



Jogi prioritások  
projekt

---

# A mesterséges bírói döntéshozatal kihívásai a liberális demokrácia számára

---

Christoph Winter

## LPP MUNKADOKUMENTUM-SOROZAT 3-2021

In: Bystranowski, P. Janik, & M. Próchnicki (szerk.), *Igazságügyi döntéshozatal: (megjelenés alatt): Empirikus és elméleti perspektívák integrálása.*

# A mesterséges bírői döntéshozatal kihívásai a liberális demokrácia számára

*Christoph Winter\**

március 17,

2021 ABSTRACT

A mesterséges intelligencia (AI) alkalmazása a bírósági döntéshozatalban már a világ számos joghatóságában megkezdődött. Miközben a mesterséges intelligencia látszólag nagyobb méltányosságot, az igazságszolgáltatáshoz való hozzáférést és jogbiztonságot ígér, felmerültek a megkülönböztetés és az átláthatóság kérdései, és nyomás alá helyezték a liberális demokratikus elveket, különösen az óvadékkal kapcsolatos döntésekkel összefüggésben. Ennek ellenére nem készült szisztematikus elemzés arról, hogy a mesterséges intelligencia alkalmazása a bírósági döntéshozatalban milyen kockázatokat jelent a liberális demokratikus értékekre nézve. Ez a cikk ezt az űrt kívánja betölteni a mesterséges igazságügyi döntéshozatalból eredő kihívások azonosításával és kezelésével, a liberális demokrácia három pillérére, nevezetesen a polgárokkal való egyenlő bánásmódra, az átláthatóságra és a bírói függetlenségre összpontosítva. Módszertanilag a munka az emberi és a mesterséges döntéshozatal összehasonlító szemléletét alkalmazza, az előbbit normatív viszonyítási alapként használva az utóbbi értékeléséhez.

A fejezet először is azzal érvel, hogy a polgárokkal szembeni egyenlő bánásmódot javító mesterséges intelligenciát már kifejlesztették, de még nem fogadták el. Másodszor, miközben a mesterséges intelligenciával kapcsolatos döntéshozatal átláthatóságának hiánya komoly kockázatokat rejt magában, amelyekkel foglalkozni kellene,

---

\*Adjunktus , Instituto Tecnológico Autónomo de México (ITAM), Jogi Kar Vendégkutató,  
Harvard Egyetem, Pszichológia Tanszék  
Igazgató, Jogi prioritások projekt  
E-mail: christoph\_winter@fas.harvard.edu

kapcsolódó előadások résztvevőinek a cikkhez fűzött megjegyzéseikért és inspirációikért. Szeretnék köszönetet mondani a könyv szerkesztőinek is a nagyon konstruktív visszajelzésekért.

A mesterséges intelligencia növelheti a lehetőségek és kompromisszumok átláthatóságát is, amelyekkel a politikai döntéshozóknak szembesülniük kell, amikor a mesterséges igazságügyi döntéshozatal következményeit mérlegelik. A lehetőségek ilyen átláthatósága demokratikus szempontból óriási előnyökkel jár. Harmadszor, a hatalomnak az emberi intuíciótól a fejlett mesterséges intelligencia felé történő általános eltolódása veszélyeztetheti a bírói függetlenséget, és ezzel együtt a hatalmi ágak szétválasztását. Bár a diszkrimináció és az átláthatóság tekintetében rendelkezésre állnak vagy a láthatáron vannak fejlesztések, továbbra sem világos, hogyan védhető meg a bírói függetlenség, különösen a fejlett mesterséges igazságügyi intelligencia (AAJI) potenciális fejlődése esetén. A mesterséges igazságügyi intelligencia gyümölcseit biztonságosan és stabilan learatni képes politikai és jogi infrastruktúra kidolgozása a jövőbeni kutatások egyik prioritása kell, hogy legyen ezen a területen.

*Kulcsszavak: Mesterséges intelligencia, liberális demokrácia, diszkrimináció, átláthatóság, bírói függetlenség, hatalmi ágak szétválasztása, fejlett mesterséges igazságügyi intelligencia (AAJI).*

## TARTALOM

ABSZTRAKT1.....	
I. AZ EMBERI ELFOGULTSÁGTÓL AZ ALGORITMIKUS MÉLTÁNYOSSÁGIG6.....	
II. AZ ELJÁRÁSOK ÁTLÁTHATÓSÁGÁTÓL A LEHETŐSÉGEK ÁTLÁTHATÓSÁGÁIG13.....	
III. A HATALMI ÁGAK SZÉTVÁLASZTÁSÁTÓL A BÍRÓSÁGI FÜGGŐSÉGIG18.....	
1. A FEJLETT MESTERSÉGES IGAZSÁGÜGYI INTELLIGENCIA LEHETŐSÉGE19.....	
2. FEJLETT AI ÉS IGAZSÁGÜGYI FÜGGETLENSÉG21.....	
IV. KÖVETKEZTETÉS24.....	
BIBLIOGRÁFIA26.....	

A mesterséges intelligencia (AI) alkalmazása az igazságügyi döntéshozatalban már megkezdődött. Jelentős előrelépés történt, nemcsak az Egyesült Államokban, elsősorban az óvadékkal kapcsolatos döntések tekintetében<sup>1</sup>, hanem Oroszországban<sup>2</sup> és Mexikóban is.<sup>3</sup> Kína több mint 100 robotot helyezett el a bíróságokon, amelyek jogi tanácsadást nyújtanak a nyilvánosságnak,<sup>4</sup> Észtország pedig kísérleti jelleggel olyan programot indít, amelyben a kisebb polgári perekben algoritmusok döntenek.<sup>5</sup> Egy nemrégiben készült felmérés szerint továbbá jogászok úgy vélik, hogy mindössze 25 éven belül a bírósági döntéshozatal átlagosan csaknem 30%-át a mesterséges intelligencia fogja végezni, ami megháromszorozza azt a becslésüket, miszerint a mesterséges intelligencia jelenlegi szerepe a bírósági döntéshozatal kevesebb mint 10%-át teszi ki.<sup>6</sup> Ennek fényében hihetőnek tűnik az a feltételezés, hogy a mesterséges intelligencia jövőbeli fejlődése forradalmasítani fogja az igazságszolgáltatást. Sokak számára a mesterséges intelligencia nemcsak nagyobb méltányosságot, igazságosságot és jogbiztonságot ígér, hanem végre megfelelhet a *tisztességes eljárás* nemzetközileg elfogadott fogalmának jogi követelményeinek, különösen az igazságszolgáltatáshoz való hozzáférés tekintetében, ahogyan azt többek között az Emberi Jogok Európai Egyezményének (EJEE) 6. cikke, az Emberi Jogok Egyetemes Nyilatkozatának (EJEE) 10. cikke, valamint az Emberi Jogok és Népek Jogainak Afrikai Chartájának (ACHPR) 7. és 25. cikke elismeri.<sup>7</sup>

Ugyanakkor az óvadéki határozatokkal összefüggésben felmerültek a megkülönböztetés és az átláthatóság kérdései, ami nyomás alá helyezi a liberális demokratikus elveket. Bár a büntető igazságszolgáltatási rendszerrel kapcsolatos intenzív viták és az "algoritmikus igazságosság" kifejezéssel illetett követelések már megjelentek, az igazságszolgáltatási rendszeren belül az emberi intuícióktól a mesterséges intelligenciára való átállásból eredő további kihívások továbbra is tisztázatlanok és elhanyagoltak a diskurzusban. Eddig még nem készült szisztematikus elemzés arról, hogy milyen kockázatokat jelent a liberális demokráciára nézve a mesterséges intelligencia bevezetése a bírósági döntéshozatalba. Ez a cikk ezért ezt az úrt kívánja betölteni a mesterséges bírói döntéshozatalból eredő kihívások azonosításával és azokkal való foglalkozással.

---

<sup>1</sup>Angwin et al. (2016).

<sup>2</sup>Zavalyova V (2018).

<sup>3</sup>Mexikóban az Expertius rendszer tanácsot ad a bírának és a hivatalnokoknak annak megállapításával kapcsolatban, hogy a felperes jogosult-e nyugdíjra vagy sem. *Lásd* Carneiro D et al (2015).

<sup>4</sup>World Government Summit (2018). A Xiao Fa nevű robot homályosan humanoid külsejű, és egyszerű jogi tanácsokat ad, például hogyan lehet pert indítani, vagy hogyan lehet lekérdezni az ügyek történetét, az ítéleteket és a törvényeket.

<sup>5</sup>Niiler, E (2019).

<sup>6</sup>Martinez és Winter (2021). 2021. január 8-ig 307, az Egyesült Királyságban, Indiában, Új-Zélandon, Bangladesben, Ausztráliában, Kanadában és Dél-Afrikában élő jogtudós válaszolt a felmérés vonatkozó kérdéseire.

<sup>7</sup>A tisztességes eljáráshoz kapcsolódó különböző jogokról *lásd még* az Egyesült Államok

---

alkotmányának hatodik módosítását.



Így ez az elemzés nem csupán egy első kísérletet tesz e kihívások rendszerezésére, hanem számos olyan kulcsfontosságú kérdést is felvázol, amelyek további kutatást igényelnek. Az elemzés során az emberi gondolkodási folyamatok és a mesterséges intelligencia<sup>8</sup> összehasonlító perspektíváját fogom alkalmazni, hangsúlyt fektetve a (I.) nagymértékű diszkriminációból és (II.) az igazságszolgáltatáson belüli átláthatóság esetleges hiányából eredő etikai kérdésekre. Valójában azzal fogok érvelni, hogy ezek a kérdések rövid és középtávon veszélyt jelentenek. Ugyanakkor - leginkább Kleinberg és munkatársai által az óvadékkal kapcsolatos döntésekkel kapcsolatban nemrégiben végzett kutatásokra<sup>9</sup> támaszkodva - biztos vagyok abban, hogy a kapcsolódó technikai és etikai kérdések megoldhatók úgy, hogy az emberi döntéshozatalnál jobb eredmények szülessenek, a megkülönböztetéssel kapcsolatos filozófiai szempontok széles skálájából. Míg a mesterséges intelligencia szűken értelmezve a döntéshozatal átláthatóságának hiánya valóban kockázatokat rejt magában, amelyekkel foglalkozni kellene, a mesterséges intelligencia egy másik - vitathatatlanul fontosabb - területen is növelheti az átláthatóságot, amelyet a továbbiakban a *lehetőségek átláthatóságának* nevezek. A liberális demokratikus értékeket fenyegető sokkal nagyobb hosszú távú veszélyt (III.) a hatalomnak az emberi intuíciónál a fejlett mesterséges intelligencia felé történő általános eltolódása jelenti, pontosabban a bírói függetlenséget és a hatalmi ágak szétválasztását fenyegető, esetlegesen ezzel együtt járó veszélyek. Ennek megfelelően az AI-alapú igazságszolgáltatás jelentősen hozzájárulhat a digitális tekintélyelvűség erősödéséhez.

Mielőtt ezekkel a kérdésekkel foglalkoznánk, tisztázni és korlátozni kell az elemzés hatókörét. A diskurzuson belüli közös keret az a különbségtétel, hogy a mesterséges intelligencia *kiegészíti* vagy *helyettesíti-e* az embert.<sup>10</sup> Vajon az AI segíteni fogja az emberi bírákat az elfogultságok felismerésében vagy a jobb döntések meghozatalában, vagy teljesen felváltja az emberi bírákat? A jogi és politikai viták továbbá általában az úgynevezett mesterséges intelligencia *szűk körű* alkalmazásaira összpontosítanak, amelyek gépi tanulást vagy egyszerű döntési fákat használnak az emberi döntéshozatal kiegészítésére vagy részleges helyettesítésére. E fejezet első két részében e részleges helyettesítésre, valamint a diszkrimináció és az átláthatóság megfelelő kérdéseire fogok összpontosítani. Az utolsó szakaszban azonban áttérek a mesterséges intelligencia fejlettebb formáira, amelyek az emberi igazságügyi döntéshozatal túlnyomó többségét helyettesíthetik, és ezzel túllépek a jelenlegi alkalmazásokon, lehetőségeken és a szakirodalom fő fókuszán. Bár ez a mesterséges intelligencia fejlett formáiról szól, amelyek jelentősen meghaladják a jelenlegi képességeket, meg kell jegyeznünk, hogy ehhez nem lenne szükség olyan szuperintelligenciára, amely "gyakorlatilag minden érdekes területen jelentősen meghaladja az emberek kognitív teljesítményét"<sup>11</sup>.

---

<sup>8</sup>Megjegyezzük, hogy ez a megközelítés, bármennyire is intuitívnak hangzik, nem annyira gyakori. A mesterséges intelligencia igazságügyi ágazatban való alkalmazásának legtöbb kritikusja ugyanis a jelenlegi mesterséges intelligencia hátrányaira összpontosít,

---

miközben figyelmen kívül hagyja az emberi döntéshozatal óriási hiányosságait.

<sup>9</sup>Kleinberg et al. (2018); Kleinberg et al. (2019).

<sup>10</sup>Vö. többek között Sourdin (2018).

<sup>11</sup>Vö. *általában*: Bostrom (2014); hasonlóan Bostrom (2006), 11. o.

Később kifejtem, hogy ehhez egy *fejlett mesterséges igazságügyi intelligencia* (AAJI) kifejlesztésére lenne szükség, amely korlátozottabb területi ismeretekkel és képességekkel rendelkezik.

A kutatók és a politikusok körében egyaránt uralkodó érvnek tűnik, hogy inkább a jelen problémáira kell összpontosítani, mintsem a jövőben felmerülő problémák megoldására törekedni.<sup>12</sup> A jövőbeli fejleményekkel - például a mesterséges intelligencia fejlettebb formáinak az igazságszolgáltatásban való alkalmazásával és az ezzel járó kockázatokkal - kapcsolatos kutatás azonban kulcsfontosságú, mivel a jövőbeli kockázatok felvázolása értékes információkkal szolgálhat arról, hogy milyen fejlesztéseket kell különös körültekintéssel követni. Ha például a jövőben a mesterséges intelligencia alkalmazása nagy kockázatot jelenthet a liberális demokráciára nézve, akkor már a jelenben meg lehet próbálni megtalálni az ilyen kockázatok mérséklésének módjait, ha a liberális demokráciát kívánatos politikai rendszernek tartjuk. Egyszerűen fogalmazva: hasznos lehet eleve elkerülni a ragadozóval való találkozást, ahelyett, hogy akkor próbálnánk elmenekülni, amikor az már előttünk van.

## I. AZ EMBERI ELFOGULTSÁGTÓL AZ ALGORITMIKUS IGAZSÁGOSSÁGIG

Ez az elemzés először a minden polgárral szembeni egyenlő bánásmódra összpontosít, amely vitathatatlanul *a liberális* demokráciák megkülönböztető jegye.<sup>13</sup> Az igazságszolgáltatásban a mesterséges intelligenciáról folytatott tudományos<sup>14</sup> és népszerű<sup>15</sup> vita nagy részét a nemi vagy faji alapon történő jogellenes megkülönböztetés kérdései uralták. Ez érthető, tekintve, hogy a Google arcfelismerő algoritmus a feketéket gorillának bélyegezte,<sup>16</sup> a munkahelyi kiválasztási döntésekben alkalmazott algoritmusok a fehér férfiakat részesítették előnyben,<sup>17</sup> és a fehérek jobban részesültek, mint a

---

<sup>12</sup> *Lásd pl.* Reiling (2018): "Vö. még Sourdin és Cornes (2018), 113. o., akik úgy tűnik, hogy az ilyen megfontolásokat "nem hasznosnak" minősítik. Miközben egyetértek azzal, hogy a belátható jövőben "az emberi intelligenciát (inkább) a technológiai fejlődés egészítheti ki", szinte naivnak tűnik, ha csak a valószínűbb forgatókönyveket vesszük figyelembe, és elutasítjuk annak mérlegelését, hogyan kellene cselekednünk abban az esetben, ha kevésbé valószínű forgatókönyvek bontakoznak ki. Ehelyett az ilyen aggodalmak és a kapcsolódó kutatások nemcsak a várható érték szempontjából tűnnek indokoltnak, hanem az elővigyázatosság elvének bármely ésszerű változata alapján is. Vö. még a mesterséges intelligenciával kapcsolatos közeli és hosszú távú aggodalmak áthidalásáról szóló, kialakulóban lévő érveket: Baum, 2018, 2020; Cave és Ó hÉigeartaigh, 2019; Prunkl és Whittlestone, 2020.

<sup>13</sup> Mukand és Rodrik (2020).

<sup>14</sup> *Vö.* többek között Chander (2017); Chen (2019); Hacker (2018); Kleinberg et al. (2019); Sourdin (2018), *Supra note* 5, 1128-1129. o.

<sup>15</sup> *Vö. pl.* Levin (2016); Smith (2016).

<sup>16</sup> Zhang (2015).

---

<sup>17V6.</sup> *pl. az Egyesült Királyságban az orvosi egyetemre való jelentkezés jól dokumentált esetét:*  
Lowry és Macpherson (1988).

többek között a COMPAS<sup>18</sup> visszaesési algoritmus<sup>19</sup> fekete bőrűek esetében. Egyértelmű, hogy javításra van szükség. Az azonban már más kérdés, hogy ezek a problémák indokolják-e a mesterséges intelligencia igazságügyi ágazatba való beépítésének megszüntetését vagy mellőzését. Végül is az emberi intelligencia (HI) sokkal hosszabb története az igazságszolgáltatásban azt látszik jelezni, hogy az emberi döntéshozatal sem szolgáltatott igazságot minden csoportnak. Bár lehetséges lehet, hogy bizonyos mértékig enyhíteni lehet az implicit emberi elfogultságokból eredő megkülönböztetést,<sup>20</sup> meg kell határoznunk, hogy a mesterséges intelligencia mennyire képes legyőzni az ilyen elfogultságokat a HI-hoz képest. Ha kellően jó esély van arra, hogy a mesterséges intelligencia hosszú távon még jobban is képes erre, akkor nem szabad teljesen elvetni a diszkrimináció miatti mesterséges igazságügyi döntéshozatal gondolatát, hanem inkább arra kell koncentrálni, hogy miként lehet a mesterséges intelligencia által a diszkriminációt enyhíteni. Ahhoz, hogy az AI fontos szerepet játsszon a jövő igazságszolgáltatásában, nem kell tökéletesnek lennie, csak *jobbnak*, mint a HI. Sőt, ilyen esetekben talán még kötelességünk is lehet elfogadni.

A diszkriminációval kapcsolatban a joghatóságok között figyelemre méltóan hasonlóan működik a *közvetlen* és a *közvetett* megkülönböztetés megkülönböztetése, amelyet megkülönböztetett *bánásmódnak*, illetve megkülönböztetett *hatásnak* is neveznek.<sup>21</sup> Közvetlen megkülönböztetés alatt általában azt értjük, hogy egy személyt bármely védett ok - például nem, faji vagy etnikai származás, vallás, fogyatékoság, életkor vagy szexuális irányultság - alapján kedvezőtlenebbül kezelnek, mint egy másik, hasonló helyzetben lévő személyt.<sup>22</sup> A hátrányos bánásmód tehát meghatározott jellemzők birtoklásán alapul. Itt az esetek a védett ok és a kedvezőtlenebb bánásmód közötti okozati összefüggés<sup>23</sup> elemzésén, valamint a hasonló helyzetben lévő személyek összehasonlíthatóságának vizsgálatán alapulnak.

---

<sup>18</sup> A COMPAS a "Correctional Offender Management Profiling for Alternative Sanctions" rövidítése, amelyet a Northpointe (ma Equivant) fejlesztett ki. A COMPAS visszaesési algoritmust az amerikai bíróságok számos államban használják annak felmérésére, hogy a vádlott milyen valószínűséggel válik visszaesővé.

<sup>19</sup> Lásd Larson et al. (2016).

<sup>20</sup> Lásd többek között Carnes et al. (2015); Lai et al. (2014); Devine et al. (2012).

<sup>21</sup> Lásd például a faji egyenlőségről szóló uniós irányelv 2. cikkének (2) bekezdését; EJE, Biao kontra Dánia [GC], 38590/10. sz. ügy, 2016. május 24., 89-90. pont. Az amerikai megközelítéshez lásd a McGinley (2011), 626. o. által közölt meghatározásokat.

<sup>22</sup> Lásd pl, az uniós jogban a faji egyenlőségről szóló uniós irányelv 2. cikkének (2) bekezdése által kínált nagyon hasonló fogalom meghatározásokat, az EJEB ítélezési gyakorlatát az EJE, Biao kontra Dánia [GC], No. 38590/10, 2016. május 24., (2) bek. 89.; és EJE, Carson és társai kontra Egyesült Királyság [GC], 42184/05. sz. ügy, 2010. március 16.; valamint az Egyesült Államok megközelítése a Bolling kontra Sharpe, 347 U.S. 497, 499, 74 S. Ct. 693, 694, 98 L. Ed. 884 (1954); Brown v. Board of Education, 347 U.S. 483 (1954); Washington v. Davis, 426 U.S. 229, 239, 96 S. Ct. 2040, 2047, 48 L. Ed. 2d 597 (1976). Egyes jogrendszerek, pl. a német Alaptörvény 3. cikke, ezen okok nem kimerítő felsorolását részesítik előnyben, ami azt jelenti, hogy szükség esetén további okok is hozzáadhatók.

<sup>23</sup> Az ilyen okozati összefüggés megállapításához a következő kérdésre kell választ adni: "Vajon a

---

kevésbé kedvező banásmódban részesült volna, ha más nemű, fajú, korú, vagy más nemzetiségű,  
vagy

helyzetek. Egyes jogrendszerekben a közvetlen megkülönböztetés igazolható,<sup>24</sup> máshol nem.<sup>25</sup>

Közvetett megkülönböztetésről akkor van szó, ha egy látszólag semleges kritérium, rendelkezés vagy gyakorlat hátrányos helyzetbe hozná a megkülönböztetés általános tilalma által védett személyeket (pl. a védett tulajdonsággal rendelkezőket) másokkal szemben.<sup>26</sup> Ha ilyen megkülönböztetés történik, annak igazoltnak kell lennie, azaz a szóban forgó rendelkezés csak akkor lesz helytálló, ha annak legitim célja van, és e cél elérésének eszköze szükséges és megfelelő. A közvetlen megkülönböztetéssel ellentétben a közvetett megkülönböztetés látszólag semleges kritériumokon alapul, amelyek formálisan nem tiltottak. Mindazonáltal mind a közvetlen, mind a közvetett megkülönböztetés következménye lényegében ugyanaz: a védett csoporthoz tartozó egyén hátrányos helyzetbe kerül.

Annak ellenére, hogy e meghatározások alkalmazása az eset pontos tényállásától és a rendelkezésre álló bizonyítékoktól függően bonyolult lehet, az algoritmusokkal összefüggésben meglehetősen jól működnek, mivel kifejezetten meghatározzák, hogy mely kritériumokat veszik közvetlenül figyelembe, és melyeket nem. Például az óvadékkal kapcsolatos döntésekkel összefüggésben az olyan algoritmusok, amelyek nem veszik közvetlenül figyelembe a faji, nemi, vallási vagy egyéb tiltott tényezőket az egyén szökési kockázatának értékelésénél, nem jelentenek közvetlen megkülönböztetést.<sup>27</sup> Ugyanakkor, ahogyan azt korábban említettük, a COMPAS visszaesési algoritmus rosszabb hatással volt a színesbőrűekre a fehérekhez képest. Így az algoritmus közvetett módon diszkriminálhat azáltal, hogy a menekülési kockázat előrejelzésekor olyan látszólag semleges kritériumokat vesz figyelembe, mint az irányítószám, az iskolai végzettség vagy a rossz hitelminősítés.

---

a többi védett ok bármelyike alapján?" Ha a válasz "igen", akkor a kedvezőtlenebb bánásmódot a szóban forgó okok okozzák.

<sup>24</sup> *Lásd például az Egyesült Királyság esetében: Professor John Pitcher kontra Chancellor, Masters and Scholars of the University of Oxford and Saint John the Baptist College in the University of Oxford [2019] UK Employment Tribunals 3323858/2016. Az Egyesült Államokban még az olyan fokozottan védett kategóriák esetében is, mint a faj és a vallás, a közvetlen megkülönböztetés akkor igazolható, ha az átmegy a szigorú vizsgálaton (strict scrutiny). Megjegyzendő azonban, hogy ez ritkán van így, Fisher v. Univ. of Texas at Austin, 570 U.S. 297, 310, 133 S. Ct. 2411, 2419, 186 L. Ed. 2d 474 (2013).*

<sup>25</sup> Például az Egyesült Királyságban a közvetlen megkülönböztetés csak az életkor és a fogyatékoság tekintetében igazolható.

*Lásd Egyesült Királyság, 2010. évi egyenlőségi törvény, 2. rész, 2. fejezet, 13. és 19. szakasz.*

<sup>26</sup> *Egyesült Királyság, Egyenlőségi törvény 2010. Vö. még az EJEB ítélkezési gyakorlatát: EJEB, D.H. és társai kontra Cseh Köztársaság [GC] (57325/00. sz.), 2007. november 13., (1) bekezdés. 184; EJEB, Opuz kontra Törökország (33401/02. sz. ügy), 2009. június 9., para. 183; EJEB, Zarb Adami kontra Málta (17209/02. sz. ügy), 2006. június 20., 80. pont.*

<sup>27</sup> Ha nem is ez az egyetlen meghatározó tényező az előzetes letartóztatás előtti szabadlábra helyezés engedélyezése szempontjából, a szökés veszélye számos joghatóságban döntő

---

fontosságú tényező. *Lásd* például New Hampshire államot.

v. Christina A. Hill (2019) Supreme State Court NH 2018-0637; az USA más államaiban annak valószínűsége is számít, hogy a vádlottat letartóztatják vagy elítélik bűncselekmény miatt, *vö.* Dabney et al. (2017), 408. o.; Karnow (2008), 1. o..



kockázat. Mivel a fent említett tényezőket gyakran befolyásolta a korábbi diszkrimináció,<sup>28</sup> jogos az az általános aggodalom, hogy a mesterséges intelligencia *állandósíthatja* a diszkriminációt, még akkor is, ha az algoritmus nem veszi kifejezetten figyelembe a faji vagy nemi hovatartozást.

A COMPAS-ügy azonban, amely erősen alakította a mesterséges intelligenciáról szóló nyilvános vitát az igazságszolgáltatásban, korlátozott értékkel bír a hosszú távú szakpolitika alakítása szempontjából, mivel a pontos algoritmust nem hozták nyilvánosságra, és üzleti titokként tartják nyilván.<sup>29</sup> Míg az eset azt mutatja, hogy minden bizonnyal lehetséges olyan algoritmusokat készíteni, amelyek közvetve hátrányosan megkülönböztetik a védett csoportokat (sőt, a fejlesztés és alkalmazás során alkalmazott biztosítékok hiányában talán még valószínű is a diszkrimináció), a mi elsődleges kérdésünk az, hogy lehetséges-e olyan algoritmust vagy mesterséges intelligenciát készíteni, amely az emberi döntéshozatalhoz képest különböző filozófiai szempontok alapján előnyös, még a diszkriminációs kérdések tekintetében is. Ebben az értelemben Kleinberg és munkatársai nemrégiben végzett kutatásai azt mutatják, hogy az ilyen fejlesztések nem pusztán utópisztikus vízió, hanem ma már megvalósíthatóak.<sup>30</sup> Ismét az óvadékkal kapcsolatos döntésekre összpontosítva, a Kleinberg és munkatársai által kifejlesztett gépi tanulási algoritmust, amely gradiens-növelő döntési fákat használt, 758 027 olyan vádlottból álló nagy adathalmazon képezték ki, akiket 2008 és 2013 között tartóztattak le New Yorkban. Az adatkészlet tartalmazta a vádlott korábbi bűnlajstromát, a jelenlegi bűncselekményt és a bírák számára a döntés meghozatalához rendelkezésre álló egyéb tényezőket. A több mint 100 000 különböző ügyön végzett tesztelés során az algoritmus az emberi bíráknál lényegesen jobbnak bizonyult annak megjóslásában, hogy a vádlottak nem jelennek-e meg, vagy a szabadlábra helyezést követően újra letartóztatják-e őket, ami - a politikai preferenciáktól függően - eltérő, de erős előnyökkel járhat. Pontosabban, a szimulációk azt mutatták, hogy a megjelenés elmulasztása és az újbóli letartóztatás ("bűnözés") akár 24,7%-kal, de legalább 14,4%-kal csökkent, a börtönbüntetési arányok változatlanul hagyása mellett, illetve a börtönbüntetési arányok akár 41,9%-kal, de legalább 18,5%-kal csökkentek a bűnözési arányok növekedése nélkül.<sup>31</sup> A megkülönböztetés megelőzése szempontjából döntő fontosságú, hogy az algoritmus képes volt a bűnözés csökkentésére, miközben egyidejűleg *csökkentette* a faji egyenlőtlenségeket valamennyi bűncselekmény-kategóriában.<sup>32</sup> Az algoritmus elfogadása tehát a politikai döntéshozókat abba a meglehetősen kényelmes helyzetbe hozná, hogy választhatnak aközött a lehetőség között, hogy több ezer embert engednek szabadon előzetes letartóztatás előtt anélkül, hogy a bűnözési ráta növekedne, *vagy pedig* több ezer bűncselekményt akadályoznak meg anélkül, hogy akár egyetlen további személyt is börtönbe zárnának - miközben a faji egyenlőtlenségek csökkennének. Mondanom sem kell, hogy nem ezek az egyetlen lehetőségek, de jól szemléltetik, hogy a politikai döntéshozóknak milyen kompromisszumokat kell majd kötniük a következőkkel kapcsolatban

<sup>28</sup>Vö. a strukturális rasszizmusról szóló vitákat többek között Wallace et al. (2017), Hammer (2018).

---

<sup>29</sup>Vö . Chohlas-Wood A (2020). Az átláthatóság nyilvánvaló hiányát részletesebben a II. szakaszban tárgyalom.

<sup>30</sup>Kleinberg et al (2018).

<sup>31</sup>Kleinberg et al. (2018), 241. o.

<sup>32</sup>Kleinberg et al. (2018), 237-238. o.

a bűnözés és a fogva tartási arányok kiegyensúlyozása, valamint a kisebbségek, különösen a fekete és a spanyolajkú férfiak aránytalan bebörtönzése az Egyesült Államokban, ha a mesterséges intelligencia lép a HI helyébe. Sunstein ugyanis azt állítja, hogy a Kleinberg és munkatársai által kifejlesztett algoritmus "sokkal jobban teljesít, mint a valós bírák (...) *minden olyan dimenzióban, ami számít*".<sup>33</sup>

Fontos, hogy ezek az eredmények nem csak New York városára jellemzőek, mivel Kleinberg és munkatársai egy országos adathalmazban is hasonló minőségi eredményeket tudtak elérni.<sup>34</sup> Azonban megkísérelhetnénk azzal érvelni, hogy továbbra is bizonytalan, hogy ezek az eredmények az Egyesült Államokon kívüli joghatóságokban is elérhetőek-e. Lehet, hogy egyszerűen csak valami nem stimmel az Egyesült Államokban a bírói döntéshozatalban? Lehet, hogy az óvadékkal kapcsolatos döntések tekintetében van egy olyan hiba a rendszerben, amely más joghatóságokban nem létezik? Végül is a common law és a polgári jogi rendszerek különböznek egymástól,<sup>35</sup> és az Egyesült Államokban a jogi oktatás alapvetően különbözik a legtöbb helytől, még más common law rendszerektől is, mint például Ausztrália, az Egyesült Királyság vagy India. Ennek ellenére nem látom, hogy ezek a különbségek miért vezetnének jelentősen eltérő eredményekhez, mert először is a jog az óvadékkal kapcsolatos döntések tekintetében figyelemre méltóan hasonlóan működik.<sup>36</sup> Másodsor, Sunstein rámutat, hogy az adatok elemzése arra utal, hogy a kognitív torzítások magyarázatot adhatnak arra, hogy az AI miért teljesít olyan egyértelműen jobban a HI-nál.<sup>37</sup> Ez releváns és fontos felismerés, tekintettel arra, hogy az elfogultságok világszerte előfordulnak, függetlenül a joghatóságtól.<sup>38</sup> Pontosabban, Sunstein szerint a bírák két alapvető hibát követnek el, amelyek jelentősen befolyásolják az általános teljesítményüket.<sup>39</sup> Úgy kezelik a nagy kockázatú vádlottakat, mintha alacsony kockázatúak lennének, amikor az aktuális vádjuk viszonylag csekély, és úgy kezelik az alacsony kockázatúakat, mintha magas kockázatúak lennének, amikor az aktuális vádjuk különösen súlyos.<sup>40</sup> Úgy tűnik tehát, hogy a bírák minden ilyen esetben túl nagy jelentőséget tulajdonítanak a jelenlegi bűncselekménynek más releváns tényezőkhöz képest, beleértve a vádlott korábbi büntetett előéletét is,

---

<sup>33</sup>Sunstein (2019), p. 2.

<sup>34</sup>Kleinberg et al (2018), 241. o.

<sup>35</sup>Még ha fenntartom is, hogy ezeket a különbségeket még mindig nagyon eltúlozzák; e témában lásd Pejovic (2001) élelátó elemzését.

<sup>36</sup>Baughman (2017), 15. o.

<sup>37</sup>Sunstein (2019), 501. o.

<sup>38</sup> Természetesen nem minden azonosított torzítás fordul elő globálisan ugyanúgy. Azonban, mint látni fogjuk, nagy az esélye annak, hogy a kérdéses torzítás nem csak az Egyesült Államokban fordul elő. Lásd még Dhami és Ayton (2001) tanulmányát, amely bemutatja, hogy az emberi bírálók az Egyesült Királyságban egyszerű heurisztikákat követnek.

---

<sup>39</sup> Sunstein (2019), 502. o.

<sup>40</sup> *Ib.*

életkor és foglalkoztatási előzmények. <sup>41</sup> Sunstein ezt a jelenséget az *aktuális bűncselekményekre vonatkozó előítéletnek* nevezi. <sup>42</sup>

Ezután összekapcsolja az aktuális bűncselekményre vonatkozó előítéletet a jól ismert *elérhetőségi előítéletekkel*, vagyis azzal a tendenciával, hogy túlbecsüljük egy jövőbeli esemény bekövetkezésének valószínűségét, ha könnyen eszünkbe jutnak példák. <sup>43</sup> Még ha kissé szkeptikusak is lehetünk az aktuális bűncselekményi torzítás és a rendelkezésre állási torzítás közötti összefüggés mértékét illetően,<sup>44</sup> az általános pont áll: a különböző joghatóságok bírái valószínűleg ugyanazokat a hibákat követik el, mivel nincs okunk feltételezni, hogy az ázsiai, afrikai vagy európai bírák nem szenvednek az aktuális bűncselekményi torzításoktól. Ez különösen azért van így, mert a kapcsolódó rendelkezésre állási torzítás az elme általános tulajdonságának tűnik, amely az asszociatív távolság alapján próbál hozzáférni egy esemény bekövetkezésének valószínűségéhez. <sup>45</sup>

Mostanra láttuk, hogy a bűnözés és a fogva tartás aránya jelentősen csökkenthető anélkül, hogy a faji megkülönböztetés állandósulna vagy akár növekedne. Ez önmagában üdvözlendő fejleménynek tekinthető, mivel csökkenne a börtönben lévő kisebbségek abszolút száma. De vajon ez azt jelenti, hogy ha elfogadjuk Kleinberg és munkatársai algoritmusát, akkor az afroamerikaiak és más kisebbségek börtönben lévő aránya ugyanannyi marad? Ha feltételezzük, hogy az emberi döntéshozatal előbb-utóbb csökkenteni fogja a diszkriminációt, és ezáltal nemcsak az abszolút, hanem a börtönben lévő afroamerikaiak relatív számát is, akkor felmerül a kérdés, hogy az algoritmusok egy idő után lemaradnának-e a HI mögött. Plauzibilisnek tűnik az a feltételezés, hogy mind az abszolút, mind a relatív fogvatartási arány a jövőben csökkenni fog a kisebbségek esetében, ha az emberközpontú igazságszolgáltatás követi az emberiség erkölcsi fejlődéstörténetét. <sup>46</sup> Azonban időbe fog telni, amíg a szükséges és döntő fontosságú kognitív funkciókat fejleszteni fogják, hogy ez teljesebb mértékben megvalósuljon, például az implicit előítéletek felülbírálásának képességeinek fokozásával. <sup>47</sup> Ugyanakkor lehetséges egy algoritmust úgy instruálni, hogy az

---

<sup>41</sup>Ib .

<sup>42</sup>Ib .

<sup>43</sup>Tversky és Kahneman (1982).

<sup>44</sup>Sunstein (2019), 502. o. "közeli rokonoknak" tekinti őket.

<sup>45</sup> Tversky és Kahneman (1973). Mindezek után a jelenlegi bűncselekményi torzítás joghatóságokon átívelő alkalmazása végső soron empirikus kérdés, amely empirikus megerősítést igényel.

<sup>46</sup> Pinker (2011; 2018).

<sup>47</sup> Félreteszem az egyéb jelentős problémákat, amelyekkel ezek a megközelítések járhatnak, például a magánélethez és a gondolatszabadsághoz kapcsolódó aggályokat. Félreteszem azt a kérdést is, hogy az amerikai büntető igazságszolgáltatási rendszerben a diszkriminatív eredményeket jelenleg milyen mértékben befolyásolja az implicit faji előítéletesség, vagy más tényezők, például az oktatás, a szegénység és a nagyobb rendőri jelenlét az afroamerikai közösségekben a fő mozgatórugók

---

ALGORITMIKUS MÉLTÁNYOSSÁGIG  
(miközben elismerem, hogy az ilyen tényezőkre viszont hatással vannak a következők)

hogy erkölcsileg vagy társadalmilag kívánatos eredményeket hoz, bármi legyen is az. Például az egyik forgatókönyvben Kleinberg és munkatársai arra utasították az algoritmust, hogy tartsa fenn ugyanazt a fogva tartási arányt, miközben minden faj esetében kiegyenlíti a szabadon bocsátási arányt. Ezt figyelembe véve az algoritmus még mindig képes volt 23%-kal csökkenteni a meg nem jelenés és az újbóli letartóztatás ("bűnözés") arányát.<sup>48</sup> Egy másik szimulációban az algoritmust arra utasították, hogy ugyanazt a bűnözési arányt produkálja, mint amit a bírák jelenleg elérnek. Ilyen körülmények között megdöbbentő 40,8%-kal kevesebb afroamerikai és 44,6%-kal kevesebb spanyolajkú került volna börtönbe.<sup>49</sup>

Elgondolkodhatunk azon, hogy ezek az eredmények valóban arra utalnak-e, hogy az emberi döntéshozatalt a megkülönböztetés szempontjából *ki* kellene *egészíteni*, vagy akár (részben) *helyettesíteni* kellene a mesterséges intelligenciával. Ha az egész a kognitív előítéletekről szól - akár az implicit faji előítéletekről, akár az aktuális támadási előítéletekről -, akkor talán megpróbálhatjuk javítani az emberi döntéshozatalt, mielőtt teljesen megszabadulnánk tőle. Valóban hihető, hogy bizonyos fokig enyhíthetünk bizonyos előítéleteket. Például a rendelkezésre állási torzítás erőssége akár azzal is csökkenthető, ha egyszerűen tudatában vagyunk.<sup>50</sup> Emellett korábban már javasoltam különböző intézményi és eljárási változtatásokat, valamint a bírói kar tagjai számára kötelező képzést a viselkedési közgazdaságtan és a kognitív torzítások témakörében az elfogultságok mérséklése érdekében.<sup>51</sup> Ezek teljes kiküszöbölése azonban ilyen módon nem tűnik megvalósíthatónak.<sup>52</sup> Egyszerűen fogalmazva, ahhoz, hogy napjainkban teljesen megszabaduljunk a kognitív elfogultságoktól, meg kell szabadulni az emberi döntéshozataltól.<sup>53</sup>

Összefoglalva, Kleinberg és munkatársai megállapításai a diszkrimináció csökkentése szempontjából jelentősen megerősítik a mesterséges intelligencia mellett szóló érveket az igazságszolgáltatásban. Különösen azért, mert ezzel egyidejűleg csökkenthető a bűnözés és a fogva tartás aránya. Bár az emberi döntéshozatalról még sok mindent el lehet mondani, a kognitív

---

strukturális megkülönböztetés). Az implicit előítéletesség szerepéről lásd például Oswald et al. (2013) kritikai megjegyzését.

<sup>48</sup> Kleinberg et al. (2018).

<sup>49</sup> Kleinberg et al. (2018).

<sup>50</sup> Gigerenzer (1991). Az implicit előítéletek elleni küzdelemről *lásd* Teal et al (2012), Boscardin (2015), Ingriselli (2015), Shaked-Schroer (2008).

<sup>51</sup> Tél (2020).

<sup>52</sup> *Ib.*

<sup>53</sup> *Ib.* Megjegyzendő azonban, hogy ez nem jelenti azt, hogy a mesterséges intelligencián alapuló igazságszolgáltatásnak nem kell szembenéznie az ilyen kockázatokkal. Természetesen magának a mesterséges intelligenciának a kialakítása, valamint a képzési adatok kiválasztása ismét az emberi döntéshozataltól függ. Ennek ellenére

---

reálisabb lehet az említett tervezési és kiválasztási folyamaton belüli torzítások elkerülése, mint egy emberközpontú bírói rendszer lebutítása, ahogy azt Kleinberg és munkatársai tanulmányának eredményei jelzik.



elfogultságok és az empátia szerepe ebben,<sup>54</sup> a jelenlegi közhiedelemmel ellentétben a mesterséges intelligencia "a (...) méltányosságot szolgáló erő" lehet.<sup>55</sup> Bár fontos figyelmeztetni arra, hogy Kleinberg és munkatársai algoritmusai csupán óvadékkal kapcsolatos döntésekkel foglalkozott, ésszerűnek tűnik feltételezni, hogy általában könnyebb lenne bizonyítani, hogy közvetlen megkülönböztetés történt-e az algoritmikus, mint az emberi bírói döntéshozatal során. Valójában, ha a döntéshozatali eljárás átlátható, akkor ez annyira nyilvánvaló lenne, hogy eleve nem valószínű, hogy előfordulna. A bűnözés, a fogva tartás és a megkülönböztetés közötti előnyös kompromisszumokat továbbra is meg kellene kötni, de függetlenül attól, hogy hogyan döntünk, az eredmény a HI-hez képest előrelépés lenne. Mindezek után fontos megjegyezni - amint azt a COMPAS-ügyben is tapasztaltuk -, hogy a pontos kompromisszumok nem mindig láthatók. Tehát el kell-e utasítani az AI alkalmazását, még ha az képes is csökkenteni a diszkriminációt, az átláthatóság hiánya miatt?

## II. AZ ELJÁRÁSOK ÁTLÁTHATÓSÁGÁTÓL AZ ELJÁRÁSOK ÁTLÁTHATÓSÁGÁIG OPCIÓK

A demokrácia és az átláthatóság közötti kapcsolat erősen vitatott. A széles körben elterjedt intuícióval ellentétben még az is előfordulhat, hogy a tekintélyelvűség nagyobb átláthatósághoz vezet, mint a demokrácia. Ez azért lehet így, mert a közvélemény rosszallásával szembeni nagyobb sebezhetőség miatt a demokratikusan megválasztott tisztviselők hajlamosabbak lehetnek az átláthatatlanság előmozdítására vagy az információ visszatartására autokratikus társaikhoz képest, akiknek kevésbé kell aggódniuk a közvélemény megítélése miatt.<sup>56</sup> Azt, hogy a demokratikus kormányokat ösztönzi politikájuk elkendőzésére, jól szemléltethetik Horst Seehofer német belügyminiszter közelmúltbeli megjegyzései, aki nemrégiben kijelentette, hogy "a törvényeket bonyolulttá kell tenni".<sup>57</sup> Annak ellenére, hogy maga a kijelentés politikai szempontból vitathatóan nem volt jól átgondolt, és Seehofer nem sokkal később azzal próbált érvelni, hogy "kissé ironikusan" fogalmazott, a kijelentés mégis a demokrácia és az átláthatóság közötti feszültség lényegét érinti. Ha egy adott törvény elfogadása a választott tisztségviselők szempontjából *jobb* a polgárok számára, mint bármelyik alternatíva, de a szóban forgó törvény nagyon népszerűtlen, akkor mégis el kell-e fogadni? És ha igen, szabad-e azt kommunikálni?

<sup>54</sup> Az empátiát gyakran hozzák fel érvként a mesterséges intelligencia ellen. Megjegyzendő azonban, hogy ez nem mindig az az együttérző, kedves fogalom, aminek gyakran gondoljuk, sőt, a diszkrimináció nagy részéért is felelős lehet. *Lásd* általánosságban Bloom (2016).

<sup>55</sup> Kleinberg et al (2018) 241. o.

<sup>56</sup> Hollyer et al. (2011).

<sup>57</sup> A megjegyzéseivel kapcsolatos további információkat *lásd* Das Gupta és Fried (2019).

úgy, hogy az újraválasztás valószínűbbé váljon, annak ellenére, hogy a polgárok nem értenek egyet a meghozott intézkedéssel, még akkor sem, ha végső soron hasznot húznak az új törvényből?

Az az érv, hogy a demokrácia *valójában* nem mindig az átláthatóság ideális motorja, nem jelenti azt, hogy az átláthatóság *normatív* szempontból nem döntő tényező. Természetesen a tágan értelmezett átláthatóság segít a korrupció elleni küzdelemben, a közintézményekbe vetett bizalom előmozdításában és a közbeszédhez való hozzájárulásban. <sup>58</sup> Ahhoz azonban, hogy az átláthatóság helyzetére vonatkozó konkrétabb kérdésekre választ adhassunk, amikor a mesterséges igazságügyi döntéshozatalról van szó demokratikus szempontból, először is meg kell vizsgálnunk az átláthatóság demokrácián belüli funkcióját. Érdekes módon, annak ellenére, hogy az átláthatóság mint fogalom láthatóvá vált a közbeszédben, meglepően kevés figyelmet fordítottak a mögöttes céljára. <sup>59</sup> Gupta még "túlhasznált, de alulelemzett fogalomnak" is nevezi. <sup>60</sup> Az alapvető érv amellet, hogy az átláthatóság *minden* demokratikus rendszer alapvető része, a következő vonalak mentén fut:<sup>61</sup>

1. A demokrácia megköveteli, hogy a polgárok aktívan részt vegyenek a politikában.
2. Ahhoz, hogy a polgárok részt tudjanak venni a politikában, képesnek kell lenniük arra, hogy releváns információk alapján politikai döntéseket hozzanak.
3. Ahhoz, hogy a polgárok releváns információkon alapuló politikai döntéseket hozhassanak, hozzá kell férniük a releváns információkhoz.
4. Ahhoz, hogy a polgárok hozzáférjenek a releváns információkhoz, ezeknek az információknak átláthatónak kell lenniük.

A központi kérdés ebben a keretrendszerben az, hogy pontosan mi számít *releváns* információnak. A polgároknak tudniuk kell, *hogy ki* és *hogyan* hozta meg a kérdéses döntést, vagy tudniuk kell arról is, hogy milyen alternatív útvonalakat lehetett volna választani? Milyen információknak kell a polgárok rendelkezésére állniuk, ha nem a végrehajtó vagy a törvényhozó, hanem az igazságszolgáltatás döntéseiről van szó?

---

<sup>58</sup> Megjegyzendő azonban, hogy egyes kutatások abba az irányba mutatnak, hogy a nagyfokú átláthatóságnak negatív következményei is lehetnek. Lásd többek között Fox (2007), Licht (2011), Licht (2013) és Moore (2018). E dolgozat terjedelme arra kényszerít, hogy kizárólag az átláthatóság liberális demokratikus perspektívájára összpontosítsak, és mellőzöm a kérdéssel kapcsolatos egyéb elméleteket, perspektívákat és megközelítéseket, amelyek megváltoztathatják vagy nem változtathatják meg az ember véleményét, különösen az átláthatóság kívánatos mértékéről és a figyelemre méltó kivételekről.

<sup>59</sup> Moore (2018).

<sup>60</sup> Gupta (2008).

<sup>61</sup> Lásd többek között Dahl (1971), aki szerint a demokrácia bármely felfogása megköveteli az információ szabad áramlását a megalapozott döntések meghozatalához, szemben a demokrácia minimalista megközelítésével, ahogyan azt Schumpeter (1942) vagy - újabban - Przeworski és társai (2000) vallják, akik a demokráciát olyan rendszerként határozzák meg, amelyben a végrehajtó és a törvényhozó hatalmat egyaránt "megmérettetéses választások" útján töltik be.

Ha a demokrácia - mint fentebb említettük - megköveteli a polgároktól, hogy aktívan részt vegyenek a politikában, akkor a releváns információ bármilyen információ lehet, amely segíthet a politikai döntések meghozatalában, például abban, hogy kire vagy mire szavazzunk, mi ellen tiltakozunk, és általában véve mikor és hogyan vegyünk részt a nyilvános vitában. Ahogy Bellver és Kaufmann érvel, a nyújtott információnak figyelembe kell vennie a közintézmények teljesítményét, mivel az átláthatóság a közintézmények értékelését megkönnyítő eszköz. <sup>62</sup> Ennél is fontosabb, hogy nemcsak a közvetlenül vagy közvetve megválasztott személyek által hozott döntéseknek, hanem általában véve az állami intézkedéseknek, beleértve az igazságszolgáltatás által hozott döntéseket és ítéleteket,<sup>63</sup> átláthatónak kell lenniük ahhoz, hogy a polgárok kellőképpen tájékozottak legyenek.

A közintézmények körültekintő értékeléséhez a polgároknak ideális esetben nemcsak a döntés eredményét, az érintett személyeket és a folyamatban betöltött szerepüket kell ismerniük (vagy legalábbis megismerniük), hanem - ami fontos - az alternatív lehetőségeket is, amelyeket meg lehetett volna választani. Ha például egy döntés nem túl népszerű, de az alternatívák lényegesen rosszabbak, akkor ezt ideális esetben ugyanúgy kommunikálni kellene, mint ahogyan átláthatónak kellene lennie annak is, ha egy olyan döntés született, amely intuitívan jól hangzik, de a polgárok szemszögéből vitathatatlanul sokkal jobb alternatívák álltak rendelkezésre. Ennek megfelelően annak biztosítása érdekében, hogy a polgárok megfelelő információkat kapjanak a közintézmények értékeléséhez, meg kell tudniuk, hogy milyen döntést hoztak, milyen eljárások vezettek hozzá, és milyen más lehetőségek álltak rendelkezésre. <sup>64</sup> Leegyszerűsítve, a liberális demokrácia szempontjából *a lehetőségek átláthatósága ugyanolyan* fontos lehet, mint az eljárások és a szereplők átláthatósága.

Mivel liberális demokratikus szempontból az jelentene előrelépést, ha a választási lehetőségek gyakrabban lennének átláthatóak, bizonyos mesterséges intelligenciák - amint azt Kleinberg és munkatársai egyértelmű kompromisszumokat produkáló algoritmusa is mutatja - e tekintetben komoly előnyökkel rendelkeznek a HI-val szemben. Míg manapság a polgároknak és a politikai döntéshozóknak egyaránt a (gyakran megbízhatatlan) megérzéseikre kell hagyatkozniuk a helyzet értékelésében, addig az AI bevezetésével a kompromisszumok egyértelműsége - például a közbiztonság, a diszkrimináció és a fogva tartási arányok között, mint az óvadékkal kapcsolatos döntések esetében - egyértelműbbé válik. Ez például azt jelenti, hogy az állampolgárok nem csak azt látják majd, hogy milyen faji összetételű az óvadék ellenében elutasított személyek száma, hanem azt is, hogy milyen következményekkel járna ennek az aránynak a csökkentése vagy növelése. Összefoglalva, a polgárok jobban hozzáférnek majd a magasan

---

<sup>62</sup> Bellver és Kaufmann (2005, 5. o.); vö. még Florini (1999, 5).

<sup>63</sup> Ahogy Liptak (2008) megjegyzi, az Egyesült Államok az egyetlen olyan ország, amelyik *jelentős számú* bírának egy része közvetlenül.

<sup>64</sup> Vö. még Licht és Licht (2020), akik különbséget tesznek a) a végleges döntésekről vagy politikákról tájékoztató átláthatóság, b) a döntésekhez vezető folyamatra vonatkozó

---

átláthatóság és c) a döntés alapjául szolgáló indokok átláthatósága között.

a politikai folyamatban való részvételhez szükséges információk.<sup>65</sup> Valóban, jobb híján sokkal nehezebb lesz a Seehoferhez hasonló politikusoknak bonyolult törvényekkel elhomályosítani a politikai döntéshozatali és értékelési folyamatot. Ez még akkor is így van, ha a programok nem kifejezetten programozott döntéshozatalt (pl. egy egyszerű döntési fát vagy egy összetettebb algoritmust), hanem gépi tanulást használnak, beleértve a mély tanulást is, mivel a releváns kompromisszumok a tényleges és a szimulált hatások értékelésével a gépi tanulás összetettségétől függetlenül is átláthatóvá válhatnak.

A mesterséges intelligenciáról és az átláthatóságról szóló vita legtöbb kommentátora azonban más álláspontot képvisel és másra összpontosít. Ahelyett, hogy az átláthatóság új szintjeire koncentrálnának az alkalmazott kompromisszummal kapcsolatban, joggal mutatnak rá arra, hogy maga az algoritmus gyakran ismeretlen.<sup>66</sup> A COMPAS kockázatértékelési algoritmus, amelyet ma már több mint 20 joghatóságban használnak az óvadékról és az ítélethozatalról szóló döntések meghozatalakor az Egyesült Államokban, védett üzleti titok, és így továbbra is fekete doboz marad.<sup>67</sup> Ez már csak azért is problematikus, mert az algoritmus vagy a gépi tanulóval kapcsolatos képzési készlet ismerete nélkül nem lehet megérteni és megtámadni a döntést, amennyiben a tervezés hibás. Liberális demokratikus szempontból ez még problémásabb. Ha a polgárok nem jutnak hozzá az igazságszolgáltatás működésének és a kapcsolódó eljárási normáknak az értékeléséhez elengedhetetlen információkhoz, akkor nem tudnak érdemben részt venni a demokratikus folyamatban, azon túl, hogy rámutatnak arra, hogy az algoritmust nyilvánosságra kellene hozni. Lehet, hogy *önmagában* nem etikátlan az igazságszolgáltatás tipikus feladatainak kiszervezése magáncégeknek, de az már erősen megkérdőjelezhető, ha ez a folyamat az átláthatóság hiányához vezet, mint aminek a COMPAS-ügyben tanúi lehettünk.

Ezért, bár a mesterséges intelligencia komoly potenciális előnyökkel járhat a kompromisszumok tisztázása tekintetében, az átláthatóság (az eljárások) jelenlegi hiánya sok esetben, illetve annak lehetséges hiánya a jövőben egyre szkeptikusabbá teszi a tudósokat.<sup>68</sup> Ahhoz azonban, hogy ez elegendő ok legyen az AI bevezetésének elutasítására, be kellene mutatni, hogy egyrészt a HI valójában jobban alkalmas e kérdések kezelésére, másrészt pedig az ilyen jellegű átláthatóság hiánya inkább

---

<sup>65</sup>Sunstein (2019), 7. o. még azzal is érvel, hogy a visszaesési algoritmus tekintetében a kompromisszumok egyértelműsége "lehet a legfontosabb pont".

<sup>66</sup>Ezzel kapcsolatban többek között Floridi et al. (2018); O'Neil (2016); Wachter et al. (2017).

<sup>67</sup>Piovesan és Ntiri (2018).

<sup>68</sup>Az AI Now Institute éves jelentésében például az első számú ajánlás szerint "az alapvető közintézmények, például a büntető igazságszolgáltatásért, az egészségügyért, a jólétért és az oktatásért felelős szervek (azaz a "nagy tételek" jelentő területek) nem használhatnak többé "fekete dobozos" mesterséges intelligenciát és algoritmikus rendszereket". AI Now (2017), 1. o. A jelentés hangsúlyozza, hogy a "fekete doboz" rendszerek az átláthatóság

---

hiánya miatt különösen érzékenyek a torzított képzési adatokra és az emberi elfogultságok megismétlődésére (15. o.). *Vö. még Završnik (2020).*

a liberális demokrácia szempontjából károsabb, mint az átláthatóság hiánya a fent említett kompromisszumok tekintetében. Ezeket sorra veszem.

Azzal lehet érvelni, hogy az emberi bírói döntéshozatal sokkal átláthatóbb, mint a mesterséges bírói döntéshozatal, mivel a polgárok azonosítani tudják az emberi bírót. Elolvashatják érveiket, és maguk dönthetik el, hogy az ítélet ésszerű volt-e vagy sem. Röviden: az emberi bírák *megmagyarázzák*, hogy miért jutottak arra a döntésre, amelyre jutottak. Bár a mesterséges intelligencia magyarázhatóságával kapcsolatban rengeteg kutatás folyik, a jelenlegi alkalmazások minden bizonnyal elmaradnak az emberi magyarázó képességektől. <sup>69</sup> Ezzel az érveléssel azonban vannak figyelemre méltó problémák.

Először is, az emberi magyarázatok megadása nem egyenlő az eljárások teljes átláthatóságával. Az ítéletek megmagyarázása fontos, és alapot nyújthat a fellebbezéshez, de a magyarázat nem feltétlenül tárja fel a valódi mögöttes kognitív folyamatokat. Az emberi (bírósági) döntéshozatal folyamatai éppúgy ismeretlenek, mint a mesterséges intelligencia egyes alkalmazásainak eljárásai. A megismerés számítógépes elméletét <sup>követve</sup><sup>70</sup> akár azzal is érvelhetnénk, hogy a mesterséges döntéshozatal tekintetében legalább az algoritmust vagy az adathalmazt ismerhetjük, még ha nem is értjük a működését, míg az emberi döntéshozatalt működtető algoritmus ismeretlen marad. Lényeges megjegyezni, hogy bár a megismerés számítógépes perspektívája kétségtelenül erősíti ezt az érvet, de nem támaszkodik rá. Bármilyen elmefilozófiát is preferáljunk, az emberi döntéshozatal is egy fekete doboz marad.

Másodszor, az emberi magyarázatok nem csak a hibákra és a kognitív torzításokra hajlamosak, hanem a mögöttes motivációk és preferenciák racionalizálásának kockázatát is magukban hordozzák. <sup>71</sup> Intuitív módon azt várhatnánk, hogy ezt az általános tendenciát enyhítheti a szakértői tudás és képzés, azonban, ahogyan azt a tanulmányok újra és újra kimutatták, a szakértők ritkán teljesítenek jobban, mint a laikusok. <sup>72</sup> Ennek megfelelően aggodalomra ad okot, hogy még ha ismerjük és értjük is a megadott hivatalos indokokat, azok félrevezetőek lehetnek, és fennáll a veszélye annak, hogy a végső soron döntő okok továbbra is ismeretlenek maradnak. Ez nemcsak a demokratikus értékelések szempontjából fontos, hanem csökkentheti egy esetleges sikeres fellebbezés esélyeit is.

---

<sup>69</sup>Vö. Walzl és Vogl (2018), Barredo Arrieta et al. (2020); a magyarázható mesterséges intelligencia (XAI) technikai kihívásait *értelmezhetőségi problémaként* is emlegetik.

<sup>70</sup>Piccinini (2016), Chalmers (2011).

<sup>71</sup>Cushman (2020) további hivatkozásokkal.

<sup>72</sup> Vö. a katonai vezetőkkel, mérnökökkel, könyvelőkkel, orvosokkal, ingatlanértékelőkkel, opciós kereskedőkkel, pszichológusokkal és jogászokkal kapcsolatos kutatásokat tartalmazó átfogó listát, amelyet Guthrie et al. (2001) mutat be; lásd még Meadow és Sunstein (2001). Lásd általában Kahneman és Tversky (1983). Kifejezetten a bírói döntéshozatalra összpontosító kutatások Englich et al. (2006); Wistrich et al. (2015); Rachlinski és Wistrich (2017); Wistrich és Rachlinski (2018); Struchiner et al. (2020); Winter (2020).

Harmadszor, az egyes ítéletek összetettségétől függően megkérdőjelezhető, hogy a laikusok egyáltalán képesek-e teljes mértékben megérteni az ítéleteket. Bizonyára ellenérvként felhozható, hogy az ügyvédek képesek lesznek megérteni az ítéletet, és ennek megfelelően tanácsot adni az ügyfelüknek. Liberális demokratikus szempontból azonban nem sok haszna van annak, ha az ügyvédek eligazodnak az ügyekben, ha a polgárok túlnyomó többsége nem képes megérteni és - ennek alapján - helyesen értékelni a bírói normákat és magatartást. Vitathatóan a túlzott magabiztosság még azt is eredményezheti, hogy a laikusok azt hiszik, hogy megértették az ítéletet, és hamis érzést kelt bennük.

Ezek az elméleti és gyakorlati aggályok megkérdőjelezzik azt a feltételezést, hogy a HI valóban nagyobb liberális-demokratikus átláthatóságot biztosít, mint a mesterséges intelligencia. Tegyük azonban fel, hogy ez az elemzés vagy téves, vagy pedig vannak más nyomós érvek, amelyek e tekintetben a HI-transzparencia mellett szólnak.<sup>73</sup> Következésképpen még mindig meg kellene vizsgálni, hogy az egyes ítéletek emberi magyarázatának hiánya károsabb-e, mint az átláthatóság hiánya a fent említett kompromisszumok tekintetében. *Lehetőséget* kell-e adni az embereknek arra, hogy megértsék az egyedi esetekben hozott ítéleteket, vagy inkább tisztán kell látniuk a szabályok mögötti szükséges kompromisszumokat, amelyek nemcsak a társadalmat, hanem éppen azokat az egyéni ítéleteket is alakítják, amelyek megértése érdekli őket? Meg kell-e érteniük a szabályok célját, a mögöttes kompromisszumokat és az általuk megszavazottak (a törvényhozás) által adott magyarázatokat, vagy a gyakran nem demokratikusan elszámoltatható bírák által kiválasztott forgatókönyvekben történő alkalmazásukat? Bár nem szabadna lebecsülni az emberi bírói magyarázatokhoz való hozzáférés előnyeit, attól tartok, hogy a szóban forgó kompromisszumok átláthatósága egyszerűen még fontosabb lehet.<sup>74</sup>

### III. A HATALMI ÁGAK SZÉTVÁLASZTÁSÁTÓL A BÍRÓSÁGI FÜGGŐSÉGIG

Míg a diszkrimináció (I.) és az átláthatóság (II.) tekintetében lehetségesnek tűnik a javulás, és az ezzel kapcsolatos kérdések megfelelő tudományos figyelmet kaptak az elmúlt néhány évben, addig a fejlett mesterséges intelligencia igazságszolgáltatáson belüli alkalmazásából eredő közép- és hosszú távú veszélyek a liberális demokrácia számára szinte teljesen elhanyagoltak.<sup>75</sup> Ezért az elemzés ezen részének célja, hogy felhívja a figyelmet ezekre a kockázatokra. Pontosabban arra fogok rámutatni, hogy a fejlett mesterséges igazságügyi intelligencia

---

<sup>73</sup> Érvelhetünk például azzal, hogy a mesterséges intelligencia gondolkodása eléggé idegen az emberi megismeréstől ahhoz, hogy az emberek mindig jobban megértik (azt gondolják, hogy megértik) egy másik HI gondolkodási folyamatait, mint egy mesterséges intelligencia műveleteit.

<sup>74</sup> Feltételezve, hogy az emberi jogi magyarázatok fontosabbak, mint az átlátható kompromisszumok, még mindig azt kellene bizonyítani, hogy ez indokolja a magasabb bebörtönzési arányokat is - mind általánosságban, mind pedig különösen a kisebbségek tekintetében.

<sup>75</sup> Az egyik kivétel, amelyet alább tárgyalunk, Michaels (2020) munkája.



veszélyeztetheti a bírói függetlenséget és általában a hatalmi ágak szétválasztását. Mielőtt folytatom az elemzést annak felvázolásával, hogy a liberális demokrácia szempontjából miért kellene szorosán figyelemmel kísérni a mesterséges igazságügyi intelligencia fejlődését, röviden meg kell magyarázni ezt a sci-finek tűnő forgatókönyvet.

### 1. A fejlett mesterséges bírói intelligencia lehetősége

A *fejlett mesterséges igazságügyi intelligencia* (AAJI) olyan mesterségesen intelligens rendszerként definiálható, amely az igazságügyi döntéshozatal szempontjából releváns valamennyi területen megfelel vagy felülmúlja az emberi döntéshozatalt. Ami döntő fontosságú, hogy ehhez nem szükséges mesterséges általános intelligencia (AGI) kifejlesztése, miközben elkerülhető az ember és az AI közötti hibrid igazságügyi rendszerek szükségessége. Egy ilyen rendszer jelentősen túlmutat a technika jelenlegi szintjén; segítségével az embereknek lehetőségük lenne arra, hogy az igazságügyi ágazaton belüli döntéshozatalt teljes mértékben kiszervezzék, ha ezt kívánják. Bár egy ilyen fejlődést a jogi kutatók gyakran rendkívül valószínűtlennek tartanak a közép- vagy akár hosszú távú jövőben,<sup>76</sup> a gépi tanulás kutatói úgy vélik, hogy még egy AGI, azaz egy olyan mesterséges intelligens rendszer kifejlesztése, amely *minden* releváns területen megfelel vagy felülmúlja az emberi döntéshozatalt,<sup>77</sup> is sokkal közelebb van, mint azt a jogászok körében elterjedt józan ész sugallja.<sup>78</sup> A szakértők szerint például 50%-os esély van arra, hogy a mesterséges intelligencia 45 éven belül minden feladatban felülmúlja az embert.<sup>79</sup> Konkrét tevékenységek tekintetében pedig azt jósolják, hogy az AI viszonylag hamar felül fogja múlni az embert. Ide tartozik a nyelvek fordítása (2024-re), a középiskolai esszéírás (2026-ra), a teherautóvezetés (2027-re), a kiskereskedelmi munka (2031-re), a bestsellerek írása (2049-re) és a sebészi munka (2053-ra).<sup>80</sup>

Bizonyára lehet vitatkozni arról, hogy lehetséges-e egyáltalán, hogy egy mesterséges intelligencia jogi érvelést folytasson. Crotoof például rámutat arra, hogy "az ítékezés, amelyet a szokásjogi eljárásban értékelünk, kifejezetten emberi készség".<sup>81</sup> Ezzel kapcsolatban továbbá azzal érvel, hogy "[g]éltükben a változó társadalmi normákra való érzékenységük miatt az emberi bírák egyedülálló módon képesek felügyelni a jogi fejlődést, és biztosítani, hogy igazságszolgáltatási rendszerünk "lépést tartson[s] a korrall.". Ha azonban elsősorban a (változó) társadalmi (változó) társadalmi

---

<sup>76</sup> Lásd Martinez és Winter (2021).

<sup>77</sup> Az AGI fogalmáról lásd Wang és Goertzel (2006). Ugyanezen vagy nagyon hasonló technológiák leírására más terminológiák is használatosak: "erős mesterséges intelligencia", "emberi szintű mesterséges intelligencia", "valódi szintetikus intelligencia" és "általános intelligens rendszer".

<sup>78</sup> Az AGI idővonalakkal kapcsolatos szakértői felméréseket lásd Baum et al. (2011); Müller és Bostrom (2016); Grace et al. (2018); Gruetzemacher et al. (2019).

<sup>79</sup> Grace et al. (2018).

<sup>80</sup> Grace et al. (2018).



normák, a mesterséges intelligencia vitathatatlanul jobban fel van erre szerelve, mint a HI. Míg a mesterséges intelligencia különböző forrásokból származó hatalmas adathalmazokkal képezhető, és könnyebben és szélesebb körben igazítható a változó normákhoz, addig a HI a médiából származó, viszonylag kevés és erősen szelektált információ személyes értelmezésére és az emberek egy nagyon kiválasztott csoportjával való interakciókra támaszkodik. Ezért a mesterséges intelligencia nemcsak nagyobb mennyiségű releváns adat feldolgozására képes, hanem kisebb valószínűséggel esik bele a mindenütt jelenlévő megerősítési torzításba és más félrevezető szelekciós mechanizmusokba is, amelyek a meglévő társadalmi normákról valótlán kép kialakításához vezetnek.<sup>82</sup>

Mindezek után a jog dworkiniánus értelmezését követve kétségtelenül megkérdőjelezhető, hogy még egy fejlett mesterséges intelligencia is alkalmas lenne-e a "morális érvelés" feladatára a bírói döntéshozatali folyamat részeként.<sup>83</sup> Ugyanakkor az igazságszolgáltatási rendszer kiszervezhetőségének esélye vitathatóan növekedne, ha Oliver Wendell Holmes jogról szóló előrejelzési elméletét követnénk.<sup>84</sup> Holmes híres kijelentése szerint "[a] jóslatok arról, hogy mit fognak a bíróságok valójában tenni, és semmi más mesterkélt dolog, ez az, amit én a jog alatt értek."<sup>85</sup> E szakasz célja nem az, hogy meggyőzze az olvasót egy konkrét jogelmélet érdemeiről, hanem inkább az, hogy hangsúlyozza a jogtudományi bizonytalanság magas fokát.<sup>86</sup> Ha egyetértünk abban, hogy a mesterséges intelligencia *bizonyos* jogelméleti szempontok alapján képes lenne átvenni az igazságszolgáltatás irányítását - még akkor is, ha személy szerint nem osztjuk ezeket a nézeteket, vagy nem tartjuk őket valószínűnek -, akkor komoly okunk van arra, hogy elkezdjünk gondolkodni a kísérő következményekről és a potenciálisan releváns biztosítékokról. Más szóval, az AAJI pusztán lehetősége is arra kellene, hogy késztesse bennünket, hogy komolyan vegyük a következményeit.

---

<sup>82</sup> Ismét feltételezve, hogy a mesterséges intelligencia tervezési és adatválasztási folyamatában könnyebb elkerülni az elfogultságokat, mint egy emberközpontú bírói testület lebutításában. *Lásd még a fenti* 53. mozzanatot.

<sup>83</sup> Megjegyzendő azonban, hogy még ez sem egyértelmű. Érvelhetünk például azzal, hogy Dworkin (2011) leírása, miszerint az erkölcsi gondolkodás "az erkölcsi fogalmak értelmezése" (102. o.), nem feltétlenül zárja ki a fejlett mesterséges intelligenciát, amely feltehetően ugyanolyan vagy még alkalmasabb lenne az embernél a "háttérértékek és a konkrét értelmező meglátások integrálására" (135. o.).

<sup>84</sup> *Lásd* Holmes (1897), 457. és 461. o.

<sup>85</sup> Holmes (1897), 461. o.

<sup>86</sup> A "jogelméleti bizonytalanságról", amelyet "a jogelmélettel kapcsolatos normatív bizonytalanságként" (Winter et al., 2021, 97. o.) értelmeznek, csak most kezdődtek el a viták. *Lásd* Winter et al. (2021); Winter (2021). A kapcsolódó etikai vita az utóbbi időben nagyobb figyelmet kap. *Lásd* többek között Gustafsson és Torpman (2014); Lockhart (2000); MacAskill (2014); MacAskill et al. (2020), Tarsney (2018). *Lásd még* Barry és Tomlin (2019), akik a morális bizonytalanságot a büntetőjogelmélet különböző kérdéseire, többek között a büntetéskiszabási és kriminalizációs elméletekre

---

alkalmazzák.

## 2. Fejlett mesterséges intelligencia és igazságügyi függetlenség

Bár vitatható, hogy a liberális demokratikus rendszer elméletileg megvalósítható és fenntartható a hatalmi ágak szétválasztása nélkül, a hatalmi ágak szétválasztása világszerte a liberális demokratikus rendszerek sarokkövévé vált. A bírói, a törvényhozói és a végrehajtó hatalmi ágak szétválasztása létfontosságú célt szolgál a hatalom koncentrációjának minimalizálásában és a kormányzaton belüli fékek és ellensúlyok fenntartásában. Bár a fékek és ellensúlyok konfigurációja joghatóságokként eltérő, egy adott hatalmi struktúrára (pl. kormányzati ágra) gyakorolt bármilyen jelentős befolyás destabilizálja a rendszert. Ehhez kapcsolódóan ez a destabilizáció a különböző rendszerekben nagyon különböző formákat és kimeneteleket ölthet. Például a nyugat-európai parlamentáris demokráciák egészen másképp reagálhatnak az AAJI bevezetésére, mint az amerikai kontinens elnöki rendszerei. Emellett azok a jogrendszerek és kultúrák, amelyek úgy vélik, hogy a bírói hatalomnak van némi jogalkotáshoz hasonló jogalkotói hatásköre, mint például az Európai Unió<sup>87</sup> és az Egyesült Államok<sup>88</sup>, egészen másként értékelhetik az AAJI által jelentett veszélyeket és lehetőségeket, mint azok, amelyek elsősorban politikailag semlegesnek tekintik a bírói hatalmat (még akkor is, ha ez valójában nem így van, ahogyan azt a jogi realizmus mondja), mint például az Egyesült Királyság<sup>89</sup> vagy Németország.<sup>90</sup> Ezt a tényt jól szemlélteti a legfelsőbb bírósági bírák kiválasztási eljárása során zajló viták intenzitása. Míg az Egyesült Államokban a kinevezések hetekig lekötik a média figyelmét,<sup>91</sup> addig Németországban vagy az Egyesült Királyságban viszonylag kevés vita zajlik, ami nem meglepő, ha az igazságszolgáltatásról azt gondoljuk, hogy semlegesen alkalmazza a törvényeket. Végezetül a döntő kérdés nem az, *hogy a kormányzati ágak közötti kölcsönhatást és hatalmi dinamikát veszélyeztetheti-e, hanem az, hogy hogyan veszélyezteti az AAJI. Vajon a bírói hatalom nagyobb vagy kisebb lesz? Hogyan befolyásolná ez az interakciót a*

---

<sup>87</sup> Barnard (2013), 377. o.

<sup>88</sup> Vö. Michaels (2020), 1098. o. érvelése ezzel kapcsolatban: "A bíróságok fontos jogalkotói és szakpolitikai funkciót töltenek be, amikor a jogot értelmezik a jogi kérdések megoldása érdekében, és hasznos, ha ez az értelmezés konkrét ténybeli viták keretében történik".

<sup>89</sup> Lásd például Denning (1963) máig releváns megállapítását: "A brit felfogás szerint létfontosságú, hogy a bírák kívül maradjanak a politikai viták körén. A jogot értelmezniük kell, és a kor igényeinek megfelelően kell alakítaniuk, de a politikában nem hozhatnak jelentős változásokat. A bírák csak így maradhatnak kívül a politika szféráján" (300. o.).

<sup>90</sup>Kischel

(2013)  
összehasonlító szempontból elemzi a német alkotmánybírák választási eljárását, és megállapítja, hogy az "szükséges a megfelelő politikai□ semleges működéshez".

<sup>91</sup>A 2018-as Kavanaugh-meghallgatások áttekintéséhez lásd Grynbaum (2018), Bowden (2018).

és a törvényhozás és a végrehajtó hatalom hatáskörét? Ezek a kérdések minden liberális demokratikus szempontból fontosak, mert mindenekelőtt a tekintélyelvűség hajlamos a gyenge igazságszolgáltatásnak kedvezni. Másodszor, a független igazságszolgáltatás fontos szövetséges a kisebbségi jogokért folytatott küzdelemben és a "többség zsarnokságának" megakadályozásában.

Michaels amellet érvel, hogy a jelenlegi emberközpontú igazságszolgáltatás szükséges a hatalmi ágak szétválasztásának fenntartásához, mert csak egy emberközpontú igazságszolgáltatás emeli fel a bírák, ügyvédek, jogászprofesszorok stb. részvételéből fakadóan a jogra irányuló bőséges figyelmet.<sup>92</sup> Michaels számára valószínűnek tűnik, hogy a jogi közösség csökkenne, ha a mesterséges bírói intelligencia felváltaná az emberi bírákat,<sup>93</sup> és jogi közösség nélkül az emberek kevés figyelmet fordítanának a jogra.<sup>94</sup> Továbbá azzal érvel, hogy a jogi közösség kellő figyelmének ilyen mértékű hiánya pusztító következményekkel járhat. Először is, nehéz lenne elképzelni, hogy a nyilvánosság reagáljon a hatalommal való visszaélésekre vagy a hatalom három ágra kiterjedő koncentrációjára.<sup>95</sup> Másodszor, általában "kevésbé lenne ösztönző a magas színvonalú jogi érvek megfogalmazására, ha nem lenne lehetőség arra, hogy ezzel alakítsák az eredményt"<sup>96</sup>.

Azonban még ha feltételezzük is, hogy a jogi közösség valóban csökkenne vagy jelentősen lecsökkenne, nem egyértelmű, hogy ez *önmagában* szükségszerűen nagy hatalmi egyenlőtlenségekhez vezetne a csökkent emberi figyelem miatt. Éppen ellenkezőleg, a hangsúlyeltolódásból adódóan jelentős kísérő előnyökkel járhat. A manapság konkrét esetekre fordított figyelem a jogalkotás folyamatára és a szóban forgó törvények általános következményeinek elemzésére terelődhet át. Míg a társadalmat gyakran olyan rendkívüli egyedi esetek foglalkoztatják, amelyek hosszú távon kevésbé rendszerszintű értékkel bírnak, addig az AAJI segítségével az emberek az absztrakt elvek és szabályok értékelésére összpontosíthatnának, nem pedig a konkrét, gyakran erősen átpolitizált esetekre vonatkozó intuitív ítéleteinkre. Mivel az ilyen érzékelési intuíciók különösen hajlamosak a kognitív torzításokra,<sup>97</sup> kevés normatív iránymutatást nyújtanak, és korlátozott értékkel bírnak a társadalom egésze számára, ez akár még a közbeszédet is hasznosabb útra terelheti. Ahelyett, hogy a rendkívüli egyedi esetekre összpontosítanánk, a figyelem a sokkal fontosabb kompromisszumokra terelődhetne a fogva tartás, a bűnözési arányok és a

---

<sup>92</sup> Michaels (2020), 1096. o.

<sup>93</sup> Michaels (2020), 1096. o.; vö. még Volokh (2019); Re és Solow-Niederman (2019).

<sup>94</sup> Michaels (2020), 1096. o. Meg kell jegyezni, hogy Michaels nem határozza meg, hogy mennyire fejlettnak kell lennie a mesterséges intelligenciának ahhoz, hogy egy ilyen forgatókönyv bekövetkezzen. Az azonban egyértelmű, hogy egy kevésbé alkalmas mesterséges intelligenciára gondol, mint amit ebben a cikkben AAJI-ként definiált. Vélhetően az érvelése még inkább érvényes lenne az AAJI esetében.

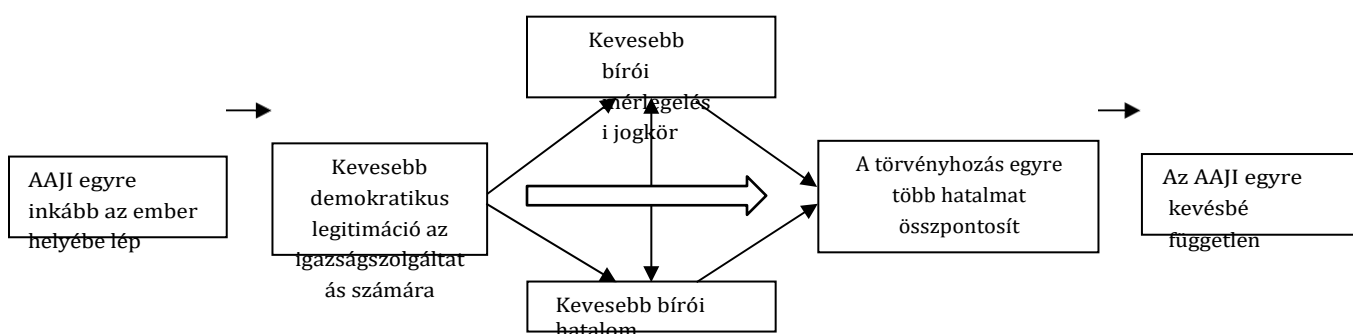
<sup>95</sup> Michaels (2020), 1096. o.

<sup>96</sup> Michaels (2020), 1097. o., lásd még: 1084. o. Bár ez a pont az alábbi érvelés szempontjából nem döntő, lehetséges, hogy az egyes felek nevében kontradiktórium AI-k érveljenek, és egy AAJI ítélkezzen.

<sup>97</sup> A "perceptuális intuíció" terminológia Sidgwickre (1907) vezethető vissza.

a cikk első részében tárgyalt diszkriminációt - ami potenciálisan alacsonyabb bebörtönzéshez, nagyobb biztonsághoz és kevesebb diszkriminációhoz vezethet. Továbbá, a hangsúly áthelyezése a bírói hatalomról a jogalkotóra (és a végrehajtó hatalomra) azzal az előnnyel járhat, hogy a jogalkotó nagyobb nyilvános elszámoltathatóságot biztosít, amely ismét profitálna a fent említett kompromisszumok átláthatóságából.

Bár lehet vitatkozni arról, hogy az AAJI elfogadása valóban az igazságszolgáltatásra irányuló figyelem csökkenéséhez vezetne-e, és hogy ez a fejlődés szükségszerűen együtt járna-e a hatalom elvesztésével, az igazságszolgáltatás mégis elveszítheti hatalmának jelentős részét és végső soron függetlenségét. Ez végső soron azért van így, mert ha az embert kivonjuk az egyenletből, az a demokratikus legitimitás kockázatával jár. Ez önmagában a bírói hatalom csökkenéséhez vezethet. Ráadásul a demokratikus legitimitáció csökkenésével nehéz elképzelni, hogy a bírói mérlegelés sok esetben döntő tényező maradjon. ~~ma már~~ Ellenkezőleg, feltételezem, hogy a jogalkotók egy demokratikusan gyenge igazságszolgáltatást használnának fel arra, hogy egyre több és konkrétabb döntést indokoljanak, kevés mérlegelési lehetőséget hagyva az AAJI-nek. A hatalomnak a bírói hatalomból a törvényhozásba való átkerülésével veszélybe kerülne a bírói függetlenség és általában a hatalmi ágak szétválasztása. Az alábbi ábra felvázolja a bírói függetlenséget fenyegető veszélyt, amely az AAJI bevezetéséből ered.



**1. ábra.** AAJI, legitimitás és bírói függetlenség.

Ha a demokratikus legitimitás ilyen módon veszélybe kerülne, az emberek talán még a nagyobb hatalmú törvényhozókat (és végrehajtókat) is szívesen látnák. Ezt a fejlődést felgyorsíthatná a törvényhozás jobb elszámoltathatósága, amely a fent említett kompromisszumok nagyobb átláthatóságából adódna. Végső soron a törvényhozást egyre inkább a bíróságok döntéseikért felelősnek tekinthetik. Ha például az AAJI nem a közvélemény kívánságának megfelelően dönt, az emberek a jogalkotót hibáztathatják a kérdéses törvények - szerintük - rossz megtervezése vagy az elfogult AAJI alkalmazásával kapcsolatos rossz döntés miatt, ahelyett, hogy az AAJI-t hibáztatnák az ilyen törvények rossz végrehajtásáért vagy értelmezéséért. Ebből a szempontból még a bírói függetlenséget határozottan támogató kormánypartok is engedhetnek a nyilvánosság nyomásának, hogy idővel korlátozzák az AAJI mérlegelési jogkörét.

Ez nem biztos eredmény, és nagyon bizonytalan, hogy ez a hatalomváltás végül hová vezetne. Az AAJI elfogadása azonban nagy kockázatot jelenthet a bírói függetlenségre nézve, amit nem szabadna megtenni, amíg a szükséges biztonsági mechanizmusok nem jönnek létre. Miközben hatalmas erőfeszítéseket fektetnek a fejlettebb mesterséges intelligencia, köztük az AGI és az AAJI kifejlesztésébe, kevés figyelmet fordítottak a szükséges politikai infrastruktúra megteremtésére, amely lehetővé tenné a társadalom számára, hogy kihasználja a mesterséges intelligencia potenciálisan hatalmas előnyeit anélkül, hogy kockáztatná a bírói függetlenség összeomlását. Bár nem értek egyet Michaelsszel abban a tekintetben, hogy miért várható a hatalomváltás (figyelem vs. legitimitás), ez az elemzés megerősíti az AAJI-nek a hatalmi ágak szétválasztására gyakorolt hatását. Pontosabban azt állítom, hogy nagy veszélyt jelent a bírói függetlenségre, amit nem szabad félvállról venni.

#### IV. KÖVETKEZTETÉS

A jogi közösség részeként hajlamosak lehetünk arra, hogy a lehető legnagyobb mértékben fejlesszük az igazságszolgáltatási rendszert, és megvédjük azt a potenciális fenyegetésektől, még akkor is, ha ezek a fenyegetések más területeken történő rendkívüli fejlesztésekkel járnak együtt. Akadémikusokként szeretnénk teljesen megérteni a mesterséges intelligencia érvelését, mielőtt javasolnánk annak alkalmazását, még akkor is, ha az emberi érvelést sem értjük teljesen. Emberként pedig hajlamosak vagyunk a status quo-t előnyben részesíteni, még akkor is, ha a változás nettó pozitív lenne.<sup>98</sup> Mindezek a motivációk megmagyarázhatják, hogy miért van olyan sokunkban erős ellenérzés a mesterséges intelligencia igazságszolgáltatásban való bevezetésével szemben, de sajnos a motivációk nem ideális igazságkövetési folyamatok.<sup>99</sup> Legalábbis a diszkrimináció és az átláthatóság kérdéseit illetően ez az elemzés világossá teszi, hogy a jelenlegi alkalmazások nagy előnyökkel járhatnak. Míg ez a tétel a diszkriminációval kapcsolatos normatív perspektívák széles skálájából védhető, az átláthatósággal kapcsolatos javulások attól függnek, hogy valaki a lehetőségek átláthatóságát előnyben részesíti-e az eljárások és a szereplők átláthatóságával szemben, vagy sem. Mindezek után szem előtt kell tartani, hogy az ilyen fejlesztések lehetőségének megléte nem jelenti azt, hogy nem áll fenn a nagymértékű megkülönböztetés vagy az átláthatóság hiánya. Mint látható, az a tény, hogy az ilyen mesterséges intelligencia *lehetséges*, nem jelenti azt, hogy a *gyakorlatban* alkalmazott mesterséges intelligencia átláthatóbb vagy kevésbé diszkriminatív lesz.

Míg az eddigi elemzés a meglévő technológiai képességekre összpontosított, az utolsó fejezet bevezette az AAJI fogalmát, amelyet mesterségesen intelligens

---

<sup>98</sup> Vö. Samuelson és Zeckhauser (1988) elemzését a status quo torzításról. Eidelman és Crandall (2012) továbbá megjegyzi, hogy a status quo torzítás "akadályokat állít a kognitív és társadalmi változások elé" (270. o.).

<sup>99</sup> Éppen ellenkezőleg, tekintettel e motivációk intuitív vonzerejére, megfontolhatjuk az érvek leleplezését.



---

olyan rendszer, amely az emberi döntéshozatalt minden, a bírósági döntéshozatal szempontjából releváns területen felülmúlja, vagy meg is haladja azt. A bírói hatalomnak a törvényhozás felé történő esetleges elmozdulására összpontosítva azzal érveltek, hogy az AAJI elfogadása veszélyezteti a bírói függetlenséget - és ezzel együtt a liberális demokrácia egyik központi alapját.

Mondanom sem kell, hogy ez a kísérlet a mesterséges bírói döntéshozatal kihívásainak felvázolására a liberális demokrácia számára messze nem teljes, és sok tekintetben további kutatásokra van szükség. Először is meg kell határozni, hogy milyen egyéb kihívásokat hozhat magával az AAJI elfogadása, például azt a kockázatot, hogy az algoritmusok újfajta állandóságot biztosítanak a törvényeknek, és ezáltal további akadályt állítanak a jogfejlődés elé.<sup>100</sup> Másodsor, döntő fontosságú lesz annak vizsgálata, hogy miként lehet fenntartani a liberális demokratikus értékeket általánosabban, és különösen a hatalmi ágak szétválasztását egy AI-bírósággal. A mesterséges intelligencia rövid távon olyan hatalmas előnyöket kínálhat a HI-vel szemben az igazságszolgáltatáshoz való hozzáférés, az átláthatóság és a méltányosság tekintetében, hogy az ember figyelmen kívül hagyhatja az AAJI által jelentett hosszú távú veszélyeket. A jövőbeni kutatások egyik prioritásává kell tenni a politikai és jogi infrastruktúra kidolgozását, hogy a mesterséges igazságügyi intelligencia gyümölcseit biztonságos és stabil módon lehessen learatni.

---

<sup>100</sup> Crootof (2020) ezt a jelenséget "technológiai-jogi zártságnak" nevezi.

## BIBLIOGRÁFIA

- AI Now (2017) AI Now 2017 Report. New York University, New York. [ainowinstitute.org/AI\\_Now\\_2017\\_Report.pdf](http://ainowinstitute.org/AI_Now_2017_Report.pdf)
- Aletras et al. (2016) Predicting Judiciting Decisions of the European Court of Human Rights: a Natural Language Processing Perspective. PeerJ Computer Science. <https://doi.org/10.7717/peerj-cs.93>.
- Angwin J et al. (2016) Machine Bias. ProPublica. [www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing](http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing). Hozzáférés 2020. június 25.
- Armstrong S, Sotala K, hÉigeartaigh SSÓ (2014) A híres mesterséges intelligencia előrejelzések hibái, meglátásai és tanulságai - és mit jelentenek a jövőre nézve. Journal of Experimental & Theoretical Artificial Intelligence, 26, 317-342. <https://doi.org/https://doi.org/10.1080/0952813X.2014.895105>. <https://doi.org/https://doi.org/10.1080/0952813X.2014.895105>
- Barnard C (2019) Az EU anyagi joga, 6. kiadás. Oxford University Press, Oxford, New York
- Barredo Arrieta A et al. (2020) Explainable Artificial Intelligence (XAI): Fogalmak, taxonómiák, lehetőségek és kihívások a felelős mesterséges intelligencia felé. Information Fusion 58:82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Barry C, Tomlin P (2019) Morális bizonytalanság és a büntetőjog. In: Alexander L, Ferzan KK (szerk.) Palgrave Handbook of Applied Ethics and the Criminal Law. Palgrave Macmillan/ Springer Nature, Cham, Svájc.
- Baum S (2018) A közel- és hosszú távú mesterséges intelligenciára összpontosító frakciók közötti megbékélés. AI & Society 33(4):565-572.
- Baum S (2020) Mesterséges intelligencia és társadalom. Information 2020, 11(6):290-305. <https://doi.org/10.3390/info11060290>. <https://doi.org/10.3390/info11060290>
- Baum SD, Goertzel B, Goertzel TG (2011) How long until human-level AI? Egy szakértői értékelés eredményei. Technological Forecasting and Social Change 78:185-195. <https://doi.org/10.1016/j.techfore.2010.09.006>. <https://doi.org/10.1016/j.techfore.2010.09.006>
- Bloom P (2016) Against Empathy: The Case for Rational Compassion. Ecco Press, New York
- Boscardin C (2015) Reducing Implicit Bias Through Curricular Interventions. Journal of Általános belgyógyászat 30(12):1726-1728
- Bostrom N (2006) How Long Before Superintelligence. Nyelvészeti és filozófiai vizsgálatok 5:11-30.
- Bostrom N (2014) Superintelligence: Szuperintelligencia: Utak, veszélyek, stratégiák. Oxford University Press, Oxford
- Bostrom N, Ord T (2006). A fordított teszt: a status quo elfogultság megszüntetése az alkalmazott etikában. Ethics 116(4):656-679. <https://doi.org/10.1086/505233>.
- Bowden J (2018) Idővonal: Brett Kavanaugh jelölése a Legfelsőbb Bíróságra. The Hill. <https://thehill.com/homenews/senate/410217-timeline-brett-kavanaughs-nomination-to-the-supreme-court>. Hozzáférés 2020. június 26.
- Carneiro D et al. (2015) Online vitarendezés: An Artificial Intelligence Perspective. Artificial Intelligence Review 41:227-228
- Carnes M et al. (2015) A nemi előítéletesség szokásának megtörésére irányuló beavatkozás hatása egy intézmény oktatói körében: klaszteres, randomizált, kontrollált kísérlet. Academic Medicine 90(2):221- 230

- Cave S, Ó hÉigeartaigh S (2019) A mesterséges intelligenciával kapcsolatos közeli és hosszú távú aggodalmak áthidalása. *Nature Machine Intelligence* 1:5-6
- Chalmers D (2011) A Computational Foundation for the Study of Cognition. *The Journal of Cognitive Science* 12(4):325-359. <https://doi.org/10.17791/jcs.2011.12.4.325>.  
<https://doi.org/10.17791/jcs.2011.12.4.325>
- Chander A (2017) The Racist Algorithm? *Michigan Law Review* 115(6):1023-1045
- Chen D (2019) A gépi tanulás és a jogállamiság. In: Livermore M, Rockmore D (szerk.) *Law as Data*. Santa Fe Institute Press, Santa Fe, pp 433-441.
- Chohlas-Wood A (2020) A kockázatértékelési eszközök megértése a büntető igazságszolgáltatásban. *Brookings*. [www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice](http://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice). Hozzáférés: 2020. június 26.
- Crootof R (2019) "Cyborg Justice" and the Risk of Technological-Legal Lock-In. *Columbia Law Review Forum* 119:233-251
- Cushman F (2020) A racionalizálás racionális. *Behavioral and Brain Sciences* 43. Cambridge University Press (43): E28.  
<https://doi.org/10.1017/S0140525X19001730>
- Dabney D et al. (2017) Az amerikai óvadék és a büntető igazságszolgáltatás színezése. *The Harvard Journal of Crime and Justice* 56(4):397-418.
- Dahl R (1971) *Poliarchia; részvétel és ellenzékiesség*. Yale University Press, New Haven
- Das Gupta O, Fried N (2019) Seehofer redet über Gesetzestrick - hinterher spricht er von Ironie. *Süddeutsche Zeitung*. [www.sueddeutsche.de/politik/seehofer-datenaustauschgesetz-1.4479069](http://www.sueddeutsche.de/politik/seehofer-datenaustauschgesetz-1.4479069). Hozzáférés 2020. június 26.
- Denning L (1963) Az igazságszolgáltatás funkciója a modern demokráciában. *Pakisztáni Horizont* 16(4):299-305.
- Devine P et al. (2012) Long-term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention. *Journal of Experimental Social Psychology* 48(6):1267-1278.
- Dhami MK, Ayton P (2001), Az óvadék letétele és a börtönbüntetés gyors és takarékos módon. *Journal of Behavioral Decision Making* 14:141-168.  
<https://doi.org/10.1002/bdm.371>. <https://doi.org/10.1002/bdm.371>
- Dworkin R (2011) *Justice for Hedgehogs*. Harvard University Press, Cambridge
- Eidelman S, Crandall C (2012) Bias in Favor of the Status Quo. *Social and Personality Psychology Compass* 6(3):270-281. <https://doi.org/10.1111/j.1751-9004.2012.00427.x>.  
<https://doi.org/10.1111/j.1751-9004.2012.00427.x>
- Englich B et al. (2006) Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin* 32(2):188-200.  
<https://doi.org/10.1177/0146167205282152>.  
<https://doi.org/10.1177/0146167205282152>
- Floridi L, Cowls J, Beltrametti M et al. (2018) AI4People-An Ethical Framework for a Good AI Society: Lehetőségek, kockázatok, elvek és ajánlások. *Minds & Machines* 28:689-707. <https://doi.org/10.1007/s11023-018-9482-5>. <https://doi.org/10.1007/s11023-018-9482-5>
- Florini A (1999) Kell-e a láthatatlan kéznek átlátszó kesztyű? Az átláthatóság politikája. Előadás a Világbank éves fejlesztésgazdasági konferenciáján, Washington, D.C., 28-30.
- Fox J (2007) Government Transparency and Policymaking. *Public Choice* 131(1/2):23-44.
- Gigerenzer G (1991) Hogyan tüntessük el a kognitív illúziókat: Túl a "Heurisztikákon és előítéleteken". *European Review of Social Psychology* 2:83-115.
- Goertzel B, Pennachin C (szerk.) (2007) *Artificial General Intelligence*. Springer-Verlag, Berlin, Heidelberg
- Grace K, Salvatier J, Dafoe A, et al (2018) When Will AI Exceed Human Performance? Bizonyítékok a mesterséges intelligencia szakértőitől. *Journal of Artificial Intelligence Research*



- Gruetzemacher R, Paradise D, Lee KB (2019) Forecasting Transformative AI: An Expert Survey. arXiv:190108579 [cs]
- Grynbaum MM (2018) Kavanaugh meghallgatások a televízióban Riveting Drama to a Captive Nation. *TheNewYorkTimes*. <https://www.nytimes.com/2018/09/27/business/media/kavanaugh-blasey-ford-hearing-tv.html>. Hozzáférés 2020. június 26.
- Gupta A (2008) Átláthatóság a vizsgálat alatt: Information Disclosure in Global Environmental Governance. *Global Environmental Politics* 8(2):1-7.
- Gustafsson JE, Torpman O (2014) In Defence of My Favourite Theory. *Pacific Philosophical Quarterly* 95:159-174. <https://doi.org/10.1111/papq.12022>  
<https://doi.org/10.1111/papq.12022>
- Guthrie C et al. (2001) Inside the Judicial Mind. *Cornell Law Review* 86(4):777-830.
- Hacker P (2018) Teaching Fairness to Artificial Intelligence; Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law. *Common Market Law Review* 55:1143-1185.
- Hammer P (2018) Detroit 1967 és ma: Spatial Racism and Ongoing Cycles of Oppression. *Journal of Law in Society* 18(2):227-235
- Hollyer et al. (2011) Democracy and Transparency. *The Journal of Politics* 73(4):1191-1205. <https://doi.org/10.1017/s0022381611000880>. <https://doi.org/10.1017/s0022381611000880>
- Holmes, O (1897) A jog útja. *Harvard Law Review*, 10:457-ADD Oldaltartomány vége
- Huq AZ (2020) Az emberi döntéshez való jog. *Virginia Law Review* 106:611-688
- Ingriselli E (2015) Mitigating Jurors' Racial Biases: A tartalom és az időzítés hatásai Esküdtszéki utasítás. *Yale Law Journal* 124(5):1690-1745.
- Kahneman D, Tversky A (1982) 30 - Intuitív előrejelzés: Az előrejelzés: torzítások és korrekciós eljárások. In: Kahneman D, Slovic P, Tversky A (szerk.) *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, pp 414-421.
- Kahneman D, Tversky A (1983) Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90(4):293-315. <https://doi.org/10.1037/0033-295X.90.4.293>. <https://doi.org/10.1037/0033-295X.90.4.293>
- Kaminski M (2019) Binary Governance: A GDPR algoritmikus megközelítésének tanulságai. *Elszámoltathatóság. Southern California Law Review* 92:1529-1616.
- Karnow C (2008) Az óvadék megállapítása a közbiztonság érdekében. *Berkeley Journal of Criminal Law* 13(1):1- 30
- Kaufmann D, Bellver A (2005) Transparenting Transparency: Kezdeti empirikus adatok és politikai alkalmazások. MPRA Paper 8188, Münchener Egyetemi Könyvtár, Németország.
- Kischel U (2013) Párt, pápa és politika? A német alkotmánybírák megválasztása összehasonlító perspektívában. *International Journal of Constitutional Law* 11:962-980. <https://doi.org/10.1093/icon/mot040>. <https://doi.org/10.1093/icon/mot040>
- Kleinberg J et al. (2018) Human Decisions and Machine Predictions. *The Quarterly Journal of Economics* 133(1):273-293. <https://doi.org/10.1093/qje/qjx032>. <https://doi.org/10.1093/qje/qjx032>
- Kleinberg J et al. (2019) Discrimination in the Age of Algorithms. *Journal of Legal Analysis* 10:113-174.
- Lai C et al. (2014) Reducing Implicit Racial Preferences: I. 17 beavatkozás összehasonlító vizsgálata. *Journal of Experimental Psychology General* 143(4):1765-1785.

- Larson et al. (2016) How We Analyzed the COMPAS Recidivism Algorithm. ProPublica. [www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm](http://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm). Hozzáférés 2020. június 26.
- Levin S (2016) A szépségversenyt az AI bírálta, és a robotok nem szerették a sötét bőrt. The Guardian. [www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people](http://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people). Hozzáférés 2020. június 26.
- Licht JD (2011) Do We Really Want to Know? Az átláthatóság potenciálisan negatív hatása a döntéshozatalban az észlelt legitimitásra. *Scandinavian Political Studies* 34:183-201. <https://doi.org/10.1111/j.1467-9477.2011.00268.x> <https://doi.org/10.1111/j.1467-9477.2011.00268.x>
- Licht JD (2014) A szakpolitikai terület mint az átláthatósági hatások lehetséges moderátora: Egy kísérlet. *Közigazgatási Szemle* 74(3):361-371
- Licht KD, Licht JD (2020) Mesterséges intelligencia, átláthatóság és nyilvános döntéshozatal: Miért kulcsfontosságúak a magyarázatok, amikor megpróbáljuk előállítani a vélt legitimitást. *AI & Society* 35(4):917-926. <https://doi.org/10.1007/s00146-020-00960-w> <https://doi.org/10.1007/s00146-020-00960-w>
- Liptak A (2008) Az amerikai bíróválasztás zavarba ejt más nemzeteket. The New York Times. [www.nytimes.com/2008/05/25/world/americas/25iht-judge.4.13194819.html](http://www.nytimes.com/2008/05/25/world/americas/25iht-judge.4.13194819.html). Hozzáférés 2020. június 26.
- Lockhart T (2000) Morális bizonytalanság és annak következményei. Oxford University Press, Oxford
- Lowry S, Macpherson G (1988) A blot on the profession. *British Medical Journal* 296(6623):657-658. <https://doi.org/10.1136/bmj.296.6623.657>
- Macaskill W (2014) Normatív bizonytalanság. Dissertáció, Oxfordi Egyetem
- MacAskill W, Bykvist K, Ord T (2020) Moral Uncertainty. Oxford University Press, Oxford
- Martinez E, Winter CK (2021) Mesterséges intelligencia az igazságszolgáltatásban: A Survey of Expert Vélemény. [Kézirat előkészítés alatt]
- McGinley A (2011) Ricci kontra DeStefano: Diluting Disparate Impact and Redefining Disparate Treatment. *Nevada Law Journal* 12(3):626-639.
- Meadow W, Sunstein C (2001) Statistics, Not Experts. *Duke Law Journal* 51:629-646
- Michaels AC (2019) Mesterséges intelligencia, jogi változások és a hatalmi ágak szétválasztása. *University of Cincinnati Law Review* 88:1083-1103
- Moore S (2018) Az intézményi átláthatóság szociológiája felé: Nyitottság, megtévesztés és a közbizalom problémája. *Sociology* 52(2):416-430. <https://doi.org/10.1177/0038038516686530>
- Müller VC, Bostrom N (2016) Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In: Müller VC (szerk.) *Fundamental Issues of Artificial Intelligence*. Springer International Publishing, Cham, pp 555-572.
- Mukand S, Rodrik D, (2020) A liberális demokrácia politikai gazdaságtana. *The Economic Journal* 130(627):765-792. <https://doi.org/10.1093/ej/ueaa004> <https://doi.org/10.1093/ej/ueaa004>
- Niiler E (2019) Lehet-e az AI igazságos bíró a bíróságon? Észtország szerint igen. *Wired*. [www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so](http://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so). Hozzáférés 2020. június 25.
- O'Neil C (2016) *Weapons of math destruction: how big data increases inequality and veszélyezteti a demokráciát*. Broadway Books, New York

---

Oswald F et al. (2013) Predicting ethnic and racial discrimination: A meta-analízis az IAT kritériumvizsgálatokról. *Journal of Personality and Social Psychology* 105(2):171-192. <https://doi.org/10.1037/a0032734>. <https://doi.org/10.1037/a0032734>

- Pejovic C (2001) Civil Law and Common Law: Két különböző út vezet ugyanahhoz a célhoz. *Victoria University of Wellington Law Review* 32(3):817-842
- Piccinini G (2016) The Computational Theory of Cognition. In: Müller V. (szerk.) *Fundamental Issues of Artificial Intelligence. Szintézis Könyvtár, 376. kötet.* Springer, Cham, pp 203-221
- Pinker S (2011) *The Better Angels of Our Nature: Why Violence Has Declined.* Viking Press, New York
- Pinker S (2018) *Enlightenment Now: Az ész, a tudomány, a humanizmus és a haladás mellett.* Viking Press, New York
- Piovesan C, Ntiri V (2018) Adjudication by algorithm: A mesterséges intelligencia kockázatai és előnyei a bírósági döntéshozatalban. *The Advocates' Journal* 44:42-45.
- Prunkl C, Whittlestone J (2020) Túl a közeli és a hosszútávon: arXiv:2001.04335v2 [cs.CY].
- Przeworski A (2000) *Demokrácia és fejlődés: Politikai intézmények és jólét a világban, 1950-1990.* Cambridge University Press, Cambridge
- Rachlinski J, Wistrich A (2017) Judging the Judiciary by the Numbers: Empirical Research on Judges. *Annual Review of Law and Social Science* 13:203-229. <https://doi.org/10.1146/annurev-lawsocsci-110615-085032>. <https://doi.org/10.1146/annurev-lawsocsci-110615-085032>
- Reiling D (2018) Milyen szerepe lehet a mesterséges intelligenciának a bírói döntéshozatali folyamatban? Előadás az Európai Bizottság az igazságszolgáltatás hatékonyságával foglalkozó konferenciáján: "Artificial Intelligence at the Service of the Judiciary"
- "CouncilofEurope.www.coe.int/en/web/cepej/justice-of-the-future-predictive-justice-and-artificial-intelligence. Hozzáférés 2020. június 25.
- Richard M, Solow-Niederman (2019) Developing Artificially Intelligent Justice. *Stanford Technology Law Review*, 22(2):242-289.
- Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. *Journal of Risk and Uncertainty* 1:7-59.
- Schumpeter (1942) *Kapitalizmus, szocializmus és demokrácia.* Harper & Brothers, New York
- Shaked-Schroer N (2008) Reducing racial bias in the penalty phase of capital trials. *Behavioral Sciences and the Law* 26(5):603-617
- Sidgwick H (1907), *The Methods of Ethics*, 8. kiadás.
- Smith M (2016) In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures. *The New York Times*. [www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html](http://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html). Hozzáférés 2020. június 26.
- Sourdin T (2018) Judge v Robot? Mesterséges intelligencia és a bírói döntéshozatal. *University of New South Wales Law Journal* 41(4):1114-1133
- Sourdin T, Cornes R (2018) Do Judges Need to Be Human? A technológia hatásai a rezponzív bíraskodásra. In: Sourdin T, Zariski A (szerk.) *The Responsive Judge: International Perspectives (Nemzetközi perspektívák).* Springer, Singapore, pp 87-119.
- Struchiner N, Almeida G, Hannikainen I (megjelenés 2020) Legal Decision-Making and the Abstract/Concrete Paradox. *Cognition*
- Sunstein (2019) Algorithms, Correcting Biases. *Társadalomkutatás: An International Quarterly* 86(2):499-511.
- Tarsney C (2018) Moral Uncertainty for Deontologists. *Ethic Theory Moral Prac* 21:505-520. <https://doi.org/10.1007/s10677-018-9924-4>. <https://doi.org/10.1007/s10677-018-9924-4>



- Teal C et al. (2012) Az orvostanhallgatók segítése az egyes betegcsoportokkal szembeni tudattalan előítéletek felismerésében és kezelésében. *Medical Education* 46(1):80-88
- Tversky A, Kahneman D (1973) Availability: A heurisztika a gyakoriság és a valószínűség megítéléséhez. *Cognitive Psychology* 5:207-232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9).
- Volokh E (2019) Chief Justice Robots. *Duke Law Journal*, 68:1135-1192.
- Wachter S, Mittelstadt B, Floridi L (2017) Miért nem létezik az automatizált döntéshozatal magyarázatához való jog az általános adatvédelmi rendeletben. *International Data Privacy Law* 7(2):76-99.
- Wallace M et al. (2017) Separate and unequal: Structural racism and infant mortality in the US. *Health & Place* 45(3):140-144.
- Waltl B, Vogl R (2018) Increasing Transparency in Algorithmic- Decision-Making with Explainable AI. *Datenschutz und Datensicherheit* 42:613-617
- Wang P, Goertzel B (2007) Bevezetés: Bert B. In: *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms, Proceedings of the AGI Workshop 2006*.
- Winter CK (2021) *Metamoralisches Strafrecht*. [Kézirat előkészítés alatt]
- Winter CK (2020) A viselkedési közgazdaságtan értéke az uniós igazságügyi döntéshozatalban. *Német jogi folyóirat* 21(2):240-264
- Winter CK, Schuett J, Martinez E, Van Arsdale S, Araújo R, Hollman N, Sebo J, Stawasz S, O'Keefe C, Rotola G (2021) *Jogi prioritások kutatása: A research agenda*. Jogi prioritások projekt. [https://www.legalpriorities.org/research\\_agenda.pdf](https://www.legalpriorities.org/research_agenda.pdf). Hozzáférés 2021. január 10.
- Wistrich A et al. (2015) Heart Versus Head: Do Judges Follow the Law or Follow Their Érzelmek? *Texas Law Review* 93:855-923
- Wistrich A, Rachlinski J (2018) Implicit Bias in Judicial Decision Making, How It Affects Judgment and What Judges Can Do About It. In: Redfield S (szerk.) *Enhancing Justice, Reducing Bias*. ABA Book Publishing, Chicago, pp 87-130.
- World Government Summit (2018) Vajon egy mesterséges intelligencia valaha is helyettesítheti a bírót a bíróságon? [www.worldgovernmentsummit.org/observer/articles/could-an-ai-ever-replace-a-judge-in-court](http://www.worldgovernmentsummit.org/observer/articles/could-an-ai-ever-replace-a-judge-in-court). Hozzáférés 2020. június 25.
- Završnik A (2020) Büntető igazságszolgáltatás, mesterséges intelligencia rendszerek és emberi jogok. *ERA Forum* 20:567-583. <https://doi.org/10.1007/s12027-020-00602-0>
- Zavyalova V (2018) Pénzt takarít meg a jogi tanácsadáson: A mesterséges intelligencia helyettesíti az ügyvédeket Oroszországban. *Russia Beyond*. [www.rbth.com/science-and-tech/327585-free-legal-advice-robotlawyer](http://www.rbth.com/science-and-tech/327585-free-legal-advice-robotlawyer). Hozzáférés 2020. június 25.
- Zhang M (2015) Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software. *Forbes*. [www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/#61fdae0c713d](http://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/#61fdae0c713d). Hozzáférés 2020. június 26.