*, HARV52. L. REV. 582, 584-85 (1939) (hangsúlyozva az esküdtek széles körű döntési jogkörét mind ténybeli, mind jogi kérdésekben az alapító korszakban).

⁷³ Jenny E. Carroll, *Nullification As Law*, GEO102. L.J. (2014) lásd még JEFFREY ABRAMSON, WE, THE JURY: THE JURY SYSTEM AND THE IDEAL DEMOCRACY 22-24, 68-75 (1994) (az esküdszék mint politikai intézmény kialakulásának történetét tárgyalja a gyarmati Amerikában).

hogy a kimenő teljesítményt az esküdtszék nem ~~ha~~ Ezen túlmenően a hatodik módosítás a konfrontációs klauzula formájában további súrlódást (ha nem is tilalmat) jelenthet a gépi döntésekre.⁷⁵

Másodszor, a tisztességes eljárás eszméje is lehet alapja egy emberi, nem pedig egy gépi ítéletnek. Az eljárásjogi tisztességes eljárás eszméje alapvetően "értésítést és valamilyen meghallgatást" feltételez.⁷⁶ Van némi vita a meghallgatás időzítéséről és tartalmáról, legalábbis ami az Alkotmány Due Process garanciáját illeti.⁷⁷ De nem nehéz belátni, hogy felmerülhet a kérdés, hogy a megfelelő eljárást egy gépi határozat biztosítja-e. Valójában vitathatóan nehéz értelmet adni a "meghallgatás" fogalmának, ha nincs olyan természetes személy, aki vagy fizikailag jelen van a szóbeli érvelésnél, vagy aki elolvassa és értékeli az írásos beadványokat.⁷⁸

Eddig még nem került sor a hatodik módosítás vagy a tisztességes eljárás alapján az algoritmikus eszközök frontális megtámadására. Ez talán azért van így, mert az ilyen eszközöket eddig inkább az emberi döntéshozatal támogatására használták, mintsem annak formális kiszorítására. A Wisconsin Legfelsőbb Bíróság azonban a *State v. Loomis* ügyben hozott határozatában 2016 megoldotta a büntetőeljárással kapcsolatos kihívást, amely egy büntetőjogi vádlottnak az ítélelhozatali algoritmus kritériumainak (nem pedig magának az algoritmusnak a használatának) a megtámadásán alapult.⁷⁹ Ez a kifogás részben a vádlott korlátozott lehetőségén alapult, hogy megtámadhassa az algoritmus feltételeit, részben pedig azon (inkább általános, mint személyre szabott) adatokon, amelyekre az algoritmus támaszkodott az ajánlása elkészítéséhez.⁸⁰ A wisconsini bíróság elutasította az alperes alkotmányos érveit. Azzal érvelt, hogy az algoritmus csak nyilvánosan hozzáférhető adatokat használt fel, valamint olyan adatokat, amelyeket az alperes szolgáltatott, és hogy az alperes tagadhatta vagy magyarázhatta volna az előrejelzésének kialakításához felhasznált bármely információt.⁸¹ A Bíróság rámutatott, hogy az említett ügyben az ítélelhozatali algoritmus képzéséhez használt adatok inkább csoportalapúak, mint egyénre szabottak, de elutasította, hogy alkotmányellenesnek tekintse azt.

⁷⁴ Booker kontra Egyesült Államok, USA543 (2005). Az úgynevezett bizonyítékokon alapuló eszközöket már széles körben használják az ítélelhozatal során, bár szigorú irányítás és biztosítékok nélkül. Erin Collins, *Punishing Risk*, GEO107. L.J. (57,662018) (leírva a biztosításmatematikai alapú büntetészkiszabási eszközök alkalmazását, gyakran "arra vonatkozó iránymutatás nélkül, hogy milyen célokra használhatók [ezek]"). Általánosságban úgy tűnik, hogy azért alkalmazzák őket, hogy a hatodik módosítás által meghatározott kereteken belül irányítsák a bírák mérlegelési jogkörének gyakorlását a büntetés hossza tekintetében. John Monahan & Jennifer Skeem, *Risk Assessment in Criminal Sentencing*, ANNU12. REV. CLIN. PSYCH. 489, 500-01 (2016).

⁷⁵ Andrea Roth, *Machine Testimony*, YALE126 L. J. 972, 973.

⁷⁶ Richard H. Fallon, Jr., *Some Confusions About Due Process, Judicial Review, and Constitutional Remedies*, COLUM93. L. REV. 309,330 (1993). Ennek a gondolatnak az osztályos kifejtéséhez: Henry J. Friendly, "Some Kind of Hearing", 123U. PA. L. REV. 1267 (1975) (a tisztességes tárgyalás jellemző elemeinek tárgyalása és azok relatív fontosságának értékelése).

⁷⁷ *Vö.* United States v. Florida East Coast Ry. Co., 410 U.S. (1973)224,239 ("A "meghallgatás" kifejezésnek jogi kontextusában kétségtelenül számos jelentése van.").

⁷⁸ Friendly, *Supra* note at76, ("Bár1270 meghallgatás" szóbeli jelentéssel bír, nem látom okát annak, hogy bizonyos körülmények között miért ne lehetne "meghallgatást" tartani kizárólag írásbeli anyagok alapján.").

⁷⁹ State v. Loomis, N.W 881.2d (7492016); lásd még Huq, *supra* note at 45,-- (a Wisconsinban használt büntetészkiszabási eszköz tárgyalása).

⁸⁰ Ahogy az alperes szakértő tanúja, Thompson kifejtette: "A bíróság nem tudja, hogy a COMPAS hogyan hasonlítja össze az adott egyén előéletét azzal a populációval, amellyel összehasonlítja. A bíróság még azt sem tudja, hogy ez a populáció egy wisconsini, egy New York-i vagy egy kaliforniai populáció-e. Ez az egész

olyan információkkal, amelyekkel a bíróság nem ~~is~~mit mi teszünk, az az, hogy félretájékoztatjuk a bíróságot, amikor eléjük tesszük ezeket a grafikonokat, és hagyjuk, hogy ezeket használják az ítélethozatalhoz." *Loomis*, N881.W.2d at 756-57.

⁸¹ *Id.* 761-62.

tekintettel arra, hogy ez csak egy volt a vádlott büntetésének számos közül.⁸² Az ilyen adatok, figyelmeztetett a Bíróság, "nem tekinthetők meghatározó tényezőnek annak eldöntése során, hogy az elkövető biztonságosan és hatékonyan felügyelhető-e a közösségben", összhangban a tisztességes eljárással.⁸³ Továbbá, figyelmeztetett a Bíróság, a gépi előrejelzés "nem vehető figyelembe meghatározó tényezőként annak eldöntése során, hogy az elkövető biztonságosan és hatékonyan felügyelhető-e a közösségben" a tisztességes eljárásnak megfelelően.⁸⁴

Végül áprilisban két 2019. demokrata törvényjavaslatot terjesztett elő, amely előírja a Szövetségi Kereskedelmi Bizottság számára, hogy hozzon olyan szabályozást, amely előírja az "automatizált döntéshozatali rendszerek hatásvizsgálatát".⁸⁵ Bizonyos algoritmusok esetében az értékelésre "a bevezetés előtt" van szükség.⁸⁶ A Szövetségi Kereskedelmi Bizottságnak (FTC) rendeleteket kell kihirdetnie ennek a követelménynek az érvényesítésére, és mind az FTC, mind a különböző államok főügyészei érvényesíthetik azt.⁸⁷ Bár ez nem tartalmaz egyéni kereseti jogot, ez a törvényjavaslat lehetővé teheti az algoritmusok elleni pereket, amelyek kizárhatják azok használatát.

Az amerikai jog, összefoglalva, nem teremt jogot az emberi döntéshez ennyi szóval. Inkább az esküdtszéki tárgyaláshoz való jog alkotmányos védelmének váratlan következményeiként jelenik meg egy ilyen jog. Amennyiben az amerikai szabályozók követik a GDPR példáját, az ilyen jog alapjai léteznek.

C. Egy újszerű jog kísérleti formája

Túl korai lenne arra a következtetésre jutni, hogy az európai vagy az amerikai jogban gyakorlati szempontból létezik az emberi döntéshez való jog. Ugyanilyen korai lenne tagadni, hogy egy ilyen jog mind a büntetőjogi, mind a polgári jogi szabályozás területén kezd megvetni a lábát. Az általános adatvédelmi rendelet 22. cikkének (1) bekezdése valószínűleg az ilyen jog zsengéje lesz. E jog valamilyen változata azonban azóta is létezik a nemzeti vagy szupranacionális európai jogban, anélkül, hogy 1978, sok gyakorlati hatása lenne. A 2000-es évek eleje óta a gépi tanulás, és különösen a mélytanulási eszközök kapacitásában bekövetkezett technikai fejlődés azonban⁸⁸ új terhet róhat erre a status quo-ra. Az új félelmek váratlan szabályozási szempontok mentén konfliktust válthatnak ki

⁸² *Id.* at (765)[A] tisztességes eljárás következményei bennünket, hogy figyelmeztessük a körzeti bíróságokat, hogy mivel a COMPAS kockázatértékelési pontszámok csoportos adatokon alapulnak, képesek a magas kockázatú elkövetők csoportjait azonosítani - nem pedig egy adott magas kockázatú egyént."

⁸³ *Az Id.* at külön760., az ötödik kiegészítésnek az önvádra kötelezés elleni kiválthatják azok a kihallgatások, amelyek célja, hogy a vádlottból információkat csikarjanak ki abból a célból, hogy hosszabb büntetéssel járó algoritmikus besorolást kapjon. Cassie Deskus, Note, *Fifth Amendment Limitations on Criminal Algorithmic Decision-Making*, N 21.Y.U. J. LEGIS. & PUB. POL'Y 259-66237, (2018). Ez az alkotmányos kérdés az emberi döntéshez való jog kiegészítéseként nem jól értelmezhető, ezért itt félretersem.

⁸⁴ *State v. Loomis*, N.W881.2d 749,760

⁸⁵ Lásd Algoritmikus Elszámoltathatóság Törvény, S. --, §3(1)(A), <https://www.gov/imo/media/doc/Algorithmic%Accountability%Act%of%-%-%Bill%Te20202020192020> xt. pdf.

⁸⁶ *Id.* §3(1)(A)(ii).

⁸⁷ Id. § (a)(1) és (d).

⁸⁸ *Lásd a fenti szöveg kísérő megjegyzés*113cit 115.

árrés.⁸⁹ Az ebből következő jogi kihívás megújítására.⁹⁰ A gépi döntések tiltó szabályozása valószínűleg alázza a gépi tanulás valamilyen formáját alkalmazó gazdasági egységek üzleti stratégiáját.⁹¹ A gépi tanulásból való kilépés joga által nem akadályozott, az adatok extenzív felhasználását támogató, előreláthatóan erős lobbik mind a jogalkotási, mind a szabályozási vitákban az adatvédelem szószólóinak erőteljes ellenfelei lesznek. Ugyanakkor ezek a lobbik a GDPR-hez és a CCPA-hoz hasonló szabályozásban a megfeleléssel kapcsolatos előnyös méretgazdaságosságot érzékelhetnek. Ahelyett, hogy egy balkanizált szabályozási terep kialakulását kockáztatnák, ezek az érdekcsoportok talán elfogadják valamiféle emberi döntéshozatalhoz való jogot, mint kisebbik rosszat. Nyilvánvaló, hogy a büntetőjogi térben egészen más politikai gazdaságtan jellemző, ami eltérő szabályozási utakat eredményez, eltérő technológiai egyensúlyokkal.

E politikai-gazdasági újabb jogi fejleményekből levonható általános következtetés az emberi döntéshozatalhoz való jog formájára és erejére vonatkozóan. Itt kettőt érdemes kiemelni.

Először is, az emberi döntéshozatalhoz való jog szóbeli megfogalmazása igen változatos formában van az általános adatvédelmi rendelet 22. cikkének (1) bekezdésében foglalt kifejezett kötelezettségvállalástól az esküdttszék (emberi) döntésére vonatkozó kötelezettségvállalásig a hatodik módosításban. Alternatív megoldásként a nem emberi döntéshozatalhoz való kötelezettség minimalizálható az adatáramlás pusztán mennyiségének szabályozásával is, például az adatmegosztásról való egyéni lemondás alkalmazásával.⁹² A jogi formák ezen változatossága azt jelenti, hogy nem világos, hogy az emberi döntéshozatalhoz való jogot (amennyiben az kívánatosnak bizonyulna) hogyan lehetne a legjobban megfogalmazni. Az emberi döntéshozatalhoz való jog korábbi, az európai jogban létező változatainak történelmileg elmaradott volta, valamint a GDPR cikkének továbbra is fennálló kétértelmősége mind arra utal, hogy egy ilyen jellegű, jogilag érvényesíthető jog létrehozása jelentős működési nehézségekkel járna.

A második következtetés a következőre vonatkozik. Úgy tűnik, hogy az amerikai és az európai jog eltérő prioritásokat követ az emberi döntéshozatalhoz való jog tekintetében. Az Egyesült Államokban ez az elv a büntetőjogi ügyekben bír a legnagyobb befolyással. Ezzel szemben a GDPR kiveszi a bűnözéssel és biztonsággal kapcsolatos funkciókat a hatálya alól. Valójában a nem állami szereplők által végzett adatfeldolgozásra összpontosít. Talán ez az eltérés úgy is felfogható, mint az európai aggodalom újabb példája az alábbiakkal kapcsolatban.

⁸⁹ De nem minden félelem alaptalan. Edward Geist és Andrew J. Lohn, Rand Institute, *How Many Artificial Intelligence Affect the Risk of Nuclear War?* 2 (2018), <https://www.rand.org/pubs/perspectives/PE296.html> ("A mesterséges intelligencia még a technikai fejlődés szerény üteme mellett is 2040 képes lehet arra, hogy a nukleáris stratégiai stabilitás előtt álló kihívásokat súlyosbítsa. ").

⁹⁰ Ez nyilvánvalóan nem a technológiai változások előrejelzésének miatt van így. Vö. Jon ELSTER, EXPLAINING TECHNICAL CHANGE: A CASE STUDY IN THE PHILOSOPHY OF SCIENCE 7-9 (1983) (a "technikai változás" definíciója: "az eszközök gyártása és módosítása", és az ilyen változások lehetséges különböző útjainak megvitatása). Az emberi döntéshozatalhoz való jog magában foglalja a technológiai változás elutasításának lehetőségét. A technológiai alkalmazkodásban és fejlődésben azonban nincs semmi kényszerítő, és így nincs ok arra, hogy a technológiákról ne lehessen lemondani.

⁹¹ MIT Technology Review Insights, *Gépi tanulás: MIT TECH. REV.*, március 16, 2017, <https://www.technologyreview.com/s//machine-learning-the-new-proving-ground-for-603872-competitive-advantage/> (a vállalkozások körében végzett felmérés szerint a vállalkozások egy százaléka 60 már alkalmazza a gépi tanulást valamilyen módon, és csak egy százalékuknak 5 nem áll szándékában ezt tenni a jövőben).

⁹² Lásd pl. Ch 375, "55. Törvény a Polgári 3. Törvénykönyv magánéletre vonatkozó 4. részének (az 1798.100. szakasszal 1.81.5 kezdődő) címével való kiegészítéséről," június Meg kell vizsgálni, hogy az ilyen opt-out a feldolgozásból

való kilépés hatásos-e. Ez a törvényjavaslat a továbbiakban 28,2018.is érvényes. Az online magánélet védelmét szolgáló, hozzájáruláson alapuló stratégiák eredménytelensége nem sok jót ígér az ilyen jellegű, személyre szabott megközelítés számára.

"tisztelet és személyes méltóság", ellentétben a "szabadság értékeire" koncentrázó amerikaiakkal.⁹³ A 22. cikk etiológiájának homályossága a francia 1978 adatvédelmi törvényben és az európai 1994 irányelvben azonban a magabiztos diagnózis ellen szól. Bármilyen legyen is az eredete, a területek közötti különbség nettó hatása az, hogy az emberi döntéshez való jog a két kontinensen nagyon eltérő körülmények között kerül először tesztelésre, ami talán egészen eltérő jogi rendszerekhez vezet.

Az emberi döntéshez való amerikai és európai formáját egy közös vonás köti össze. A jogalkotók vagy a bírúk egyik kontextusban sem dolgoztak ki szilárd elméleti magyarázatot arra vonatkozóan, hogy az egyes technológiákat és az azokkal járó következtetési érvelés sajátos módozatait miért kell kifogásolhatónak tekinteni. Azt sem mondták meg, hogy miért kellene előnyben részesíteni az emberi döntést. Egy ilyen indoklás nem vezethető le mechanikusan az emberi és a gépi döntések közötti különbségek empirikus számbavételéből. Végül pedig az általános adatvédelmi rendelet (vagy annak elődei), illetve a CCPA sem tartalmaz szilárd elméleti magyarázatot a jog indoklásáról. Következésképpen jelenleg elméleti szakadék tátong a kialakulóban lévő jog és annak empirikusan és normatív szempontból meggyőző igazolása között. E cikk ezt a hiányt vizsgálja.

II. A gépi és az emberi döntések közötti különbség

Ahhoz, hogy megértsük az emberi döntéshez való jog gondolatát, és hogy normatív szempontból miért lehet vonzó, világosan meg kell értenünk, hogy milyen nem emberi (gépi) döntésekről van szó; miben különböznek az emberi döntéshozatali folyamatoktól; és hogy a gyakorlatban az emberi és a gépi döntések hogyan különülnek el vagy keverednek egymással, hogy funkcionálisan elválaszthatók vagy elválaszthatatlanok legyenek. E célból ez a rész a vonatkozó technológiák rövid és nem technikai jellegű bemutatását tartalmazza. Különösen a gépi tanulási eszközök (néha mesterséges intelligenciának⁹⁴ is nevezik) azon csoportjára összpontosít, amelyek a legközvetlenebbül ígéretesek az emberi döntések kiszorítására. E technikai különbségek normatív következményeit a III. rész részletesebb normatív vizsgálatának előzményeként elemzem. Ennek a kontextualizáló beszámolóknak közvetlen normatív hozadéka van: segít tisztázni az emberi döntéshez való jog technikai plauzibilitásának feltételeit. Azt javaslom, hogy ezt a jogot leginkább a gépi döntések utólagos felülvizsgálatának követeléseként lehet értelmezni. E rész beszámolója egyben a III. rész bővebb normatív vizsgálatát is előkészíti.

A. A gépi tanulás mint az emberi döntések helyettesítője

A gépi tanulás algoritmusai legáltalánosabb megfogalmazásban egy olyan "tanulási problémát old meg, amely valamilyen feladat végrehajtása során a teljesítmény bizonyos mértékének ...javítására irányul, valamilyen képzési tapasztalaton keresztül".⁹⁵ Az ilyen algoritmusoknak két formája van: felügyelt és felügyelet nélküli.⁹⁶ A felügyelt gépi tanulási algoritmusok általában egy $f(x)$ függvény definiálására ösztönöznek, amely egy

⁹³ James Q. Whitman, *A magánélet két nyugati kultúrája. Dignity Versus Liberty*, **N.J.** (1151,20041161).

⁹⁴ Azért kerülöm ezt a kifejezést, mert a félreérthetőbb a hatálya. Vö. STUART RUSSELL & PETER NORVIG, *ARTIFICIAL INTELLIGENCE: A MODERN APPROACH* 2-14 (3d ed. 2013) (a mesterséges intelligencia egy sor alternatív definícióját kínálja, amelyek a racionális mellett az emberi gondolkodást és cselekvést is magukban foglalják).

⁹⁵ M.I. Jordan & T.M. Mitchell, *Gépi tanulás*: SCIENCE349 (2015)255,255

⁹⁶ LeCun et al., *Supra* note at 33, (felügyelt eszközök viszonylag nagyobb gyakoriságát 436,442⁹⁶ élküli eszközökkel szemben).

Kimenetei⁹⁷ tehát az x -nek az y ⁹⁸ kategóriákba való besorolását jelentik: például a képeket az "arc" és a "nem arc" osztályokba; a gyanúsítottakat a "veszélyes" és a "nem veszélyes" osztályokba; vagy a vásárlókat az "impulzusvásárlók" és a "nem impulzusvásárlók" osztályokba. Ezek az osztályozások korrelációs és nem oksági jellegűek. Egy algoritmus teljesítményét azon mérik, hogy mennyire jól ragadja meg x és y kapcsolatának erősségét, nem pedig azon, hogy képes-e felismerni x és y tényleges oksági kapcsolatát. E tekintetben ⁹⁹a sikere attól függ, hogy milyen mértékben képes "optimalizálni egy teljesítménykritériumot példaadatok vagy múltbeli tapasztalatok felhasználásával".¹⁰⁰ Ez a feladat alapvetően analóg azzal a funkcióval, amelyet az olyan ismert eszközök, mint a közönséges legkisebb négyzetek és a logisztikus regresszióelemzés végeznek.¹⁰¹ A gépi tanulás eszközei azonban jellemzően nagyságrendekkel felülmúlják ezeket az eszközöket az előrejelzési pontosság tekintetében; emellett általában kisebb torzítású és kisebb szórású eredményeket generálnak, mint a közönséges regresszió.¹⁰² Ezt azért érik el, mert algoritmusaik dinamikusan frissítik az adatokon belüli kapcsolatok feltérképezésére használt modelleket az új példák megjelenésével. Így "szabályokat tanulnak[]" az adatokból" arról, hogy miként hajtsák végre jobban a feladatukat a végrehajtás során.¹⁰³

A felügyelet nélküli gépi tanulási megközelítést a számítási eszközöket használóval kissé más céllal. Címkezetlen képzési adatokkal kezd, majd a programozó előzetes útmutatása helyett az adatok immánens struktúrája alapján alakítja ki az osztályozásokat.¹⁰⁴ Egy online képekből álló készletet egy felügyelet nélküli algoritmus például tetszőleges számú kategóriába sorolhatja, amelyek közül egyiket sem határozta meg a priori: macskák vs. nem macskák; emberek vs. tárgyak stb. Ezek a kategóriák

⁹⁷ *Id.* Ez a folyamat inkább egy "osztályozó", mint egy függvény leírható, amely "jellemzőértékekkel" rendelkező bemeneteket vizsgál, és egy osztályváltozót ad ki. Pedro Domingos, *Néhány hasznos dolog, amit a gépi tanulásról tudni kell*, COMM. ACM, Oct. at 2012,78-80 ("Az osztályozó olyan rendszer, amely bemenetként (jellemzően) diszkrét és/vagy folytonos jellemzőértékek vektorát adja meg, és egyetlen diszkrét értéket, az osztályt adja ki").

⁹⁸ Az y értékeit a programozónak meg kell határoznia. Comm. on the Analysis of Massive Data et al., *Frontiers in Massive Data Analysis* (1042013), http://www.nap.edu/catalog.php?record_id=18374 (megjegyezve, hogy a felügyelt tanulásban az elemzőnek aktívan meg kell határoznia az érdeklődésre számot tartó változót); PETER FLACH, MACHINE LEARNING: THE ART AND SCIENCE OF ALGORITHMS THAT MAKE SENSE OF DATA (142012) (megjegyezve, hogy a "többosztályos osztályozás" "saját jogú gépi tanulási feladat").

⁹⁹ Jordan & Mitchell, *Supra* note, 95,255-57. o. A teljesítményt a pontosság szempontjából lehet meghatározni, a hamis pozitív és hamis negatív arányok különböző súlyokkal vannak ellátva).

¹⁰⁰ ALPAYDIN, *Supra* note at 3,3.

¹⁰¹ IAN GOODFELLOW, YOSHUA BENGIO & AARON NEUR, MACHINE LEARNING 34-35 (2016); lásd még Coglianese & Lehr, *supra* note at 4, (a párhuzam 1158kidolgozása). Egy tekintetben a párhuzam talán pontatlan. A gépi tanulás egyik úttörője, Leo Breiman ezért szembeállítja az "adatmodellező" megközelítést, amely abból a feltételezésből indul ki, hogy egy sztochasztikus adatmodell írja le a rendelkezésre álló adatokat, majd folytatja annak paramétereinek becslését, és az algoritmikus modellező megközelítést, amely nem tesz feltevést az adatok szerkezetéről, majd keres egy olyan függvényt, amely illeszkedik az adatokhoz. Leo Breiman, *Statistikai modellezés: A két kultúra (a szerző megjegyzéseivel és ellenvéleményével)*, *Stat16. Sci.* (2001)199,199

¹⁰² Jon Kleinberg et al., *Prediction policy problems*, AM105. NOV. 491, 493-94 (2015). Azokról a technikai eszközökről, amelyekkel a gépi tanulás eszközei ezt az előrelépést a közönséges regresszióhoz képest elérik, lásd Esteban Alfaro, Matias Gamez és Noelia Garcia, *Adabag: An R package for classification with boosting and bagging*, J54. STAT. SOFTWARE 1-21, (2013) (a "boosting" és "bagging" technikák működésének leírása).

¹⁰³ Ziad Obermeyer & Ezekiel J. Emanuel, *Predicting the future: data, machine learning, and clinical medicine*, NEW13 ENGLAND J. MED. 1216,1217 (2016); PEDRO DOMINGUES, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 6-7, (232015).

¹⁰⁴ FLACH, *fenti* 14-15. o3,.

úgy képzelhetők el, mint az adatokban olyan példányok klaszterei, amelyek jobban hasonlítanak egymáshoz, mint más példányokhoz. Az algoritmus ezeket a klasztereket több rétegű reprezentáció létrehozásával azonosítja, és minden réteg más absztrakciós szintet képvisel.¹⁰⁵ Az adatokon belül egy adott ábrázolási rétegben azonosított különböző klaszterek határainak iteratív frissítésével az algoritmus növeli a klaszteren belüli hasonlóságot és a klaszterek közötti eltérést.¹⁰⁶ A felügyelet nélküli gépi tanulás célja más szóval az, hogy "megnézzük, mi történik általában, és mi nem".¹⁰⁷

A felügyelt és nem felügyelt általában különböző számítási architektúrákon alapulnak.¹⁰⁸ A felügyelet nélküli tanulási módszerek kiemelkedő formái közé tartozik az asszociációs tanulás, a klaszterelemzés, a főkomponens-elemzés és a többdimenziós skálázás.¹⁰⁹ A felügyelt tanulás számos különböző számítási stratégiára támaszkodik, amelyek közül két gyakorlati érdekes megemlíteni. Először is, néhány felügyelt tanulás úgynevezett "neurális hálózatokat" alkalmaz.¹¹⁰ Ez utóbbiak "neuronok" vagy csomópontok rétegeit alkalmazzák. A bemeneteket a neuronok első rétege kapja, amelyek egy függvényt alkalmaznak, amely átalakítja ezeket a bemeneteket. Az így kapott kimenetek ezután a hálózat további rétegeibe kerülnek, ahol különböző átalakításoknak vetik alá őket, amíg el nem érik a kimeneti réteget.¹¹¹ A neuronok közötti kapcsolatokat folyamatosan újralibrálja a Hebb-féle tanulási algoritmus, amely megerősíti az egyidejűleg aktivált neuronok közötti kapcsolatokat.¹¹² A másik megközelítés a "véletlen erdők" algoritmusok iskolája, amelyek az adatokat leképező döntési fák ezreinek létrehozásával generálnak előrejelzéseket.¹¹³ Minden egyes "fát" a képzési adatok egy véletlenszerű mintáján képeznek ki, és a modell egy olyan előrejelzést ad vissza, amely a többségi előrejelzés a

¹⁰⁵ Geoffrey E. Hinton, *Learning Multiple Layers of Representation*, *IEEE* 428 (2007).

¹⁰⁶ JOHN D. KELLEHER & BRENDAN TIERNEY, *DATA SCIENCE* (2018). Az algoritmus úgy értelmezhető, hogy maximalizálja ezeket a paramétereket.

¹⁰⁷ ALPAYDIN, *Supra* note at 3,11.

¹⁰⁸ Egyes eszközöket mindkettőre használnak. ALPAYDIN, *supra*, at 114 (megjegyezve, hogy a támogató vektor gépeket strukturált és strukturálatlan tanulásra egyaránt használják). A támogató vektorgép (SVM) olyan összefüggések azonosítására szolgál a változók között, amelyek az ilyen adatok grafikus ábrázolásának emberi szemmel történő vizsgálata során nem lennének nyilvánvalók. Lásd Isabelle Guyon, *Data Mining Történelem: The Invention of Támogatás Vektor Gépek*, *KDNuggets* (július 2016), <http://www.kdnuggets.com//guyon-data-mining-history-svm-support-vector-machines201607.html> [<https://perma.cc/N-CRUY459>] (az SVM történetének leírása az algoritmust az 1990s-ban módosító egyik tudós által).

¹⁰⁹ Trevor Hastie, Robert Tibshirani & Jerome H. Friedman, *Unsupervised Learning*, in *THE ELEMENTS OF STATISTICAL LEARNING* (4852009). A klaszterezés során például "egy fa struktúrát hozunk létre különböző szemcseméretű klaszterekkel, és a fán feljebb lévő klasztereket kisebb klaszterekre osztjuk", hogy "korábban ismeretlen struktúrát találjunk az adatokban". ALPAYDIN, *Supra* note at 3,117.

¹¹⁰ Az alábbiakban ALPAYDIN, *supra* note, 3,88-103. o., és Bengio, *Supra* note, 88-103. o.33., világos beszámolójára támaszkodunk. 48-49.

¹¹¹ For a useful graphical representation, ALPAYDIN, *supra* note at 33,49.

¹¹² KELLEHER & TIERNEY, *supra* note at 106, (2018). Neurális hálózatok általában foglalja a helyes súlyok megtalálását.

a hálózatban belüli kapcsolatokra."). A semleges hálózat képzésének eszköze a backpropagation algoritmusnak nevezett algoritmus. Ez úgy működik, hogy véletlenszerű súlyokat rendel a hálózat kapcsolataihoz, majd ezeket a súlyokat minden egyes alkalommal frissíti, amikor egy képzési példányra kerül sor. Azért kapta ezt a nevet, mert az algoritmus a kimeneti rétegből a bemeneti rétegbe továbbítja (vagy visszaterjeszti) a hibákat. *Id.* 129-30; James Somers, *Is AI Riding a One-Trick Pony?*, *MIT 120TECH. REV.* 29,31 (2017) (a backpropagation nem technikai jellegű bemutatása, amely hangsúlyozza annak központi szerepét a gépi tanulásban). Az alapvető technikai beszámolót lásd: David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, *Learning representations by back-propagating errors*, *NATURE* 323

(5331986).

¹¹³ Leo Breiman, *Random forests*, MACHINE LEARNING 5, 3

fák az erdőben.¹¹⁴ A véletlenszerű erdők a döntési famodellek szélesebb kategóriája, amelybe tartoznak, különösen hasznosak nominális vagy ordinális adatok esetében; ezzel szemben a neurális hálózatok jól működnek numerikus adatokkal.¹¹⁵

Újabb fejlemény a mélytanulás. A mélytanulásban egy algoritmus több szintű reprezentációval épül fel, amelyek mindegyike egyre bonyolultabb. A mélytanulás "kulcsfontosságú szempontja", hogy "a jellemzők rétegeit nem emberi mérnökök tervezik: azokat az adatokból tanulják meg egy általános célú tanulási eljárással".¹¹⁶ A mélytanulási eszközök különösen alkalmasak a felügyelet nélküli feladatokra, a jellemzők meghatározása nélkül. Kevés "kézi beavatkozást" igényelnek, így a tervezők "csak várnak, és hagyják, hogy a tanuló algoritmus magától felfedezzen mindent, ami szükséges".¹¹⁷

A gépi tanulási felhasználási korlátozottak.¹¹⁸ A legtöbbet a felügyeleti adatok esetében az eredmények (nem oksági) előrejelzésére, illetve a klaszterek vagy asszociációk azonosítására használják (az olyan ajánlórendszerek esetében, mint amilyeneket a Netflix és az Amazon alkalmaz). Kelleher és Tierney valóban hasznosan négy nagy kategóriára redukálja a gépi tanulással megoldható "valós világbeli" problémákat: a klaszterek vagy asszociációk azonosítása egy populáción belül; a populáción belüli kiugró értékek azonosítása; asszociációs szabályok kidolgozása; valamint az osztályozási és regressziós előrejelzési problémák.¹¹⁹ Ezzel szemben a gépi tanulás eszközei korlátozottabban használhatók, ha egy probléma "egy beavatkozás alkalmi hatásának becslését" is igényli.¹²⁰ Erre példa egy nemrégiben készült tanulmány, amely a csípőprotézis műtétek elosztását vizsgálta az egyébként jogosult betegek között.¹²¹ A tanulmány gépi tanulási eszközöket használt annak megállapítására, hogy mely betegek élnek elég sokáig ahhoz, hogy részesüljenek a műtét előnyeiből. Ezek az eszközök azonban nem tudták megbecsülni a műtétnek a betegek jólétére gyakorolt oksági hatását, hogy megkönnyítsék a rangsorolást azok között, akik valószínűleg elég sokáig élnek ahhoz, hogy részesüljenek a műtét előnyeiből.¹²² Általánosabban, a gépi tanulás alkalmazása során "az oksági következtetés csak akkor lehetséges, ha az elemző az előrejelzési módszerekhez szükséges feltételezéseken túlmutató feltételezéseket tesz".¹²³

¹¹⁴ KELLEHER & TIERNEY, *fentebb*, 106,141-42. lábjegyzet.

¹¹⁵ *Id.* 136.

¹¹⁶ LeCun et al., *supra* note at 33,436; *id.* at ("A mélytanulási architektúra egyszerű modulok többrétegű halmaza, amelyek mindegyike (vagy többsége) tanuláshoz van kitéve, és amelyek közül sokan nemlineáris bemeneti-kimeneti leképezéseket számolnak."); KELLEHER & TIERNEY, *supra* note at 106,131-32 ("A mélytanulási hálózatok egyszerűen neurális hálózatok, amelyek több réteg rejtett egységgel rendelkeznek"). (kiemelés és lábjegyzet kihagyva).

¹¹⁷ ALPAYDIN, *fenti* megjegyzés, 3,107-08.

¹¹⁸ Judea Pearl, *Theoretical impediments to machine learning with seven sparks from causal revolution* PREPRINT ARXIV:1801.04016, AT 1-2 (2018) (azzal érvelve, hogy a gépi tanulás képtelensége az ellentett tényállások elemzésére az ok-okozati összefüggések kikövetkeztetése érdekében jelentős akadály).

¹¹⁹ KELLEHER & TIERNEY, *fenti* 106,151-80. lábjegyzet különböző feladatokra). A felhasználások egy másik tipológiája a gépi tanulás négy felhasználási módját határozza meg: priorizálás, osztályozás, asszociáció és szűrés. Bruno Lepri et al., *Az adatok zsarnoksága? A társadalmi jólétet szolgáló adatvezérelt döntéshozatal fényes és sötét oldalai*, in TRANSPARENT DATA MINING FOR BIG AND SMALL DATA (3,42017).

¹²⁰ Athey, *Supra* note at 3,483.

¹²¹ Jon Kleinberg et al., *Supra* note. 102.

¹²² *Id.* 493.

¹²³ Athey, *Supra* note, 3,484-85. o. (az "ösztönzők és a manipulálhatóság" miatti kiemelve).

B. A gépi és az emberi döntések megkülönböztetése

A gépi tanulás nyilvánvalóan **emberi** döntéshozataltól. De hogyan? És miért lehet ez a különbség normatív szempontból fontos? Az alábbiakban e különbségek három nézőpontját vizsgálom meg. Először is, a legnyilvánvalóbb különbségek valószínűleg a gépi tanulás léptékével, kapacitásával és a mögöttes mechanizmusokkal kapcsolatosak. Másodszor, néha azt is mondják, hogy a gépi döntések átláthatatlanabbak, mint az emberi döntések. Elfogadom a gépek és az emberek közötti hatékonysági különbséget, de megkérdőjelezem annak jelentőségét a jogok vizsgálata szempontjából. Kétkedem azonban abban, hogy létezik-e átláthatósági szakadék. Sőt, még ha az algoritmus átláthatatlansága az algoritmus bonyolultságának és működési kiszámíthatatlanságának függvénye is, nem egyértelmű, hogy a gépek emberhez képest *nagyobb* átláthatatlanságára vonatkozó panaszok megalapozottak, legalábbis ha általánosításként fogalmazzuk meg, nem pedig esetenkénti összehasonlításként.¹²⁴ Végezetül megvizsgálom az emberi és a gépi cselekvés közötti összefonódás szükséges mértékét.

1. *Hogyan térnek el a gépi és az emberi döntések a működés során?*

A gépi és az emberi **alapvető** működésükben három szempontból **az** érvelés alapvető felépítésében; a hibázási hajlamban; és az azonosítási és előrejelzési feladatok elvégzéséhez szükséges puszta kapacitás tekintetében. Mind a második, mind a harmadik különbségnek van normatív jelentősége - de nem feltétlenül olyan módon, amely az emberi döntéshez való jogot érinti.

Nagyon elemi szinten gépi tanulás **eltér** az emberi döntéshozatalétól. Csábító azt gondolni, hogy a számítógépes architektúra egyes formái, különösen a neurális hálózatok, valamilyen módon követik az emberi agy kognitív folyamatát.¹²⁵ De az általánosság nagyon magas szintjét kivéve, ennek az analógiának ellen kell állni. Igaz, hogy az emberi megismerés tudománya "befolyásolta a mesterséges neurális hálózatok kialakulását".¹²⁶ A verbális visszhang ellenére "a neurális hálózatot ugyanúgy az agy ihlette, mint a pekingi olimpiai stadiont a madárfészek", és a gyakorlatban inkább hasonlít a regresszióelemzéshez, mint az emberi neuronok működéséhez.¹²⁷ Az első neurális hálózatok között volt a "perceptron", amely egyetlen "neuronból" vagy csomópontból állt, amely egyetlen nemlineáris függvényt hajtott végre.¹²⁸ Ez azonban nyilvánvalóan kevésbé hasonlít az emberi agyra. És ma, míg az emberi megismerés tanulmányozása a "sejttípusok, molekulák, sejtszintű állapotok, valamint a számítási és információátviteli mechanizmusok" taxonómiáját tűzte ki célul, addig a gépi tanulás "nagy részt egyetlen elv: a funkcióoptimalizálás instanciáira összpontosított".¹²⁹ Ez

¹²⁴ A **pu** fajtájára példaként **W.** Nicholson Price II, *Black-Box Medicine*, HARV28. J.L. & TECH. 419,421 (2015) (egy olyan "orvostípus azonosítása, [amely] mindenki számára "fekete doboz" a kifejlesztésének természete miatt... nem pedig ... azért, mert a működését szándékosan elrejtik a szem elől").

¹²⁵ Davide Castelvecchi, *Kinyithatjuk-e a mesterséges intelligencia fekete dobozát?*, **MIT**5,2016, (azt állítva, hogy a neurális hálózatok az agyról mintázódnak).

¹²⁶ Bengio, *Supra* note at 33,49.

¹²⁷ Jayesh Bapu Ahire, *Mesterséges neurális hálózat: Adam H. Marblestone, Greg Wayne & Konrad P. Kording, Toward an integration of deep learning and neuroscience*, FRONTIERS10 IN COMPUTATIONAL NEUROSCIENCE94, (942016) (megjegyezve a matematikai előrelépések, és nem az idegtudományi áttörések fontosságát a gépi tanulásban).938034

¹²⁸ Az alapvető tanulmány Frank Rosenblatt, *The **pu** probabilistic model for information storage and organization in the brain*, PSYCH65. REV. 386 (1958).

¹²⁹ Marblestone et al., *supra* note at 127,94.

ezért hiba azt gondolni, hogy gépi tanulás egyszerűen egy ismert szénalapú folyamat szilíciumalapú változata, még akkor is, ha a gépi tanulás menetrendjének nagy része bizonyos értelemben az ismert emberi képességekre van kalibrálva.¹³⁰

A másik terület, ahol az emberi döntések eltérnek, a minőség. Az emberi döntéshozatalt még a feltételezhetően magas színvonalú környezetben is a heurisztikák, az ¹³¹implicit elfogultság ¹³²és a pusztán zaj torzító hatása sújtja.¹³³ A gépi tanulás természetesen továbbra is érzékeny a szennyezett képzési adatokra, és elkerülhetetlenül normatív döntéseket igényel az osztályozási szabályok kialakításakor.¹³⁴ Például a nagyméretű, például a világhálóról lekérdezhető szöveges adattömegek elemzését befolyásolják a (például a faji és nemi hovatartozással kapcsolatos) előítéletek, amelyek befolyásolták az adattömegek létrehozását.¹³⁵ Az ilyen eszközök óvatlanul alkalmazva súlyosbíthatják az erkölcsileg gyanús társadalmi rétegződéseket. Ugyanakkor azonban olyan pontos előrejelzéseket is képesek létrehozni, amelyekkel a gyakorlatban egyetlen ember sem képes felvenni a versenyt.¹³⁶ Ha van is határa az emberi viselkedés kiszámíthatóságának, azt még nem sikerült azonosítani.¹³⁷ Még ha van is emberi helyettesítő egy döntéshez, a különböző területeken végzett tanulmányok arra utalnak, hogy az emberi jólétben nagy nyereség érhető el, ha ember helyett gépi tanulási eszközt használunk.¹³⁸ Például a vezető nélküli autók korai bevezetése szinte biztosan drámaian csökkenteni fogja a közlekedési balesetek társadalmi áldozatait.¹³⁹ Amint az alábbiakban érvelni fogok, ¹⁴⁰éppúgy hiba lenne figyelmen kívül hagyni a gépek tervezésével járó kockázatokat.

¹³⁰ Vegyük például a problémáját, amikor az ember egy nehéz feladattal szembesül, és a környezet hatékony reprezentációit kell levezetnie a nagydimenziós érzékszervi bemenetekből, hogy új kihívásokat oldjon meg. Ezt a feladatot a gépi tanulás még csak most kezdi kezelni. Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, NATURE (518529,2015529).

¹³¹ Lásd pl. Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, U65. CHI. L. REV. 570,571 (1998) (a lehorgonyzási torzítás megállapítása).

¹³² Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?* 84 NOTRE DAME L. REV. 1195,1195 (2009) (dokumentálva az implicit faji előítéletességet a tárgyaló bírák döntéseiben).

¹³³ Daniel L. Chen & Jess Egel, *Can Machine Learning Help Predict the Outcome of Asylum Adjudications?*, PROCEEDINGS OF THE 16TH EDITION OF THE INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW (237,2372017) (megállapítva, hogy az esettel kapcsolatos tényezők csak a menekültügyi határozatok eredményeinek körülbelül egyharmadát magyarázzák).

¹³⁴ Lásd Huq, *Supra* note at 45,- (a hibás adatok és a háttérben lévő faji egyenlőtlenségek által létrehozott erkölcsi döntések megvitatása).

¹³⁵ Aylin Caliskan, Joanna J. Bryson, and Tom Griffiths, *Semantics derived automatically from language corpora contain human-like biases*, SCIENCE356 (183,1832017).

¹³⁶ A legújabb alkalmazások közé tartozik például a közlekedési jelzések programozása az összesített forgalom minimalizálása érdekében. Rose Yu et al.: *A generic approach for extreme condition traffic forecasting*, Proceedings of the SIAM2017 International Conference on Data Mining 777-785 (2017).

¹³⁷ Jake M. Hofman et al., *Prediction and explanation in social systems*, 6187-88486, (2017) (e korbátokat "nyitott" kérdésként írja le).

¹³⁸ Kleinberg et al., *supra* note at 102,494. Lásd Kleinberg et al., *Human decisions and machine predictions*, Q133. J. ECON. 237 (2017) (óvadék). Véleményem szerint az óvadékre vonatkozó példa kétértelmű, mivel a jóléti nyereség attól függ, hogy a büntető igazságszolgáltatási rendszer más szereplői hogyan reagálnak a tárgyalás előtti kényszerítés célzottabb eszközeinek kilátására. Egyáltalán nem egyértelmű, hogy válaszuk az lesz, hogy ritkábban vagy bölcsebben alkalmazzák majd a kényszerítő eszközöket.

¹³⁹ A vezető nélküli autók gyorsított mellett érvelő, társadalmi jóléti okokra hivatkozva lásd Nidhi Kalra és David G. Groves, *The Enemy of Good: Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles* (2017) https://www.rand.org/pubs/research_reports/RR2150.html (a vezető nélküli járművek gyors bevezetését ajánlja).

biztonsági okokból).

¹⁴⁰ *Lásd a* III.C. részt.

átláthatóság mellett 9érvel); *lásd még* Alyssa Carlson, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, IOWA103 L. REV. 303, 344-46 (2017) (amellett érvelve, hogy az információs szabadságról szóló törvények terjedjenek ki a nem állami szolgáltatókra is).

nagy dimenziójú adatok, összetett ~~h~~változtatható döntéshozatali logika."¹⁴⁶ Úgy tűnik, hogy az átláthatósággal kapcsolatos aggodalom részben az emberi döntéshez való jog iránti igényt motiválja.¹⁴⁷ Ennek az igénynek az empirikus predikátumai ezért külön kezelést érdemelnek.

Először is ki kell zárunk az algoritmikus átláthatósággal kapcsolatos ~~h~~Ez azon alapul, hogy a gépi tanulási eszközt birtokló vállalati egységek a gazdasági károktól való félelem miatt nem hajlandók nyilvánosságra hozni az adataikat.¹⁴⁸ Az ilyen titkolózás nem különbözteti meg egyértelműen a gépi és az emberi döntéseket. Nincs nagy különbség az üzleti titokra vonatkozó jogszabályok vagy a szellemi tulajdon védelmének egyéb formái és a szerződéses záradékok, például a versenytilalmi záradékok alkalmazása között, amelyek célja a technikai információk terjesztésének megakadályozása.¹⁴⁹ Nincs semmi megkülönböztető jegy abban a tényben, hogy nem egy személyt, hanem egy gépet véd a törvény a vizsgálattól.

Valószínűbb, hogy a ~~h~~átláthatatlanság gondolata a számítási algoritmusok összetett, rekurzív és programozatlan működési módján alapul. A gépi tanulást jellemzően olyan "problémákra alkalmazzák, amelyeknél a döntéshozatal explicit logikájának kódolása nagyon rosszul működik".¹⁵⁰ A gépi tanulási eszköz által azonosított osztályozási szabály lehet (például) egy neurális hálózat dinamikus függvénye, inkább egy számítási sorozat eredménye.¹⁵¹ A hálózat emergens tulajdonságainak megismétlése vagy megértése azonban megterheli az emberi képzelőerőt. A "bonyolult vagy homályos" forráskód pusztán vizsgálata keveset árul el arról, hogyan működik a program a való világban.¹⁵² Az ex post szemszögből nézve tehát van egy olyan érzés, hogy az algoritmusok nem lehetnek átláthatóak, mert lehetetlen rekonstruálni, hogy milyen alapon született egy adott döntés.

De az ilyen jellegű, utólagos kizáró ~~h~~nem kell technológiai adottságnak tekinteni. Az informatikusok felvetették, hogy lehetséges ex ante garantálni "egy olyan manipulálható nyilvántartást, amely nem cáfolható bizonyítékot szolgáltat minden csomópont cselekedetéről".¹⁵³ Ebből a nyilvántartásból a "hibás" csomópontok egy

algoritmikus eszközök); Danielle Keats Citron & ~~h~~*The Scored Society*: WASH89.

L. REV. 1,25 (2014) (a felügyelet és az ~~h~~mellett érvelve)

¹⁴⁶ Brent Daniel Mittelstadt et al., *Az algoritmusok etikája*: BIG3 DATA & SOC. 1,3 (2016).

¹⁴⁷ *Lásd a fenti szöveg kísérő megjegyzés* 13.

¹⁴⁸ Erre az aggodalomra lásd például Rebecca ~~h~~*Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, STAN70. L. REV. 1343, (20181350) (dokumentálva "az üzleti titokra vonatkozó bizonyítékok bevezetését a büntetőügyekbe"); FRANK PASQUALE, *THE BLACK BOX SOCIETY* 12-15 (2013).

¹⁴⁹ *Lásd pl.* Orly Lobel, *Enforceability Tbd: From Status to Contract in Intellectual Property Law*, BU. L. Rev. (869,2016871) (azzal érvelve, hogy a munkaszerződések "a cégek számára eszközként szolgálnak arra, hogy a hagyományos szellemi tulajdon határain túli információkat megfelelő értesítés vagy vita nélkül magukba zárják").

¹⁵⁰ Jenna Burrell, *Hogyan "gondolkodik" a gép*: BIG DATA & SOC. 3,
6

¹⁵¹ *Id.* 8.

¹⁵² Kroll et al., *supra* note at 2,647; Joshua ~~h~~*The fallacy of inscrutability*, PHIL TRANS. SOC. A376 (2018) (az ilyen közzétételt "sem nem szükséges" a megértés javításához).

¹⁵³ Andreas Haeberlen, Petr Kuznetsov & Peter ~~h~~*PeerReview*: ACM41 SIGOPS Operating Sys. Rev. (175,1752007); Kroll et al., *supra* note at 2,662-65 (ugyanaz); Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, Harv31. J.L. & Tech. 1, 10-11(2017) (ugyanaz).

gépi tanuló rendszerrel. ¹⁵⁴ Az algoritmus architektúrájának kulcsfontosságú elemeinek következményei elkülöníthetők és elemezhetők. Alternatív megoldásként lehetőség van arra is, hogy "több különböző ellentényt" kínáljunk fel egy algoritmusnak, így tesztelve a bemeneti eredményekben bekövetkező fokozatos változások hatását még a tényállás után is.¹⁵⁵ Ennek van egyfajta emberi analógiája a kocsis probléma kísérleti tesztjeiben.¹⁵⁶ Végül, az algoritmikus tervezés bizonyos fajtái jobban értelmezhetők, mint mások.¹⁵⁷ Az egyszerűen tanulható modellek általában egyszerűbben ábrázolhatók.¹⁵⁸ A tervező tehát bizonyos korlátok között kiválaszthatja, hogy a számítási architektúra mennyire magyarázható. Alternatív megoldásként lehetőség van a modell "egyszerűbb formában történő közelítésére" még azután is, hogy azt már létrehozták és a természetben alkalmazzák.¹⁵⁹ A szükséges magyarázat létrehozására való képesség hiánya, amely lehetővé teszi egy gépi tanulási eszköz világra gyakorolt hatásának releváns értékelését, "csupán tervezési döntés, nem pedig a nagy rendszerek komplexitásának elkerülhetetlen velejárója".¹⁶⁰

Tekintettel a gépi tanulással kapcsolatos döntések vizsgálatára mechanizmusok elérhetőségére - amelyek közül néhány párhuzamos az emberi döntéshozatal megértésére szolgáló módszerekkel -, nem lehet eleve azt állítani, hogy az előbbieket átláthatatlanabbak lennének, mint az emberek.¹⁶¹ Igaz, az algoritmikus eredmények vizsgálatához speciális eszközökre van szükség. De a bíróságok által az emberi tanúvallomások értékelésére kidolgozott bonyolult bizonyítási szabályok azt sugallják, hogy a szakértőkre ugyanúgy szükség van az emberi tanúvallomások megértésének feladatához. Az az egyszerű tény, hogy a gépi döntések megértésére szolgáló diagnosztikai eszközök idegenebbek, mint az emberi döntések megértésére szolgáló eszközök, nem teszi őket sem jobban, sem kevésbé áthatóvá.

Ha nincsenek rendszerszintű okok, hogy gépi döntések mindig átláthatatlanabbak, mint az emberi döntések, akkor fontoljuk meg az emberi és a gépi átláthatóság közötti nagyjából egyenlő esélyt, legalábbis az esetek bányászata során. Kezdjük azzal, hogy vegyük észre, hogy más elmék éppúgy fekete dobozok, mint a gépi tanulás eszközei. Egyszerűen "nincs általánosan elfogadott elmélet a

¹⁵⁴ Haeberlen et al., *Supra* note at 153,175.

¹⁵⁵ Sandra Wachter, Brent Mittelstadt & ~~Counterfactual explanations without opening the black box~~: 14(132017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3063289. ("Finding a closest possible world to x such that the classification changes is, under the right choice of distance function, the same as finding the smallest change to x"). A mélytanulási vizuális felismerő eszközökkel végzett tanulmányok azt mutatják, hogy a bemenetek kis perturbációi is képesek kategorikus osztályozási változásokat generálni. *Lásd* S. Huang, N. Papernot, I. Goodfellow, Y. Duan, & P. Abbeel, *Adversarial attacks on neural network policies*, arXiv preprint arXiv:1702.02284 (2017). Mindkettő inkább "alanyközpontú", mint "modellközpontú" megközelítés. Lillian Edwards & Michael Veale, *Slave to the Algorithm? Why A 'Right to an Explanation' Is Probably Not the Remedy You Are Looking for*, DUKE L. & TECH. REV., december, 4at18, 56, 2017.

¹⁵⁶ *Lásd pl.* Alessandro Lanteri, Chiara Chelini & Salvatore Rizzello, *An experimental investigation of emotions and reasoning in the trolley problem*, J83. BUS. ETHICS 789-90789, (2008).

¹⁵⁷ Desai & Kroll, *supra* note at 153, n11.61 ~~h~~ döntési fa, a naiv Bayes és a szabálytanulók könnyebben értelmezhetőek, mint a neurális hálózatok vagy a támogató vektor gépek).

¹⁵⁸ Michael Gleicher, *A Framework for Considering Comprehensibility Modeling*, ~~DATA~~ (72,822016).

¹⁵⁹ Andrew D. Selbst & Solon Barocas, *The Intuitive ~~And~~ Explainable Machines*, FORDHAM 87 L. REV. 1085, 1110-12 (2018).

¹⁶⁰ Kroll, *supra* note at 152, Hasznos **is**, hogy miért különböznek a különböző magyarázatok, és hogyan lehet hatékony válaszokat adni, lásd Menaka Narayanan et al., *How do Humans Understand Explanations from Machine Learning Systems? An Evaluation of the Human-Interpretability of Explanation*, arXiv preprint arXiv:1802.00682 (2018).

¹⁶¹ Anthony J. Casey & Anthony Niblett, *A Framework for the New Personalization of Law*, U86. CHI. L. REV. 333, 355-56 (2019).

hogyan keletkeznek a kognitív jelenségek az agykéregben,¹⁶² vagy hogy a tudat szolgál-e bármilyen "jelentős funkciót" a szervezet számára.¹⁶³ Még egy általánosan elfogadott népi elmélet sem létezik, amely kitöltené ezt a hiányt. Ahogy egy korábbi biofizika professzor megjegyezte, "minden nap döntéseket hozunk olyan területeken, amelyeket nem értünk teljesen", és "nem tudjuk megmagyarázni azt a komplex állapotot, amely alapján hogyan jutottunk egy adott következtetésre".¹⁶⁴ Ráadásul minél nagyobb a döntés, annál kevésbé alkalmas az ésszerű megoldásra.¹⁶⁵

Az emberi döntések ~~h~~ kutat, amikor más elmék megértéséről van szó. Nem tartunk a fejünkben világos és egyértelmű (nemhogy pontos) magyarázatot arra, hogyan gondolkodnak más emberek... És nincs közvetlen hozzáférésünk az ő kognitív folyamataikhoz sem. Ennek ellenére képesek vagyunk sikeresen interakcióba lépni velük. Még arra is képesek vagyunk, hogy értelmesen tulajdonítsunk nekik hiedelmeket és más mentális állapotokat olyan módon, amely nem nyilvánvalóan téves. Továbbá, egy másik személy hiedelmei és mentális állapotai utólag is kikérdezhetők (ahogy a mi hiedelmeink is az ő hiedelmeikről és mentális állapotukról). Sőt, sok esetben az emberek "előállítják és tárolják a [egy] döntés megmagyarázásához szükséges információkat", és megkérdezhetők, hogy ezeket az információkat a tényállás után előállítsák.¹⁶⁶ Az őszinteséggel és az őszinteséggel kapcsolatos problémák bőven akadnak. Ezeket azonban a jog nem kezeli megoldhatatlannak. Az átláthatóság, amikor más elmékről van szó, más szóval, elméleti vagy gyakorlati szempontból nem mindig lehetséges. De az is lehet, hogy nem is mindig szükséges.

A szociális cselekvés, amely annak ~~n~~ alapul, ~~h~~ más embereknek vannak meggyőződései és mentális állapotai, a mentalizálásnak nevezett készséget igényli. Ez egy olyan képesség, amely megelőzi a kifinomult megismerést. A legtöbb ember (legalábbis négyéves kor felett, akik nem jogászprofesszorok) a társadalmi világban úgy tájékozódik, hogy feltételezi, hogy a többi embernek van elméje.¹⁶⁷ Az idő bizonyos ésszerű arányában a más elmékről alkotott elképzeléseink még elég közel is vannak a valósághoz ahhoz, hogy lehetővé tegyék a hatékony társadalmi interakciókat.¹⁶⁸ Továbbá, úgy tűnik, hogy a megismerés biokémiai folyamataira vonatkozó, általánosan hozzáférhető vagy széles körben megértett magyarázatok hiánya nem ássa alá a mentalizáció képességét, illetve azt, hogy ítéleteket alkossunk mások hiteiről és mentális állapotáról. A hatékony átláthatóság hiánya

¹⁶² Leslie G. Valiant, *Mit kell megmagyarázania a kéreg globális elméletének?*, ~~15~~ NEUROBIOLOGY (15,201415).

¹⁶³ David M. Rosenthal, *Consciousness and its function*, ~~16~~ (331,8392008) (azzal érvelve, hogy megkönnyíti a "racionális gondolkodást szándékos cselekvést, végrehajtó funkciót vagy komplex gondolkodást").

¹⁶⁴ Vijay Pande, *Artificial Intelligence's 'Black Box' Problem is Nothing to Fear*, N.Y. ~~11~~ 201825; lásd még Saul Levmore & Frank Fagan, *The Impact of Artificial Intelligence on Rules and Standards* (72018) (tervezet a szerzőnél) ("A legjobb bírák, akárcsak a legjobb sportolók és tanárok, gyakran nem tudják azonosítani sikerük okait").

¹⁶⁵ Edna Ullmann-Margalit, *Nagy döntések*: ROYAL ~~18~~ INST. PHIL. SUPP. 157, 158-59 (2006) (a "nagy" döntések egy olyan kategóriájának leírása, amelyeket nem lehet a szokásos költség-haszon elemzéssel megoldani).

¹⁶⁶ Finale Doshi-Velez et al., *A mesterséges intelligencia jogi elszámoltathatósága: The role of explanation*, arXiv preprint arXiv:1711.01134 at (2017).

¹⁶⁷ Rebecca Saxe, Susan Carey & Nancy ~~11~~ *Other minds: linking developmental psychology and functional neuroimaging*, ANN55. REV. PSYCHOL. 87, 94-95 (2004). Úgy tűnik, hogy az agy bizonyos régiói felelősek ezért a képességért. *Id.* 99-100. o.

¹⁶⁸ Chris D. Frith & Uta Frith, *Interacting minds--a biological basis*, SCIENCE ~~169~~ 21999) (a "szociális intelligencia" fontos összetevőjeként jellemzi "a képességet, hogy megértsük és manipuláljuk más emberek mentális

állapotait, és ezáltal megváltoztassuk viselkedésüket").

amikor más emberek mentális **h**an szó, röviden: ez a filozófusok problémája.¹⁶⁹ A többiek számára nem feltétlenül probléma.¹⁷⁰

Tehát korántsem egyértelmű, hogy **z**szisztematikusan hozzáférhetetlenebb a gépi, mint az emberi beszélgetőpartnerek számára.¹⁷¹ Mindkét területen az előzetes szabályozás inkább jobb, mint rosszabb egyidejű feljegyzéseket eredményezhet utólag. A bizonyító erejű feljegyzések értelmezéséhez mindkét területen technikai készségekre van szükség. És mind a gépek, mind az emberek esetében van néhány okunk azt gondolni, hogy az átláthatóság legalábbis az esetek bizonyos csoportjaiban elmarad.¹⁷² Általánosságban véve a két terület között nehéz kiszámítható és stabil különbséget felfedezni.¹⁷³ Egy egészséges adag szkepticizmus a gépek és az emberek közötti feltételezett átláthatatlansági szakadékkal kapcsolatban empirikus alapon leveszi az asztalról az emberi döntéshez való jog egyik lehetséges indoklását. Ha az emberi és a gépi döntések hasonlóan átláthatatlanok, bár eltérő módon, vagyis az előbbihez való jog nem magyarázható pusztán az olvashatósággal. A III. részben vizsgált érveket ezért gondosan úgy kell megfogalmazni, hogy elkerüljük az átláthatatlanság szükséges különbségeire vonatkozó feltételezéseket.

C. Az emberi és a gépi döntések összefonódása

Van még egy másik zavarba ejtő **h**nkább empirikus, mint jogi jellegű: milyen mértékben empirikusan hihető egyáltalán elképzelni, hogy létezhet olyan "gépi döntés", amely akusztikailag elkülönül az emberi döntéstől? A22 GDPR cikke előírja a "kizárólag automatizált feldolgozáson alapuló" döntések lehetőségét.¹⁷⁴ De vajon tényleg létezik-e ilyen fenevad?

Talán nem - három okból. Először is, **g**épi tanulási eszköz eredetileg emberi tervezési és mérnöki döntések eredménye. Nem létezik teljesen endogén algoritmus.¹⁷⁵ És egy gépi tanulóeszköz tervezése nem mechanikus feladat. Ez tele van

¹⁶⁹ A probléma klasszikus megfogalmazása **k**nowledge of Other Minds, J55. PHIL. 969, 969-70 (1958) (Millre hivatkozva arra a kérdésre, hogy honnan lehet tudni, hogy valóban léteznek más elmék). Ugyancsak régi hagyomány, hogy az ilyen alapos szkepticizmus öncáfoló. *Lásd* ANITA AVRAMIDES, OTHER MINDS 5-6 (2001) (ezt a szkepticizmust a józan ész filozófusáig, Thomas Reidig vezeti vissza).

¹⁷⁰ Emlékezzünk itt Samuel Johnson híres **B**erkeley püspök **h**

¹⁷¹ Más szóval, ha tudjuk, hogy egy géppel vagy egy emberrel **h**az nem sokat mond arról, hogy mennyire lesz átlátható a döntés. Doshi-Velez et al., *supra* note at166, ("[I]t is lehetnek olyan helyzetek, amelyekben az emberektől többet lehet követelni, és más helyzetek, amelyekben a mesterséges intelligencia rendszereket magasabb szintű magyarázatra lehet kötelezni.").

¹⁷² Sőt, egyesek szerint **h**hibás megoldás, mert "elfedheti a valódi problémákat, amelyek a társadalmi hatalmi viszonyokban és intézményekben éppúgy jelennek, mint az alkalmazott szoftvereszközökben". Edwards & Veale, *Supra* note at 155,68; *accord* Mike Ananny & Kate Crawford, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, NEW20 MEDIA & SOCIETY (9732018) (az átláthatóság mint "neoliberális" megoldás kritikája).

¹⁷³ Paradox módon az **h**öntések ellenőrzésének több módja lehet, mint az emberi döntéseknek. Vegyük például a titkosítási algoritmusok robusztusságát, amelyek nem "magyarázatot", hanem formális matematikai bizonyításokon alapulnak. Doshi-Velez et al., *supra* note at166, Nehéz párhuzamot vonni az emberekkel.¹¹

¹⁷⁴ GDPR 22. § (1) bekezdés (kiemelés hozzáadva).

¹⁷⁵ A Google későn 2017.jelentette be **A**utóMűs amelynek célja "egy olyan gépi tanuló algoritmus létrehozása, amely megtanul más gépi tanuló algoritmusokat építeni". Cade Metz, *Building A.I. That Can Build A.I.*, N.Y. TIMES, Nov. at5,2017, B1. Hasonló kutatások folynak a Carnegie Mellon Egyetemen. *Lásd* Renato Negrinho & Geoff

Gordon,

normatív döntések. A tervezőnek kell bizonyos algoritmikus architektúrát - neurális hálózatot, döntési fán alapuló modellt, például véletlen erdőt, vagy valami mást. A választás nehéz és szükségszerűen értékelő jellegű. A gépi tanulás minden formája rendelkezik továbbá jellegzetes tanulási torzításokkal - vagyis olyan sajátos funkciókkal, amelyeket az elemzés során nagyobb valószínűséggel alkalmaznak. Az algoritmus tanulási torzítása és az adathalmaz közötti "legjobb egyezés" megtalálása emberi ítélőképességet igényel.¹⁷⁶

Az emberi ítélőképességre való felhívás itt nem együk például egy mélytanulási eszköz indítási folyamatát. A gépi tanulásnak ebben a formájában egy többrétegű neurális hálózatot kell létrehozni. A műszer tervezőjének meg kell határoznia, hogy hány réteget építsen be, és hány neuron kerüljön minden egyes rétegbe.¹⁷⁷ A tervezőnek ezután el kell döntenie, hogyan kapcsolja össze a hálózat különböző elemeit. Választhatja a rekurrens neurális hálózat létrehozását, amelyben a hálózat topológiája hurkolt. Minden egyes neuron a bemeneteket az előzőleg feldolgozott bemenetek kontextusában dolgozza fel, egyfajta "memóriát" létrehozva.¹⁷⁸ Alternatív megoldás lehet egy konvolúciós neurális hálózat, amelyben a neuronok lokalizált csoportjait arra képzik ki, hogy felismerjenek bizonyos mintákat, függetlenül attól, hogy azok hol jelennek meg az adatokban (például egy szem vagy egy orr a vizuális felismerő rendszerben).¹⁷⁹ A hálózati topológia megválasztása ismét emberi döntés, amely az emberi viselkedés kviddicseiben gyökerezik. A tényleges algoritmusban semmi sem adhat választ erre a kérdésre.

Másodszor, az emberi szerep a gépi nem korlátozódik az algoritmus kezdeti megtervezésére. A tervezőnek ki kell választania azokat az adatokat is, amelyeken a gépi tanulási eszközt kezdetben betanítják. Ezt a képzési adatot ráadásul általában nem egy algoritmus állítja elő. Ezek az emberi tevékenység függvényei. Ennek eredményeképpen megismételheti az azt létrehozó egyének elfogultságait és vakfoltjait.¹⁸⁰ Például a rendőri tevékenységgel összefüggésben aggodalomra ad okot, hogy a letartóztatási adatok, ha a jövőbeni bevetési döntések alapjául szolgálnak, tükrözni és reprodukálni fogják a bűncselekményekkel szembeni faji hajlamra vonatkozó aggasztó feltételezéseket, amelyek a múltban a rendőrök körében elterjedtek voltak.¹⁸¹ Az adatgyűjtés után a tervezőnek elő kell dolgoznia és át kell alakítania az adatokat, az adatok fényében ki kell igazítania az algoritmus paramétereit, és az eredmények minősége alapján finomhangolnia kell az algoritmust.¹⁸² Saul Levmore és Frank Fagan, akik az ember-gép partnerség elkerülhetetlensége mellett érvelnek, továbbá azt sugallják, hogy a hiányzó adatok problémái azt jelentik, hogy elkerülhetetlenül szükség lesz emberi "gondolkodásra az adatok kiválasztásáról és a célok kitűzéséről".

Deep Architect: Automatically Designing and Training Deep Architectures, arXiv preprint. Még mindig szükség van egy emberi mérnökre, aki a különböző potenciális architektúrák feletti keresési folyamatok egy részét irányítja. *Id.* 1.

¹⁷⁶ KELLEHER & TIERNEY, *Supra* note at 106,99-100; ~~King~~ Casey & Niblett, *Supra* note at 161, ("354[H]umans are involved in all stages of setting up, training, coding, and assessing the merits of the algorithm.").

¹⁷⁷ Bengio, *Supra* note at 33,50.

¹⁷⁸ KELLEHER & TIERNEY, *Supra* note at 106,133.

¹⁷⁹ *Id.*; lásd még LeCun et al., *supra* note at 33, hálózatok általános 439.

¹⁸⁰ Az elfogultságokról lásd: Kroll et al., *supra* note 2, [A]lgoritmusok, amelyek valamilyen típusú gépi tanulást tartalmaznak, diszkriminatív eredményekhez vezethetnek, ha az algoritmusokat olyan történelmi példákon képezték ki, amelyek múltbeli előítéleteket vagy implicit elfogultságot tükröznek"). A vakfoltokról lásd Kate Crawford, *Artificial Intelligence's White Guy Problem*, N. Y. TIMES, June ,25 , h2016thttps://www. nytimes. com////opinion/sunday/artificial-intelligences-white-guete-guy-problem20160626.

gyezve a hiányzó adatokból adódó problémákat).

¹⁸¹ Huq, *Supra* note at 45,--.

¹⁸² Saleema Amershi et al., *Hatalmat a népnek: AI 35MAGAZINE (105,2014105): Az ember szerepe az interaktív tanulásban.*

még a rendszer üzembe helyezése után is.¹⁸³ hangsúlyozzák, a gépi tanulás megvalósítási és tervezési folyamatát az ember és a gép közötti interakció "hosszadalmas és aszinkron iterációi" szakítják meg, amelyek biztosítják, hogy a gépi tanulást mindig kritikusan emberi kéz formálja.¹⁸⁴

Harmadszor, ha egyszer már működnek, a gépi eszközöknek még mindig szükségük van "emberi gondozókra", akiknek feladata a látás szimulálására használt mélytanulási algoritmusok eredményeinek moderálásától kezdve a karbantartóig, akik az algoritmusok működéséhez szükséges nagy mennyiségű információ tárolására használt adatközpontokat tisztítják és javítják.¹⁸⁵ A neurális hálózatokon belüli tanulási folyamat szintén "megköveteli az ember szoros közreműködését", akinek címkéket kell készítenie a képzési adatokhoz, majd hipotéziseket kell létrehoznia, amelyek az optimalizálási folyamatot irányítják.¹⁸⁶

Egyszóval, a magától működő gép elképzelése távolról sem tűnik pontosnak. Ez persze változhat. Egyes mélytanuló eszközök már a harmadik szakaszban "nagyon kevés kézzel történő tanulást" igényelnek (bár tervezésük még mindig jelentős kihívásokat jelent).¹⁸⁷ De legalábbis egyelőre egy gépi tanuló eszköz emberi tervezési és teljesítményspecifikus döntéseken nyugszik. A tisztán gépi döntés lehetősége, ismét csak egyelőre,¹⁸⁸ a technológiai eseményhorizonton túl van.

Ez az empirikus tény felveti a kérdést, hogy pontosan hogyan¹⁸⁹ kell értelmezni a "kizárólag automatizált feldolgozáson alapuló határozat" fogalmát. Ez még akkor is így van, ha egyetlen emberi tervező sem tudná megismételni az algoritmus kognitív munkáját, vagy alternatívaként megjósolni az algoritmus által végül meghozott minden egyes osztályozási döntést. Ebből következik, hogy az emberi döntéshez való jog szó szerinti értelmezése nem hihető. Ez azonban nem jelenti azt, hogy a kizárólag autonóm gépi döntések gondolatát el kellene vetni. Az emberi döntéshez való jog egy másik lehetséges értelmezése nem az emberi részvétel abszolút mértékére, hanem annak időzítésére és minőségére összpontosítana. A gépi tanulás tervezése és képzése, amelyet itt hangsúlyoztam, nagyrészt ex ante történik. Az algoritmus működésével egyidejűleg történő emberi beavatkozás jelentéktelen vagy nagyrészt szolgálati tevékenységet jelent. Az emberi döntés iránti igényt úgy is fel lehet fogni, hogy az szűkebb értelemben attól az időszaktól függ, amelyben a gép egy adott ügyben eljár. Ez egy olyan igény, amely szerint a

¹⁸³ Levmore & Fagan, *Supra* note at 164,8.

¹⁸⁴ Amershi et al., *Supra* note at 182,105.

¹⁸⁵ AINow, AINOW 2017 REPORT 12
https://assets.ctfassets.net/wprhvhvnpfc/AcZTCZaKEYMWsca/cfbfbbecc/_AI80193264286365575142742643078_Now_Institute_2017_Report_.pdf

¹⁸⁶ Bengio, *supra* note at 33,50-51 (rámutatva hogy egy "tanuló algoritmus megrekedhet egy úgynevezett lokális minimumban, amelyben nem tudja csökkenteni a neurális hálózat előrejelzési hibáját a paraméterek kismértékű módosításával"); lásd még Amershi et al., *supra* note at 182, (106leírva a "modell tér iteratív felfedezését" a szakemberek által, amint egy tanuló algoritmus elindult és működik).

¹⁸⁷ LeCun et al., *Supra* note at 33,436.

¹⁸⁸ Egy nemrégiben készült felmérés szerint egy a kettőhöz az esélye hogy 20402050, körül magas szintű gépi intelligenciát fejlesztenek ki - ez az esély kilenc a tízhez V.C2075.. Müller & N. Bostrom, *Future progress in artificial intelligence: FUNDAMENTAL ISSUES OF ARTIFICIAL INTELLIGENCE* 555-56555, (2016).

¹⁸⁹ GDPR 22. § (1) BEKEZDÉS.

Interakció közvetlen "tranzakciós kerete"¹⁹⁰, az embereknek nem szabadna csak az interakció egyik időbeli oldalán állniuk. Időbeli szemellenzőn keresztül nézve ez azt feltételezi, hogy az a tény, hogy egy emberi mérnök egyszer már kalibrálta a gépet, amellyel az ember éppen együttműködik, nem vigasztal, vagy egyáltalán nem vigasztal.

Ebben az értelmezésben az emberi döntésre igény családilag hasonlít egy "normához", szemben a szabályokkal, ahogyan a jog- és közgazdaságtani szakirodalomban definiálják: egy olyan norma, amely a szabályozott alanyok cselekvése után kap tartalmat, nem pedig előzetesen.¹⁹¹ Ahogyan a szabályok és a normák megkülönböztetése a norma időzítésén (és bizonyos mértékig a specifikáció¹⁹² mértékén) múlik, úgy az emberi döntéshez való jog gondolata is azt a feltételezést vezeti be, hogy végső soron egy cselekvés normatív indoklását csak egy ember, nem pedig egy gépi szereplő szolgáltathatja, aki az algoritmus számítási folyamatainak lezárása után avatkozik be.¹⁹³ Ez azt sugallja, hogy amennyiben a munkacsoport iránymutatása az Art. 22. cikkének (1) bekezdése nem egyértelmű, úgy kell értelmezni, hogy az "értelmes" és "utólagos" felülvizsgálatot ír elő.¹⁹⁴ Ez azt sugallja, hogy a 22. cikk (1) bekezdését csak akkor kell úgy értelmezni, hogy az csak akkor sérül, ha hiányzik az utólagos, nem pedig az előzetes emberi részvétel. A GDPR azt sem magyarázza meg, hogy az ilyen felülvizsgálat minőségét hogyan kell értékelni.¹⁹⁵

Összefoglalva, az emberi való jog nem érhető úgy egyszerűen egy olyan döntéshez való jogot jelent, amely valamilyen emberi elemmel bír, bármilyen időzített és bármilyen érdemi is legyen az: A jog ily módon történő meghatározása a jelenlegi technológiai korlátok mellett megfosztaná azt hatékony tartalmától. Ha azonban nem értelmes olyan gépi döntésekről beszélni, amelyekben nem vesz részt ember, akkor felmerül a kérdés, hogy a szükséges emberi közreműködés időzítése miért jelent gyakorlati különbséget. Ha az emberi döntéshez való jog nem merülhet ki valamiféle esztétikai preferenciában, akkor az *ilyen módon időzített* jog normatív alapjait kell megadni. Ezt a kérdést a III. rész tárgyalja.

¹⁹⁰ Vö. Aziz Z. Huq, *The Consequences of Disparate Policing: Evaluating Stop and Frisk As A Model of Urban Policing*, MINN101. L. REV. 2397,2402 (2017) (megjegyezve, hogy "egy szűk, individualista "tranzakciós keret" ... jelenleg uralja mind a jogot, mind a közgazdaságtant).

¹⁹¹ Louis Kaplow, *Szabályok kontra normák*: DUKE42 L.J559-635(1992).

¹⁹² *Lásd pl.* Anthony J. Casey & Anthony Niblett, *The Law Rules and Standards*, IND92. L.J. (2017)1401,1407 ("A szabályok pontosak és ex ante a természetben.... A szabványok viszont pontatlanok, amikor hatályba lépnek.").

¹⁹³ A vizsgáló egyetemes iránymutatás-tervezet kínosan elismeri az időbeliség jelentőségét, amikor kimondja, hogy ha "nem lehetséges vagy nem praktikus az automatizált döntést megelőzően emberi döntést hozni", akkor az embernek az eredményekre kell összpontosítania. The Public Voice, *Draft Universal Guidelines on Artificial Intelligence*, at 1, <https://epic.org/international/AIGuidelinesDRAFT20180910.pdf>.

¹⁹⁴ A munkacsoport iránymutatásai, *fenntebb*, 59. §. 22. cikk (1) bekezdéséből eseteket, amikor az ember "a végső döntés meghozatalakor más tényezőket vizsgál és vesz figyelembe"); *lásd még a fenti*, a lábjegyzethez tartozó szöveget. 195.

¹⁹⁵ Nem világos, hogy mit értelmessé. *Lásd* Michael Veale & Lilian Edwards, *Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling*, COMP34. L. & SEC. REV. 398, 401(2018) ("Az, hogy a "kizárólagosság" e kibővített fogalmát hogyan lehetne gyakorlatilag értékelni az adatkezelő vagy az érintett szempontjából, az egyik jelentős szürke terület, amelyet az iránymutatás maga után hagy.")). Ezen túlmenően, ha egy automatizált folyamat nem változtatja meg a "jogokat", vagy nem jár "egyenértékű vagy hasonlóan jelentős" hatással, a munkacsoport szerint az nem tartozik a 22. cikk (1) bekezdésének hatálya alá. A munkacsoport iránymutatásai,

fenti megjegyzés, 59,21-22. o. (megjegyezve, hogy bár a célzott reklám jellemzően nem tartozik a 22. cikk (1) bekezdésének hatálya alá, a célzott reklámozás "tolakodó jellege", az egyén elvárásai és az üzemeltetőnek a személy "sebezhető pontjairól" való tudása miatt a 22. cikk (1) bekezdésének hatálya alá tartozhat).

D. Következtetés

A gépi tanulás heterogén és fejlődő számítási formák nagy felőli. Az itt leképezettek nyilvánvalóan alig karcolják a felszínét. Néhány központi rendszertani vonalat eléggé hangsúlyoztam ahhoz, hogy lehetővé tegyem az emberi döntéshez való jog értelmes megvitatását. Feltártam továbbá azokat a dimenziókat, amelyek mentén egy gépi döntés kategorikusan megkülönböztethetőnek tekinthető emberi megfelelőjétől. Ennek a vitának tisztáznia kell, hogy mi a tétje annak, hogy az emberi döntést a gépi döntéstől megkülönböztetve követeljük.

Nagyon általános szinten a lépték mechanizmusok szisztematikusan különböznek az emberek és a gépek között, de óvatossá kell lennünk azzal, hogy ebből a következtetésből a helyesre vonatkozó állítást vonjunk le. A gépek kevesebb hibával képesek osztályozni és jósolni, mint az emberek. Legalábbis dinamikus szempontból ez azt sugallja, hogy a jogi szabályoknak inkább a jobb gépek létrehozását kellene ösztönözniük, mintsem az emberekkel való helyettesítésüket. Ezzel szemben az átláthatóság, amely a jogi tudományok nagy részének középpontjában áll, nem nyújt értelmes pontot az emberek és a gépek közötti különbségtételre. Végezetül felvettem, hogy a jelenleg (és valószínűleg a következő húsz évben is) működő algoritmusok többségét emberi tevékenység fogja alakítani és irányítani. Ha tehát az emberi döntéshez való jognak most értelmezhető tartalommal kell bírnia, akkor azt úgy kell értelmezni, hogy egy adott pillanatban emberi ítéletet igényel: miután egy gépi tanulási eszköz a természetben találkozik egy emberi szereplővel és osztályozza azt. Bármely más meghatározás vagy technológiailag megvalósíthatatlan, vagy túl könnyen teljesíthető ahhoz, hogy a feltételezett jog koherens, független entitásként álljon fenn.

III. Az emberi döntéshez való jog normatív alapjai

E technológiai kontextus ismeretében rész az emberi döntésekhez való jog legjobb rendelkezésre álló normatív beszámolóját vizsgálja. Nem abból a feltételezésből indulok ki, hogy az emberi döntéshez való jog indokolt. Éppen ellenkezőleg, e jog elismerésének legjobb normatív indokainak bemutatásával remélem, hogy ítéletet tudok mondani arról, hogy érdemes-e a jogot a jogba beágyazni, és ha igen, milyen okokból. A gépi tanulási eszközök állami alkalmazásaira összpontosítok. Amint azt már említettem, a büntető igazságszolgáltatás és a jóléti igazgatás jellemezte eddig az ilyen eszközök leggyorsabb elterjedését annak meghatározásában, hogy kit állítson meg a rendőrség, kit börtönözzenek be a tárgyalás előtt (vagy után), és ki részesül állami juttatásokban.¹⁹⁶ Annak a személynek az esete, akit egy hibás gépi döntés miatt kényszerítésnek vetnek alá, vagy akitől megtagadnak egy juttatást, a legmeggyőzőbb példája annak, hogy miért lehet indokolt az emberi döntéshez való jog. Feltételezem tehát, hogy valami lényeges dolog forog kockán, amikor a gép (vagy az ember) dönt, és ez a feltételezés nem mindig teljesül a gépi tanulás magánhasználatának kontextusában.

A kifejtés kedvéért¹⁹⁷ háromféle indokot is emberi döntéshez való jogra (a II. rész szerint az utólagos emberi felülvizsgálathoz való jogként értelmezve). Az első az *alanyra összpontosít*, mivel az algoritmikus osztályozási szabály alanyának cselekedeteire vagy potenciális cselekedeteire összpontosít. Az indokok ezen csoportja arra az intuíción épül, hogy egy nem emberi eszköz kizár bizonyos lehetőségeket az értelmes emberi cselekvés gyakorlása elől. Az okok második csoportja az *osztályozóra összpontosít*. Ezek arra az intuíción épülnek, hogy az állam különösen az egyéneknek tartozik egy bizonyos fajta döntéssel, tekintet nélkül arra, hogy az egyén tud-e, vagy valószínűsíthetően tudna-e reagálni, vagy megpróbálná-e megváltoztatni ezt a döntést. Az indokok harmadik csoportja a *rendszer szintű hatásokra* összpontosít. Míg az első és a második típusú indokok

szűk értelemben vett

¹⁹⁶ *Lásd a fenti szöveg kísérő megjegyzését* 31.

¹⁹⁷ Azaz, a kategóriák eszközei, és nem arra szolgálnak, hogy megragadjanak bármilyen mögöttes erkölcsi tényt, ha egyáltalán léteznek ilyenek.

az új osztályozó közötti kölcsönhatáson alapuló vonat, az okok ezen osztálya kitágítja a lencsét, megragadva a csak dinamikus és kumulatív módon kibontakozó tovagyűrűző következmények lehetőségét. Azzal fogok érvelni, hogy az okok e csoportjai közül egyik sem nyújt biztos normatív alapot az emberi döntéshez való joghoz. Bár különösen a rendszerszintű hatásoknak van némi erejük, ezek nem alkalmasak arra, hogy egy egyéni jog alapjaként működjenek.

Megközelítem nem támaszkodom egyetlen konkrét jog szem, hanem a normatív elméletek lehető legszélesebb körét veszem figyelembe. Ezért lehetséges (sőt valószínű), hogy a hatodik kiegészítés esküdtzéki tárgyaláshoz való joga a Legfelsőbb Bíróság értelmezése szerint mechanikusan kizárja a bűnösséggel vagy ártatlansággal kapcsolatos tényekről szóló gépi döntést.¹⁹⁸ Az én projektem azonban nem a jog mechanikus alkalmazására összpontosít: Az érdekel, hogy van-e olyan plauzibilis normatív elmélet, amelyre ez az alkalmazás támaszkodhat. (És ha a válasz nem, annál rosszabb a hatodik módosítás joggyakorlatának). Ezen túlmenően az itt kínált indokok taxonómiája nem kizárólag olyan kétértelmű, szóösszetételes fogalmakra támaszkodik, mint az "autonómia" és a "méltóság".¹⁹⁹ Ezek a fogalmak, az biztos, hogy vezető szerepet játszottak az állam kötelezettségeiről szóló normatív elméletalkotásban a különböző döntéshozatali eljárások tekintetében a "tisztességes eljárás" rovatában.²⁰⁰ Az olyan fogalmak azonban, mint az "autonómia" és a "méltóság", pontosítást igényelnek. Széles és változatos intellektuális történettel rendelkeznek a jogon belül és kívül. Például a Legfelsőbb Bíróság által a "méltóság" szó használatának módját vizsgáló felmérés²⁰¹¹ csak a bírák részéről öt különböző használatot talált: "státusz mint méltóság, egyenlőség mint méltóság, szabadság mint méltóság, személyes integritás mint méltóság és kollektív erény mint méltóság".²⁰¹ A filozófiai hagyományban pedig Schopenhauer egyszer úgy jellemezte a méltóságot, mint "az összes tanácstalan és üresfejű moralista sibbolethjét, akik eme impozáns kifejezés mögé rejtik az erkölcs minden valódi alapjának hiányát".²⁰² Az olyan elvont és homályos normatív kifejezések, mint az "autonómia" és a "méltóság" általában csak azután hasznosak, hogy egy teljes értékű normatív elmélet megidézésével kiszínezték és behatárolták őket.²⁰³

¹⁹⁸ Lásd a fenti szöveg kísérő megjegyzés⁷¹eit 72.

¹⁹⁹ Ezeket a gondolatokat az algoritmikus eszközökkel más döntő normatív alapként hozták fel - nem mindig a legbiztosabb módon. Lásd pl. Tal Z. Zarsky, *The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making*, SCI41. TECH. & HUMAN VALUES 118-19118, (2016) (hivatkozva "az autonómiával kapcsolatos aggályokra, amelyek az egyéni méltóság sérelmével járnak"); lásd még Margaret Hu, *Algorithmic Jim Crow*, FORDHAM86 L. REV. 633,696 (2017) (hasonló érveléssel).

²⁰⁰ Lásd pl. Obergefell v. Hodges, S135. Ct. (2584,2015) (az "egyéni méltóságot és autonómiát" a "Due Process Clause által védett alapvető szabadságjogok" között azonosítja); accord Jerry L. Mashaw, *Administrative Due Process: The Quest for a Dignitary Theory*, B.U61. L. REV. 885, 901-04 (1981) [a továbbiakban: Mashaw, *Administrative Due Process*] (a tisztességes eljárás méltóságon alapuló meghatározásának tárgyalása).

²⁰¹ Leslie Meltzer Henry, *The Jurisprudence of Dignity*, U160. REV. 169,190 (2011); lásd még Vicki C. Jackson, *Constitutional Dialogue and Human Dignity: States and Transnational Constitutional Discourse*, MONT65. L. REV. 16 (2004) (a "méltóság" kifejezés használatát vizsgálja az állami jogban és a nemzetközi jogban). Az "autonómia" különböző felhasználásainak hasonló taxonómiájához lásd John Christman, *Constructing the Inner Citadel: Recent Work on the Concept of Autonomy*, ETHICS99 (109)1988 (az autonómia közelmúltbeli filozófiai irodalomban alkalmazott koncepcióinak áttekintése); lásd még Richard H. Fallon, Jr., *Two Senses of Autonomy*, STAN46. L. REV. 875, 876-77 (1994) (megkülönböztetve az autonómia "leíró" és "leíró" értelmét az első módosítási vitákra való hivatkozással).

²⁰² MICHAEL ROSEN, DIGNITY: ITS HISTORY AND MEANING (az idézeteket nem idézzük).

²⁰³ Martha Nussbaum például az emberi jólét képességekkel kapcsolatos megközelítését használta fel arra, hogy a méltóság gondolatának értelmes tartalmat adjon. Martha C. Nussbaum, *Human Dignity and Political Entitlements*, in HUMAN DIGNITY AND BIOETHICS: ESSAYS COMMISSIONED BY THE PRESIDENT'S COUNCIL ON BIOETHICS (351Adam Schulman & Martha C. Nussbaum, *Human Dignity and Political Entitlements*). Nussbaum, szerk., 2008). Jeremy Waldron az emberi jogi jog használja fel arra, hogy a méltóság olyan felfogását ismerje

fel, amely a társadalmi rangsor bizonyos fajtáinak elutasításán alapul. JEREMY WALDRON, MÉLTÓSÁG, RANG ÉS JOGOK (2012). Egyik sem

Hogy elkerüljük a kérdéseket feszegető vitákat, úgy gondolom, hasznos, ha azzal kezdjük, hogy a gépi döntések milyen konkrét módon érinthetnek bizonyos normatív értékeket. Azt is hasznosnak tartom, ha feltételezzük, hogy a szóban forgó algoritmus nem támaszkodik szennyezett képzési adatokra, nem manipulálták rosszindulatúan, és hozzáértő módon tervezték. Vagyis nem fogom a paklit azzal megterhelni, hogy feltételezem, hogy a gépi döntés eleve hibás. Továbbá, következtetésemben visszatérek a jog törvényére az államosztályozás pontosságának javulását katalizáló törvényre.

Mielőtt rátérnék az emberi döntéshez való jog három ~~indok~~ indokára, először egy olyan vonzó indokot vizsgálok meg, amelyről közelebbről megvizsgálva kiderül, hogy kevés teherbíró képességgel bír. Ez a pontosság fogalma. Miután ezt elintézem, ezután azt elemzem, hogy az emberi döntéshez való jog alátámasztására alkalmas normatív értékek három csoportja közül bármelyik alátámasztja-e az emberi döntéshez való jogot. Arra a következtetésre jutok, hogy e jog alapjai korántsem biztosak.

A. A döntéshozatal pontossága

Először is, úgy gondolom hasznos egy olyan normatív szálát, amely nem fonódik be hihetően az emberi döntéshez való jog alapjába. A Legfelsőbb Bíróság időnként azt sugallta, hogy a tizennegyedik módosítás Due Process Clause-je jogot teremt a pontos döntéshez.²⁰⁴ Jellemző a Bíróság e tekintetben tett kijelentéseire az állítása, hogy "[a] jogi eljárás funkciója, ahogyan ez a fogalom az Alkotmányban és a ténymegállapítás területén megtestesül, a téves döntések kockázatának minimalizálása".²⁰⁵ Csábító ezt úgy definiálni, mint a pontos és igaz döntéshez való jogot. Ha egy döntés eltér az alapigazságtól, akkor tévedek. De még az olyan nagy tétet jelentő kontextusokban is, mint a büntetőügyek vagy a halálbüntetés elítélés utáni felülvizsgálata, a Legfelsőbb Bíróság elzárkózott a *valós* döntéshez való személyes jogtól.²⁰⁶ A pontossághoz való jogot ehelyett valószínűségi fogalmakkal értelmezik, mint például "minden kétséget kizáróan".²⁰⁷ Ezt továbbá az ítélkezési eljárás egészének attribútumaként tekintik. Alkotmányos szempontból elegendő egy olyan rendszer, amely "az esetek összességében elfogadható szintre képes csökkenteni a hibakockázatot".²⁰⁸

a méltóság e fogalmai nyilvánvalóan ~~az~~ algoritmikus kontextusban. Margaret Radin hegelianus szellemében írva azt javasolta, hogy "az autonómia [a legjobb] az egyénnek tulajdonított absztrakt racionalitásként és felelősségként értelmezhető". Margaret Jane Radin, *Property and Personhood*, STAN34. L. REV. 957,960 (1982).

²⁰⁴ Lásd *Honda Motor Co. v. Oberg*, U.S 512. (415,9994) (megjegyezve, hogy az "önkényes és pontatlan ítélkezés" sértheti a Due Process-et); Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, YALE95 L.J. (455,4761986) ("Az olyan tisztességes eljárásbeli védelmek, mint az értesítés, a meghallgatás és a védelemhez való jog értékes, mert hozzájárulnak a pontosság céljához.").

²⁰⁵ *Greenholtz v. Inmates of the Neb. Penal Complex*, U.S.442. (1,131979); *accord* *Heller v. Doe ex rel. Doe*, U.S.509. (312,3321993) (a Due Process meghatározása a "bíró előtt lévő ügyek pontos meghatározásához" fűződő érdek szempontjából); Jerry L. Mashaw, *The Supreme Court's Due Process Calculus for Administrative Adjudication in Mathews v. Eldridge: Három tényező az értékelmélet keresésében*, U44. CHI. L. REV. 9(1976) [a továbbiakban: Mashaw, *Due Process Calculus*] ("Az Eldridge-bíró az eljárási védelem egyetlen célját a pontosság fokozásában látja..., és ezért számítását a helyes vagy helytelen döntésekből származó előnyökre vagy költségekre korlátozza.").

²⁰⁶ Lásd *Dist. Attorney's Office v. Osborne*, 52 Ct. 2308,2321 (2009) (kifejtve, hogy a "tényleges ártatlanság" mint szövetségi jog létezése továbbra is "nyitott kérdés"); *Herrera v. Collins*, U506.S. (390,1993404) (kifejtve, hogy a "tényleges ártatlanságot" soha nem tekintették önálló alkotmányos igénynek).

²⁰⁷ Lásd *pl. In re Winship*, U.S 397. (1970358,372) (nyitótörté) ("[I]t sokkal rosszabb elítélni egy ártatlan embert, mint szabadon engedni egy bűnös embert.").

²⁰⁸ Patrick Woolley, *Rethinking the Adequacy of Adequate Representation*, 15 *REV.* 571,630 (1997)

Az ilyen rendszerszintű és valószínűségi szempontból értelmezett pontosító jog biztosíthat-e normatív alapot az emberi döntéshez való jog számára? Tegyük zárójelbe azt az elképzelést, hogy az emberi döntéshez való jog megalapozható azon az elképzelésen, hogy egy gép képtelen új bizonyítékot venni egy szabályozott alanyról, ²⁰⁹és koncentráljunk arra a puszta elképzelésre, hogy az ember *eo ipso* jobb, mint a gép. A II. részben összegyűjtött bizonyítékok azt sugallják, hogy a válasz általában nemleges. Amint azt a II. rész kifejtette, a gépi tanulás olyan feladatokat végez, amelyek átfedésben vannak az emberi döntések egy csoportjával. Mivel a jelenlegi algoritmikus eszközök inkább korrelációs, mintsem oksági kapcsolatokat azonosítanak, az empirikus kérdések egy csoportja nem alkalmas arra, hogy megválaszolja őket.²¹⁰ Bár egy egyénnek jogos panasza lehet, ha egy algoritmikus döntésnek van alávetve egy ok-okozati következtetés kérdésében, a panasza valójában nem annyira a pontosságra, mint inkább az alkalmazott módszer alkalmatlanságára vonatkozik.

Azon feladatok azon csoportja esetében, akár ember, akár gépi tanuló eszköz végezhet, a rendelkezésre álló bizonyítékok azt sugallják, hogy az utóbbiak gyakran összességében kevesebb pozitív és negatív eredményt produkálnak, mint a legtöbb emberi döntéshozatal.²¹¹ Ez olyan összefüggésekben igaz, mint például a tárgyalás előtti óvadék²¹² letétele és a családon belüli erőszakkal kapcsolatos vádemelések.²¹³ Így empirikusan nem hihető azt állítani, hogy a pontos döntéshozóhoz való jog (ebben az értelemben) ezekben az esetekben emberi döntéshozóhoz való jogot von maga után.²¹⁴ Sőt, ahogyan a következtetés feltárja, talán éppen az ellenkezője igaz, és a jogi szabályokat úgy lehetne megalkotni, hogy a meglévő gépi döntések javítása, nem pedig helyettesítése legyen a cél.

Vegyük fontolóra két lehetséges ellenvetést állásponttal. Az első a következőképpen hangzik: Az algoritmikus osztályozás azért nem megfelelő, mert az államnak egyáltalán nincsenek meg a szükséges képzési adatai az érdeklődésre számot tartó változó méréséhez, és az e célból felhasznált rendelkezésre álló adatok félrevezetőek. Elképzelhető, hogy ilyen ellenvetést fogalmaznak meg (például) olyan tanárértékelési eszközökkel szemben, amelyek olyan tesztelési módszerekre támaszkodnak, amelyek nem képesek a gyermekek fejlődésének hasznos mérésére. A probléma itt azonban nem az, hogy az emberi döntés helyett a gépi döntést választjuk. Sokkal inkább az az előzetes (emberi) döntés, hogy egy kétségtelenül releváns paraméter méréséhez irreleváns adatokra támaszkodnak. Nem nyilvánvaló eleve, hogy a megoldás ilyen esetekben az *emberi* döntéshez való jog, szemben a feladathoz jól igazított gépi döntéshez való joggal.

Másodszor, ahol az algoritmus képesség társadalmilag rétegzett (mondjuk faj vagy nem szerint), és ahol a hibák eloszlása követi és erősíti a hierarchikus törésvonalakat, úgy gondolom, hogy komoly normatív aggályok merülnek fel, amelyek gondos vizsgálatot igényelnek. De a feloldásukról kiderül, hogy

²⁰⁹ Lásd az alábbi III.B.2. részt.

²¹⁰ Lásd Mullainathan & Spiess, *supra* note 144, (88megkülönböztetve az előrejelzést - az általánosítható minták feltárását - és a paraméterbecslést, és megjegyezve, hogy a gépi tanulás az előbbit jól végzi, az utóbbit nem). A gépi tanulás eszközeinek erénye, hogy "rugalmas funkcionális formákat használnak, [amelyek] lehetővé teszik számunkra, hogy az adatok változatos struktúráihoz illeszkedjünk". *Id.* at 91-92

²¹¹ Lásd a *fenti* szöveg kísérő megjegyzés 136-138.

²¹² Kleinberg et al., *fentebbi* jegyzet, 138,237-38. o.

²¹³ Richard A. Berk, Susan B. Sorenson & *Forecasting domestic violence: A machine learning approach to help inform inform vád alá helyezési döntések*, J13. EMP. L. STUD. 94,110 (2016) (megállapítva, hogy a családon belüli erőszakos vádlottak egy csoportjában a családon belüli erőszakos vádlottak százalékos²⁰ visszaeső elkövetők szabadulási aránya

százalékosra¹⁰ csökkenthető a bírói döntésekről a gépi vezérlésű döntésekre való áttéréssel).

²¹⁴ Előfordulhat, hogy egy algoritmikus eszközt úgy, hogy a hamis pozitív eredmények aránya megnő. De akkor a kifogás a hibás emberi megvalósítás, és nem maga az algoritmus.

elég nehéz. Az algoritmikus eszközökkel kapcsolatos tanulmányok azt mutatják, hogy a hamis pozitív arányok leggyakoribb populációsintű mérőszámait nem lehet egyszerre kiegyenlíteni.²¹⁵ Ehelyett a hamis pozitívumok egyik mérőszámának kiegyenlítése elkerülhetetlenül a hamis pozitívumok egy másik mérőszámának egyenlőtlenségéhez vezet. Ezt a rejtélyt, ahogy máshol már javasoltam, inkább a méltányosság, mint a pontosság problémájaként kellene jellemezni.²¹⁶ Mindenesetre nem oldható meg azzal, hogy visszatérünk a hibásabb emberi döntéshozatali protokollhoz. Ugyanazok a hibaeloszlások fordulhatnak elő, csak nagyobb számú téves pozitív eredménnyel, mint az emberi döntések esetében. Következésképpen ezt a problémát az emberi döntéshez való jogtól elkülönítve kell vizsgálni.

B. Tantárgyi szempontú indokok

Az emberi döntéshez való jog normatív alapjainak sokkal hihetőbb csoportja hozható létre, ha a gépi döntésnek alávetett személyre összpontosítunk. Ő ugyanis úgy érezheti, hogy megfosztja a hatalmától és elerőtleníti az, hogy kizárják minden hatékony szerepvállalásból abban a folyamatban, amelyben részt vesznek. Vagyis hedonikus veszteséget tapasztal a hatékony *részvétel* hiánya miatt. Másodsorban, egy automatizált döntés látszólag áthatolhatatlan az ember által felajánlott, a döntés szempontjából releváns indokokkal szemben. A gép, más szóval, kioltja az egyén által keresett *indoklás lehetőségét*. Az egyén közvetlenül a kezelésével és a gépi döntés pontosságával kapcsolatos indoklást is adhat, vagy pedig olyan magyarázatot is adhat, amely túlmutat az irányadó szabály formális hatályán. Vagyis előfordulhat, hogy az érintett személy a pontossághoz fűződő érdekét akarja igazolni egyedi információkkal, vagy pedig kivételt lát a szabály alól, még akkor is, ha az az ő esetére általában alkalmazandó lenne.

Itt azt vizsgálom meg, hogy az *alá* jog alapulhat-e a pusztán részvételhez való érdeken, vagy pedig a (tágra értelmezett) indokláshoz való érdeken.²¹⁷ Arra a következtetésre jutok, hogy mindkét alapon több okunk van a szkepticismusra, mint a reményre az emberi döntéshez való joggal kapcsolatban.

1. *Részvétel*

A *részvételhez* való pusztán jogot gyakran akkor is értelmesnek tekintik, ha azt nem lehet instrumentális vagy pontossági szempontból indokolni vagy magyarázni. E jogról azt mondhatjuk, hogy "abból a mélyen gyökerező történelmi hagyományból fakad, hogy mindenkinek meg kell kapnia a saját napját a bíróságon"²¹⁸, még akkor is, ha ez a részvételi jogosultság nem feltétlenül változtatja meg az eljárás kimenetelét. Ezt az érdeket megalapozhatja az az érzett emberi "szükséglet, hogy megmagyarázzuk és igazolni tudjuk cselekedeteinket", úgy, hogy "a

²¹⁵ Nagyjából különböző *idő* mérni a hamis pozitív és hamis negatív eredmények arányát, és a különböző mérőszámok szinte elkerülhetetlenül különböző irányokba mutatnak. Ennek matematikai bizonyítására lásd Sam Corbett- Davies et al., *Algorithmic decision making and the cost of fairness*, in PROCEEDINGS OF THE 23RD ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (797,7982017); Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Trade-offs in the fair determination of risk scores* (4,92016), <https://arxiv.org/abs/1609.05807>.

²¹⁶ *Lásd pl.* Huq, *Supra* note at 45,- (elemzi, hogy a népességben belüli pontosság *fogalmi* hogyan alkalmazhatók a büntető igazságszolgáltatás kontextusában, és azt javasolja, hogy a faji egyenlőséget a faji kisebbségek összesített költségeinek minimalizálása segíti elő).

²¹⁷ Ezt a kifejezést úgy használom, hogy az indoklásnak számító okokat és a kifogásnak számító okokat is

magában foglalja. Ebben az összefüggésben semmi sem fordul meg a kettő közötti különbségen.

²¹⁸ Martin kontra Wilks, Egyesült Államok 490 (755,7621989).

az erre való lehetőség elvesztése ~~in~~értékelésünket."²¹⁹ A részvétel e nézet szerint azonnali hedonikus nyereséget hoz létre.²²⁰ Alternatívaként a részvétel úgy is megfogalmazható, mint "az [egyenrangú állampolgárként]" való "erkölcsi felelősség gyakorlása", amelyen a jog "erkölcsi igénye az állampolgár hűségére" nyugszik.²²¹ Ebben a nem instrumentális értelemben a részvételt az autonómia erkölcsileg fontos megnyilvánulásaként ²²²vagy a méltóság megnyilvánulásaként szépitették meg.²²³ E két általános fogalom alkalmazási körének átfedése utal a közös és egymást átfedő kétértelműségükre.²²⁴

A ~~indoklás~~ erősségének értékeléséhez megvizsgálom, hogy mikor és hogyan ismeri el ezt az érdeket a jelenlegi jog. Ez nem jelenti azt, hogy a mai jog az erkölcsi érték pontos mércéje. Nem is az. De durva iránymutatást nyújt arra vonatkozóan, hogy egy érdeknek általában mekkora súlyt tulajdonítanak. Ha a jog jelenleg nem ismer el egy részvételi érdeket olyan helyzetekben, amelyek hasonlóak azokhoz, amelyekben az emberi döntéshez való jog működne, és ha nincs komoly kifogás e mulasztás ellen, akkor ez némi ok arra, hogy szkeptikusak legyünk e joggal szemben. Ezzel szemben, ha a részvételi jogot elismerik, vagy erős érvek szólnak az elismerése mellett, ez némi ok arra, hogy úgy gondoljuk, a jogot komolyabban kell venni.

Az egyik hely, ahol a pusztán ~~való~~ jog ~~de~~ ha az ellentétes a vádlott érdekeivel - a joggyakorlatban megtestesül, az a hatodik módosítás szerinti ügyvédhez való jog.²²⁵ Ez utóbbi alkotmányos rendelkezés kiterjed az önképviselethez és a választott ügyvédhez való jog megteremtésére a büntetőeljárással összefüggésben, annak ellenére, hogy nincs okunk azt gondolni, hogy különösen az előbbi jobb eredményt hozna a vádlottak számára.²²⁶ Az autonómiával kapcsolatos indoklás erőssége az ilyen

²¹⁹ Mashaw, *Administrative Due Process*, *supra* note at 200,903.

²²⁰ *V.ä.* Amershi et al., *Supra* note at (182,111).~~az~~ algoritmus működésének átláthatósága összefügg a nagyobb felhasználói elégedettséggel).

²²¹ T.R.S. Allan, *Procedural Fairness and the Duty of Respect*, OXFORD 18 JEUUD. 497,509 (1998). Nem vagyok biztos abban, hogy a következőkben teljes választ tudok adni Allan állítására. Ha ez az állítás az a nyers állítás, hogy az állam erkölcsi legitimitása a jogi eljárásokban való személyes részvétel egy bizonyos formáján nyugszik, akkor valószínűleg nem cáfolható az általam alább kifejtett jogi és doktrinális példákkal. Mivel itt nem kínálok teljes körű elméletet az állam legitimitásáról, nem tudok teljes mértékben reagálni az állításnak erre a változatára. Elég, ha annyit mondok, hogy nem tartom általánosságban véve meggyőzőnek azt az elképzelést, hogy az államnak ahhoz, hogy legitim legyen, konkrét, személyes részvételt kell generálnia a tanácskozásában.

²²² Jane Rutherford, *The Myth of Due Process*, B.U72. L. 17 (1992) ("Amikor az egyén részt vesz a kormányzati döntéshozatalban, nemcsak arra van lehetősége, hogy befolyásolja a döntés pontosságát és növelje annak legitimitását, hanem arra is, hogy autonómiát gyakoroljon").

²²³ Mashaw, *Due Process Calculus*, *supra* note at 205,491.~~in~~ méltóságot a tisztességes eljárás fontos szempontjaként ~~in~~ Sanford H. Kadish, *Methodology and Criteria in Due Process Adjudication--A Survey and Criticism*, YALE66 L. J. (1957) (a méltóságot a tisztességes eljárás alapvető értékeként azonosítja)319,347

²²⁴ A ~~jog~~ egy másik megfogalmazásában Lawrence Solum a részvételnek az ítélezési rendszer legitimitására gyakorolt hatására összpontosított. Lawrence B. Solum, *Procedural Justice*, S 78. CAL. L. REV. 181,191,274 (2004) (azzal érvelve, hogy "a részvételi jog olyan okokból is igazolható, amelyek nem redukálhatók sem a részvétel pontosságra gyakorolt hatására, sem az ítékezés költségeire gyakorolt hatására").

²²⁵ *Lásd az Egyesült Államok Alkotmányának VI. módosítását* ("Minden ~~vádlottnak~~... a védelem érdekében ügyvédi segítséget kell kapnia.").

²²⁶ *Faretta v. California*, U.S 422806,819.~~hatodik~~ módosítás nem csupán arról rendelkezik, hogy a vádlott számára védelmet kell biztosítani; a vádlottnak személyesen biztosítja a jogot, hogy védekezzen."). A választott ügyvéd doktrínájához lásd: *Egyesült Államok kontra Gonzalez-Lopez*, U548.S. (140,1462006) ("[A hatodik módosítás] nem azt parancsolja, hogy a tárgyalás legyen tisztességes, hanem azt, hogy a tisztességesség egy bizonyos garanciáját biztosítsák -

nevezetesen, hogy a vádlottat az általa legjobbnak tartott ügyvéd védje meg."); *egyetértésben* Flanagan kontra Egyesült Államok, US465 (259,2681984) (azzal érvelve, hogy ez a jog "tükrözi

jogok továbbra is vitatottak a között. Egy nemrégiben ügyvédválasztási ügyben azonban a Bíróság többsége "alapvetőnek" minősítette a vádlottnak az ügyvédválasztáshoz való jogát.²²⁸

Az ezekben a határozatokban megdő részvételi érdek azonban nem feltétlenül a személyes részvétel, és semmiképpen sem az alperes által megfelelőnek tartott formában való részvétel kérdése. A hatodik módosítás változatlanul kiterjed mind az önképviselőre való jogra, mind pedig a saját ügyvéd megválasztásához való jogra. Ha a Hatodik módosítás az utóbbi formában nyilvánul meg (ami a leggyakoribb eset), akkor nem áll fenn a személyes részvételhez fűződő érdek. A tét ehelyett az, hogy az egyén egy adott eljárásban szabadon választhassa meg azokat az ügynököket (ügyvédeket), akik az egyén képviselőjét ellátják. Ez részvételi jogként írható le. Ez azonban elhallgatja az ügyvédek és az ügyfelek közötti ügynöki kapcsolatot, valamint a szakemberek által az ítélkezési kontextusban gyakorolt ellenőrzés mértékét. Ésszerű azt gondolni, hogy a hatodik módosítás által védett releváns autonómia a saját ügyvéd megválasztása; az önképviselőre való jog e nézet szerint a saját ügyvéd megválasztásához való általánosabb jog mellékterméke. Ez nem egy kifejezetten védett részvételi érdek megvalósulása.

Még ha a büntetőjogi vádlott önképviselőre való jogával, ezt a jogot rendkívül formalizált módon korlátozzák, ami "bizonyos távolságot" teremt a bírák és a felek között.²²⁹ Az önképviselőre való jogát gyakorló büntetőjogi vádlott nem gyakorolja azt a jogot, hogy saját magát bármilyen módon képviselje. Éppen ellenkezőleg, egy erősen korlátozott jogosultságra hivatkozik olyan körülmények között, amelyek között valószínűleg nem rendelkezik a tájékozódáshoz szükséges episztemikus kompetenciákkal. A hatodik módosításban foglalt, a saját büntetőeljárásban való részvételre való jog tehát erősen korlátozott. Nem egyértelmű, hogy ezt a jog normatív magjaként kell-e értelmezni, szemben az általános ügyvédválasztási rendszer létrehozásának nem várt mellékhatásával, amely nem szűkíthető le úgy, hogy kizárja az önképviselőt némi bonyolult szöveges vagy doktrinális lábmunka nélkül. E nézet szerint a büntetőeljárásban a hatodik módosításra hivatkozva való részvételre való jog nem úgy értendő, mint a részvételre való érdek önmagában, szemben a vádlott szabad ügyvédválasztásának sajátos (és marginális) megnyilvánulásával.²³⁰ Ami a büntetőjogi kontextusban érvényesül, az a polgári kontextusban is érvényesül. A polgári eljárásban való részvételre való jog valószínűleg leginkább úgy értelmezhető, mint a részvételre való jog egy meghatalmazott, például egy ügyvéd révén. (Egy pillanat múlva rátérünk a tisztességes eljáráshoz fűződő érdekekre.

a vádlott választásának alkotmányos védelme, függetlenül az eljárás objektív tisztességességével kapcsolatos aggodalmaktól"). Megjegyzendő, hogy ugyanez elmondható a választásokról általánosságban.

²²⁷ Vö. John Rappaport, *The Structural Function of the Sixth Amendment Right to Call Choice*, SUP2016. CT. REV. 117,118 (2016) (arra a következtetésre jutva, hogy "a méltányosság és az autonómia fenséges hangzású fogalmi, illetve - küszködnek a választási jogtanácsosi doktrína magyarázatával"), Erica J. Hashimoto, *Resurrecting Autonomy: The Criminal Defendant's Right to Control the Case*, B90.U. L. REV. 1147 (2010) (a doktrína védelme az autonómia szempontjából).

²²⁸ Luis kontra Egyesült Államok, S136. Ct. 1083, többségi vélemény).

²²⁹ Emily Buss, *The Missed Opportunity in Gault*, U70. CHI. L. REV. 47 (2003) ("Az eljárás formalitása és a döntéshozóknak a jog semleges döntőbírójává minősítése bizonyos távolságot biztosít a döntéshozó és a felek között, amelynek célja a meghozott döntések megbízhatóságának növelése, még akkor is, ha az méltóságteljes a felek és a szóban forgó érdekek szempontjából)."

²³⁰ Vö. Rappaport, *Supra* note at 227, ügyvédválasztási szabályt "gyenge, rendszerszintű biztosítékként

118jellemzi a büntetőjogi védelem szocializációja ellen").

értésítés és rendelkezés) az a különbség nem kívánatos.²³² De aligha támasztja alá azt az intuíciót, hogy a puszta részvételi jog eléggé széles körben elfogadott ahhoz, hogy igazolja az emberi döntéshez való jogot.

Az ítélezés legtöbb formája ráadásul kellően összetett és formális vállalkozással is, hogy a közvetlen, személyes hatásokkal járó döntésekben való személyes részvételhez való jog nem hihető. Amint azt a hatodik módosításhoz fűződő jogok közelmúltbeli elismerése a vádalku kontextusában mutatja, ²³³még az informális vitarendezés a bírósági kontextusban is megköveteli a technikai szakértelem olyan csekély mértékét, amellyel kevés nem jogász egyéni pereskedő rendelkezik. Az ítélezésben való *személyes* részvételhez való jog nem egyeztethető össze ezzel az összetettséggel.

Mindezek fényében nehéz hogy a nem instrumentális, részvételen alapuló érv milyen erővel szól az emberi döntéshez való jog mellett, még akkor is, ha az élet vagy a szabadság forog kockán. A gépi tanulás eszközei, emlékeztetőül, talán nem sokkal átláthatatlanabbak, mint az emberi döntéshozók.²³⁴ Hasonlóképpen, nem feltétlenül van nagy fenomenológiai távolság a zavartalanság között, amelyet egy tanulatlan büntetőjogi vádlott ésszerűen érezhet, amikor szembesül a büntető igazságszolgáltatási rendszer hálószerű és összetett formáival, és az algoritmikusan levezetett eredmény által kiváltott zavarodottság között. Minden gyakorlati szempontból mindkettő fekete doboz. Ráadásul, amennyire meg tudom ítélni, nincs olyan valódi társadalmi mozgalom, amely a büntetőeljárás egyszerűsítését célozná - ami nem egyenlő azzal, hogy igazságosabb vagy kedvezőbb lenne a vádlottak számára - a saját *érdekében*. Ha a nem instrumentális részvételhez fűződő érdek rövidre zárul olyan körülmények között, ahol az emberi szabadság gyakran közvetlenül veszélyeztetve van, meglepő lenne, ha kiderülne, hogy a személyes részvételhez fűződő erős emberi érdek átteődik a gépi tanulás kontextusába.

A emberi döntéshez való *g*zembeni ellenállás utolsó oka az állítás elégtelensége. Az utólagos emberi döntés gyakran nem az optimális válasz az érzett pszichológiai veszteségre. Gyakran egy diszkrét döntés "magyarázatának" legjobb módja nem az emberi felülvizsgálat, amely korlátozottan világhat rá egy összetett algoritmus működésére. Ehelyett legalább egy algoritmikus eszközt fejlesztettek ki, amely "bármely osztályozó vagy regresszor előrejelzéseinek hűséges megmagyarázására" szolgál.²³⁵ Amikor a gépi tanulás kimeneteinek érthető magyarázatát legvalószínűbben adó eszköz maga is egy gép, akkor az érintettek sérelmeit diagnosztizáló vagy "csillapító"²³⁶ ember iránti igénynek nem a magyarázat iránti igényen, hanem inkább az emberi beszélgetőpartner iránti nyers és indokolatlan igényen kell alapulnia. Ez az emberi interakció iránti igény azonban a társadalmi tapasztalat esetleges jellemzője lehet, és ami ma dehumanizálónak vagy érzéketlennek tűnik számunkra, az a jövőben

²³¹ Például a "távollétében" *h*utasítási eljárások a C8.íme 8 U.S.C.A. § 1229a(5)(A) (2018).

²³² Ennek ellenére hajlok arra, hogy a *h*örténő *h*nmagában kifogásolható.

²³³ *Lásd* Padilla v. Kentucky, U.S.559. (356,2010) (a hatodik módosításnak a védőügyvédhez való jogát alkalmazza a bevándorlási következményekkel járó vádalkura). A Padilla-ügynek a bevándorlási kontextuson túli kiterjesztéséről lásd Missouri v. Frye, S132. Ct. 1399,1408 (2012); Lafler kontra Cooper, S132. Ct. 1376,1388 (2012).

²³⁴ *Lásd a fenti szöveg kísérő megjegyzés*145eit 173.

²³⁵ Marco Tulio Ribeiro, Sameer Singh & *M*ért *h*ézzak benned?: *Explaining the predictions of any classifier*, PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (1135,11352016).

236 Matthew J.B. Lawrence, *Kötelező eljárás*, IND90. (20151432).

a gyermekeink számára csupán ésszerűnek és ~~h~~Jelenleg az emberi felülvizsgálat iránti igény a várható költségek és a rendelkezésre álló alternatív válaszok tükrében aligha tűnik többnek, mint esztétikai preferencia az állami szereplőkkel való érintkezés módjára vonatkozóan. Nem vagyok benne biztos, hogy ez elég ahhoz, hogy az emberi döntéshez való jog elinduljon.

Talán van mód ~~h~~pszichológiai érvelésnek a ~~h~~Talán egy gépi döntésnek alávett egyén nem azért kérhet utólagos emberi felülvizsgálatot, mert a döntést helytelennek tartja, hanem mert reméli, hogy a külső tényezők a döntés következményeinek enyhítésére készítenek. Más szóval, kegyelmet keres - vagy a hatósági mérlegelési jogkör gyakorlását egy olyan jogi következmény enyhítésére, amely egyébként az ember törvényes sorsa.²³⁸ A kegyelemre való remény azonban gyenge alapja egy emberi döntéshez való "jognak". Az irgalom alatt általában olyan *mérlegelési jogkörrel járó* cselekedetet értünk, amelyre az embernek nincs jogosultsága. Mint ilyen, az irgalom egyre nehezebben megfogható tulajdonság még a büntetőjogi kontextusban is, ahol valószínűleg a legmeggyőzőbb hellyel rendelkezik.²³⁹ Lehet, hogy a jelenlegi büntetőjogi megítélése indokolatlan, de még így is nehéz belátni, hogy az irgalom felélesztése miért az algoritmikus területen kezdődne.

2. *Indoklás Adás*

De talán hiba lenne ~~h~~részvételre nem instrumentális szempontból gondolni. Ehelyett talán jobb lenne inkább azokra a kézzelfogható módokra összpontosítani, amelyekkel a részvétel megváltoztathatja az eredményeket. Erre több lehetséges érv is kínálkozik. A legnyilvánvalóbb, hogy a részvétel azért fontos, mert az egyén tényszerűen hozzájárulhat ügyének elbírálásához. Az ilyen hozzájárulás legfontosabb formája az egyén azon képességén múlik, hogy képes-e információt szolgáltatni, és általánosabban szólva indokokat felhozni a számára kedvező döntés meghozatalához.²⁴⁰ Emlékezzünk vissza, hogy ez az ereje Cathy O'Neill anekdotájának a jóléti kérelmet benyújtó Catherine Taylorról, amelyet korábban²⁴¹ elmeséltem: Feltételezzük, hogy csak Taylor rendelkezett a saját ügyének helyes eldöntéséhez szükséges információkkal, és csak az emberi felülvizsgálat tudta volna ezeket az információkat kicsikarni. Ez az érvelés az észérvek megadásából származó érv szorosán kiegészíti az emberi döntéshez való jog leghihetőbb technológiai értelmezését. Amint azt a II. rész feltárta, az emberi döntések áthatják és strukturálják a gépi tanulási eszközöket.²⁴² A további emberi közreműködésre való igény leghihetőbb glosszája a gépi tanulóval kapcsolatos döntés meghozatalát követő utólagos emberi szerepvállaláson kell, hogy alapuljon. Ez azt jelentené, hogy a gépi döntésnek alávett egyén közvetlenül beszélhetne

²³⁷ A technológiai eseményhorizonton belül vannak azok a robottársak, amelyek képesek felismerni ~~h~~magálni védenekik mutatív érzelmi állapotára. A tudomány jelenlegi állásának áttekintését lásd Giulio Sandini et al., *Social Cognition for Human-Robot Symbiosis-Challenges and Building Blocks*, FRONTIERS IN NEUROBOTICS (342018).

²³⁸ Aziz Z. Huq, *A demokratikus irgalom nehézségei*, CAL103. L. ~~h~~681 (2015). Az irgalom általában "a megérdemelt büntetés részben vagy egészben történő elengedését jelenti a bűncselekményt elkövetők számára olyan jellemzők alapján, amelyek együttérzést vagy szimpátiát váltanak ki, de amelyek erkölcsileg nem kapcsolódnak az elkövető kompetenciájához és ahhoz a képességéhez, hogy a bűncselekmény elkövetését válassza". Dan Markel, *Against Mercy*, MINN88. L. REV. 1421,1436 (2004). Lesznek olyan esetek, amikor az így meghatározott kegyelem még azután is gyakorolható, hogy egy algoritmikus osztályozót alkalmaztak a határozat meghozatalához.

²³⁹ A kegyelem hanyatlásáról lásd Rachel ~~h~~*The Ascent of the Administrative State and the Demise of Mercy*, HARV121. L. REV. 1332, 1348-49 (2008). Az esküdszék mint enyhítő intézmény hanyatlásáról lásd: Nancy J. King, *Silencing Nullification Advocacy Inside the Jury Room and Outside the Courtroom*, U65. CHI. L. REV. 433, 492-94 (1998).

²⁴⁰~~h~~fejezetben az indokoláshoz való jog tárgyalásakor nem teszek különbséget az egyén azon képessége között, hogy empirikus bizonyítékokat tud felajánlani, és azon képessége között, hogy olyan jogi vagy normatív állításokat tud megfogalmazni, amelyek nem a világról szóló információkon alapulnak. Gyanítom, hogy a gyakorlatban az érvelés a tényszerű és a normatív állítások valamilyen keverékét foglalja magában. Ezek megkülönböztetése itt nem szolgál hasznos célt.

²⁴¹ *Lásd a fenti szöveg kísérő megjegyzés17eit* 18.

²⁴² *Lásd a fenti szöveg kísérő megjegyzés175eit* 188.

a döntéshez azáltal, hogy felhívja rá az ügy ilyen jog reagálna például arra a lehetőségre, hogy az algoritmus képzéséhez használt "jellemzőértékek" kizárnak valamilyen paramétert, amely az egyének egy részhalmaza számára releváns, de az általános populáció számára nem.²⁴³

Az által jog külön-külön is védhető az eredménytől független, de instrumentális okokból. Hosszabb távon például a részvétel balzsamként működhet azok számára, akiknek az ügye megingott.²⁴⁴ Ezzel a nézettel összhangban a Tom Tylerhez kapcsolódó eljárás igazságossággal foglalkozó szakirodalom azt az empirikus állítást hangoztatta, hogy a hivatalos meghallgatás lehetősége az interakciót követően a jogkövetés magasabb arányával jár együtt.²⁴⁵ Az eljárás igazságosság nézete szerint a rendszerszintű legitimitás az indoklás lehetőségének a jogrendszerrel való állampolgári interakciók szerkezetébe való beágyazásából ered.²⁴⁶

Az érvek bármelyik formája lehet alapot a Due Process Clause-ban.²⁴⁷ A meghallgatás követelménye természetesen nem terjed ki *minden* olyan állami döntésre, amely közvetlenül és közvetlenül érinti az egyéneket. Az ügynökségek által kihirdetett jogszabályok és törvényszerű rendeletek például drámaian és azonnal megváltoztathatják az egyén jogait, kötelezettségeit és a kényszerítő kockázatnak való kitétségét. Az egyéneknek nincs joguk az egyénre szabott részvételre velük kapcsolatban.²⁴⁸ Nem teljesen egyértelmű a határvonal a megengedhető jogalkotói döntések (amelyekre a tisztességes eljárás nem vonatkozik) és a kormányzati intézkedések azon formái között, amelyekre az vonatkozik. ²⁴⁹ De létezik. Egyes kommentátorok például azt sugallják, hogy "a tisztességes eljárás szóbeli meghallgatást követel meg, ha kisszámú embert érintő, bírói tényeken alapuló, egyedi megfosztásokról van szó²⁵⁰", függetlenül attól, hogy az eredményt jogszabályként nevezik-e meg. Legalábbis a büntetőjogi és szociális jóléti döntések egy része, amelyekhez jelenleg algoritmikus eszközöket használnak, egyértelműen ebbe a területbe tartozik. Az indokoláshoz fűződő egyéni érdek ráadásul különbözik a fentiekben²⁵¹ tárgyalt, a pontossághoz fűződő általánosabb érdektől:

²⁴³ Domingos, *Supra* note, 97,78. o.; lásd még LeCun et al. note,33, (a funkció kiválasztását 436tárgyalja az algoritmikus tervezésben).

²⁴⁴ Egy hisselkedésgazdaságtani szempontból hangzó instrumentális érvert lásd Matthew J.B. Lawrence, *Mandatory Process*, IND90. L.J. (1429,20151432) ("A vitarendezési eljárásban való részvétel eredendő értéke részben abból a hatalmából fakad, hogy megnyugtatja az ilyen sérelmeket, ha azok bekövetkeznek, akár győznek, akár veszítenek.").

²⁴⁵ TOM R. TYLER & YUEN J. HUO, TRUST IN THE LAW 11 Számos bizonyíték van arra, hogy ha az emberek úgy tekintenek a jogrendszer azon szereplőire, akikkel személyesen találkoznak, mint akik az általuk igazságosnak vélt módon és olyan motívumok által vezérelve cselekszenek, amelyekről úgy vélik, hogy megbízhatóak, akkor hajlandóbbak engedelmeskedni az utasításaiknak"). A ennek a hatásnak a bizonyítékait nemrégiben Daniel S. Nagin & Cody W. Telep, *Procedural Justice and Legal Compliance*, ANN13. REV. L. & SOC. SCI. 5 (2017).

²⁴⁶ Stephen J. Schulhofer et. al., *American Policing at A Crossroads: Fenntarthatatlan politikák és az eljárási igazságszolgáltatás alternatívája*, J101. CRIM. L. & CRIMINOLOGY (335,3452011) ("Az empirikus kutatások azt mutatják, hogy ezt a fajta legitimitást nem az egyéni jogokat alárendelő agresszív stílus tartja fenn, hanem inkább valami, ami közelebb áll ennek ellentétéhez - az eljárási igazságosság címszó alá sorolható gyakorlatok.").

²⁴⁷ A legfontosabb esetek a *Goldberg v. Kelly*, U.S397; *Bell v. Burson*, U402.S. (5351971); és *Arnett v. Kennedy*, U.S416. (1974134).

²⁴⁸ Lásd *Bi-Metallic Inv. Co. kontra State Bd. of Ind.* (4411915); *Londoner kontra City of Denver*, U210.S. (3731908).

²⁴⁹ Henry Friendly, "*Some Kind of Hearing*"123, U. PA. L. 267, 1276-77 (1975) ("[I]t tűnik jelenleg lehetetlennek megíjósolni, hogy a Bíróság milyen szinten fogja meghatározni azt a küszöböt, amely alatt nincs szükség meghallgatásra.").

²⁵⁰ Adrian Vermeule, *Conventions of Agency Independence*, COLUM113. 1163,1213 (2013).

²⁵¹ *Lásd fentebb a III.A. részt.*

Az állítása nem egészen az, hogy a gépi összességében pontatlan, hanem inkább az, hogy az emberi felülvizsgálat hozzáadása kiküszöbölheti a hamis negatív (pozitív) eredmények egy osztályát a szabályozott alanyok birtokában lévő magáninformációk felhasználásával.

Véleményem szerint ez a legerősebb érve emberi döntéshez való jog ellenében. Az a tény, hogy az állami ügynökségek ösztönözhetik a mélyen hibás gépi tanulási rendszerek elhamarkodott bevezetését, csak növeli a vonzerejét. Anélkül azonban, hogy elvetném a Catherine Tayloréhoz hasonló esetek által felvetett normatív aggályokat, szeretném visszautasítani azt az elképzelést, hogy az emberi döntéshez való jog kényszerítő válasz lenne. Különösen szeretnék ellenállni annak a kísértésnek, hogy a Tayloréhoz hasonló esetekben az emberi felülvizsgálatból származó rövid távú nyereséget összemossuk a dinamikus optimalitás kérdésével: Azaz, hogy egy ilyen jog elfogadása hosszú távon hogyan alakítja ki a kívánatos és nem kívánatos eredményeket, és (implicit módon) van-e valamilyen jobb alternatív eszköz a normatív aggályok kezelésére.

Az első, kétségtelenül gyenge ok a távlatokra doktrinális jellegű. Nem egyértelmű, hogy a Due Process Clause kiegészítő emberi cselekvést követel meg a kellően pontos gépi döntések biztosításához. Az ítélkezési formák, amelyeket a megfelelő eljárás felvehet, gyakran felületesebbek.²⁵² Egyáltalán nem nyilvánvaló, hogy az algoritmikus eszközök - a Hatodik Alkotmánymódosítás esküdtstészi tárgyaláshoz való jogának sajátos kontextusán kívül - nem képesek biztosítani az alkotmányosan szükséges tisztességes eljárást (kivéve, ha az eredményeik a mért minőséghez képest teljesen ortogonálisak). Ez a jogi következtetés azonban vitathatóan viszonylag kis súlyt érdemel. A tisztességes eljárás joggyakorlata ugyanis lehet, hogy egyszerűen téves, és a technológiai változások fényében frissítésre szorul.

A második oknak nagyobb súlya van. Az emberi döntésnek a gépi döntést helyettesítő eszközként olyan perverz és nemkívánatos következményekkel járhat a szabályozott népességre nézve, amelyek meghaladják a részvétellel kapcsolatos előnyöket.²⁵³ Sőt, a helyzet empirikus jellemzőitől függően az emberi döntés hozzáadása az Adam Samaha által "indokolatlan eljárásnak" nevezett formává válhat, amely "a kormányzati eljárás alkotmányosan előírt felső határait" sérti.²⁵⁴ Ez a második érv azon az előfeltevésen alapul, hogy nem minden eljárás "megfelelő". Ahogy Henry Friendly bíró a meghallgatás követelményéről szóló kánoni elmékedésében megjegyezte, az eljárás minden további fokozata költséggel jár, mivel "az eljárási követelmények korlátozott erőforrások ráfordításával járnak, [így] egy bizonyos ponton az egyének számára egy további biztosítékból származó előnyöket lényegesen ellensúlyozzák az ilyen védelem biztosításának költségei".²⁵⁵ Amennyiben egy eljárási lépés hozzáadása várhatóan rendszerszintűen növeli a hibaarányok általános gyakoriságát, vagy valamilyen más költséget generál, indokolt lenne megállni és jogilag és közrendi szempontból felülvizsgálni a felhatalmazást.

²⁵² Goss v. Lopez, U.S.419. (565,5791975) "Valamiféle értesítést és valamiféle meghallgatást" a fegyelmezett állami iskolai tanulók számára, anélkül, hogy sokkal részletesebben részleteznék).

²⁵³ A jogoknak gyakran a végrehajtással kapcsolatos káros hatásuk van, amelyek másokra is áttérnek, akik nem gyakorolják a jogot. Például a büntetőeljárási jogok költségesebbé tehetik a bűnüldözést, és így csökkentik azt a mértéket, amennyire az állam képes a közbiztonságot mindenki számára megteremteni. Az érvelés itt egy másik lehetőségre összpontosít: arra, hogy egy jog gyakorlása olyan költségekkel jár, amelyek a jogból feltételezhetően hasznot húzó népességre terjednek ki.

²⁵⁴ Adam M. Samaha, *Undue Process*, STAN59. L. REV. (2006). Samaha gondosan elismerte a lehetőség újszerűségét és a fennálló doktrínával való feszültségét. Itt nem annyira azért hivatkozom az ötletére, hogy azt sugalljam, hogy létezhet felismerhető "indokolatlan eljárásra" vonatkozó igény, hanem inkább azért, hogy rámutassak az

emberi részvételhez való ragaszkodás perverzítésára bizonyos esetekben.

²⁵⁵ Friendly, *Supra* note at 76,1276.

Vegyük először a gépi döntéstanulási folyamat összes kimenete esetében egy ~~mag~~ döntéshozó hatását a nettó hamis pozitív és hamis negatív arányok szempontjából.²⁵⁶ Természetesen "még a jól megtervezett" algoritmikus eszközök is követnek el hibákat.²⁵⁷ De egy jól megtervezett gépi döntés emberi háttérintézkedéssel való kiegészítése nem feltétlenül növeli a pontos ítéletek általános arányát. Mint említettük, a gépi döntések gyakran kevésbé hibásak, mint a közeli emberi helyettesítők.²⁵⁸ Nem biztos, hogy az ember képes lesz azonosítani és kijavítani az összes olyan esetet, amelyben a gép hibázott. Éppen ellenkezőleg, fennáll annak a reális lehetősége, hogy az emberi beavatkozás *magasabb* hibaarányhoz vezet. Továbbá, ha az egyik osztályozás elérésekor emberi joggal számolunk, a másiknál pedig nem, fennáll a veszélye annak, hogy az ebből eredő hibák egyenlőtlenül oszlanak el a populációban. Ha nincs okunk azt gondolni, hogy a gép maga is elfogult volt, nehéz belátni, hogy a magasabb, aszimmetrikusan eloszló hibaarány hogyan lenne kívánatos. Továbbá, ha egy algoritmikus eszköz hibás, ebből nem következik, hogy ezért "jár" az utólagos emberi felülvizsgálat. Sőt, minden okunk megvan azt hinni, hogy ami "jár", az inkább egy jobb gépi döntés, mint egy megbízhatóan megbízhatatlan emberi döntés. Azaz, még ha a gépi döntés hibás is, a várható pontosság és a részvételi kereskedelem közötti várható kompromisszum valószínűleg az emberi döntés ellen fog dönteni.

Az elemzés motiválására tekintsünk ~~is~~ esetet, amelyben egy nem gépi algoritmust kiegészítettek az emberi felülvizsgálat utólagos jogával. A 2007. Legfelsőbb Bíróságtól kezdve a szövetségi bíróságokon kiterjesztette a bírák mérlegelési jogkörét a büntetékiszabás tekintetében, elutasítva egy nagyon elemi algoritmus, a szövetségi büntetékiszabási irányelvek kötelező erejét.²⁵⁹ A bírák közötti ítélezési különbségek ezt követően meredeken megnöttek, egyes joghatóságokban majdnem megduplázódtak.²⁶⁰ A faji egyenlőtlenségek is megnöttek.²⁶¹ Talán van itt egy szerencsés történet arról, hogy a bírák hajlamosak az ítéleteket az elkövetők számos olyan jellemzője alapján összevetni, amelyek túlmutatnak az ítélethozatali jelentésben foglaltakon. Ezt kétlem. A mai napig minden okunk megvan arra, hogy úgy gondoljuk, a bírói mérlegelésnek megdőbbséget és társadalmilag romboló hatásai vannak.²⁶² Hozzáadás:

²⁵⁶ Emlékeztetőül, ez volt a ~~B~~íróság álláspontja az algoritmikus ítélethozatali eszközökkel kapcsolatban. Állam

v. Loomis, N.W881.2d (749,7602016); *lásd az* alábbi⁷⁹ szöveghez ~~is~~.

²⁵⁷ Kroll, *Supra* note at 152,11.

²⁵⁸ *Lásd a fenti* szöveg kísérő megjegyzés¹³⁶eit 138.

²⁵⁹ United States v. Booker, U543.S. (220,2005~~3~~) (a szövetségi büntetési irányelvek kötelező erejének érvénytelenítése); *lásd még* Rita v. United States, U551.S. 350-55338, (2007) (az irányelveken belüli ítéletek ésszerűségének fellebbviteli véelme); Gall v. United States, U552.S. 48-4938, (2007) (az irányelveken kívüli ítéletek fokozott fellebbviteli felülvizsgálatának elutasítása); Kimbrough v. United States, U.S552. 109-1085, (2007) (az ítélethozatal törvényi korlátozásának elutasítása).

²⁶⁰ Crystal S. Yang, *Növekedtek-e a bírák közötti ítélezési különbségek a tanácsadói iránymutatások ~~az~~ Evidence from Booker*, N.Y89.U. L. REV. 1268,1333 (2014) (megállapítva, hogy "a bírák közötti egyenlőtlenségek megduplázódtak a kötelező iránymutatás szerinti ítélezés időszakától a Booker utáni ítélezésig, ahol a vádlott potenciálisan hat hónappal hosszabb büntetést kaphat, ha véletlenül szigorú bíróhoz kerül"); *accord* Ryan W. Scott, *Inter-Judge Sentencing Disparity After Booker: A First Look*, STAN63. L. REV. 1,3, 52-53 (2010) (hasonló eredmény egy massachusettsi kerületi bíróság esetében).

²⁶¹ Crystal S. Yang, *Végre szabadon? Judicial Discretion and Racial Disparities in Federal ~~J~~ 44. LEGAL STUD. 75, 77 (2015)* (a Booker utáni "jelentősen ~~negatív~~ egyenlőtlenségek megállapítása a széleskörű elkövetői és bűnügyi jellemzők ellenőrzése után"); *accord* Max Schanzenbach & Emerson H. Tiller, *Strategic Judging under the United States Sentencing Guidelines: Positive Political Theory and Evidence*, J23. L & ECON. & ORG. 23 (2007) (hasonló megállapítás).

²⁶² Nem akarom azt állítani, hogy a Booker előtti ~~hibátlan~~ volt. A bírák közötti, következetlen büntetési

preferenciákból eredő egyenlőtlenségek azonban nem tűnnek annak. Az egyenlőtlenségek pusztán ténye önmagában nem bizonyítja a rendszer hibás voltát. Annak értékelése, hogy mikor indokolatlan az eltérés, "megköveteli, hogy tudjuk, miért büntetünk". Kevin Cole, *The Empty Idea of Sentencing Disparity*, NW91. U. L. REV. 1336,1337 (1997). Legalábbis a hasonló vádlottak közötti faji egyenlőtlenségekre vonatkozó adatok jelentős kérdéseket vetnek fel azzal kapcsolatban, hogy azonosíthatók-e megfelelő indokok.

az emberi input egy (egyszerű, nem gépi) hibát ártott, mint használt. Ezen a példán keresztül látható, hogy az emberi döntéshez való jog mellett szóló érvek, amelyek egy konkrét személyre összpontosítanak, akitől jogtalanul megtagadtak egy juttatást, félrevezetőek.²⁶³ Nem egy személy esetével kell foglalkoznunk, hanem a rendszer által hozott helytelen emberi vagy gépi döntések összességével.

Tekintsük most ennek az egyik változatát: ~~A~~ hogy minden eredmény utólagos emberi felülvizsgálatát engedélyoznánk, csak azok az esetek váltanának ki emberi beavatkozást, amikor a szabályozott alany hátrányt szenved. Vagyis a jog csak a világ Catherine Taylorjait illetné meg, és nem azokat, akiknek előnyököt biztosítanak. Ezt az aszimmetriát ki lehetne élesíteni. Az algoritmus az emberi felülvizsgálatot az utóbbi költségeinek korlátozására szolgáló eszközként titrálhatná azáltal, hogy elkülöníti az egyének azon részhalmazát, amelyek valószínűsíthetően olyan kiemelkedő, új tényeket kínálnak fel, amelyek más eredményt eredményezhetnek. Ezért egy olyan algoritmus, amely egy kockázati paramétert folytonos változóként generált, egy numerikus küszöbértékkel rendelkezik majd osztályozási szabályként.²⁶⁴ Az e küszöbértéket kis mértékben meghaladónak minősített egyének fellebbezhetnek. Azoknak, akik a küszöbértéket nagymértékben átlépték, nem lenne joguk emberi döntéshez. Röviden, magát az algoritmust úgy terveznék meg, hogy kiválassza a negatív kimenetek azon részhalmazát, amelyek esetében az utólagos emberi felülvizsgálat kiegészítése a pontosság növelése szempontjából indokolt lehet. Miért *ártana* az emberi felülvizsgálathoz való jog ebben a szűk, aszimmetrikus formában? Vagyis hogyan nem lenne indokolt?

Továbbra is szkeptikus vagyok. Először is, ~~az~~ aszimmetrikus és szűk körű utólagos emberi felülvizsgálat mellett) azt feltételezi, hogy az emberi felülvizsgálat korrigálja a téves pozitív eredményeket, és csakis a téves pozitív eredményeket. De nincs okunk feltételezni, hogy csak a kedvezőtlen döntés érdemi alanyai fognak fellebbezni. Ezért lehetséges, hogy még a szűken kalibrált emberi felülvizsgálat is növeli a hibák nettó mennyiségét azáltal, hogy több valódi pozitív, mint téves pozitív döntést változtat meg. Vagyis a Catherine Taylors osztálya nem fog mindannyian könnyítést kapni, míg az érdemtelen kedvezményezett egy osztálya, akiknek a kérelmét helyesen utasították el - nevezzük őket Elizabeth Taylorsnak -, érvényesülni fog. Hogy ez ne hangozzék valószínűtlennek, gondoljunk bele, hogyan fog működni a szelekció az emberi fellebbezésekben. Nincs okunk a priori azt gondolni, hogy csak és kizárólag a relevánsan helyes magáninformációval rendelkezők fognak utólagosan emberhez fordulni. Sokkal valószínűbb azt gondolni, hogy a vagyon, az ismeretszerzési erőforrások vagy a társadalmi osztály megjósolja a fellebbezésre való hajlamot.²⁶⁵ Hogy ez összességében több vagy kevesebb hibát eredményez-e, azt csak találgatni lehet. Valójában az Elizabeth Tayloroknál más hamisítási arányokat kellene várni, mint a Catherine Tayloroknál. Az eredmény inkább egy összevonó egyensúly lesz, mint egy szétválasztó egyensúly, amelyben az emberi döntéshozó a valós és hamis magáninformációk keverékével találkozhat. Megint csak meglehetősen fantasztikus azt gondolni, hogy az emberi döntéshozónak költség- és súrlódásmentes mechanizmusa van arra, hogy kiválassza a komoly Catherine-eket a machiavellista Elizabeth-ektől. Ami O'Neill küszöbpéldájából egyszerű, viszonylag költségmentes lépésnek tűnik, az a vizsgálat során tehát rendkívül problematikusnak és korántsem súlytalannak bizonyul.²⁶⁶ Az utólagos emberi felülvizsgálat ezen optimális eseteinek empirikus és technológiai feltételezései tehát a valószínűtlenségig igényesek.

²⁶³ Lásd a fenti megjegyzést és a 17kísérő szöveget.

²⁶⁴ Camelia Simon, Sam Corbett-Davies & ~~the~~ *The Problem of Infra-Marginality in Outcome Tests for Discrimination*, ANN11. APP. STAT. 1193,1194 (2017) (példát mutat be).

²⁶⁵ A jelzések ~~is~~ a pooling és a szeparáló egyensúly lehetőségeire vonatkozó eredeti felismerés Michael Spence, *Job Market Signaling*, Q87. J. ECON. 355, 362-63 (1973); lásd még MARTIN J. OSBORNE & ARIEL

RUBINSTEIN, A COURSE IN GAME THEORY (2381994) (a Spence-féle jelzőjáték és az ebből eredő pooling és separating egyensúlyok leírása).

²⁶⁶ A probléma technikai megoldása költséges. Képzeljünk el egy optimális gépi döntést, amelyet úgy terveztek meg, hogy emberi értelmezéssel is értelmezhető legyen. *Lásd* Desai & Kroll, *Supra* note, 153, 61. pont 11. Ez a gép "egy manipulálhatóan egyértelmű

Másodszor, ~~h~~ük körű emberi felülvizsgálat felszínes vonzereje (talán hallgatólágoosan) azon a feltételezésen alapul, hogy a valódi pozitív eredmények visszafordítása nem jár költségekkel. Ez azonban ritkán lesz így. Az óvadékkal kapcsolatos kontextusban például egy valódi pozitív eredmény visszafordítása súlyos erőszakos bűncselekmény elkerülhető elkövetéseként jelentkezhet. A jóléti kontextusban ez azt jelenti, hogy egy érdemtelen személy olyan juttatáshoz jut, amelyet egyébként egy rászoruló kedvezményezett kaphatott volna. Bármely rendszerben kívánatos gondolkodás lenne a visszafordított pozitív eredmények költségeit elhárítani, bár gyanítom, hogy az emberi felülvizsgálathoz való jog vonzereje e költségek bizonyos mértékű elfojtásával jár.

Harmadszor, gyakran előfordulhat, hogy ~~g~~z aszimmetrikus, szűkre szabott emberi joghoz való jog ~~v~~agy kezdettől fogva haszontalan lesz, vagy a megerősítő tanulással irrelevánssá tehető. Az egyén lehetősége, hogy indokokat szolgáltatson egy emberi döntéshozónak, csak akkor releváns, ha ezek az indokok bizonyos valószínűséggel befolyásolják a folyamat kimenetelét. Számos olyan döntés esetében azonban, amelyeknél algoritmusokat lehet alkalmazni hivatalos kezekben, például az ellátásokra való jogosultság vagy a feltételes szabadlábra helyezés visszavonása esetében, a törvény a releváns paraméterek zárt körét határozza meg. Vagyis a jog gyakran alkalmaz olyan szabályokat, amelyek a normákkal szemben a releváns tények egy rögzített halmazát választják ki előzetesen. Az utóbbiak lehetővé teszik a felsorolt és nem előre nem látható tényezők nyitottabb mérlegelését.²⁶⁷ Amennyiben egy algoritmus ilyen értelemben inkább jogi szabályt, mint szabványt alkalmaz, nem világos, hogy miért számítanak azok az újszerű okok vagy tények, amelyekre a besorolás alá vont személy utólagosan szeretne rámutatni. Maga az a tény, hogy a vonatkozó jogként egy szabályt és nem egy normát választanak, kizárja azt az állításukat, hogy a váratlan tények relevánsak. Még ha az algoritmust úgy tervezték is, hogy egy normát alkalmazzon (pl. a veszélyesség az óvadékkal összefüggésben), előfordulhat, hogy e norma rutinszerű alkalmazása esetek százain vagy ezerein keresztül olyan alszabályokat hoz létre, amelyek paraméterek zárt és kiszámítható csoportjain alapulnak.²⁶⁸ Ezek az alszabályok lefedik a lehetséges tényeknek azt a területét, amelyet a gépi döntés szempontjából kiemelkedőnek állítanak. Az utólagos emberi felülvizsgálat így ritkán ad hozzá bármi értékeset.²⁶⁹ A zárt jogi szabályok és a funkcióválasztás kölcsönhatása ezért feleslegessé teheti bármelyik

rekord, amely megdönthetetlen bizonyítékot szolgáltat ~~h~~selekedetéről." Haeberlen, Kuznetsov & Druschel, *supra* note at153,

175. Ezt a rekordot ezután ~~n~~em lehet vizsgálni megállapítására, hogy történt-e hiba, vagy az osztályozási szabályban használt adatparaméterek nem rögzítettek-e egy adott paramétert, amely az osztályozó által rangsorolt egyén szempontjából releváns.

²⁶⁷ Louis Kaplow, *Szabályok kontra normák*: L.J559-635(1992) (a szabály meghatározása szerint a szabály olyan jogi norma, amely a szabályozott alanyok cselekvése előtt kap tartalmat, míg a szabvány olyan jogi norma, amely a szabályozott alanyok cselekvése után kap tartalmat).

²⁶⁸ Az ilyen szabály-szabvány "ciklikusság" számos jogi területen ~~n~~eg. Lásd Aziz Z. Huq & Jon D. Michaels, *The Cycles of Separation-of-Powers Jurisprudence*, YALE126 L.J. (3462016) (a ciklusok feltérképezése a strukturális alkotmányjogban); Carol M. Rose, *Crystals and Mud in Property Law*, STAN40. L. REV. 577, ~~e~~zt a megállapítást teszi a tulajdonjogi szabályokról); Adrian Vermeule, *The Cycles of Statutory Interpretation*, U68. CHI. L. REV. 149,150 (2001) (a jogszabály-értelmezés ciklikusságának azonosítása).

²⁶⁹ Mi a helyzet "a társadalmi elkötelezettséggel, ~~h~~ megpróbáljuk egymást" és "a kapcsolat és a közösség lehetőségével" - azaz az "empátiával" és az "etikai fejlődéssel"? EUBANKS, *Supra* note at 8, Figyelembe168. véve az állam algoritmusok használatát a jóléti és állami juttatások kontextusában, Eubanks azzal érvel, hogy empátia hiányában a kisebbségekkel és nőekkel szembeni előítéletesség sokkal valószínűbb. *Id.* Egyetértek azzal, hogy az ellenségeskedés néha az empátia hiányosságai. Az emberi empátia intézményi terének kiterjesztése azonban azonnali, sőt középtávú megoldást jelent az együttérzés kudarcaira. Az emberi ítélőképesség utcai szintű minőségének drámai javulása nélkül -

és Eubanks nem ad okot arra, hogy az utcai szintű tisztviselők egyik napról a másikra jobb és igazságosabb döntéshozókká váljanak - az ilyen bővítések pontosan az általa szorgalmazott ellenkező hatást fogják kiváltani: Lehetővé teszi az elfogult vagy motivált érvelést, és még nehezebbé teszi az ebből eredő torzulások orvoslását. Ezzel szemben a szerszámgépek reformja lehetővé teszi a környezeti torzítások központi mérséklését. Ezért az az érv, hogy a jobb állami döntéshozatalhoz azonnali emberi kapcsolatra van szükség, egyszerűen tévhit. Ennek a tévhitnek ellentmond a huszadik század néhány nagy szociáldemokrata vívmánya, a brit Nemzeti Egészségügyi Szolgálattól az amerikai Great Society-ig. Röviden, az empátia nem elég: stratégiai és átgondolt módon kell cselekedni vele, nem pedig mankóként kell használni felületes és végső soron eredménytelen válaszokhoz.

emberi döntésre való igényt (vagy legalábbis a statikus algoritmus helyett a dinamikus algoritmushoz való jogot hihetőbbé teszi).

Negyedszer, emlékezzünk arra, hogy eddig a képzési adatok paramétereit pontosan azért választottuk ki, mert lehetővé teszik a célváltozó előrejelzése szempontjából releváns, mögöttes tulajdonság "hűség mérése".²⁷⁰ A részvételen alapuló érvelés az emberi döntés mellett valójában feltételezheti, hogy ez nem így van. Feltételezi, hogy a képzési adatok paramétere nem elegendőek a pontos eredmények maximalizálásához, így az utólagos indoklásnak korrekciós értéke van. De ha ez így van, akkor a megfelelő válasz a képzési adatok javítása vagy az algoritmus finomhangolása. A kiskereskedelmi válaszok, mint például az utólagos emberi döntéshez való jog, perverz módon fenntartják a hiányos status quo-t, és még késleltethetik is a jobb, rendszerszintű javítások végrehajtását, amelyek jobb gépi döntéshozatal révén érhetőek el.

Ez utóbbi pont általánosítható, hogy utolsó okot is hogy minden esetben ellenálljunk a részvételen alapuló emberi döntéshez való jognak. Amennyiben az emberi felülvizsgálatot úgy értelmezzük, mint egy algoritmikus eszköz integritásának ellenőrzésére szolgáló eszközt egy egyedi esetben, aligha egyértelmű, hogy a kiskereskedelmi beavatkozások az egyes besorolási döntések tekintetében hosszú távon megfelelő megoldást jelentenek (még ha jó diagnosztikai eszköz is lehet a rendszerszintű problémák kiszűrésére). Éppen ellenkezőleg, számos statikus felügyeleti eszköz - amelyek az algoritmus alapjául szolgáló forráskódra összpontosítanak - és dinamikus felügyeleti eszköz - amelyek az algoritmus viselkedését vizsgálják a természetben - áll rendelkezésre, ²⁷¹és idővel olyan szisztematikus javulást eredményezhetnek, amely valószínűleg elkerülné az emberi felülvizsgálathoz való jogot. A szűkös erőforrások miatt az algoritmus integritásának szisztematikusabb tesztelése helyett az egyedi utólagos felülvizsgálatra fordított idő valószínűleg gyakran az általános pontosság csökkenését eredményezi. Ezért, tekintettel a tisztességes eljárás elemzés marginális jellegére - amely az eljárási változtatások bármely adott lépcsőjének diszkrét pozitív vagy negatív hozzájárulására összpontosít -, az egyes döntések utólagos emberi felülvizsgálata ritkán jelent optimális stratégiát.

Gyanítom, hogy néhány maradnak kételyei az olyan esetekkel kapcsolatban, mint Catherine Taylor jóléti elutasítása.²⁷² Mít lehet mondani annak a személynek, akit tévesen rangsorolnak? Először is, érdemes megismételni, hogy a téves besorolás pusztán ténye önmagában még nem alapoz meg jogi vagy erkölcsi hibát. Az ésszerű kétség mércéjének alkalmazása a büntetőeljárásokban azt jelenti, hogy hajlandóak vagyunk eltérni bizonyos számú téves ítéletet. A kérdés az, hogy Taylort egy hibás rendszer minősítette-e, és nem az, hogy az ő esetében történt-e hiba.

A rendszer szintjén továbbá korántsem egyértelmű, hogy egy emberi döntéshozó hozzáadása csökkentené a hibák nettó mennyiségét. Lehet, hogy mind az ellátások odaítélésének, mind az elutasításoknak az emberi felülvizsgálattal magasabb hibaarányt vagy a hibák faji szempontból torzabb eloszlását lehetne elérni, mint a gépi döntéshozatal esetében. És végül, hogy megelőlegezzem azt a pontot, amelyre a következőkben még visszatérek, a gépi döntés hibás minősége nem jelenti azt, hogy egy emberi döntéshozó jobban teljesítene: Lehet, hogy a világ összes Taylorja számára az optimális rendszer jobb gépi döntést jelent, mint az emberi döntés.

²⁷⁰ David Lehr & Paul Ohm, *Játék az adatokkal*: U.C51. Davis L. Rev. (653,6792017): *What Legal Scholars Should Be About Machine Learning*.

²⁷¹ Kroll et al., *Supra* note at 2,547-52 és dinamikus tesztelési protokollok; lásd még Wachter, Mittelstadt &

Russell, *Supra* note at 155,13-14 (az algoritmikus integritás tesztelésének elvégzéséhez használt kontrafaktuális módszerek leírása).

²⁷² Lásd a fenti szöveg kísérő megjegyzés17eit 18.

felülvizsgálat. Tekintettel a gépi döntések nagyobb képességére, minden okunk megvan azt gondolni, hogy ez a feltétel fennáll - amint azt a következtetésben röviden kifejttem. Elemzésemben tehát semmi sem utal arra, hogy az ilyen számítási eszközök utólagos ellenőrzése vagy auditálása indokolatlan lenne. Szűkebb értelemben azt jelenti, hogy elutasítom a *kiskereskedelmi emberi*, nem pedig a *rendszerszintű* felülvizsgálatot, mint indokolt beavatkozást. Mivel azonban az emberi döntéshez való jog a GDPR cikk szerinti formájában²², valamint kevésbé artikulált változataiban is leginkább olyan kiskereskedelmi jogként írható le, amely az algoritmikus besorolás alá vont különböző egyénekre vonatkozik, a rendszerszintű felülvizsgálat folyamatos szükségessége nem adhat neki normatív alapot.

C. Az osztályozóval területek

Most rátérek az emberi döntéshez való jog mellett szóló érvekre, amelyek az állami cselekvés jellegén alapulnak. Az amerikai jogban ismerős az a gondolat, hogy egy cselekvés jogi vagy erkölcsi szempontból megengedhetetlen lehet az állam magatartásának módja miatt, nem pedig azért, mert behatol az egyéni érdekek valamely védett zónájába. A Legfelsőbb Bíróság például a közelmúltban megerősítette, hogy az egyének jogosultak olyan kereseteket benyújtani, amelyekben azt állítják, hogy a hatósági fellépés olyan strukturális alkotmányos elvek megsértésén alapul, mint a föderalizmus vagy a hatalmi ágak szétválasztása,²⁷³ még akkor is, ha ugyanazt az intézkedést, amely ugyanolyan hatással van az egyén érdekeire, megfelelően megalkotott kormányzati intézkedéssel is el lehetett volna érni. E jelenség egy másik változata akkor merül fel, amikor az állam megengedhetetlen megfontolások, például faji vagy vallási identitás alapján cselekszik.²⁷⁴ Legalábbis elméletileg a megengedhetetlen indokokon alapuló állami fellépés akkor is jogellenes, ha ugyanezt a fellépést más motivációs alapon nem tiltanák meg.²⁷⁵ Mindkét esetben az állami cselekvés nem a hatásai, hanem az állami cselekvés módja miatt minősül hibásnak.

Emberi döntéshez való jog minősített indoklása két érv egyikére támaszkodhat. Először is, azzal érvelhetnénk, hogy az egyénnel szembeni jogszerű állami fellépés egyik jellemzője, hogy az indokolt. Az algoritmikus döntések - érvelhetnénk - nem felelnek meg az ésszerűség e minimális küszöbkritériumának. Másodszor, ezek a döntések csoportalapú általánosítás alapján osztályozzák az egyéneket. Mint ilyenek, nem kezelik őket egyénként. Ez utóbbi pont értelmezhető a "profilalkotással" kapcsolatos aggodalomként, de értelmezhető a méltósággal kapcsolatos aggodalomként is. Bár mindkét érv mély normatív forrásokból táplálkozik, és az amerikai alkotmányjogban lehorgonyozott intuíciókra támaszkodik, végül azt javaslom, hogy egyik sem nyújt plauzibilis alapot az emberi döntéshez való joghoz.

1. Indokolt állami fellépés

Az angol-amerikai jogban mélyen gyökerezik az az elképzelés, hogy az állami cselekvésnek, és állami cselekvésnek szilárdan meg kell alapozódnia az indokokon. Az indokolás ebben az értelemben "az indoklás egy módja annak, hogy

²⁷³ Bond kontra Egyesült Államok, U.S.564(2011),2230 megállapítva, hogy "az egyének ... a hatalmi ágak szétválasztásának működése által védettek". *De lásd* Aziz Z. Huq, *Standing for the Structural Constitution*, VA99. L. REV. 1435, 1490-514 (2013) (kétségsbe vonva ezt az állítást).

²⁷⁴ Példák olyan döntésekre, amelyek olyan tiltanak meg, amelyek megengedhetetlen megfontolások hiányában megengedettek lennének, lásd pl. *Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. U. 1,551S. (701,2007720)*; *lásd még* Gratz v. Bollinger, U539.S. (2003244,270) (az ilyen osztályozások használatát "károsnak" nevezte (idézet kihagyva)).

²⁷⁵ Egy közelmúltbeli kivételt lásd Trump v. ~~USA~~ Ct. (2392,24212018) (egy bevándorlással kapcsolatos, nyilvánosan diszkriminatív alapon indokolt végrehajtási rendelet fenntartása, "mivel meggyőző bizonyíték van arra, hogy a beutazás felfüggesztésének legitím alapja a nemzetbiztonsági aggályok, teljesen függetlenül bármilyen vallási ellenségeskedéstől").

a téma iránti tiszteletet mutatva, és beszélgetés megnyitása, ahelyett, hogy megelőzné azt."²⁷⁶ "Értéket tulajdonít annak, hogy az egyének elmondják, miért bánik vele kedvezőtlenül az ügynök, és hogy részt vett a döntésben".²⁷⁷ Egy határozat e nézet szerint csak akkor áll összhangban a jogállamisággal, ha az "a határozat alanyai számára érthető".²⁷⁸ A század közepi liberális szemlélet szerint az államhatalom akkor legitim, ha "megfelelő semlegesség és általánosság indokán" alapul.²⁷⁹ Egy lelkes libertárius hozzátehetné, hogy az indokolás az állami cselekvés üdvös súrlódásaként működik, olyan tranzakciós költségeket generálva, amelyek kizáró erejűek lehetnek, ha egy cselekvésnek nincs közérdekű alapja. Valami ebből az intuícióból látszik működni a negyedik módosítási jog alkalmazásában, amely kizárja a "puszta gyanú" alapján történő utcai megállítást.²⁸⁰

Tekintsünk el attól a lehetőségtől, hogy a bíró lehet megtervezni, hogy megmagyarázott döntéseket hozzanak.²⁸¹ Az indokolt állami cselekvés iránti igény még mindig nem nyújt megfelelő alapot az emberi döntéshez való joghoz, azon egyszerű oknál fogva, hogy ez önmagában csak alkalmanként és hiányosan teljesül. Sok állami döntés mindenféle alátámasztó indoklás nélkül születik. Ahogy Lon Fuller az ítélezésről szóló híres esszéjében¹⁹⁷⁸ megjegyezte, az "ítélkezés integritása" "nem feltétlenül" követeli meg, hogy "a meghozott döntést meg kell indokolni".²⁸² Az utcai megállítástól a certiorari elutasításig számos olyan diszkrét állami beavatkozás létezik az ítélezési kontextuson belül és azon túl, amelyek jellemzően nem rendelkeznek kifejezett indoklással. Ezen túlmenően a jogszabályok a társadalmi realitások széleskörű megváltoztatását alakíthatják anélkül, hogy bármit is kínálnának megfelelő normatív indoklással.²⁸³ Valójában ma már hagyományos bölcsességnek számít, hogy a jogalkotók gyakran anélkül fogadnak el jogszabálysöveget, hogy konszenzusos véleményt alakítanának ki a jogszabálysöveg valamely elemének jelentéséről. Kétértelmű törvényszöveg - a fellebbviteli bíróságok mindennapi munkája - akkor merülhet fel, ha "a Kongresszusnak nem volt különösebb szándéka a kérdéssel kapcsolatban".²⁸⁴ Az sem hihető, hogy minden fontos ítélezési cselekményt e kifejezés teljes értelemben vett érveléssel indokolnak. Amikor egy tárgyaló bíró elutasít egy bizonyítási kifogást, amikor egy fellebbviteli bíróság mérlegelési jogkörében engedélyezi a nem kötelező fellebbezést, vagy amikor a Legfelsőbb Bíróság elutasítja a certiorari-t, aligha egyértelmű, hogy a cselekvésnek vannak kifejezhető, nemhogy jól megalapozott indokai.²⁸⁵ A nem kellően indokolt állami intézkedések pusztán mértéke is sokatmondó: Ha ezt a követelést névértéken vennénk, az valójában megfojtaná a modern államot, mielőtt az alapvető kötelezettségeinek eleget tehetne.

²⁷⁶ Frederick Schauer, *Giving Reasons*, STAN47. L. REV. 658 (1995).

²⁷⁷ Frank I. Michelman, *Formai és társulási célok az eljárásjogi tisztességes eljárásban*, NOMOS (826,1977127)

²⁷⁸ Mireille Hildebrandt, *Algoritmikus szabályozás és a jogállamiság*, PÉ. SOC. 1,3 (2018)

²⁷⁹ Herbert Wechsler, *Toward Neutral Principles of Constitutional Law*, MBL. REV. 1,15 (1959); *egyértelműsítés*: LON L. FULLER, *THE MORALITY OF LAW* (341964). Az indokolt általánosság követelményének állítólagosan liberális eredetéről lásd Mark V. Tushnet, *Following the Rules Laid Down: A Critique of Interpretivism and Neutral Principles*, HARV96. L. REV. 781, 783- (851983).

²⁸⁰ United States v. Arvizu, U.S.534. 266. (idézet és idézőjelek kihagyva).

²⁸¹ *Lásd a fenti szöveg kísérő megjegyzés* 174eit 194.

²⁸² Lon L. Fuller, *The Forms and Limits of Adjudication*, HARV92. L. 53,387 (1978).

²⁸³ Schauer, *Supra* note at 276,636.

²⁸⁴ Antonin Scalia, *Judicial Deference to Agency Interpretations of Law*, 1089 L.J. 516511.; *lásd még* ROBERT A. KATZMANN, *JUDGING STATUTES* 15-22 (2014) (a tudatosság, a megegyezés, az előrelátás, a pontos megfogalmazás és a gondosság hiányosságait a jogalkotói kétértelműség tipikus forrásaként írja le).

²⁸⁵ David Enoch meggyőzően érvelt amellett, hogy állami intézkedéseket kizárólag az előre látható (pozitív vagy negatív) következmények alapján kell értékelni, nem pedig a tervezett vagy indokolt célok alapján. David Enoch, *Intending, Foreseeing, and the State*, LEG13. THEORY 91-9269, (2007) (ezt a következtetést az egyéni és

De tekintsük át megfontolt döntéshozatal érvének egy szűkebb változatát. Talán amikor az állam bizonyos kényszerítő intézkedéseket hoz - beleértve a büntetőjogi és szociális jóléti döntéshozatalt, amely ma már algoritmikus eszközöket használ -, nem cselekedhet "puszta megérzés alapján".²⁸⁶ Az esetek szűkebb kategóriájában a tisztviselőknek megalapozott okkal kell rendelkezniük intézkedéseikhez. Még ezekben az esetekben is félrevezető és hiányos az a benyomás, hogy a gépi döntések nem, vagy nem lehetnek indokoltak. Nem arról van ugyanis szó, hogy a gépi döntések nélkülözik az igazoló okokat. Sokkal inkább arról van szó, hogy ezeket az indokokat az egyénre ható állami cselekvéstől távol eső időpontban szolgáltatják. Emlékezzünk arra, hogy a gépi tanulási eszközök tervezése, tesztelése és megvalósítása alaposan összefonódik a célzott emberi döntésekkel és szándékossággal.²⁸⁷ Az emberi szándékok irányítják a felügyelt és nem felügyelt modellek közötti választást; a jellemzők kiválasztásának folyamatát; a képzési adatok kiválasztását; és a folyamatos finomítási és kalibrálási folyamatot az optimális osztályozó felé.²⁸⁸ E szándékos emberi cselekvés nagy részét szükségszerűen a gép által szolgálendő célok megértése határozza meg. Ezért a gépi döntésekben a megfontolt ítéletek hiánya csak optikai csalódás. Nem annyira arról van szó, hogy hiányoznak az ilyen ítéletek. Sokkal inkább az, hogy mire egy algoritmus a világban dolgozik, már beépültek egy osztályozóba. Ezek a kódolt ítéletek ráadásul ugyanazt a célt szolgálják, mint az utólagos indoklás igénye: Az általánosságra való előzetes kötelezettségvállalásként és a személyeskedő vagy önkényes állami cselekvés elleni védelemként működnek.²⁸⁹ Tekintettel formalizált - szó szerint kódként meszesedett - természetükre, az algoritmusokba ágyazott indokok ellenállóbbak lehetnek az utólagos manipulációval szemben, mint azok az indokok, amelyekre a bíróságok ítéletei támaszkodnak.

Összefoglalva, még ha szilárd állunk is, amikor ésszerű állami cselekvést követelünk - különösen, ha fontos emberi érdekek megfosztásáról van szó -, követelésünknek nem kell az emberi döntéshez való joggal végződnie. Éppen ellenkezőleg, a követelés alapjául szolgáló aggályok akár az ellenkező irányba is mutathatnak: Egy robusztusan megtervezett algoritmus, amelyet átgondoltan, elfogulatlan és tanulságos képzési adatokkal látnak el.

2. *Az egyéni döntéshez való jog*

Az emberi döntéshez való jogot lehet indokolni, hogy az algoritmus az osztályozási döntés meghozatalakor az indokok általános jellegére hivatkozik. Durván fogalmazva, az intuíció itt az, hogy az államnak kizárólag a saját viselkedése vagy érdemei alapján kellene fellépnie egy személy ellen. Úgy kellene kezelnie őket, vagyis "egyénként".²⁹⁰ A nagyobb társadalmi csoport által osztott tulajdonságok alapján történő fellépés ipso facto nem veszi komolyan az adott személyt egyénként.

állami felelősség). Lehet, hogy az ő beszámolójából az hogy az állami cselekvés ésszerű minősége vagy nem minősége irreleváns.

²⁸⁶ *Arvizu*, U 534.S. at De 274.lásd Craig S. Lerner, *Policing Hunches*, J4.L. ECON. & POL'25,25Y (2007) (a megérzéseket "nélkülözhetetlen heurisztikus eszközökként védi, amelyek lehetővé teszik az emberek számára, hogy feldolgozzák a környezetükre vonatkozó diffúz, összetett információkat és értelmet adjanak a világnak").

²⁸⁷ *Lásd a fenti szöveg kísérő megjegyzés* 174eit 194.

²⁸⁸ Az algoritmikus iteratív folyamatáról Bengio, *Supra* note, 33,50-51; Amershi et al., *Supra*. jegyzet a 182,106

²⁸⁹ *Vö.* Schauer, *Supra* note, 276,651-52. o. olyan előzetes kötelezettségvállalási mechanizmusként amely általánosságot eredményez).

²⁹⁰ Kasper Lippert-Rasmussen, "Mindamnyian mások vagyunk": J15. ETHICS (47,201149): *Statistikai diszkrimináció az egyénként való bánásmódhoz való jog*.

Az22. indokolás jellegével kapcsolatos aggályokhoz hasonlóan a GDPR cikke is tartalmazza a "profilalkotást", mint az automatizált adatkezelés egyik formáját.²⁹¹ A rendelet máshol a profilalkotást tágran határozza meg: "a személyes adatok felhasználása ... a természetes személy munkahelyi teljesítményére, gazdasági helyzetére, egészségére, személyes preferenciáira, érdeklődésére, megbízhatóságára, viselkedésére, tartózkodási helyére vagy mozgására vonatkozó szempontok előrejelzésére".²⁹² E fogalom meghatározás szélessége és a cikkben szereplő 22 automatizált feldolgozással való kapcsolata azt jelenti, hogy az algoritmikus rendszerekkel kapcsolatos aggodalomra ad okot, amelyek nagy mennyiségű adatot vesznek fel annak érdekében, hogy előrejelző osztályozásokat hozzanak létre. Az automatizált és nem automatizált profilok közötti hallgatóságos megkülönböztetés továbbá arra utal, hogy a GDPR beavatkozásának alapja az egyes gépi döntések személytelen általánosságával és az egyéni részletektől való korrelatív elszakadással kapcsolatos aggodalom. Ezzel párhuzamos gondolat fedezhető fel abban, hogy a bíróságok szkeptikusan viszonyulnak bizonyos típusú statisztikai bizonyítékok felhasználásához, amelyekkel a kártérítési ügyekben az alperes felelősségének valószínűségét lehet bizonyítani.²⁹³ Itt a "személyre szabott bizonyítékok" iránti felhívás jelenik meg.²⁹⁴ Mind ez, mind az emberi döntés iránti igény az állami fellépés partikuláris (és nem a populáció egészére kiterjedő) bizonyítási alpra való felhíváson nyugszik.

Az emberi való jognak az egyénként való bánásmód iránti igényen intuíciója számos filozófiai hagyományra hivatkozva igazolható. Ezt először is a kanti felfogás indokolhatja, miszerint az egyén nem kezelhető "puszta" eszközként egy cél eléréséhez.²⁹⁵ A német Alkotmánybíróság például a "verdinglicht und zugleich entretlicht" ("reifikálja, majd megfosztja a jogaitól") állami intézkedések megengedhetetlenségére hivatkozott, hogy érvénytelenítse a légi 2004 közlekedés biztonságáról szóló törvény azon rendelkezését, amely a német alaptörvény emberi méltóság iránti elkötelezettsége alapján bizonyos körülmények között lehetővé tette egy eltérített repülőgép lelövését.²⁹⁶ Alternatívaként ez a követelés összekapcsolható a szerencse-egyenlőségi követeléssel, hogy az ember arra törekedjen, hogy "amennyire csak lehetséges, kiküszöbölje az emberek életére gyakorolt olyan balszerencse hatását, amely önhibájukon vagy választásukon kívül érte őket".²⁹⁷ Mivel a gépi döntések gyakran olyan tulajdonságokon alapulnak, amelyekre az embernek nincs ráhatása, a gépi döntések ütköznek ezzel az igénnyel. Ez utóbbi indok azonban jelentős nehézségekkel szembesül, tekintettel a következők közötti eltérésre

²⁹¹ GDPR art. 22. Ez a cikk nem tilos bizonyos profilalkotáson alapuló "döntéseket".

²⁹² GDPR art. 4(4).

²⁹³ A bíróságok néha azt mondták, hogy pusztán "Smith v. Rapid Transit Inc., Mass317. 469,470 (1945), vagy "mennyiségi valószínűség", Day v. Boston & Me. R.R., Me96. 207,217 (1902), soha nem elegendőek a deliktális felelősség megállapításához. A különböző jogrendszerek válaszainak áttekintését lásd Rebecca Haw, Note, *Prediction Markets and Law: A Skeptical Account*, HARV122. L. REV. 1217,1229 (2009).

²⁹⁴ Judith Jarvis Thompson, *Felelőség és egyedi bizonyítás*, 49& CONT. PROB. 199,205 (1996)

²⁹⁵ I. KANT, FOUNDATIONS OF THE METAPHYSICS OF MORALS Beck trans. 1959) (1. kiadás, Riga 1785) ("Act

hogy az emberiséget, akár a akár más mindig célként és soha nem csak eszközként kezeld."). Ez a kanti fogalom, meg kell jegyezni, elég képlékeny ahhoz, hogy a puszta részvételhez való jog igazolására is bevetették. Edward Pincoffs, *Due Process, Fraternity, and a Kantian Injunction*, NOMOS18 (172,1791977) ("[P]articipation is morally valuable to the degree that it makes determinate the moral principle that we should never treat a man as a mere means."). A kategorikus imperatívusz e változatának sokféle értelmezéséről lásd Thomas E. Hill, Jr., *Humanity as an End in Itself*, ETHICS91 84,849 még Alexander Somek, *German Legal Philosophy and Theory in the Nineteenth and Twentieth Centuries*, in A COMPANION TO PHILOSOPHY OF LAW AND LEGAL THEORY 343-343, (43Dennis

Patterson szerk., 1996) (elhelyezve ezt a gondolatot a német jogelmélet történetében).

²⁹⁶ Bundesverfassungsgericht (BverfG - ~~Súgó~~Alkotmánybíróság), Neue 59Juristische Wochenschrift 751m, C II. b) aa) és bb) pontok. (2006).

²⁹⁷ Richard J. Arneson, *Szerencse egalitarizmus és prioritarianizmus* ~~10~~ 10 (9392000).

a társadalmi bánásmód választás számos kritériuma, valamint a nem választott tulajdonságok és a választás tárgyát képező tulajdonságok szétválasztásának nehézsége.²⁹⁸

Első pillantásra egyáltalán nem algoritmikus döntéseket miért kell kiemelni, mint amelyek nem képesek az egyénre szabásra. Az algoritmusokat úgy lehet megtervezni, hogy figyelembe vegyenek "minden releváns, statisztikai vagy nem statisztikai információt", amely "ésszerűen rendelkezésre áll".²⁹⁹ Az egyének eltérő tulajdonságaik és viselkedésük alapján történő differenciált bánásmódja sem mindig erkölcsi hiba. Éppen ellenkezőleg, egy olyan egységes szabály, amely mindenkire egyenlő "részesedést ró a közjó fenntartásának és megőrzésének költségeiből" - gondoljunk csak az általános katonai behívóra - gyakran erkölcsileg kényszerítő erejű.³⁰⁰ Valóban, "még a józan ítélőképesség" gyakran nem téves általánosításokra épül.³⁰¹ Ezért egyszerűen nem tartható az a pusztán állítás, hogy egy döntés erkölcsileg hibás, mert az nem egyedi bizonyítékokra, hanem a népesség egészére vonatkozó adatokra épül.

Úgy gondolom, hogy egy finomabb megközelítésre van értelmet nyerni ez az érvelés. Bár nem vagyok meggyőződve arról, hogy ez általános ellenvetés a gépi döntésekkel szemben, úgy gondolom, hogy bizonyos feltételezésekkel és bizonyos feltételek mellett bevethető, hogy ellenálljon az emberi döntések gépi helyettesítésének. A kérdéses kérdések sokkal szűkebb körűek, mint a GDPR rendszere. E finomabb érvelés indoklásához azt kell feltételeznünk, hogy az emberi döntéshozónak hozzáférése van az egyedi bizonyítékokhoz, míg a gépi döntéshozó csak statisztikai vagy populációra vonatkozó információkhoz férne hozzá. Vegyük észre, hogy semmi sem kényszeríti ki az episztemikus munkamegosztást; a gépet egyedi bizonyítékokkal is elláthatják, míg az emberi döntéshozó statisztikai bizonyítékokra támaszkodhat. A feltételezés azonban úgy tűnik, hogy az emberi döntéshez való jogba van beépítve, ahogyan azt a GDPR cikkének "profilalkotás" elleni iránya is megvilágítja. 22.

Ezzel a feltételezéssel a kezünkben különbséget tehetünk a gépi döntések különböző fajtái között. Amennyiben a döntés előrejelzés, nincs nyilvánvaló ellenvetés a nem egyedi bizonyítékokra való támaszkodás ellen. Ha a népesség egészére kiterjedő bizonyítékok elegendőek, mondjuk, a biztonsági övre vonatkozó előírások bevezetéséhez az autókban³⁰² vagy az iskoláskorú gyermekek oltási rendjéhez,³⁰³ akkor miért kellene ezeket rosszabbnak tekinteni a részletesebb, előrejelző jellegű állami intézkedések, például az óvadékról és a feltételes szabadlábra helyezéstről szóló döntések alapjaként? A különböző típusú bizonyítékok abszolút ismeretelméleti minősége nem lehet a megkülönböztetés alapja. Az egyénre szabott bizonyítékok és a népesség egészére vonatkozó bizonyítékok minősége egyaránt eltérő. Nincs a priori okunk azt gondolni, hogy az egyikre alapozott döntések kevésbé pontosak lennének, mint a másokra alapozott döntések.³⁰⁴

²⁹⁸ Mindkét probléma körvonalazódik Samuel *Why egalitarianism?*, PHIL31 & PUB. AFFAIRS 17-215, (2003); lásd még Elizabeth S. Anderson, *What is the Point of Equality?*, ETHICS109 (287,2891999) (a szerencse egalitarizmusának három további kritikáját fejt ki).

²⁹⁹ Lippert-Rasmussen, *Supra* note at 290,54.

³⁰⁰ Annabelle Lever, *Miért nehéz igazolni a faji alapú profilalkotást?* PHIL38 & PUB. AFF. 94,110 (2005).

³⁰¹ FREDERICK SCHAUER, *PROFILOK, VALÓSZÍNŰSÉGEK ÉS* (2003).

³⁰² *Lásd pl. N1984.J. Laws c. §-t, amelyet 179*legutóbb2,2N 2009.J. Laws c. §318 módosított. 1.

³⁰³ Például mind a területVirginia állam törvényben írja elő a humán papillomavírus elleni védőoltást az iskoláskorú lányok számára. *Lásd* D.C. Code § 7-1651.04 (b)(1)(B)(iii) (2008); Va. Code Ann. § 32.1-46 (D)(3) (2008).

³⁰⁴ Lásd Thompson, *Supra* note, 294,200-02. o. (egy példa).

Amikor azonban a népszerűségeen alapuló bizonyítékokat egy történelmi tény alátámasztására a felelősség megállapítása céljából, finoman szólva más szempontok merülnek fel. Különösen akkor merül fel probléma, ha a döntés célja az elrettentő hatás elérése a jövőben. A Martin Smith filozófus által kidolgozott kifejezéssel élve, számos olyan eset van, amikor az egyénre szabott bizonyítékok "normálisan alátámasztják" azt a következtetést, amely mellett felhozzák őket,³⁰⁵ és így képesek elrettentő hatást kiváltani. Smith érvelésének megértéséhez képzeljük el, hogy van egy laptopom, amelynek képernyővédője az idő kilentizedében kék képernyőt mutat. Míg én dolgozom, a barátom elsétál a számítógépem mellett, és kék képernyőt lát. A barátom meggyőződése, hogy a képernyő kék, "normálisan alátámasztott" az ő észlelése által; az én analóg meggyőződésem, hogy a képernyő kék volt, nem.³⁰⁶ Ezt másképp is ki lehet fejezni. A barátom meggyőződése, hogy a képernyő kék, *kontrafaktikusan érzékeny* az igazságra, míg az én evidenciám nem.³⁰⁷ Ha később kiderül, hogy a képernyő abban a pillanatban nem volt kék, lehet, hogy egyszerűen megvonom a vállamat a szerencsétlen sejtésem miatt. A barátom számára ez a közömbösség "helytelennek" tűnne.³⁰⁸ Talán meg kellene vizsgálatnia a látását színvakság szempontjából.

Ez a megkülönböztetés a következőképpen ültethető át jogi kontextusba. Ha a barátom bizonyítékát hátrányos intézkedés alapjául használják, akkor az én bizonyítékom olyan módon érzékeny a történelmi tényekre, ahogyan az enyém nem. Ez azt sugallja, hogy míg mindkét fajta bizonyíték racionálisan felhasználható a meggyőződések és előrejelzések kialakításához, csak a tényellentétre érzékeny bizonyíték használható fel elrettentő hatás kiváltására.³⁰⁹ Ha a felelősséget nem érzékeny indokok (azaz statisztikai bizonyítékok) alapján szabják ki, az nem fog elrettenteni. Ezért az érzékenység akkor is számít az elrettentés szempontjából, ha az előrejelzés vagy esetleg a tudás szempontjából nem számít.³¹⁰ Ha tehát egy gépi döntés a populáció egészére kiterjedő bizonyítékokra támaszkodik, akkor az elrettentő hatás csökken.

Tehát lehet, hogy a nem egyedi bizonyítékok felhasználásával szembeni ellenvetés az optimális elrettentés iránti igényből fakad, és talán a vádaskodás társadalmi gyakorlatából is.³¹¹ De ha a nem individualizált bizonyítékok javítják a pontosságot, miközben nem teremtenek kívánatos ösztönzőket, akkor miért lenne ez az *egyén* ellenvetésének alapja? Az államnak, és nem a szabályozott egyénnek áll érdekében az elrettentés. Emlékezzünk továbbá arra, hogy ez az érvelés abból a (valószínűleg hibás) feltételezésből indul ki, hogy a gépek csak a populáció egészére kiterjedő bizonyítékokra támaszkodnak, míg az emberi döntéshozók mindig hozzáférnek az egyénre szabott bizonyítékokhoz. Ez az érvelés végül nem elegendő az emberi döntéshez való általános jog magyarázatához, tekintettel arra, ahogyan az algoritmikus döntéseket jelenleg alkalmazzák. A gépi tanulás állam általi jelenlegi felhasználásának többsége előrejelzésekre vonatkozik, nem pedig történelmi tények megállapítására, amelyeken az elrettentés alapul.³¹² Nem mindegyik vonatkozik a vádaskodásra, és ha

³⁰⁵ Martin Smith, *What Else Justification Could Be*, NOÛS44:30, 0

³⁰⁶ *Id.* (e egy összetettebb változatát kínálja)

³⁰⁷ David Enoch, Levi Spectre & Talia Fisher, *Evidence, Sensitivity, and the Legal Value of Knowledge*, PHIL40 & PUB. AFF. 197, 209-10 (2012).

³⁰⁸ *Id.* 209.

³⁰⁹ *Id.* 218-19.

³¹⁰ Létezik szakirodalom arról, hogy lehet-e a tudás vagy a racionális meggyőződés alapja. *Lásd pl.* HENRY KYBERG, *PROBABILITY AND THE LOGIC OF RATIONAL BELIEF* (1961). Nem hiszem, hogy az erős szkepticizmusnak ez a formája releváns lenne a jogi és intézményi tervezés kérdései szempontjából.

³¹¹ Enoch, Spectre, & Fisher, *supra* note at 307, hogy a vádaskodáshoz szükség lehet kontrafaktuálisan érzékeny

bizonyítékokra is).

³¹² *Lásd a fenti szöveg kísérő megjegyzés*133eit 134.

a hiba kiosztását tekintik egy döntési eszköz funkciójának, akkor a gépek ugyanolyan alkalmatlanok lehetnek, mint az oksági kérdések esetében. Mindezen okok miatt szkeptikus vagyok azzal kapcsolatban, hogy a kontrafaktuális érzékenységgel kapcsolatos aggodalom kiválthatja az emberi döntéshez való jogot.

Röviden, az emberi döntésekhez hasonlítva szembeállított indokai nem járnak jobban, mint azok az érvek, amelyek az egyének jogjaiból indulnak ki. Sem az indokolt állami cselekvés miatti aggodalom, sem az állami cselekvés alapjául szolgáló bizonyítékok individuális jellege miatti aggodalom nem bizonyul kielégítőnek.

D. Rendszerszintű aggályok

Egy utolsó lehetséges alap, amely alapján emberi döntéshez képest, az elemzési szempontot az egyén és a gép közötti közvetlen tranzakción túlmutatóan tárgyítja, hogy megvizsgálja a gépi döntésekre való kizárólagos hagyatkozás dinamikus következményeit az állami cselekvés szélesebb körű mintáira. Bár az emberi döntések nagyobb számának lehetnek kívánatos rendszerszintű következményei - amelyekről alább lesz szó -, ezek csak nehezen használhatók fel egy önálló egyéni jog fenntartására. Ehelyett mindezek az érdekek pontosabban megcélózhatók és előmozdíthatók olyan alternatív beavatkozások révén, amelyek nem az egyének véletlenszerű mérlegelésére támaszkodnak abban a kérdésben, hogy érvényesítsék-e jogi érdekeiket, vagy sem. Míg az itt azonosított rendszerszintű aggályok tehát a szóban forgó jognak járulékos támogatást nyújthatnak, nem lehetnek annak legfőbb támaszai.

Először is, az állami szereplőkkel általában a bíróságilag kikényszerített felelősségi küszöbértékhez közeli "csökkentett aktivitási szintekkel" jár együtt.³¹³ Az emberi döntéshez való jog sem lenne másként. A gépi döntésekre való kizárólagos támaszkodás csökkenti az adott állami hatalom gyakorlásának határkölségét. A jog tehát az állami cselekvés egyfajta enerváló súrlódásának tekinthető. De természetesen az, hogy ez kívánatos-e, nyilvánvalóan a tevékenység jellegétől függ. Gondoljunk például a távoli harctéren halálos erő gyakorlására alkalmas pilóta nélküli drónok terveinek teljes automatizálásának lehetőségére.³¹⁴ Még a megértő szemmel nézve is összetett etikai kérdéseket vet fel az emberi szerep alapos kiiktatása a hadviselésben.³¹⁵ Az egyik ilyen kérdés az emberi szerep fenntartásának az aktivitási szintekre gyakorolt hatása lehet. Ha az emberi szerep fenntartása alacsonyabb aktivitási szinthez vezetne, anélkül, hogy a háborús erőfeszítéseknek kizáró költségei lennének, akkor hihetően beszélhetnénk arról a kötelezettségről, hogy az embereket a körforgásban kell tartani, hogy megelőzzük a halálos drónok használatának gyors inflációját.

Másodszor, több kommentátor hív "automatizálási torzítás", vagy az "automatizálás használata az éber információkeresés és -feldolgozás heurisztikus helyettesítésére".³¹⁶ Az emberek az automatizált döntésekre való hagyatkozás heurisztikáját alkalmazzák, "az éberebb és figyelmesebb

³¹³ John C. Jeffries, Jr., *The Right-Remedy Gap in Constitutional Law*, 111 J. (1998) 105. A felelősségi szabály hatása vitatott. A hagyományos bölcsesség szerint mind a gondatlansági, mind a szigorú felelősségi rendszerek a túlzott tevékenység kockázatával járnak. STEVEN SHAVELL, A BALESETI JOG GAZDASÁGI ELEMZÉSE 66-71 (1987).

³¹⁴ Az akut leírást lásd HUGH GUSTAFSON, REMOTE-CONTROL WARFARE 2-25 (2017).

³¹⁵ Lásd Robert Sparrow, *Robotok és tisztesség*: ETHICS30 & INT'L AFF. 93, 94-95 (2016) (a legfontosabb etikai kérdések összefoglalása).

³¹⁶ Linda J. Skitka et al., *Automatizálási torzítások és hibák: Jobbak-e csapatok, mint az egyének?*, INTL10' J. AVIATION PSYCHOLOGY (85,862000).

rendszerfelügyelet vagy döntéshozatal."³¹⁷ Amperi döntéshez való jog megelőzésként működik az ellen, hogy az emberek túlzottan megbízzanak a gépi döntésekben, mert a szakértelem és objektivitás látszatát keltik a technológiában.³¹⁸ Bármennyire is erőteljes ez az aggodalom - empirikus bizonyítékok alig állnak rendelkezésre -, valószínűleg számos módja van annak, hogy megakadályozzuk az automatizált döntéshozókra való önelégült hagyatkozást - nem utolsósorban valamilyen gyakori ellenőrzés. Az osztályozás alá vont egyénekre való hagyatkozás lehet az egyik megoldás a megoldások közül, de aligha elkerülhetetlen tervezési döntés.

Egy harmadik érv az álló jog mellett a gépi döntéseknek a társadalmi hatalom elosztására gyakorolt hatásaira összpontosít. A gépi tanulás lehetővé teszi a társadalmi jólét növekedését az új vagy pontosabb előrejelzések eredményeként. Ezek a nyereségek azonban olyan egyenlőtlenül oszlanak el, ami mély normatív aggodalmat vált ki. Mivel a gépi tanulás eszközei nagy adathalmazt és robusztus számítási erőforrásokat igényelnek, valószínű, hogy olyan szervezeti egységek - nem utolsósorban az állam - fogják elfogadni és használni őket, amelyek már most is aszimmetrikus kapcsolatban állnak a nagyközönséggel. A gépi tanulás elfogadása nemkívánatos módon súlyosbíthatja ezeket az egyensúlytalanságokat. Ez felveti annak lehetőségét, hogy a gépi döntésekkel kapcsolatos nyugtalanság nem azok megkülönböztető minőségén, hanem az állam és az alattvalók, illetve a nagyvállalatok és az egyes piaci szereplők közötti kapcsolatra gyakorolt hatásukon alapul. Az ilyen hatalom aszimmetrikus elosztása alááshatja a részvételi demokrácia lehetséges feltételeit, ha a gépi döntéseket a politikai preferenciák alakítására használják. Vagy pedig lehetővé tehetik a szabályozás új, erősen beavatkozó formáit, amelyek összhangban vannak az egyéni szabadság normatívan megalapozott beszámolójával.

Megértem ezeket az aggodalmakat. Szkeptikus vagyok azonban azzal kapcsolatban, hogy az egyéni jog értelmes választ adna a II. részben és e rész korábbi részeiben feltérképezett technológiai realitások és normatív következmények ismeretében. Az emberi döntéshez való jog olyan technológiákra adott válasz, amelyek a személyek és a koncentrált szervezeti hatalom közötti problémás aszimmetriákat generálnak. A probléma az, hogy hősies feltételezésekre van szükség ahhoz a következtetéshez, hogy a szétszórtan élő - az állami vagy vállalati nyomásnak több margó mentén is kiszolgáltatott - egyének képesek lesznek olyan kollektív cselekvésre irányuló joggal élni, amely hatékonyan orvosolja az aszimmetrikus társadalmi berendezkedést. Vagyis a jogok pusztá biztosítása nem oldja fel a hatalmi aszimmetriát. Ez nyilvánvaló a büntető igazságszolgáltatással kapcsolatos eljárási jogosultságokkal kapcsolatos fél évszázados tapasztalatból, amely arra utal, hogy a jogok hatékonyságát erősen korlátozza az állam (vagy a hasonlóan szabályozott szereplő) azon képessége, hogy helyettesítő befolyásolási módokat találjon.³¹⁹ A magánélethez való egyéni jogokkal kapcsolatos legújabb tapasztalatok a közösségi média és az internetes platformok kontextusában szintén pesszimizmusra adnak okot.³²⁰ A hozzájáruláson alapuló adatvédelmi rendszerek egyik központi problémája az, hogy a fogyasztók a jelek szerint különböző értékeket tulajdonítanak ennek a jószágnak attól függően, hogy megkérdezték-e őket, mennyi pénzt vállalnának azért, hogy egyébként magánjellegű információkat tegyenek közzé.

³¹⁷ Linda J. Skitka, *Does Automation Bias Decision-making?*, INT51. JUD. 991,992 (1999).

³¹⁸ A közelmúltban az automatizálási torzítás hatásstrofális következményekkel járó példája a michigani Marshallban történt olajvezeték-szakadás volt, lásd David 2010.Wesley & Luis Alfonso Dau, *Complacency and Automation Bias in the Enbridge Pipeline Disaster*, ERGONOMICS25 IN DESIGN 19-2017, (2017).

³¹⁹ Ennek az aggodalomnak a klasszikus kifejezése NStuntz, *The Uneasy Relationship Between Criminal Procedure and Criminal Justice*, YALE107 L. J. (1997)1,64

³²⁰ *Lásd pl.* Alessandro Acquisti, Leslie K. John & *What is privacy worth?*, J42. LEG. STUD. 249, 250-51 (2013) (a *szélességének azonosítása a finom kontextuális jelzésekre*).

vagy hogy mennyit fizetnének az információk védelméért.³²¹ Úgy tűnik, hogy idő-inkonzisztens preferenciákkal is rendelkeznek, abban az értelemben, hogy hajlandóak elfogadni az alacsony jutalmakat most, cserébe a "lehetséges állandó negatív járadékért a jövőben".³²² Röviden, a jogok gyakran (bár nem mindig) nem képesek katalizálni a hatalom újraelosztását.

Ha a jogok nem feltétlenül a koncentrált társadalmi hatalom elleni küzdelemre, legalábbis ha önálló eszközként fogalmazzuk meg őket, van-e alternatíva? Egy közvetlenebb megközelítés a hatalmi vagy befolyásolási aszimmetriák elleni frontális támadást jelent a társadalmi hatalom koncentrációjának feldarabolásával.³²³ Elképzelhető az is, hogy miként lehetne újraosztani az aggregált episztemikus hatalomból eredő többletet. De bizonyíték hiányában, hogy az emberi döntéshez való jog elősegítheti ezt a fajta mozgósítást - és nem hiszem, hogy létezik ilyen bizonyíték -, nem szabad elhamarkodottan arra következtetnünk, hogy az egyéni jog hatékony megoldást nyújt egy strukturális és rendszerszintű diszfunkcióra.

Végül, és a hatalommal összekapcsolódóan, előfordulhat, hogy az emberi döntést előnyben részesítik egy olyan intézménytervezési cél érdekében, mint az államhatalom elterjedése vagy a jogi szabályok folyamatos fejlődése. Valójában az állítás az lenne, hogy az emberi döntési jogkörnek bizonyos pozitív tovagyrűző hatásai vannak az egyedi eseteken túl. Egy ilyen érvelés indokolhatja az emberi döntéshozatal gépi helyettesítését. De nem adna globális indokot az ilyen helyettesítésre. Sőt, még a helyi alkalmazása is az intézményi tervezés más elemeitől függne.

Az emberi döntés előnyben részesülhet például azért, mert fenntartja a jogi kritérium nyitottságát, vagy mert bizonytalansági elemet visz a jogalkalmazásba. Egy felügyelt gépi tanulási eszköz céljait teljes mértékben meg kell határozni ahhoz, hogy megvalósítható legyen. Az emberi döntéshozatalhoz való folyamodás lehetővé teszi e célok alul-specifikálását. A jogszerűség szempontjából látszólag problematikus, de bizonyos körülmények között kívánatos lehet a jog céljainak ilyen alul-specifikálása. Lehetővé teheti például a jog dinamikus időbeli aktualizálását. Az is lehet, hogy korlátozza a jogalkotóknak a jog teljes körű meghatározására vonatkozó hatáskörét azáltal, hogy megőrzi egy háttérben álló emberi döntéshozó szabad mérlegelési jogkörét. A tekintély felülvizsgálatának és másodlagos megítélésének intézményes lehetőségét a liberális alkotmányos demokrácia egyik elemeként különösen vonzóan találjuk. Természetesen semmi sem indokolja, hogy a megmásíthatóság intézményesítése emberi döntés formájában történjen. Elképzelhető az is, hogy hasonló célt az algoritmus rendszerszintű működésére összpontosító nyilvános ellenőrzési rendszerrel is elérhetünk. Valami ilyesmi máshol is telepíthető egy politikai rendszerben, ami semmissé teszi az emberi döntéshez való jogot. Alternatív megoldásként egy adott gépi döntési eszközt el lehet utasítani azon az alapon, hogy az akadályozza az új jogi szabályok dinamikus fejlődését. Egy gépi tanulási eszköz ugyanis idővel megerősítő tanulás révén finomíthatja az osztályozási szabályát, de ez nem feltétlenül eredményez részletes új jogi iránymutatást az elsődleges magatartásra vonatkozóan. Ismétlem, ez az érv az emberi döntéshozatal mellett

³²¹ Alessandro Acquisti, Leslie K. John és Boewenstein, *Mit ér a magánélet?*, J 42. LEG. STUD. 249, 249-51 (2013).

³²² Alessandro Acquisti et al., *The Economics of Privacy*, ECON. IRODALOM 442-43442, (2016). Hasonló eredményekhez lásd Kirsten Martin, *Privacy Notices as Tabula Rasa: An Empirical Investigation into How Complying with a Privacy Notice Is Related to Meeting Privacy Expectations Online*, J34. Pub. Pol'y & Marketing (2015)210.

³²³ A vállalati hatalommal ilyen irányú érvelésért lásd: TIM WU, *THE CURSE OF BIGNESS: ANTITRUST IN THE NEW GILDED AGE* (2018).

attól függően, hogy nincsenek más ³²⁴ , részletesebb elsődleges magatartási szabályok finomítására és közzétételére.

Be rendszerszintű aggályok egyike ³²⁴ ahhoz, hogy önálló egyéni jogot indokolja. A jobb esetben is a gépi döntés helyett az emberi döntés hasznos, bár nem nélkülözhetetlen eszköze az intézményi tervezésnek a pozitív tovagyrűző hatások létrehozásához. Az, hogy ez a helyettesítés kívánatos-e, röviden, az intézményi tervezés egyéb elemeitől függ. Ezt rosszul írják le "jogként".

E. Az emberi döntéshez való "jog" Dubitante

A teljesen automatizált döntéshozatal sokakban ³²⁴ Az emberi döntéshez való jog lehetséges indokainak taxonómiája pedig sokféle normatív megfontolást tár fel. Egyesek lazán besorolhatók az "autonómia" és a "méltóság" címkék alá. Némelyik az indoklás lehetőségére vonatkozik, míg mások az állami cselekvés indokolt jellegére. Megint mások az egyének és az állam vagy a vállalati szereplők közötti hatalmi egyensúlyra vonatkoznak. Mindezen indokok alapos elemzése azonban azt sugallja, hogy egyik sem igazolhatja az emberi döntéshez való jogot. Egyes esetekben, mint például a pusztán részvételhez való jog, az analóg jog elismerésének hiánya a rokon jogterületeken kizáró oknak tűnik. Más esetekben, beleértve az indokoláshoz és az indokolt döntéshez való érdekeltséget, a jog melletti érv olyan doktrinális és technológiai körülményekre épül, amelyek csak alkalmanként és csak véletlenül teljesülnek. Ennek eredményeképpen az egyéni mérlegeléshez való jog sokkal mulandóbb, mint ahogy az első pillantásra tűnik; amennyiben szilárd normatív alapokon áll, valószínűleg nem alkalmazható a gépi tanulási technológia legtöbb gyakorlati állami alkalmazására.

Számos okunk van arra, hogy ³²⁴ gyűnk az új algoritmikus technológiák avulzív előretörésével ³²⁴ kezdve a munkaerőpiacra gyakorolt hatásuktól a történelmi előítéletek terjesztéséig és a társadalmi rétegződés megerősítéséig. A normatív és jogi aggályok azonban nem irányulhatnak a GDPR cikkében foglaltakhoz hasonló emberi döntéshez való jog megfogalmazására vagy érvényesítésére. 22. Legalábbis egyelőre lehet, hogy ez egy olyan játék, amely egyszerűen nem éri meg a gyertyát.

Következtetés: Joga van a jól kalibrált gépi döntéshez?

E tanulmány célja az volt ³²⁴ tartarja az emberi döntéshez való jog lehetséges igazolásait. A normatív és technológiai terep e felmérése alapján elutasítottam ezt a jogot mint vonzó általános jogmódosítást. Bár nem zárhatom ki, hogy egy ilyen jog az államra (elemzésem fő tárgyára) alkalmazva minden körülmények között kívánatos lenne, elemzésem aláássa az emberi döntéshez való jog általános, mindenre kiterjedő rendszerként megvalósított plauzibilitását.

Végezetül hadd tegyek egy, ³²⁴ provokatív, bár óvatos gondolatot: Ahelyett, hogy az emberi döntéshez való jogról gondolkodnánk, talán jobban járnánk, ha a jól kalibrált gépi döntéshez való jogról beszélünk? Az előző elemzésben elszórtan felsejlenek egy ilyen jog alapjainak villanásai: Ezeket itt csak röviden gyűjthetem össze, hogy a lehetőségről való gondolkodásra ösztönözzek.

³²⁴ Binns et al., *supra* note at 7,385.

A gépi ~~haló~~ jogról szóló beszámoló azzal a megfigyeléssel kezdődik, hogy a legvalószínűbb körülmények között egy jól megtervezett gépi tanulási eszköz pontosabb és kevésbé valószínű, hogy hátrányos megkülönböztetést alkalmaz, mint az emberi döntéshozók. A kormányok által jelenleg alkalmazott algoritmikus eszközök közül azonban sok nagyon hibás.³²⁵ Ez azonban nem jelenti azt, hogy vissza kellene térni a hasonlóan hibás emberi döntéshozatalhoz. A jogi szabályoknak inkább az ilyen hibák kijavítását kellene ösztönözniük. A dinamikus célnak a jobb gépi döntésnek kell lennie. Amennyiben a Due Process érdekeit a pontos döntéshozatali eljáráshoz fűződő érdek igazolja, akkor ezek vitathatatlanul a több és jobb, nem pedig a kevesebb vagy rosszabb automatizálás mellett szólnak. Az emberi kiegészítések valóban egyfajta ártalmas "indokolatlan eljárás" lehetnek, ³²⁶amely a pszichológiai kényelem érzését idézi elő a téves pozitív és negatív eredmények növekedésének árán. Feltéve, hogy egy algoritmikus eszköz jól kalibrált abban az értelemben, hogy nem támaszkodik hibás képzési adatokra, és nem alkalmaz olyan előrejelző eszközt, amely megerősíti a rétegződés káros formáit, ez az eszköz jobban teljesíthet, mint egy ember. Ezért a törvényben megfogalmazható jobb jog feljogosítaná az egyéneket arra, hogy jogszerű célt szolgáló, megbízhatóan megtervezett algoritmust kapjanak, amelyet elfogulatlan és tanulságos képzési adatokkal látnak el, majd hatékonyan ellenőrzik, hogy megakadályozzák a jogtalan eredmények kialakulását.³²⁷

Az is lehetséges, hogy a ~~gépi~~ döntéshozóknak méltóságban és autonómiában hangzó, alábecsült előnyei lesznek. Gondoljunk az ilyen jogból származó méltósági előnyök lehetőségére. Az algoritmikus terapeuták, mint például a "Woebot", ma már egy-két millió emberrel lépnek kapcsolatba az interneten; amellet, hogy ingyenes, az algoritmikus eszközzel "könnyebb beszélgetni", mert a felhasználók "nem érzik magukat elítéltnak".³²⁸ Ugyanezt a szempontot lehetne felhozni a társkereső algoritmusokra utalva, amelyek egyre nagyobb részt foglalnak el a párkeresési piacon, és amelyek kiszorítják a családot és a barátokat mint közvetítőket; a deperszonalizáció új lehetőségeket könnyíthet meg, és elkerülheti a megaláztatás bizonyos formáit.³²⁹ Könnyen lehet, hogy egy algoritmikus interfész a jóléti juttatásban részesülők számára - akik köztudottan megbélyegzett csoportot³³⁰ alkotnak - szintén csökkentheti a juttatásokkal járó pszichológiai költségeket. Ezek a példák inkább szuggesztívek, mint bizonyító erejűek. De rámutatnak arra, hogy az új technológiák stratégiailag alkalmazva hogyan enyhíthetik az egyenlőtlenséget és hogyan tehetik lehetővé az emberiséget - és nem fordítva.

Összefoglalva, a gépi döntések ~~használat~~ algoritmikus technológiák még gyerekcipőben járnak. És sok szempontból hibásak lehetnek. Túl korainak tűnik azonban annak feltételezése, hogy az emberi döntések globálisan jobbak lesznek a gépi döntéseknél, olyannyira, hogy az előbbiekhöz való jog indokolt lenne. Néha az ellenkezője is igaz lehet. Ezért legalább azt a lehetőséget mérlegelnünk kell, hogy bizonyos körülmények között a jól kalibrált gépi döntéshez való jog lehet a jobb lehetőség.

³²⁵ AI Now, *fenti* megjegyzés, 39,18-22. o.

³²⁶ Samaha, *Supra* note at 254,630.

³²⁷ Nyilvánvaló, hogy ez a tömörített definíció rengeteg részletet ~~amelyeket~~ remélem, hogy a későbbi munkámban ki fogok bontani.

³²⁸ Clive Thompson, *Segíthet az A.I.?*, N.Y. TIMES, Nov. 9

³²⁹ *Hogyan a Internet a Megváltozott a Randizás,* A
ECONOMIST, **agstts** 18, 2018, <https://www.economist.com/briefing///how-the-internet-has->

changed-dating20180818.

³³⁰ Egy klasszikus feldolgozást lásd Joel F. Ellen Jane Hollingsworth, *Stigma, privacy, and other attitudes of welfare recipients*, STAN22. L. REV. 1, 4-5 (1969) (a tapasztalt megbélyegzés dokumentálása).