

Algoritmikus átláthatóság az intelligens város számára

Robert Brauneis & Ellen P. Goodman¹

20 YALE J. L. & TECH. 103 (2018)

*"Társadalomként most döntő fordulóponthoz érkezünk annak meghatározásában, hogy miként alkalmazzuk a mesterséges intelligencián alapuló technológiákat úgy, hogy az elősegítse, ne pedig akadályozza az olyan demokratikus értékeket, mint a szabadság, az egyenlőség és az átláthatóság."*²

Ahogy a mesterséges intelligencia és a nagy adatelemzés egyre inkább felváltja az emberi döntéshozatalt, az algoritmikus etikával kapcsolatos kérdések egyre sürgetőbbé válnak. Sokan aggódnak amiatt, hogy az algoritmikus társadalom túl átláthatatlan ahhoz, hogy felelősségre vonható legyen a viselkedéséért. Egy személytől megtagadhatják a feltételes szabadlábra helyezést vagy a hitelt, kirúghatják vagy nem vehetik fel olyan okokból, amelyeket soha nem fog tudni, és amelyeket nem lehet megfogalmazni. A közsférában az algoritmikus döntéshozatal átláthatatlansága különösen problematikus, egyrészt mert a kormányzati döntések különösen súlyosak lehetnek, másrészt mert a demokratikusan megválasztott kormányoknak különleges elszámoltatási kötelezettségeik vannak.

Célunk, hogy teszteljük az átláthatóság határait a nagy adatelemzés kormányzati alkalmazása körül, hozzájárulva az algoritmikus elszámoltathatósággal kapcsolatos szakirodalomhoz a kormányzati előrejelző algoritmusok átláthatatlanságának alapos tanulmányozásával. Nyílt nyilvántartási folyamatok segítségével vizsgálatunk a prediktív algoritmusok helyi és állami kormányzati alkalmazására összpontosított. Az algoritmusok által meghatározott döntések itt, a helyi önkormányzatokban lehetnek a legközvetlenebb hatással. És ez az a terület, ahol a legvalószínűbb, hogy az érintett ügynökségek az adatelemzést magánszolgáltatóknak adják át, amelyek az ügyfélként működő ügynökségek, a nyilvánosság vagy mindkettő számára láthatatlanul hozhatnak tervezési és szakpolitikai döntéseket. Hogy teszteljük, mennyire áthatolhatatlanok az így létrejövő "fekete doboz" algoritmusok, negyvenkét nyíltan hozzáférhető

¹ Robert Brauneis a George Washington Egyetem jogi karának jogászprofesszora; Ellen P. Goodman a Rutgers jogi karának jogászprofesszora. Szeretnénk köszönetet mondani Erin Daltonnak, Jeremy Heffnernek, Andrew Nicklinnek és a Cambridge-i Egyetem The Power Switch című konferenciájának

résztevőinek: How Power is Changing in a Networked World (Hogyan változik a hatalom a hálózatos világban), a MetroLab Network "Ethical Guidelines for Applying Predictive Tools within Child Welfare Services" (Etikai iránymutatások a prediktív eszközök alkalmazására a gyermekjóléti szolgáltatásokban) című workshopján, a Bloomberg Philanthropies "What Works Cities Summit" (A működő városok csúcstalálkozója), a Wharton School "Law and Ethics of Big Data Colloquium" (A nagy adatok joga és etikája) és a European Intellectual Property Institutes Network tizennyolcadik éves kongresszusán.

- ² Peter Stone et al., *Overview*, ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE (AI100) (2016. augusztus 1.), <http://ai100.stanford.edu/2016-report/overview> [<http://perma.cc/3GQY-HRWW>].

huszonhárom államban kérte a nyilvántartási kérelmeket, amelyekben hat előrejelző algoritmusprogrammal kapcsolatos alapvető információkat kért. Kiválasztottuk a legszélesebb körben használt és jól vizsgált programokat, beleértve a profitorientált vállalatok, nonprofit szervezetek és a tudományos/magánszektorbeli partnerségek által kifejlesztetteket. A konkrét cél az volt, hogy felmérjük, hogy a nyílt iratkezelési eljárások lehetővé teszik-e a polgárok számára, hogy felfedezzék, milyen politikai döntéseket testesítenek meg ezek az algoritmusok, és hogy értékeljék hasznosságukat és igazságosságukat.

E munka elvégzéséhez meghatároztuk, hogy mit jelent az "algoritmikus átláthatóság". Azt találtuk, hogy szinte minden esetben nem biztosították. Problémát jelentett az üzleti titoktartás túlságosan széles körű állítása. A közhiedelemmel ellentétben azonban nem a megfelelően értelmezett üzleti titok jelenti a legnagyobb akadályt, mivel a prediktív modellek végrehajtásához használt, üzleti titokvédelem alatt álló kód kiadása általában nem szükséges az érdemi átláthatósághoz. Arra a következtetésre jutottunk, hogy a nyilvánosan alkalmazott algoritmusok csak akkor lesznek kellően átláthatóak, ha (1) a kormányok megfelelő nyilvántartást készítenek az algoritmikus folyamatokra, valamint az azt követő végrehajtásra és validálásra vonatkozó célkitűzéseikről; (2) a kormányzati vállalkozók elegendő információt tárnak fel a közhivatalok számára arról, hogy hogyan fejlesztették ki az algoritmust; és (3) a közhivatalok és a bíróságok az üzleti titoktartási igényeket a nyilvánosságra hozatal alól a törvény által előírt korlátozott kivételként kezelik. Bemutatjuk a véleményünk szerint nyolc fő információtypust, amelyeket a nyilvánosan alkalmazott algoritmusokra vonatkozó nyilvántartásoknak tartalmazniuk kell.

TARTALOMJEGYZÉK

BEVEZETÉS	107
I. AZ ALGORITMIKUS KORMÁNYZÁS ÍGÉRETE ÉS VESZÉLYEI	111
A. <i>A klinikai előrejelzéstől az intelligens városi algoritmikus kormányzásig</i>	111
1. Klinikai kontra biztosításmatematikai előrejelzés és megítélés	111
2. Előrejelző algoritmusok és gépi tanulás	113
3. Algoritmikus kormányzás és intelligens városok	114
B. <i>Ígéreték és veszélyek</i>	115
II. AZ ÉRDEMI ÁTLÁTHATÓSÁG MEGHATÁROZÁSA: MIT KELL TUDNIA A NYILVÁNOSSÁGNAK	118
A. <i>Mi az algoritmus politikája?</i>	118
B. <i>Teljesít</i>	121
C. <i>Igazságos</i>	122
D. <i>Az algoritmus növeli vagy csökkenti a kormányzati kapacitást?</i>	126
E. <i>Jelentősegteljes átláthatóság</i>	128
III. A NYÍLT IRATOKRÓL SZÓLÓ TÖRVÉNY PROJEKTJE: AZ ALGORITMUSOK DOKUMENTÁCIÓJÁNAK MEGSZERZÉSE	132
A. <i>Projekttervezés és végrehajtás</i>	133
1. Nyílt iratokról szóló törvények	133
2. Algoritmusok, ügynökségek és a kérelmek megfogalmazása	135
B. <i>Eredmények</i>	137
1. Közbiztonsági értékelés-előzetes szabadlábra helyezés	137
2. Eckerd gyors biztonsági visszajelzés-gyermekjóléti értékelések	141
3. Allegheny Family Screening Tool - Gyermekjóléti értékelések	144
4. PredPol-Predictive Policing	146
5. HunchLab-Predictive Policing	147
6. New York City és New York állam hozzáadottérték-modellje - tanárértékelés	150
C. <i>Következtetés</i>	151
IV. AZ ÁTLÁTHATÓSÁG LEGFŐBB AKADÁLYAI	152
A. <i>Dokumentáció hiánya</i>	152
B. <i>Agresszív üzleti titok és titoktartási igények</i>	153
C. <i>Egyéb kormányzati érdekek és a nyílt iratokról szóló törvény mentességei</i>	160
V. FIXES	163
A. <i>A nyilvántartások rendelkezésre bocsátását megkövetelő és a nyilvánosságra hozatalt engedélyező szerződéses nyelvezet</i>	164

B. *Nyilvántartások létrehozása az elszámoltathatóság érdekében*166

1. Általános előrejelzési cél és alkalmazás168
2. Adatok: Releváns, rendelkezésre álló, gyűjthető168
3. Adatkizárás169
4. Speciális prediktív kritériumok171
5. Az alkalmazottelemzési és fejlesztési technikák173
6. Legfontosabb politikai döntések173
7. Validációs vizsgálatok, ellenőrzések, naplózás és átláthatatlan elszámoltathatóság174.....
8. Algoritmus és kimeneti magyarázatok174

VI. KÖVETKEZTETÉS175

BEVEZETÉS

A kormányok egyre gyakrabban használnak számítógépes algoritmusokat a közügyek intézésére. Ez különösen igaz a városokra, megyékre és államokra, amelyek kormányainak feladata az alapvető szolgáltatások biztosítása és a kényszerítő rendőri erő alkalmazása. Az "intelligens városok" mozgalma világszerte az adatok hatékonyabb gyűjtésének és felhasználásának fontosságát hangsúlyozza a helyi önkormányzatok számára.³ Az egyik cél az, hogy a nagy adathalmazokban mintákat találjanak - például a bűncselekmények előfordulásának valószínűsíthető helyeit és időpontjait -, és hogy előrejelző modelleket hozzanak létre a közszolgáltatások elosztásának irányítására - például arra, hogy hogyan és hol kell rendőri munkát végezni.⁴ A legtöbb helyi önkormányzatnak nincs meg a szakértelme és az eszközei ahhoz, hogy saját maga alkalmazza az adatelemzést. Ha "okosak" akarnak lenni, akkor vállalatokkal, egyetemekkel és nonprofit szervezetekkel kell szerződést kötniük, hogy magánfejlesztésű algoritmikus folyamatokat hajtsanak végre. Ennek eredményeképpen a magáncégek által kifejlesztett előrejelző algoritmusok alakítják a helyi önkormányzatok intézkedéseit olyan területeken, mint a büntető igazságszolgáltatás, az élelmiszerbiztonság, a szociális szolgáltatások és a közlekedés.⁵

Mivel a tervező szervezetek jellemzően nem hozzák nyilvánosságra előrejelző modelljeiket vagy algoritmusait, egyre több szakirodalom kritizálja e folyamatok "fekete doboz" átláthatatlanságát.⁶ Ezek

³ *Lásd pl.*, Rob Kitchin, *The Real-Time City? Big Data and Smart Urbanism*, 79 GEOJOURNAL 1 (2014).

⁴ *Lásd* ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT (2017); Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1115 (2017); Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 HARV. L. & POL'Y REV. 15, 38 (2016).

⁵ Az intelligens városok algoritmusairól általánosságban lásd a 27-36. megjegyzéseket kísérő *alábbi* szöveget; az algoritmusokról, amelyekkel kapcsolatban nyílt iratokat kértünk, lásd a 123-183. megjegyzéseket kísérő szöveget.

⁶ *Lásd pl.* ROB KITCHIN, THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUKTÚRÁK ÉS KÖVETKEZMÉNYEIK (2014); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); Michael Ananny, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, 41 SCI. TECH. & HUM. VALUES 93 (2015); David Beer, *The Social Power of Algorithms*, 20 J. INFO. COMM. & SOC. 1 (2016); Taina Bucher, *'Want To Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook*, 14 NEW MEDIA & SOC'Y 1164 (2014); Jenna Burrell,

How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms, 3 BIG DATA & SOC'Y 1 (2016); Danielle Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); Kate Crawford, *Can an Algorithm be Agonistic? Tíz jelenet a kiszámított nyilvánosság életéből*, 41 SCI. TECH. & HUM. VALUES 77 (2016); Nicholas Diakopoulos, *Algorithmic Accountability: A számítógépes hatalmi struktúrák újságírói vizsgálata*,

3 DIGITAL JOURNALISM 398 (2015); Tarleton Gillespie, *The Relevance of Algorithms, in MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY* 167 (Tarleton Gillespie et al. szerk., 2014); Rob

a feketedobozok megkérdőjelezhetetlenek, és sokan aggódnak amiatt, hogy diszkriminatívak,⁷ hibásak vagy más módon problémásak lehetnek. ⁸ Azok az újságírók és tudósok, akik elkezdtek részleteket keresni az állami szervektől ezekről az algoritmusokról, általában nem jutnak eredményre, mivel az információszabadsággal kapcsolatos kéréseiket elutasítják vagy megválaszolatlanul maradnak. ⁹

A kommentátorok nagyobb átláthatóságot követelnek a mesterséges intelligencia minden megvalósítása során. ¹⁰ Vannak speciális

Kitchin, *Thinking Critically About and Researching Algorithms*, 20 INFO. COMM. & SOC'Y 14 (2016) [a továbbiakban: Kitchin, *Thinking Critically*]; Christian Sandvig, *Seeing the Sort: The Aesthetic and Industrial Defense of "The Algorithm"*, 12 J. NEW MEDIA CAUCUS, 1 (2015); Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms* (2014),

<http://www.personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20-%20Sandvig%20-%20ICA%202014%20Data%20and%20Discrimination%20-%20Preconference.pdf> [<http://perma.cc/M9FK-3R2V>]; Zynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 J. ON TELECOM. & HIGH TECH. L. 203 (2015); Malte Ziewitz, *Governing Algorithms: Myth, Mess, and Methods*, 41 SCI. TECH. & HUM. VALUES 3 (2015). *Lásd általában* Tarleton Gillespie & Nick Seaver, *Critical Algorithm Studies: A Reading List*, SOC. MEDIA COLLECTIVE RES. BLOG <http://socialmediacollective.org/reading-lists/critical-algorithm-studies/> [<http://perma.cc/TBF5-DEY4>] (források összeállítása).

⁷See, e.g., FED. TRADE COMM'N, BIG DATA: A BEFOGADÁS VAGY A KIREKESZTÉS ESZKÖZE? UNDERSTANDING THE ISSUES (Jan. 2016), <http://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report> [<http://perma.cc/B922-N6U4>]; Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671 (2016); Tufekci, *supra* note 6; Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [<http://perma.cc/684J-45ZP>] (evaluating algorithmic risk assessments used by judges to set up bail amounts in Fort Lauderdale, Florida).

⁸Vö. *általában* Lee Rainie & Janna Anderson, *Code-Dependent*: PEWREP (2017. február 9), <http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/> [<http://perma.cc/MCQ8-APDG>] (példák összefoglalása az algoritmusok széles körű használatának kockázataira).

⁹Vö. pl. Nicholas Diakopoulos, *We Need To Know the Algorithms the Government Uses To Make Important Decisions About Us*, CONVERSATION (2016. május 23.), <http://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869> [<http://perma.cc/N522-46V4>] (beszámoló az ötven államhoz intézett, az algoritmusok büntető igazságszolgáltatásban való alkalmazásáról szóló nyílt iratmegismerési kérelmekről, amelyek közül kilenc "arra alapozta a büntető igazságszolgáltatási algoritmusok részleteinek közlését megtagadó kérelmét, hogy az információk valójában egy vállalat tulajdonában vannak"); Tonia Hill, *Jamie Kalven Joins Other Chicago Journalists in Lawsuit Against CPD*, HYDE PARK HERALD (2017. június 7.), <http://hpherald.com/2017/06/07/jamie-kalven-joins-chicago-journalists-lawsuit-cpd/> [<http://perma.cc/RPM7-LV3K>] (leírja, hogy újságírók perelik a Chicagói Rendőrkapitányságot, amiért az visszatartja az információkat egy olyan algoritmusról, amely egy "hólistának" nevezett stratégiai alanylistát állítólagosan fegyveres erőszakban valószínűleg részt

vevő embereket előrejelző listát készít).

- ^{10V6.} *pl.*, *23 Principles for Beneficial Artificial Intelligence*, FUTURE LIFE INST., (2017. január 17.), <http://futureoflife.org/ai-principles/> [<http://perma.cc/THU7->].

aggodalmak, amikor az önkormányzatok és más kormányok olyan előrejelző algoritmusokat használnak, amelyek fejlesztését és végrehajtását sem a nyilvánosság, sem maga a kormány nem érti igazán. Azáltal, hogy ezeket a rendszereket kifejlesztik és eladják a kormányzatnak - vagy akár elajándékozzák -, a magánvállalkozások jelentős szerepet vállalnak a közigazgatásban. Ami az intelligens városokban okos, az a magáneladók áthatolhatatlan agyában lakozik, miközben a kormányzat, amely egyedül a nyilvánosságnak tartozik elszámolással, kiüresedik, buta és sötét. A kockázat az, hogy az algoritmus átláthatatlansága lehetővé teszi a közhatalom vállalati átvételét. Amikor egy kormányzati ügynök olyan algoritmikus ajánlást hajt végre, amelyet nem ért és nem tud megmagyarázni, a kormány elveszítette a demokratikus elszámoltathatóságot, a nyilvánosság nem tudja értékelni a kormányzati folyamat hatékonyságát és igazságosságát, és a kormányzati ügynök elvesztette a kompetenciáját, hogy a nyilvánosság munkáját bármilyen kritikus módon elvégezze.

Megpróbáltuk tesztelni, hogy mennyire átláthatatlanok az önkormányzati előrejelző algoritmusok azáltal, hogy azonosítottuk a nagy adatok előrejelzésének néhány leggyakoribb felhasználási módját ebben az ágazatban. Azonosítottuk az alapítványok, magánvállalatok és kormányzati szervek által kifejlesztett algoritmusokat, valamint a büntető igazságszolgáltatásban és a polgári alkalmazásokban használt algoritmusokat. Ezután összeállítottunk egy "portfóliót" a különböző felhasználásokat és joghatóságokat megcélzó, nyílt iratokra vonatkozó kérelmekből. A MuckRock, a nyílt iratok iránti kérelmek benyújtására szolgáló nonprofit együttműködési platform¹¹ segítségével 23 államban 42 kérelmet nyújtottunk be hat előrejelző algoritmussal kapcsolatos iratokra vonatkozóan.¹² A szövetségi kormányzat és mind az ötven állam (és Washington D.C.) rendelkezik olyan nyílt iratokra vonatkozó törvényekkel, amelyek különböző mértékű közzétételt írnak elő az algoritmusok nyilvános felhasználásával kapcsolatban. Tekintettel arra, hogy a legtöbb nyílt iratokról szóló törvényt milyen tágan fogalmazták meg, a szerződések és a szállítókkal folytatott kapcsolódó levelezés szinte mindig "nyilvános iratok", amelyeket közzé kell tenni.¹³ A szoftver "nyilvántartás"

CPGM] (megjegyezve, hogy több mint 1600 aláíró, köztük Steven Hawking, Elon Musk és a mesterséges intelligenciával foglalkozó kutatók "hibák átláthatóságát" követelték, amely megmutatja, hogy egy mesterséges intelligencia rendszer miért okozhatott kárt, valamint "igazságügyi átláthatóságot", amely egy illetékes emberi hatóság által ellenőrizhető, kielégítő magyarázatot ad minden igazságügyi határozatra).

¹¹ *AboutUs*, MUCKROCK, <http://www.muckrock.com/about/> [<http://perma.cc/L8SV-TYDL>]. Egy esetben (Allegheny County Child and Family Services) külön nyújtottuk be a kérelmeket, nem használtuk a platformot.

¹² A MuckRock projektoldalunk a következő címen található: *Uncovering Algorithms*, MUCKROCK, <http://www.muckrock.com/project/uncovering-algorithms-84/> [<http://perma.cc/HNW2-ZSUA>]. Ez az oldal hivatkozik a

kéréseinkre és a kéréseinkre válaszul rendelkezésre bocsátott dokumentumok többségére. (Néhány kormány a saját szerverén lévő fájlokra mutató linkeket adott meg, ahelyett, hogy feltöltötte volna a dokumentumokat a MuckRockra). Néhány kérésünket eredetileg rossz ügynökségekhez irányították; ezeket nem számoljuk bele a szövegben közölt számokba, de a MuckRock projektoldalán szerepelnek.

¹³ *Lásd pl.*, FLA. STAT. § 119.011(12) (2017) (amely előírja, hogy a betekintésre nyitva álló "közokirat" "minden dokumentumot, iratot, levelet, térképet, könyvet, szalagot,

a szövetségi információszabadságról szóló törvény (FOIA) értelmében¹⁴, valamint számos állami törvény értelmében, de nem mindegyiket. 15 Olyan iratokat kerestünk, mint a levelezés, szerződések, szoftverek, képzési anyagok, meglévő és tervezett validációs tanulmányok és egyéb dokumentációk. Bár a helyi és állami önkormányzatokra összpontosítottunk, gyanítjuk, hogy megállapításaink általánosíthatók más kormányzati szervekre is, amelyek magánszolgáltatókkal dolgoznak együtt.

E kérések és a kapcsolódó vizsgálat eredményei alapján arra a következtetésre jutottunk, hogy három fő akadálya van annak, hogy a nagyadat-előrejelzés kormányzati felhasználása átláthatóvá váljon: (1) az algoritmikus folyamatok körüli megfelelő nyilvántartási gyakorlatok hiánya; (2) a megfelelő közzétételi gyakorlatokhoz való elégtelen kormányzati ragaszkodás; és

(3) az üzleti titoktartás vagy más bizalmas kiváltságok érvényesítése a kormányzati vállalkozók által. Ebben a cikkben megvizsgáljuk az említett akadályok mindegyikét, és javaslatot teszünk az ezek csökkentésére irányuló irányelvekre és gyakorlatokra. Ha ezeket a problémákat kezelnénk, gyanítjuk, hogy egyes esetekben még egy akadálya lenne a valódi átláthatóságnak: a rendkívül dinamikus vagy olyan modellezést alkalmazó algoritmusok használata, amelyek még akkor is nehezen értelmezhetővé teszik őket, ha a feljegyzések nyilvánosságra kerülnek. Ezt a kérdést egy másik napra tartogatjuk.

A cikk I. részében olyan alapfogalmakat mutatunk be, mint a klinikai ítélet, a biztosításmatematikai ítélet és a prediktív algoritmusok; nyomon követjük az intelligens városi algoritmikus kormányzás fejlődését; és áttekintjük az algoritmikus kormányzás ígéreteit és veszélyeit. A II. részben a prediktív algoritmusokkal kapcsolatban kidolgozzuk az "értelmes átláthatóság" fogalmát. Ez magában foglalja annak feltárását, hogy a nyilvánosságnak mit kell tudnia az e programokba ágyazott politikáról, valamint hasznosságukról, igazságosságukról és a kormányzati kapacitásra gyakorolt hatásukról. A III. rész ismerteti a különböző joghatóságokhoz benyújtott, a prediktív algoritmusok alkalmazásával kapcsolatos nyílt iratanyag-kérelmeket és a kapott válaszokat. A IV. rész azonosítja az algoritmikus folyamatok nagyobb átláthatóságának akadályait. Az V. rész enyhítési technikákat javasol az algoritmikus átláthatóság maximalizálására, és nyolc fő információtypust mutat be.

fényképek, filmek, hangfelvételek, adatfeldolgozó szoftverek vagy egyéb anyagok, függetlenül azok fizikai formájától, jellemzőitől vagy továbbítási módjától, amelyeket törvény vagy rendelet alapján vagy bármely ügynökség hivatalos ügyeinek intézésével kapcsolatban készítettek vagy kaptak").

¹⁴ Lásd általában Katherine Fink, *Opening the Government's Black Boxes*: INFO. COMM. & SOC'Y (2017) (a forráskódra vonatkozó FOIA-kérelmekre adott szövetségi ügynökségi válaszokkal kapcsolatos kutatásokat tárgyalja).

¹⁵ *Hasonlítsa össze az ARK. CODE ANN. § 25-19-103(5)(B)* (amely előírja, hogy a "közokirat" fogalmába nem tartozik bele a "vásárlással, lízinggel vagy

licenccel szerzett szoftver"), és FLA. STAT. § 119.011(1) (amely előírja, hogy a "[d]atafeldolgozó szoftver" szerepel a "közokirat" fogalommeghatározásában). *Lásd általában az alábbi 113-117. lábjegyzeteket (példák az állami nyílt iratokra vonatkozó törvények hatálya alá).*

amelyeket a nyilvánosan alkalmazott algoritmusokra vonatkozó nyilvántartásoknak tartalmazniuk kell. A VI. rész zárásaként.

I. AZ ALGORITMIKUS KORMÁNYZÁS ÍGÉRETE ÉS VESZÉLYEI

A. *A klinikai előrejelzéstől az intelligens városokig Algoritmikus kormányzás*

Az algoritmikus kormányzást a számítási teljesítmény és a hálózatépítés hatalmas növekedése teszi lehetővé, amely nagy mennyiségű adat összegyűjtését, tárolását és elemzését teszi lehetővé. A városok arra törekszenek, hogy ezeket az adatokat a közszolgáltatások és az infrastruktúra - például az egészségügyi szolgáltatások, a közbiztonság, a büntető igazságszolgáltatás, az oktatás, a közlekedés és az energia - működésének racionalizálására és automatizálására használják fel.¹⁶ A helyi önkormányzatok korlátai miatt a magánvállalkozók központi szerepet játszanak ebben a folyamatban, ami a politika kiszervezésére jellemző elszámoltathatósági problémákat vet fel. A közszükségletektől a magántechnológiai támogatás felé való elmozdulás a "biztosításmatematikai" előrejelzésnek a "klinikai" előrejelzéssel szembeni előnyben részesítésében gyökerezik.

1. **Klinikai kontra biztosításmatematikai előrejelzés és Ítéletalkotás**

A közforrásokat elosztó és a kényszerítő rendőri hatalmat alkalmazó kormánytisztviselők döntéseik meghozatalához gyakran használják az úgynevezett "klinikai előrejelzést".¹⁷ A képzés, a tanonckodás és a tapasztalat révén ezek a tisztviselők kialakítják azt az érzéküket, hogy az emberek hogyan viselkednek, és milyen következményekkel járhat egy közigazgatási döntés. Egy tapasztalt esetfelelősnek van érzéke ahhoz, hogy egy gyermeket valószínűleg rosszul bánnak-e egy háztartásban. Egy bíró megjósolja, hogy egy elítélt elkövet-e újabb bűncselekményt, ha feltételesen szabadlábra helyezik. Egy állami főiskolai felvételi tisztviselő pedig nagyjából meg tudja jósolni, hogy egy felvételt nyert diákot milyen ösztöndíjajánlatra fog ösztönözni.

¹⁶ Lásd Hafedh Chourabi et al., *Understanding Smart Cities*: PROC. 2012 45TH HAW. INT'L CONF. ON SYS. SCI. (HICSS 2012), <http://dx.doi.org/10.1109/HICSS.2012.615>[<http://perma.cc/K5GS-66N5>] (az intelligens város különböző koncepcióinak leírása és összegzése); Lilian Edwards, *Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective*, 2 EUR. DATA PROTECTION L. REV. 28, (2016); lásd még Nils Walravens & Pieter Ballon, *Platform Business Models for Smart Cities: From Control and Value to Governance and Public Value*, 51 IEEE COMM. MAG. 72 (2013) (a mobiltechnológiák szerepének tárgyalása a városi problémák kezelésében); Ellen P. Goodman, "Smart Cities" Meet "Anchor Institutions": *The Case for Broadband and the Public Library*, 41 FORDHAM URB. L.J. 1665 (2015).

- ¹⁷ *Lásd pl.* , PAUL E. MEEHL, CLINICAL VERSUS STATISTICAL PREDICTION: A THEORETICAL ANALYSIS AND A REVIEW OF THE EVIDENCE (1954); Theodore R. Sarbin, *A Contribution to the Study of Actuarial and Individual Methods of Prediction*, 48 AM. J. SOC. 593 (1943).

a beiratkozáshoz. A klinikai előrejelzés vagy klinikai ítélet nem formalizálható vagy artikulálható teljes mértékben, mivel az egyén felhalmozott bölcsességének egy adott helyzetre való alkalmazásán alapul, amelyet a releváns körülmények nyílt végű vizsgálata alapján alakítanak ki.

A klinikai előrejelzés ellentétben áll a biztosításmatematikai előrejelzéssel, amelyet mechanikus, statisztikai vagy algoritmikus előrejelzésnek vagy ítéletnek is neveznek.¹⁸ Az aktuáriusi megközelítés a vizsgálati alanyokra vonatkozó adatokat elemzi, hogy felfedezze a jellemzők vagy tulajdonságok és a kimenetel közötti összefüggéseket. Ez határozottan nem nyílt végű. A feltételesen szabadlábra helyezettékre vonatkozó adatok elemzése például hipotetikusán feltárhatná, hogy a húsz és harminc év közöttieket kétszer olyan gyakran tartóztatták le feltételes szabadlábra helyezésük alatt, mint az ötven és hatvan év közöttieket. A biztosításmatematikai előrejelzés adatmeghatározásai, elemzési módszerei és összefüggései formalizálhatók és megfogalmazhatók, ellentétben a klinikai előrejelzéssel. Az így kapott előrejelzések vagy ítéletek azonban soha nem alapulnak egy adott helyzet összes körülményén, mivel a biztosításmatematikai elemzés mindig véges számú adatmezővel dolgozik.¹⁹ Például, míg egy büntetőjogi vádlott életkorát, nemét és bűnügyi előéletét figyelembe lehet venni, addig családi kapcsolatainak és közösségi kapcsolatainak minőségét nem.

A biztosításmatematikai megítélés már régóta létezik. Az első olyan életbiztosító társaságot, amely biztosításmatematikai táblázatokon alapuló biztosítási kötvényeket értékesített, 1762-ben alapították Londonban.²⁰ Kilencven évvel ezelőtt, 1928-ban Ernest Burgess 21 tényező statisztikai elemzésén²¹ alapuló képletet alkotott a feltételesen szabadlábra helyezettékre visszaesésének előrejelzésére - egy későbbi felülvizsgálat szerint ez a képlet pontosabb volt, mint a börtönpszichiáterek klinikai előrejelzései.²² Több mint hatvan évvel ezelőtt az aktuáriusi előrejelzés használata eléggé elterjedt volt ahhoz, hogy Paul Meehl híres könyvet jelentetett meg, amelyben szembeállította a klinikai és a

¹⁸ Lásd William M. Grove et al., *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 2000 PSYCHOL. ASSESSMENT 19; Jack Sawyer, *Measurement and Prediction, Clinical and Statistical*, 66 PSYCHOL. BULL. 178 (1966).

¹⁹ A klinikai ítélet és a partikularizmus, valamint az aktuáriusi ítélet és az általánosítás közötti összefüggésről lásd FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 19-22 (2003).

²⁰ Lásd MAURICE EDWARD OGBORN, *EQUITABLE ASSURANCES: Az életbiztosítás története az EQUITABLE LIFE ASSURANCE SOCIETY, 1762-1962*, 39. o. (1962).

²¹ Lásd Ernest W. Burgess, *Factors Determining Success or Failure on Parole*, in ANDREW A. BRUCE ET AL., *THE WORKINGS OF THE INDETERMINATE SENTENCE LAW AND THE PAROLE SYSTEM IN ILLINOIS* 205 (1928); lásd még BERNARD E. HARCOURT, *AGAINST PREDICTION: PROFILING, POLICING AND PUNISHING IN AN ACTUARIAL AGE* 47-76 (2007) (Burgess munkájáról és annak jelentőségéről); Karl F. Schuessler, *Parole Prediction: J. CRIM. L. & CRIMINOLOGY* 425 (1955) (a feltételes szabadlábra helyezésről szóló döntésekkel kapcsolatos statisztikai előrejelzés történetének tárgyalása, beleértve Burgess munkáját).

²² Lásd Daniel Glaser, *A Reconsideration of some Parole Prediction Factors*, 19 AM.

statisztikai előrejelzés.²³ És természetesen a kormányok már régóta alkalmaznak - legalábbis részlegesen - formálisabb statisztikai módszereket az informális klinikai megítélés mellett, amikor olyan kérdéseket mérlegelnek, mint például, hogy bizonyos biztonsági fejlesztések milyen mértékben csökkentenék a közlekedési halálesetek számát, vagy hogy egy közösségben a jövőben milyen mértékben változna a villamos energia iránti kereslet. A biztosításmatematikai ítélőképesség kormányzati (és üzleti) alkalmazása azonban a közelmúltban óriási lendületet vett, köszönhetően az elemzéshez szükséges nagy adathalmazok felhalmozódásának, valamint a számítási teljesítmény és a gépi tanulás elméletének fejlődésének, amely lehetővé tette ezen adathalmazok sokkal összetettebb elemzését.

2. Előrejelző algoritmusok és gépi tanulás

A nagy adatok korában az aktuáriusi ítéletalkotás prediktív algoritmusok létrehozásával és használatával valósul meg. A prediktív algoritmusok nagy adathalmazok elemzésével készülnek, hogy feltárják a különböző jellemzők (egy személy, körülmény vagy tevékenység) és a kívánt vagy kifogásolt eredmények közötti összefüggéseket.²⁴ Ezt az elemzést jellemzően gépi tanulási folyamatok segítségével végzik, olyan folyamatok segítségével, amelyek során a számítógépek több ezer vagy millió összetett összefüggést tesztelhetnek, hogy kiderítsék, melyik illik a legjobban az adatokhoz. A gépi tanulási folyamatok nagy teljesítményűek - nagyon sok lehetőséget fésülnek át - és viszonylag objektívek - a számítógépnek fogalma sincs arról, hogy egy adott változó olyan jellemzőt képvisel-e, amelyet egy személy irrelevánsnak, például a cipőméretet, vagy érzékenynek, például a faji hovatartozást, hanem egyszerűen csak az adott változó és az eredményt képviselő változó közötti korreláció erősségét vizsgálja. A legjobban illeszkedő korrelációk halmaza lesz a modell, amely megbecsüli a jövőbeli viselkedés vagy események valószínűségét (a kimenet), ha releváns tényeket (a bemenet) kapunk.²⁵ Egy algoritmikus folyamat tehát jellemzően (1) egy modell felépítését foglalja magában valamilyen cél elérése érdekében, az összegyűjtött múltbeli adatok elemzése alapján; (2) a modellt megvalósító algoritmus kódolása; (3) az alanyokra vonatkozó adatok gyűjtése az algoritmus bemeneteihez; (4) az előírt algoritmikus műveletek alkalmazása a bemeneti adatokon; és (5) kimenetek a következő formában: (1) a modell megvalósítása; (2) a modell megvalósítása; (3) a modell megvalósítása; (4) a modell megvalósítása; (5) a modell megvalósítása.

²³ Lásd MEEHL, 17. lábjegyzet.

²⁴ EXEC. ELNÖKI HIVATAL, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS,

OPPORTUNITY, AND CIVIL RIGHTS, 8 (2016. május) (a gépi tanulást úgy jellemzi, mint "annak tudományát, hogy a számítógépeket rávegyük arra, hogy cselekedjenek anélkül, hogy kifejezetten programoznák őket" (idézi Andrew Ng, *Coursera Machine Learning Course*, STAN. U. 2016)).

²⁵ Robin K. Hill, *What an Algorithm Is*, 29 PHIL. & TECH. 35 (2015).

előrejelzések vagy ajánlások az adatelemzés láncolata alapján.
26

3. Algoritmikus kormányzás és intelligens Városok

A nagyméretű adatok és a prediktív algoritmusok használata a kormányzás egy formája - vagyis a hatóságok számára az egyéni viselkedés irányításának és az erőforrások elosztásának egy módja.²⁷ Az algoritmusok helyi szintű alkalmazása az adatvezérelt döntéshozatal felé való szélesebb körű elmozdulás része, és az "intelligens város" menetrenddel összefüggésben kell értelmezni.

A XXI. században a városok és megyék egyre inkább a "digitális hardverek és szoftverek felé fordultak, amelyek hatalmas mennyiségű adatot állítanak elő a városi folyamatokról".²⁸ A digitális technológiák kormányzati integrációja kezdetben kezdetleges e-kormányzati kezdeményezéseket és a kormányzati erőforrások digitalizálását jelentette.²⁹ Az elmúlt fél évtizedben a helyi önkormányzatok kiterjedtebb analitikát vezettek be, és elkezdték kihasználni az érzékelőhálózatokat, a mindenütt jelenlévő kommunikációt és a számítástechnikát.³⁰ Mindez a munka - az adatok gyűjtése, elemzése és felhasználása - olyan technikai tudást és infrastruktúrát igényel, amellyel a legtöbb önkormányzat nem rendelkezik. A városoknak kevesebb erőforrással többet kell megoldaniuk az adatalapú kormányzásra való áttérés során. A technológiai magánvállalatok olyan "megoldásokat" kínálnak, amelyeket a városvezetők nehezen tudnak értékelni.

A helyi önkormányzatok az "intelligens" városi rendszerekhez szükséges analitika kifejlesztésében a köz- és magánszféra partnerségére támaszkodnak.³¹ Ellentmondásos módon a magánszervezetek az egész intelligens város mozgalom élére álltak.³² Az IBM valóban bejegyezte a

²⁶ Lásd Tal Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1517-20; Gillespie, *supra* note 6, 167; Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 640 n.14 (2017).

²⁷ Marijn Janssen & George Kuk, *The Challenges and Limits of Big Data Algorithms in Technocratic Governance*, GOV'T INFO. Q. 33, 371-77 (2016) (arról értekezik, hogy az algoritmusok és a big data hogyan válnak a kormányzás olyan formájává, amely gyakran áthatolhatatlan a kihallgatás vagy a magyarázat számára).

²⁸ Alan Wiig & Elvin Wyly, *Bevezetés*: URB. GEOGRAPHY 485, 488 (2016); lásd még Rob Kitchin, *The Real-Time City? Big Data and Smart Urbanism*, 79 GEOJOURNAL 1 (2014).

²⁹ Lásd általában STEPHEN GOLDSMITH & SUSAN CRAWFORD, THE RESPONSIVE CITY: ENGAGING COMMUNITIES THROUGH DATA-SMART GOVERNANCE (2014) (a digitális technológia kormányzásba való integrációjának szakaszait követve).

³⁰Vö. általában ANTHONY M. TOWNSEND, SMART CITIES: NAGY ADATOK, POLGÁRI HACKEREK, ÉS AZ ÚJ UTÓPIA KERESÉSE (2013).

³¹ Lásd pl. Alberto Vanolo, *Smartmentality: The Smart City as Disciplinary Strategy*, 51 URB. STUD. 883 (2013) (a köz- és magánszféra partnerségének központi szerepét írja le az intelligens város víziójában és megvalósításában).

³² Lásd Janine S. Hiller & Jordan M. Blanke, *Smart Cities, Big Data, and the Resilience of Privacy*, 68 HASTINGS L.J. 309 (2017) (az intelligens városok

vállalati keretezésének leírása).

az "okosabb városok" kifejezést védjegyként a technológiaalapú városüzemeltetés piacra dobására irányuló kampánya részeként.³³ A Cisco is hasonlóan aktív volt.³⁴ A Sidewalk Labs, az Alphabet (amely a Google-t birtokolja) leányvállalata, a torontói vízpart újjáépítésén dolgozik.³⁵ Ennek az intelligens vízparti negyednek az irányítása egyenlő arányban oszlik meg a Sidewalk Labs és a Waterfront Toronto kormányzati ügynökség között.³⁶

B. Ígéreték és veszélyek

Az intelligens városok magántechnológiai vállalatokra való támaszkodásában, valamint az algoritmikus folyamatokon keresztül történő biztosításmatematikai ítélőképesség kormányzati alkalmazásában általában véve is van jó és rossz is.

A jó oldalról nézve az algoritmikusan megalapozott döntéshozatal nagyobb hatékonyságot és igazságosságot ígér a kormányzati szolgáltatások nyújtásában. Ahogyan az orvostudományban is bebizonyosodott, az aktuáriusi előrejelzés néha mérhetően jobb, mint a klinikai előrejelzés: az adathalmazok formalizált elemzése jobb kockázatértékelést eredményezhet, mint az informális szakmai megérzések, amelyek a gyakorlatban szerzett többéves tapasztalat során alakultak ki.³⁷ Az adatelemzés korábban nem észlelt vagy nem pontosan számszerűsített mintákat hozhat felszínre. Például a Yelp éttermi vélemények szisztematikus nyomon követése tájékoztathatja a városi egészségügyi felügyelőket a joghatóságuk alá tartozó éttermekben előforduló, élelmiszer eredetű megbetegedésekről.³⁸ Az adatok integrálása az egymástól elszigetelt közigazgatási területeken, például a

³³ Lásd SMARTER CITIES, Registration No. 4,033,245; Ola Söderström et al., *Smart Cities as Corporate Storytelling*, 18 CITY 307 (2014); lásd még Alan Wiig, *IBM's Smart City as Techno-Utopian Policy Mobility*, 19 CITY 158 (2015) (az IBM korai okosváros-kezdeményezéseinek utópisztikus retorikáját és extravagáns ígéreteit vizsgálva); Alan Wiig, *The Empty Rhetoric of the Smart City: From Digital Inclusion to Economic Promotion in Philadelphia*, 37 URB. GEOGRAPHY 535, 540 (2016) (kifejtve, hogy az IBM Smarter Cities Challenge keretében a városok partnerséget ajánlottak a vállalati "tanácsadókkal és technológiai szakemberekkel [amelyek] segítenek az önkormányzatoknak elemezni és rangsorolni igényeiket, felülvizsgálni erősségeiket és gyengeségeiket, és tanulni a más városok által alkalmazott sikeres stratégiákból").

³⁴Gordon Falconer & Shane Mitchell, *Smart City Framework: A Systematic Process of Enabling Smart + Connected Communities*, CISCO (2012), http://www.cisco.com/c/dam/en_us/about/ac79/docs/ps/motm/Smart-City-Framework.pdf [<http://perma.cc/D3ZA-K8NC>].

³⁵Innovációs és finanszírozási partneri keretmegállapodás: SIDEWALK TORONTO (Nov.1,2017),<http://sidewalktoronto.ca/wp-content/uploads/2017/10/Waterfront-Toronto-Agreement-Summary.pdf> [<http://perma.cc/6JAL-HEQ4>].

³⁶Id.

³⁷Vö. pl., Grove et al., 18. lábjegyzet.

³⁸Vö. Edward L. Glaeser et al., *Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life* (Harv. Bus. Sch. NOM Unit,

065,2015),<http://dash.harvard.edu/bitstream/handle/1/24009688/16-065.pdf?sequence=1> [<http://perma.cc/QB2R-CFLG>].

oktatás és a humán szolgáltatások, majd az adatok felhasználásával a kormányzati segítségre szoruló családok rangsorolása javíthatja a szociális szolgáltatások nyújtását.³⁹

Az algoritmikusan megalapozott döntéshozatal segíthet a kormánytisztviselőknek abban is, hogy elkerüljék a kevésbé formális, "megérzésen" alapuló döntéshozatalban esetlegesen megjelenő explicit vagy implicit elfogultságokat.⁴⁰ Például egy feltételes szabadlábra helyezési bizottság tagjai, akik a feltételes szabadlábra helyezésről szóló döntések meghozatalához egyszerűen csak meghallgatják a fogvatartottat, túlzottan összpontosíthatnak a bűncselekmény súlyosságára, arra, hogy a fogvatartott tanúsít-e megbánást, vagy kulturális vagy etnikai általánosításokra. Ezzel szemben az adatelemzés szisztematikus alkalmazása azonosíthatja azokat a jellemzőket, amelyek jelentős korrelációt mutatnak a visszaeséssel, és értékelheti ezen korrelációk erősségét, akár külön-külön, akár együttesen. Ezek az összefüggések aztán kódolhatók egy olyan algoritmusba, amely megbecsüli a visszaesés kockázatát, ha a fogvatartottra vonatkozó információkkal táplálják.⁴¹ Az algoritmus megalkotása során a kormánytisztviselők kifejezetten dönthetnek úgy, hogy kizárják az olyan érzékeny jellemzőket, mint a faji, etnikai vagy vallási hovatartozás, valamint az ilyen érzékeny jellemzők helyettesítőjeként szolgáló adatkategóriákat a figyelembe vételből, ha arra a következtetésre jutnak, hogy ez a figyelembe vétel nem lenne tisztességes.

Ugyanakkor a prediktív algoritmusok aligha tévedhetetlenek, és különleges kockázatokat rejtenek magukban, különösen akkor, ha jelentős mértékben magánpartnerek irányítják őket. Ha nem megfelelően fejlesztik ki vagy hajtják végre, a prediktív algoritmusok kevésbé pontosnak bizonyulhatnak, mint a kormánytisztviselők klinikai ítélőképessége, és formalizálhatják és elfedhetik az adatokba beágyazott torzításokat, amelyeken képzik őket. Ezen túlmenően, amint azt alább tárgyaljuk, az algoritmusok olyan politikai döntéseket hozhatnak, amelyek eltérnek a választópolgárok vagy választott képviselőik preferenciáitól.

A magánszolgáltatók bevonása az algoritmikus tervezésbe újabb veszélyeket rejt magában, többek között az átláthatatlanságot, a nyilvánosság jogfosztottságát és az elszámoltathatóság elvesztését. Azok a köztisztviselők, akik a prediktív algoritmusok fejlesztését magánszolgáltatókra bízták, nem vehetnek részt az algoritmusokba beépülő politikai döntésekben, és nem is tudhatnak azokról. Azok a közalkalmazottak, akik a prediktív algoritmusok eredményét használják döntéseikhez, nem feltétlenül értik a tervezést és a korlátokat.

³⁹ *Lásd pl.* Erika M. Kitzmiller, *IDS Case Study: Allegheny County*, ACTIONABLE INTELLIGENCE FOR SOC. POL'Y (2014. május) (elemzi, hogy a pennsylvaniai Allegheny megye hogyan használta az adatelemzést a humánszolgáltatási ügynökség reagálóképességének javítására).

⁴⁰ *Lásd* Daniel Castro, *Data Detractors Are Wrong: The Rise of Algorithms is a*

Elektronikusan elérhető a következő címen:

<https://ssrn.com/abstract=3012499>

Cause for Hope and Optimism, CTR. FOR DATA INNOVATION (2016),
<http://www.datainnovation.org/2016/10/data-detractors-are-wrong-the-rise-of-algorithms-is-a-cause-for-hope-and-optimism/> [<http://perma.cc/YBS8-74P3>].

⁴¹ Az Arnold Alapítvány PSA-Court algoritmusairól, amelyekkel kapcsolatban nyílt iratokat kértünk, lásd a 123-136. megjegyzéshez tartozó szöveget.

az algoritmusról, és nem biztos, hogy meg tudják állapítani, hogy az figyelembe vette-e azokat a tényezőket, amelyeket ők relevánsnak tartanak. Még ha képesek is lennének az algoritmus lekérdezésére, akkor is megállíthatják őket a szállító által kidolgozott szerződések, amelyek a szállítónak adják az adatok és az elemzések feletti ellenőrzést vagy tulajdonjogot.⁴²

Az algoritmikus kormányzás révén a közigazgatásban való magánrészvétel aggodalomra ad okot, hogy az adatokat a polgárok sérelmére és a közhatalom gyengítésére használják fel. "A rendfenntartás, a megfigyelés, a tömegellenőrzés, a vészhelyzetekre való reagálás mind-mind történelmileg állami funkciók, és a polgárok elvárhatják, hogy az ezzel kapcsolatos igen érzékeny adatok az állam birtokában legyenek. Egy [a köz- és magán-szféra partnerségére épülő] városban azonban az a valószínűség, hogy az adatok ... magánellenőrzés alá kerülnek."⁴³ Az intelligens városi technológiák terén élenjáró Barcelona digitális vezetője szerint a városok "egy olyan fekete dobozos operációs rendszerrel végződhetnek, amelyben a város maga veszíti el az ellenőrzést a kritikus információk és adatok felett, amelyeket jobb döntések meghozatalára kellene használni."⁴⁴ A kockázat az, hogy az adatokat és az elemzéseket ellenőrző vállalat foglalja el a városi kormányzás parancsnoki központját, míg a demokratikusan elszámoltatható tisztviselők, akik nem képesek az adatok ellenőrzésére, a perifériára szorulnak.⁴⁵

Ehhez kapcsolódóan aggodalomra ad okot, hogy e partnerségek révén a magán-szolgáltatók a kormányzatokat saját rendszerekhez kötik. Egyes intelligens városokkal foglalkozó kommentátorok arra figyelmeztetnek, hogy az "intelligens" projektek csupán arra szolgálnak, hogy az önkormányzatoknak a szállító tulajdonában lévő és általa kezelt átfogó adatkezelési rendszereket adjanak el.⁴⁶

⁴²Vö. *pl.*, Memorandum of Understanding Between the Laura and John Arnold Foundation and the Superior Court of California, County of San Francisco 3 (Aug. 4, 2015), <http://www.muckrock.com/foi/san-francisco-city-and-county-3061/san-francisco-public-safety-assessment-court-30096/#file-113830> [<http://perma.cc/73QC-QLX9>]; lásd még Angwin et al., *supra* note 7 (a COMPAS szerződés leírása).

⁴³Edwards, *supra* 16. lábjegyzet, 33. o.

⁴⁴David Meyer, *How One European Smart City is Giving Back Power to its Citizens*, ALPHR (July 10, 2017), <http://www.alphr.com/technology/1006261/how-one-european-smart-city-is-giving-power-back-to-its-citizens> [<http://perma.cc/FM7D-YVX9>] (idézi Francesca Bria, Barcelona digitális vezetőjét). Egy másik probléma, hogy "az üzleti modell nagyon kevés szolgáltatótól való függőséget teremt". *Id.* A városi szolgáltatásoknak ez a bizonyos magán-szolgáltatókhoz való kötöttsége "kiterjeszhető a város teljes városi infrastruktúrájára. Beszélünk a közlekedésről, a jobb hulladékgyűjtésről, akár a vízről, az energiáról, az elosztott zöld infrastruktúráról. Ez nagy probléma a közigazgatás számára, hogy elveszíti az infrastruktúra irányítása feletti ellenőrzést". *Id.*

⁴⁵Christine Richter & Linnet Taylor, *Big Data and Urban Governance*, in *BIG DATA AND URBAN GOVERNANCE* 175, 180 (J. Gupta et al. eds., 2015) ("A vállalatok növekvő befolyása az intelligens városi környezet kialakítására

potenciálisan a vállalatokat helyezi a demokratikus városi folyamatok középpontjába."); lásd még Kitchin, *Thinking Critically*, supra note 6).

- ⁴⁶ Lásd pl. ADAM GREENFIELD, AGAINST THE SMART CITY (2013); Donald McNeill, *Global Firms and Smart Technologies: IBM and the Reduction of Cities*, 40 TRANSACTIONS INST. BRIT. GEOGRAPHERS 562 (2015); Wiig, *The Empty Rhetoric*.

A szolgáltatási szerződések a kormányzatot a technológiai szolgáltatótól függővé tehetik a frissítések és a folyamatos fejlesztés tekintetében, és a kormányzatot olyan védett technológiákhoz köthetik, amelyek költségeit és az innováció ütemét nem tudják ellenőrizni. A kötöttség a technikai rendszereken túl kiterjedhet a fizikai infrastruktúrára is, amelybe beágyazódnak. Az Alphabet leányvállalata, a Sidewalk Labs például a Waterfront Torontóval együttműködve a semmiből "holisztikusan" épít egy 800 hektáros minivárost, amely "a digitális technológia és az adatok által továbbfejlesztett hely".⁴⁷ A Sidewalk Labs valószínűleg adatokat fog gyűjteni és a város adatait fogja felhasználni, hogy algoritmikus előrejelzéseket készítsen a kívánatos vízparti tevékenységről.⁴⁸ Nem világos, hogy a nyilvánosságnak milyen tulajdonjoga vagy hozzáférése lesz az adatokhoz vagy a kapcsolódó elemzésekhez, hogyan fogják szabályozni a fizikai infrastruktúrát, vagy hogy az állami szervek átvehetik-e az információs és fizikai eszközök feletti ellenőrzést, ha véget kívánnak vetni a magáncéggel való kapcsolatnak.

II.

AZ ÉRTELMESES ÁTLÁTHATÓSÁG MEGHATÁROZÁSA: MI A A NYILVÁNOSSÁGNAK TUDNIA KELL

Mivel a mesterséges intelligencia és az algoritmikus előrejelzés gyorsan behatol a helyi kormányzásba, kívánatos lenne, ha a nyilvánosság tudná, hogy az algoritmusok milyen politikai döntéseket tükröznek, és mennyire jól teljesítenek a kitűzött célok elérésében. Ez a rész az algoritmusok működésének fényében határozza meg az értelmes átláthatóság összetevőit. Az V. rész az értelmes átláthatóság gondolatát konkrét közzétételi gyakorlatokon keresztül operacionalizálja, amelyeket a nyilvános előrejelző algoritmusok számára ajánlunk.

A. *Mi az algoritmus politikája?*

of the Smart City, supra note 33, at 535 ("[T]he smart city acts as a data-driven logic urban change where widespread benefit to a city and its residents is suggested, maszkírozva e politikák hasznosságát a vállalkozói gazdaságfejlesztési stratégiák előmozdítására.").

⁴⁷Darrell Etherington, *Alphabet's Sidewalk Labs To Turn Toronto into Turn Toronto into a Model SmartCity*, TechCrunch(2017október17), <http://techcrunch.com/2017/10/17/alphabets-sidewalk-labs-to-turn-toronto-area-into-a-model-smart-city/> [<http://perma.cc/2BTE-ZSFY>].

⁴⁸Laura Bliss, *When a Tech Giant Plays Waterfront Developer*, CITYLAB (2018. január 9.) ,<http://www.citylab.com/design/2018/01/when-a-tech-giant-plays-waterfront-developer/549590/> [<http://perma.cc/6LKW-FTE9>] (egy "digitális réteget" javasol, amely összeköti és kezeli a fizikai infrastruktúrát és az infrastruktúra, a városi szolgáltatások és az emberek közötti kölcsönhatásokat); *vö.* SIDEWALK LABS, *Project Vision* (2017. október 17.), <http://sidewalktoronto.ca/wp-content/uploads/2017/10/Sidewalk-Labs-Vision-Sections-of-RFP-Submission.pdf>.

Az algoritmikus kormányzásnak van politikája. Az algoritmikus folyamat minden szakaszában kódolva vannak az ítéletek.⁴⁹ Ezek olyan ítéletek, amelyeket valamilyen szinten a nyilvánosságnak ismernie kell, és amelyekről beszélnie kell. A pusztán a tudományos igazsághoz vezető eszközként pozicionált algoritmusok azonban elrejthetik a mögöttes modellekbe ágyazott értékeket.⁵⁰ Ráadásul, amikor magánszolgáltatók ellenőrzik az algoritmikus irányítást, az algoritmusok politikája visszahúzódik a magánjellegű védekezés mögé. Az az elképzelés, hogy az algoritmusok politika nélküli tudományt jelentenek, elfedheti a magánirányítás tétjeit, amelyek a privatizáció más területein, például az iskolákban és a börtönökben egyértelműbbek.

Ahogy Harry Surden megjegyzi, egy előrejelző algoritmus ajánlása "valójában a rendszer tervezőinek szubjektív döntései mögött rejtőzik, hogy milyen adatokat használjanak, milyen adatokat vegyenek fel vagy zárjanak ki, hogyan súlyozzák az adatokat, és milyen információkat hangsúlyozzanak vagy csökkentsenek".⁵¹ Bármely politikai cél megvalósítása során lesznek kompromisszumok. Például még egy olyan vitathatatlan cél megvalósításakor is, mint a forgalmi várakozási idő csökkentése, a politikai döntéshozónak mérlegelnie kell, hogy a forgalomáramlás érdekében mekkora kockázatot vállalhat a gyalogosok biztonságára nézve, és hogyan lehet figyelembe venni a kipufogógáz-kibocsátás csökkentését. A várakozási idő csökkentésére vonatkozó általános irányelv nem írja elő, hogy milyen kompromisszumokat kell kötni. Sőt, egyes választási lehetőségek talán fel sem merültek a döntéshozókban, hanem csak akkor kerülnek felszínre, amikor a mérnökök az algoritmusok megtervezéséhez érkeznek, és rájuk hárul a kompromisszumok megoldása.

Egyre több szakirodalom foglalkozik az algoritmusok társadalmi, politikai és etikai dimenzióival.⁵² A III. részben konkrét, kontextusba helyezett problémákkal foglalkozunk. Egyelőre elég, ha

⁴⁹ Mittelstadt et al., *Az algoritmusok etikája: Mapping the Debate*, July-Dec. 2016 (BIG DATA & SOC'Y 1, 1 ("A működési paramétereket a fejlesztők határozzák meg és a felhasználók konfigurálják a kívánt eredmények szem előtt tartásával, amelyek egyes értékeket és érdekeket másokkal szemben előnyben részesítik.")).

⁵⁰ Lásd Rob Kitchin, *Reframing, Reimagining and Remaking Smart Cities: The Programmable City 4* (Open Sci. Framework, Working Paper No. 20, 2016) (az intelligens városokkal kapcsolatos kritikák összefoglalása).

⁵¹ Harry Surden, *Values Embedded in Legal Artificial Intelligence* 5 (Univ. Colo. Law Legal Studies, Research Paper No 17-17, Oct. 18, 2017), <http://ssrn.com/abstract=2932333> [<http://perma.cc/3RUC-54PP>].

⁵² Lásd Diakopoulos, *supra* note 6, 400. o. (az adatok rangsorolásába, osztályozásába, társításába és szűrésébe ágyazott értékválasztásokat tárgyalja); lásd még EXEC. ELNÖKI HIVATAL NAT'L SCI. & TECH. TANÁCS COMM. ON TECH, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016), http://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf [<http://perma.cc/TZ2W-PJT7>]; Bryce Goodman & Seth Flaxman, *EU Regulations on Algorithmic Decision-Making and a "Right to Explanation"*,

ARXIV(2016) ,<http://arxiv.org/pdf/1606.08813.pdf>[<http://perma.cc/4GDR-XQ5P>]; Felicitas Kraemer et al, *Létezik-e az algoritmusok etikája?* , 13 ETHICS & INFO. TECH. 251 (2011).

egy példán keresztül kiemeljük egy algoritmus politikájának egy különösen fontos megnyilvánulását: azt, hogy egy osztályozó algoritmus hogyan kezeli a hamis pozitív és hamis negatív eredményeket. Vegyünk egy algoritmust, amely egy vasútállomáson lévő tárgyakat gyanúsnak vagy nem gyanúsnak minősít. A programozóknak formalizálniuk kell az egyensúlyt a hamis riasztások kockázata és a veszélyes tárgyak kihagyásának kockázata között. A statisztikában a hamis pozitív eredményeket általában "I. típusú hibáknak", a hamis negatív eredményeket pedig "II. típusú hibáknak" nevezik. A programozóknak úgy kell "hangolniuk" az algoritmust, hogy az egyik hibafajtát előnyben részesítse a másikkal szemben, vagy hogy egyformán kezelje őket.⁵³ Nick Diakopoulos megjegyzi, hogy az algoritmikus hangolás "előnyben részesítheti a különböző érdekelt feleket egy döntésben, ami az ilyen algoritmus tervezőjének alapvető értékítéletét jelenti a hamis pozitív és hamis negatív hibák egyensúlyának tekintetében"⁵⁴.

Azon kevés algoritmusok egyike, amelyek esetében ez a hangolás kiderült, a philadelphiai felnőtt próbaidősök és feltételes szabadlábra helyezettek körében az erőszakos visszaesés előrejelzésére szolgáló, a próbaidősök kockázatát előrejelző algoritmus. Az eszköz megjósolja annak valószínűségét, hogy egy próbaidős a szabadulást követő két éven belül erőszakos bűncselekményt követ el, és a lakosságot magas, közepes és alacsony kockázatú csoportba sorolja. Az algoritmust úgy építették fel, hogy a korábbi téves negatív eseteket 2,6-szor drágábbnak tekintették, mint a téves pozitív eseteket.⁵⁵ Richard Berk kriminológus és statisztikus, aki konzultált a programmal kapcsolatban, úgy becsüli, hogy az előrejelzések 29-38%-a téves - ezt a hibaarányt az a politika indokolja, hogy "sokkal veszélyesebb Darth Vader-t szabadon engedni, mint Luke Skywalker-t bebörtönözni".⁵⁶ Kiderült azonban, hogy a próbaidősök magas kockázatúnak való túlzott besorolása problémás volt, mivel ők drágább, a visszatérés zökkenőmentesítése érdekében nyújtott szolgáltatásokban részesültek. A város visszament Berkhez, és megkérte, hogy kalibrálja újra az algoritmust, hogy csökkentse a magas kockázatú kategória méretét. Egy másik

⁵³ Lásd Daniel Neyland & Norma Mollers, *Algorithmic IF . . . THEN Rules and the Conditions and Consequences of Power*, 20 INFO. COMM. & SOC'Y 45 (2016) (olyan algoritmusokat tárgyalva, amelyek éppen ilyen jellegű osztályozást végeznek).

⁵⁴ Diakopolous, *supra* note 6, 401. o.; lásd még Matthias Spielkamp, *Inspecting Algorithms for Bias*, MIT TECH. REV. (2017. június 12.), <http://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias/> [<http://perma.cc/K243-2DH2>] (azzal érvelve, hogy egy büntetési algoritmus "igazságosan" kezelheti a különböző csoportokat a valódi pozitívumok (visszaesés) tekintetében, de nem a hamis negatívumok (előre jelzett visszaesés, amely nem következik be) tekintetében).

⁵⁵Nancy Ritter, *Predicting Recidivism Risk: New Tool in Philadelphia Shows Great Promise*, 271 NAT'L INST. JUST.

(2013), <http://www.nij.gov/journals/271/pages/predicting->

recidivism.aspx [http://perma.cc/J9SA-88LV].

- ⁵⁶ Joshua Brustein, *This Guy Trains Computers to Find Future Criminals*, BLOOMBERG TECH. (2016. július 18.), <http://www.bloomberg.com/features/2016-richard-berk-future-crime/> [http://perma.cc/46L5-4DMU]. Lásd általában RICHARD A. BERK, STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE 13, 139-45 (2008) (a gépi tanulási algoritmusok különböző típusú hibáinak pontozását tárgyalja).

projekt résztvevője szerint a modell szándékosan kevésbé pontos, "hogya a megfelelő hibát produkálja, amikor ez megtörténik".⁵⁷

Az egyik hibatípus előnyben részesítése egy másik hibatípussal szemben csak egy a több tucat vagy száz döntés közül, amelyek a prediktív algoritmus felépítésében szerepet játszanak. Ezek közül néhány triviális, néhány pedig következetes. Néhány közülük nyilvánosan megfogalmazott politikai célkitűzéseket valósít meg, míg mások politikai irányítás nélkül a programozókra maradnak. Cary Coglianese és David Lehr elismeri, hogy "[a]zoknak az ügynökségeknek, amelyek nem szoktak erkölcsi értékelést végezni semmilyen formális folyamaton keresztül, nem hogy számokat rendelni hozzájuk, a gépi tanulási algoritmusok szükségessé teszik a szervezeti és demokratikus döntéshozatal kérdéseinek kezelését"⁵⁸.

B. Teljesít az algoritmus ?

Bármilyen rejtett politikai döntést is kódoljon egy algoritmus, a kormányzatnak feltehetően van egy magas szintű explicit politikai célja egy előrejelző algoritmus számára - legyen az a forgalmi várakozási idő csökkentése vagy a feltételeken szabadlábba helyezett visszaesésének minimalizálása. A nyilvánosságnak képesnek kell lennie arra, hogy értékelje az algoritmus teljesítményét a kitűzött célok elérésében. Ez egy viszonylag egyszerű kérdés, amely a hasznosságot a statisztikai teljesítmény alapján értékeli, amely az adatoknak a kívánt eredményhez való illesztésében mutatkozik meg.

Természetesen még itt is többféle teljesítménymérés létezik, és fontos megérteni, hogy az egyes mérőszámok mit jelentenek. Például a prediktív algoritmusok egyik népszerű mérőszáma a ROC (receiver operating characteristic) görbe alatti terület. Egyetlen 0,5 és 1 közötti számban kifejezve azt mutatja meg, hogy az algoritmus mennyivel jobb a hamis pozitív és a hamis negatív eredmények elkerülésében, mint az esetek véletlenszerű hozzárendelése. Van azonban néhány korlátja - csak akkor alkalmazható, ha az algoritmus kimenete egy olyan pontszám, amely a legkevésbé valószínűtől a legvalószínűbbig rangsorolja az alanyokat valamilyen kimenetelhez való társulás szempontjából -, és csak egy nézőpontot ad az algoritmus relatív sikeréről. Más mérőszámok a teljesítmény más aspektusaira összpontosíthatnak. Például az "illeszkedés jósága" tesztek megmutathatják, hogy bár a modell összességében elég jó egy kimenetel kockázatának előrejelzésében, az előrejelzései, amelyek szerint az alanyok a legkockázatosabb tíz százalékba tartoznak, lényegesen kevésbé pontosak, mint azok az előrejelzések, amelyek szerint az alanyok a legkevésbé kockázatos tíz százalékba tartoznak.

⁵⁷Id . (idézi Geoffrey Barnes-t).

⁵⁸Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1218 (2017).

százalék.⁵⁹ Más szóval, a pontosság mérésének számos módja van, és még több fog megjelenni. ⁶⁰ Nem lehet megérteni a teljesítményre vonatkozó állításokat anélkül, hogy ne tudnánk, hogyan mérik azt:

Számos oka lehet annak, hogy egy algoritmus nem hatékony. ⁶¹ Lehet, hogy rossz bemeneti adatokon képezték ki (garbage in, garbage out).⁶² Hibák adódhatnak a hibás induktív következtetésből, az adatok kiválasztásából és a tényezők súlyozásából is.⁶³ A tágabb értelemben vett algoritmikus folyamat másik hibapontja a végrehajtási fázisban lehet. Hacsak az algoritmikus előrejelzés nem önműködő, az embernek meg kell értenie az előrejelzést ahhoz, hogy eldöntse, mekkora súlyt adjon neki a döntéshozatali folyamatban. Önkormányzati környezetben gyakran a kormányzati dolgozók feladata lesz az adatok kiválasztása és bevitele is. Bár a validációs vizsgálatok segíthetnek annak biztosításában, hogy egy algoritmus elérje a kívánt célt, a pénzszűkében lévő kormányzatok nem feltétlenül követelik meg a validációs vizsgálatokat a bevezetés előtt vagy után, vagy azokat nem megfelelően végzik el. A validálási tanulmányok eredményeit, valamint a tervezésükre vonatkozó információkat a megfelelő értékelés lehetővé tétele érdekében nyilvános vizsgálatnak kell alávetni.

C. *Igazságos-e az algoritmus ?*

⁵⁹ *Lásd pl.* Alberto Maydeu-Olivares & C. Garcia-Forero, *Goodness-of-Fit Testing*, in INTERNATIONAL ENCYCLOPEDIA OF EDUCATION 190 (Penelope Peterson et al. eds., 3d ed. 2010).

⁶⁰ *Lásd pl.*, DEAN ABBOTT, APPLIED PREDICTIVE ANALYTICS: PRINCIPLES AND TECHNIQUES FOR THE PROFESSIONAL DATA ANALYST 283-304 (2014); Ewout W. Steyerberg et al., *Assessing the Performance of Prediction Models: A Framework for Some Traditional and Novel Measures*, 21 EPIDEMIOLOGY 128 (2010); Mauno Vihinen, *How To Evaluate Performance of Prediction Methods? Measures and Their Interpretation in Variation Effect Analysis*, 13 BMC GENOMICSS2 (2012), <http://doi.org/10.1186/1471-2164-13-S4-S2>. [<http://perma.cc/8Q73-GXYC>]; Scott Fortmann-Roe, *Accurately Measuring Model Prediction Error* (2012. május), <http://scott.fortmann-roe.com/docs/MeasuringError.html> [<http://perma.cc/3YDK-QDMK>].

⁶¹ *Lásd EXEC. OFFICE OF THE PRESIDENT*, *httphttpsupra* 24. lábjegyzet (a rosszul kiválasztott adatok, a hiányos, helytelen vagy elavult adatok, a szelekciós torzítások, valamint a történelmi torzítások nem szándékos fenntartása és támogatása).

⁶² *GarbageIn, GarbageOut*, WIKIPEDIA, http://en.wikipedia.org/wiki/Garbage_in,_garbage_out [<http://perma.cc/M8CV-P7U8>] (egy informatikai kifejezés magyarázata, amely azt az informális szabályt fejezi ki, hogy a számítógép kimenete csak annyira jó, amennyire a bemenet minősége).

⁶³ *Lásd a* módosított összefoglaló ítéletet, *Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist.*, Civil Action No. H-14-1189, at *13 (S.D. Tex. May 4, 2017) (megjegyezve, hogy egy algoritmikus pontszám "tévesen kiszámítható bármilyen okból, az adatbeviteli hibáktól kezdve a számítógépes kódban lévő hibáig. Az algoritmusok emberi alkotások, és ugyanúgy ki vannak téve a hibáknak, mint bármely más emberi vállalkozás.").

Előfordulhat, hogy egy algoritmus jól teljesít a kívánt eredmények elérése szempontjából, de a méltányossági intézkedések terén alulmarad. Erős közérdek fűződik annak biztosításához, hogy a prediktív algoritmusokat igazságosan tervezzék meg és hajtsák végre, különösen akkor, ha azok hatással vannak az egyénekre. Az igazságossággal kapcsolatos aggályok általában sokkal kevésbé számítanak a fejlesztőknek, mint az algoritmus teljesítménye, és előfordulhat, hogy egyáltalán nem is szerepelnek a mérnökök feladatai között.⁶⁴

A prediktív algoritmusok kormányzati alkalmazása a méltányosság hagyományos fogalmaival szemben kihívást jelent.⁶⁵ A prediktív modellek természetüknél fogva olyan egyszerűsítések, amelyek nem képesek figyelembe venni az alanyokkal kapcsolatos összes lehetséges releváns tényt, és ezért szükségszerűen csoportok tagjaiként, nem pedig egyénként kezelik az embereket.⁶⁶ Az általánosítások, amelyek az eltérő eseteket egyformán kezelhetik, ennek a folyamatnak a velejárói. Érzékeny döntések esetében, különösen, ha az egyéni szabadság forog kockán, a döntéshozóktól, például a bíráktól és a szociális munkásoktól elvárható, hogy az algoritmikus előrejelzésekkel szemben emberi ítélőképességet alkalmazzanak, hogy a hibás előrejelzéseket észrevegyék. Elméletileg az algoritmikus rendelet csak tanácsadó jellegű. A gyakorlatban a döntéshozók nagymértékben támaszkodnak a számokra, ami megnöveli a tétet a méltányosságukat illetően.⁶⁷

A legtöbbet vitatott algoritmikus méltányossági kérdés az volt, hogy a prediktív algoritmusok valószínűleg bevezetik-e vagy fenntartják-e a faji, nemi vagy más védett tulajdonságon alapuló hátrányos megkülönböztetést.⁶⁸ A megkülönböztetés további formái is aggodalomra adnak okot, például az, hogy egy algoritmus véletlenül nem kedvez-e bizonyos viselkedési formáknak (és ezért nem ösztönöz-e ezekre). Például, ha a mentális egészségügyi rendszer igénybevétele

⁶⁴ Nick Seaver, *Knowing Algorithms 2* (2014. február) (kiadatlan kézirat), <http://nickseaver.net/s/seaverMIT8.pdf> [<http://perma.cc/AB75-JY6M>] (azzal érvelve, hogy egy algoritmus politikai következményei "szigorúan kívül esnek a kereten" az algoritmusfejlesztők számára). Ezt a kihívást az informatika most kezdi felfedezni. *Lásd pl.*, Michael Feldman et al., *Certifying and Removing Disparate Impact*, in *PROC. 21ST ACM SIGKDD INT'L CONF. DISCOVERY & DATA MINING 259* (2015) (az algoritmikus eljárásokban a megkülönböztető hatás tesztjének és egy olyan módszernek a bemutatása, amellyel az adatokat elfogulatlaná lehet tenni); Sorelle A. Friedler et al., *On the (Im)possibility of Fairness* (Sept. 23, 2016) (kiadatlan kézirat), <http://arxiv.org/pdf/1609.07236.pdf> [<http://perma.cc/DQP4-PMWU>].

⁶⁵ Magára a méltányosságra különböző definíciók vonatkoznak; a választott definíció befolyásolja az algoritmikus méltányosság értékelését. *Lásd* Friedler et al., *supra* note 64 (azt ajánlva, hogy az informatikusok tegyék egyértelműbbé, hogy a méltányosság milyen fogalmát kívánják képviselni az algoritmusokban).

⁶⁶ *Lásd pl.* O'NEIL, *Supra* Note 6, 20-23. o.; Mittelstadt et al., *Supra* Note 49, 8. o.

⁶⁷ *Lásd általában* John Danaher, *Az algoritmikus fenyegetése*: 29 *PHIL. & TECH.* 245 (2016) (aggodalmakat fogalmaz meg az algoritmikus kimenet iránti tiszteletadással kapcsolatban olyan emberi döntéshozók részéről, akik nem értik az algoritmusok működését).

- ⁶⁸ *Lásd pl. O'NEIL, supra note 6; Barocas & Selbst, supra note 7; Joh, supra note 4; Pauline T. Kim, Data-Driven Discrimination at Work, 58 WM. & MARY L. REV. 857 (2017); Andrew D. Selbst, Disparate Impact in Big Data Policing, GA. L. REV. (megjelenés előtt), <http://ssrn.com/abstract=2819182> [<http://perma.cc/KDT8-DXF5>].*

összefügg a gyermek veszélyeztetettségének fokozott kockázatával, akkor az algoritmus a mentális egészségügyi rendszer igénybevételét is bevonhatja a kockázatértékelésbe. A mentális egészségügyi rendszer használata összefügghet vagy nem összefügghet a védett csoporthoz tartozással. Egy olyan algoritmus azonban, amely bünteti azokat, akik mentális egészségügyi kezelést kérnek, méltányossági aggályokat vet fel, valamint nagyobb jóléti aggályokat, ha azok, akiknek a mentális egészségügyi kezelés segítene, a gyermekjóléti beavatkozások elkerülése érdekében inkább nem kérnek mentális egészségügyi kezelést.

A méltányosság és a teljesítmény néha összefügg egymással.⁶⁹ Klasszikus példa erre az arcfelismerő algoritmus, amelyet az azt készítő mérnökök által ismert, többnyire fehér arcokon képeztek ki.⁷⁰ Ennek eredményeként a programok nagyobb valószínűséggel nem azonosították vagy tévesen azonosították a sötét bőrű emberi arcokat, ami miatt az ártatlan sötét bőrű embereket nagyobb valószínűséggel azonosították tévesen bűncselekmény gyanúsítottjaként.⁷¹ Az algoritmus újratanítása minden bőrszínű emberi arcot használva összességében jobb teljesítményt nyújtana, valamint csökkentené a világos és sötét bőrű emberi arcok közötti pontosságbeli különbségeket. Ha egy algoritmus igazságosabbá tétele ténylegesen növelné annak hasznosságát, akkor elvárható, hogy a teljesítmény szigorú elemzése is nagyobb igazságossághoz vezessen.

Bizonyos esetekben azonban kompromisszumot kell kötni a méltányosság és a teljesítmény között. Az egyén csoporttagságának bevonása növelheti az algoritmus hasznosságát, ha a megfigyelt korrelációk nem egyszerűen megismétlik az adatokban lévő más korrelációkat. Vegyük például azt a korrelációt, amelyet egyes adatelemzések találtak bizonyos irányítószámokból származó elítélt bűnözők és a magasabb visszaesési arányok között.⁷² Ez a korreláció nem feltétlenül növeli a

⁶⁹Vö. pl., Tal Z. Zarsky, *Understanding Discrimination in the Scored Society*, 89 WASH. L. REV. 1375, 1383 (2014) ("[I]n sok esetben a megkülönböztetés nem lehet hatékony, és így fenntarthatatlan üzleti vagy társadalmi gyakorlatot jelenthet").

⁷⁰Vö. Tess Townsend, *Most Engineers Are White-And So Are the Faces They Use To Train Software*, RECODE (Jan. 18, 2017), <http://www.recode.net/2017/1/18/14304964/data-facial-recognition-trouble-recognizing-black-white-faces-diversity> [<http://perma.cc/T5PL-XNLJ>]. Lásd általában Clare Garvie & Jonathan Frankle, *Facial-Recognition Software Might Have a Racial Bias Problem*, ATLANTIC (Apr. 7, 2016), <http://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/> [<http://perma.cc/6K5Z-H3JM>] (az arcfelismerő algoritmusok faji elfogultságának leírása).

⁷¹Vö. Garvie & Frankle, 70. lábjegyzet.

⁷² Lásd Angwin et al., 7. lábjegyzet. Lásd azonban William Dieterich és mások, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, NORTHPOINTE INC. RES. DEPT (2016. július 8.), <http://go.volarisgroup.com/rs/430-MBX->

s/ProPublica_Commentary_Final_070616.pdf[<http://perma.cc/ND8Z-Y3GS>];
Anthony W. Flores et al., *False Positives, False Negatives, and False Analyses:
A Rejoinder to "Machine Bias: There's Software Used Across the Country to
Predict Future Criminals. And it's Biased Against*

egy algoritmus előrejelző ereje, ha az algoritmus mondjuk a foglalkoztatási előzményeket is használná tényezőként. A postai irányítószámok és a foglalkoztatási előzmények majdnem egyforma tényezőknél bizonyulhatnak, de a foglalkoztatási előzmények jobb előrejelzőnek bizonyulnak. Lehetséges, hogy az irányítószámok a munkanélküliség földrajzi csoportosulása miatt csak gyenge jelzésként szolgálnak a jövőbeli foglalkoztatásra, és ezért nem javítják a foglalkoztatási előzményeket közvetlenül figyelembe vevő előrejelzéseket. Ezzel szemben azonban az irányítószám-információk bevonása bizonyíthatóan növelheti az algoritmus előrejelző erejét, rámutatva valamilyen olyan összefüggésre, amelyet semmilyen más bevont változó vagy jellemző nem fed le.

A lakóhelyi szegregáció miatt azonban az irányítószámok gyakran a faji hovatartozás helyettesítői. Ennek tudatában az ügynökségek dönthetnek úgy, hogy kizárják a postai irányítószámokat a prediktív algoritmusok bemenetéből, még akkor is, ha azok javítják az algoritmus előrejelző erejét. Arra a következtetésre juthatnak, hogy a bórszín nem áll ok-okozati összefüggésben a kívánt vagy nem kívánt eredménnyel, és ezért az előrejelző erőnek valamilyen más társváltozóban kell gyökereznie. A faji hovatartozás vagy annak helyettesítője, mint rövidítés arra, ami valójában valamilyen ok-okozati összefüggést mutathat, "a célzottan egyenlőtlen bánásmód történelmét"⁷³ állandósítaná, amely "egy olyan megváltoztathatatlan jellemzőn alapul, amelyet kizárólag a születés véletlenje határoz meg"⁷⁴. Más szóval, ha a pontosság valamely csekély mértékű növekedése szinte biztosan egy védett csoporttal szembeni méltánytalanság növekedésével jár, a közhivatal a pontosság helyett a méltányosságot választhatja.

Természetesen bizonyos helyzetekben a faji hovatartozás figyelembe vétele egyszerűen megerősítheti az előítéletesség történelmi mintáit. A kisebbségi lakónegyedekben, ahol a történelem során intenzívebb rendőri felügyeletet alkalmaztak, magasabb lesz a letartóztatási és újbóli letartóztatási arány, és az algoritmus ezt követően több rendőri felügyeletet javasol, és így tovább.⁷⁵ A diszkriminatív bánásmód történelmi mintázata így magasabb megfigyelt bűnözési arányokat fog *okozni azokban* az irányítószámokban, amelyek az algoritmus szerint nagyobb bűnözési kockázatnak vannak kitéve.

A joghatóságok különböző módon kezelték az ilyen méltányossági aggályokat. Az oaklandi rendőrség úgy döntött, hogy egyáltalán nem használ előrejelző algoritmust (PredPol), mivel arra a következtetésre jutott, hogy "a rendőröket többnyire olyan, alacsonyabb jövedelmű kisebbségi negyedekbe vezényelték volna, ahol a korábbi kábítószer-bűncselekményeket

Blacks," COMMUNITY RESOURCES FOR JUST., <http://www.crj.org/page/-/publications/rejoinder7.11.pdf> [<http://perma.cc/624E-ZUE9>]; Rhema Vaithianathan et al., *Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and*

Implementation, CTR. FOR SOC. DYNAMICS (2016. szeptember) (az irányítósámok és a faji hovatartozás egyéb helyettesítőinek tárgyalása).

⁷³San Antonio Ind. Sch. Dist. v. Rodriguez, 411 U.S. 1, 28 (1973).

⁷⁴Frontiero kontra Richardson, 411 U.S. 677, 686 (1973).

⁷⁵Lásd O'NEIL, *Supra* note 6, 97-98. o.

rögzítették. ⁷⁶ Ezzel szemben olyan városok, mint Philadelphia és Chicago prediktív rendőrségi programokat alkalmaznak, de a szállítójuk (Azavea, Inc., a HunchLab fejlesztői) úgy döntött, hogy a valószínűsíthető rendszerszintű torzítások elkerülése érdekében a rendőrségi modellek létrehozásakor nem veszik figyelembe a letartóztatási adatokat, különösen a kábítószerrel kapcsolatos és a kellemetlen bűncselekményekre vonatkozó adatokat. ⁷⁷ Végső soron, ha egy előrejelző algoritmust nem tesznek kellően átláthatóvá, nem fogjuk tudni, hogy az automatizált döntéshozatal és kockázatjóslás megfelel-e a méltányossággal kapcsolatos érdemi kötelezettségvállalásainknak. ⁷⁸

D. *Növeli vagy csökkenti az algoritmus a kormányzati kapacitást?*

További veszélyt jelent, hogy az algoritmikus kormányzás, amely nem tűri a kritikai értékelést, ugyanakkor kiszorítja az emberi döntéshozatalt, kiüresíti a köztisztviselők döntéshozatali képességét. Ehhez hozzájárulhat az algoritmus iránti indokolatlan tisztelet, az algoritmikus folyamatok elégtelen megértése és az emberi ítélőképesség elsorvasztása.

A kormánytisztviselők akkor is elfogadhatják az algoritmusok eredményeit, ha azok hibásak, diszkriminatívak, vagy túl durva kategóriák vagy túl szűk körű eredmények alapján készültek. Amikor a "gép azt mondja", a sietős és túlterhelt emberi döntéshozóknak nehéz lehet ellenállni a parancsnak. Ahogy Harry Surden megjegyzi, a bírák "nagyobb tiszteletet tanúsíthatnak a számítógép-alapú ajánlások iránt, szemben a hasonló, emberi alapú értékelésekkel, tekintettel a mechanikus objektivitás aurájára, amely a számítógép által generált, analitikán alapuló elemzéseket övezi".⁷⁹ Michael Ananny szerint "az algoritmikus kategóriák ... bizonyosságot jeleznek, elriasztják az alternatív feltárásokat, és koherenciát teremtenek az eltérő tárgyak között".

⁷⁶ Emily Thomas, *Why Oakland Police Turned Down Predictive Policing*, VICE (2016. december 28.), <http://motherboard.vice.com/read/minority-retort-why-oakland-police-turned-down-predictive-policing> [<http://perma.cc/KJH5-DS93>].

⁷⁷ ⁷⁶. A *Citizen's Guide to HunchLab*, HUNCHLAB 26 (2017. július 11.), <http://robertbrauneis.net/algorithms/HunchLabACitizensGuide.pdf> [<http://perma.cc/NY6B-QNZD>]. Az Azavea azt is ajánlja, hogy egy kis fokú véletlenszerűséget vezessenek be az algoritmusba, hogy "valószínűségi helyszínválasztást" végezzenek, például járőrözés esetén, részben az elfogultság ellen. *Lásd id.* 10-11. o.

⁷⁸ ⁷⁶. Joanna Bryson, *Three Very Very Different Sources of Bias in AI, and How To Fix Them*, ADVENTURES IN NI (July 13, 2017), <http://joanna-bryson.blogspot.co.uk/2017/07/three-very-different-sources-of-bias-in.html> [<http://perma.cc/2RXD-LZPG>] ("Az [elfogultság] kezelésének módja az, hogy ragaszkodunk a magyarázathoz való joghoz, a megfelelő eljáráshoz. Minden olyan algoritmust, amely hatással van az emberek életére, auditálni kell.").

⁷⁹ *Lásd* Surden, *Supra* 51. lábjegyzet, 2. o.; *lásd még* danah boyd & Kate Crawford, *Critical Questions for Big Data*, 15 INFO. COMM. SOC'Y 662 (2012) (az objektivitásba vetett téves hitet a nagy adatelemzésekre való támaszkodás

egyik buktatójaként azonosítja).

ezáltal újrafogalmazva az algoritmikus modell döntéseit.

⁸⁰Amikor

az

algoritmikus kimenet nem értelmezhető - amikor a döntési útvonal nincs megmagyarázva -, a kormánytisztviselők nem tudhatják, hogy az előttük álló tényezők összhangban vannak-e az algoritmikus ajánlást eredményező tényezőkkel, és ha igen, hogyan. Tegyük fel, hogy az algoritmus által értékelt bűnügyi vádlott megvakult vagy gyermeket szült. Indokolhatják-e ezek a tények az algoritmus kockázati előrejelzésétől való eltérést, vagy figyelembe veszik őket? Ha az algoritmus átláthatatlan, a kormánytisztviselő nem tudja, hogyan integrálhatja az algoritmus érvelését a sajátjával, és vagy figyelmen kívül kell hagynia, vagy vakon követnie kell azt. Így, ahogy Christopher Church és Amanda Fairchild mondta: "Az algoritmus előrejelzése mögötti érvelés kritikusan fontos. Az algoritmusnak nemcsak a magas kockázatú eseteket kell pontosan azonosítania .

. . de képesnek kell lennie arra is, hogy összefüggő indoklást adjon arra vonatkozóan, hogy miért jelölnek meg bizonyos eseteket".⁸¹

Idővel az algoritmusok iránti tisztelet gyengítheti a kormánytisztviselők döntéshozatali képességét, valamint elkötelezettségük és ügynöki képességük érzését. Az emberek automatizálással történő "leképzése" széles körben vizsgált jelenséggé vált,⁸² és ez kétségtelenül át fog terjedni a közigazgatásra is. Az etikusok azt is vizsgálták, hogy a számítógépes rendszerek hogyan ássák alá az ember saját erkölcsi cselekvőképességének érzését. Amikor "az emberi felhasználók nagyrészt mechanikus szerepekbe kerülnek, akár mentálisan, akár fizikailag", és "kevésbé értik meg cselekedeteik nagyobb célját vagy értelmét".

. . az emberi méltóság csorbul, és az egyének úgy gondolhatják, hogy nagyrészt nem tudnak elszámolni a számítógéphasználatuk következményeiért. "⁸³ Ugyanez mondható el konkrétan a prediktív algoritmusokról és az azokat használó kormányzati tisztviselőkről is. Például azok a rendőrök, akiket algoritmusok utasítanak arra, hogy pontosan hol és hogyan járőrözzenek, elveszíthetik saját tudatosságukat a bűnözési kockázatokról, és képtelenek lesznek felelősségteljesen eltérni az algoritmus utasításaitól. ⁸⁴

⁸⁰ Ananny, *Supra* 6. lábjegyzet, 103. o.; lásd még Barocas & Selbst, *Supra* 7. lábjegyzet.

⁸¹ Christopher E. Church & Amanda J. Fairchild, *In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics*, 68 JUV. & FAM. CT. J. 67, 78 (2017).

⁸² *Lásd pl.*, NICHOLAS CARR, THE GLASS CAGE: HOGYAN VANNAK A SZÁMÍTÓGÉPEINK CHANGING US 106-112 (2014).

⁸³ Batya Friedman & Peter H. Kahn, Jr., *Human Agency and Responsible Computing: Implications for Computer System Design*, 17 J. SYSTEMS SOFTWARE 9 (1992); lásd Helen Nissenbaum, *Accountability in a Computerized Society*, 2 SCI. & ENGINEERING ETHICS 25 (1996).

⁸⁴ A rendőrségnek a számítógépes kockázatelemzés által előidézett "képtelenségéről" lásd RICHARD V. ERICSON & KEVIN D. HAGGERTY, POLICING THE RISK SOCIETY 447 (1997). Az algoritmusok próbaidős

döntéshozók általi használatából eredő "képzettségvesztés, a szakmai mérlegelés eróziója és ... a szakmaitlanítás folyamata" veszélyeiről lásd: Gwen Robinson, *Implementing OASys*:

153, 162 (2016) (az intelligens városi vízgazdálkodással összefüggésben azonosítja a "komplex információk érzékelésére, elemzésére és az azokra való reagálásra szolgáló automatizált folyamatok veszélyeit . . .

az algoritmikus ítélkezés által érintett egyének megfelelő eljárási jogairól, hogy megismerjék az okokat, amiért a gép "ezt mondta" ⁸⁸.

Az biztos, hogy mindig is fennállt a kormányzati ügynökök nem hatékony vagy elfogult döntéshozatalának kockázata. Nem tudhatjuk, hogy a tárgyalás előtti szökés kockázatáról döntő bíró megfelelően mérlegeli-e a kockázati tényezőket. Miért kellene felfedni előttünk az automatizált érvelést, ha az emberi érvelés nem volt ilyen? Először is, a nagyobb átláthatóság jobb, mint a kevesebb, amikor a kormányzati erő alkalmazásáról, a polgárok szabadságának megvonásáról vagy a közforrások elosztásáról szóló döntésekről van szó. A jóslatok algoritmusban történő formalizálása lehetőséget adhat arra, hogy megismerjük az emberi tudatalattiban megismerhetetlen döntési folyamatokat, és teszteljük, hogy ezek a jóslatok pontatlanok vagy igazságtalanok-e. A jóslatok nem hibásak vagy igazságtalanok. Megfelelően szemlélve ez az algoritmusok ígéretének része.

Másodszor, a prediktív algoritmusok a tisztességtelenség és a hiba új kockázatát hordozzák magukban, még akkor is, ha javítják az általános döntéshozatalt. Ennek oka, hogy ahol probléma van, ott rosszabb és tartósabb lesz. A prediktív algoritmusokat jellemzően arra használják, hogy egy kormányzati egységen belül irányítsák a döntéseket - például egy joghatóság összes büntetőbírája, sőt, számos helyi és állami önkormányzat között. ⁸⁹ Ez az a probléma, amelyet Cathy O'Neil az algoritmusok skálázhatóságaként azonosít. ⁹⁰ Ezen algoritmikus folyamatok skálázhatósága, és ezáltal a döntések egységes és átfogó befolyásolhatósága felnagyítja az általuk megtestesített hibákat vagy elfogultságokat, és növeli az átláthatóvá tételük fontosságát.

A kihívás az átláthatóság olyan fokának és formájának meghatározása, amely a nyilvánosság számára értelmes, a fejlesztők és a kormányok számára pedig praktikus. Az alábbi V. rész meghatározza, hogy milyen információkat kell nyilvánosságra hozni a nyilvánosan alkalmazott algoritmusokról. Itt az átláthatóság több rétegét fejtjük ki, és kiemeljük az átláthatóság központi szerepét a nyílt iratokra vonatkozó törvényekben.

Az algoritmikus folyamatok különböző módon lehetnek átláthatatlanok és ellenállóak a megismeréssel szemben. Frank Pasquale nyomán a kommentátorok az algoritmikus folyamatok elrejtésére összpontosítottak.

hacsak az e rendszerek elfogadására irányuló közigazgatási eljárásokat nem teszik hozzáférhetővé, átláthatóvá, valamint folyamatos és érdemi felülvizsgálat tárgyává").

⁸⁸ *Lásd pl.*, Amended Summary Judgment Op., Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist., Civil Action No. H-14-1189 (S.D. Tex. May 4, 2017) (engedélyezve a tanárok megfelelő eljárásra vonatkozó keresetét az állami iskolai körzet ellen, amiért az olyan tanárértékelési algoritmust vezetett be, amely nem engedi a vizsgálatot).

⁸⁹ Az Arnold Alapítvány közbiztonsági értékelését például három állam - Arizona, Kentucky és New Jersey - és

körülbelül harmincöt másik joghatóság használja. *Lásd: Public Safety Assessment*, LAURA & JOHNARNOLDFOUND.,
<http://www.arnoldfoundation.org/initiative/criminal-justice/crime-prevention/public-safety-assessment/> [<http://perma.cc/9XLV-N65H>].

^{90Vö.} O'NEIL, 6. lábjegyzet, 29-31. o.

formulák, bemeneti adatok és eljárási szabályok egy "fekete dobozban".⁹¹ Az algoritmus formális összetevőinek nyilvánosságra hozatala felfedheti az algoritmusban rejlő hibákat - felfedheti például, hogy az algoritmus néha az elvileg korlátozott tartományon kívüli eredményeket hoz, vagy fordítva, hogy az eredményei mindig korlátozottabbak lesznek, mint az elvileg meghatározott tartomány.

Az algoritmusoknak képesnek kell lenniük a matematikai és logikai jelölések és a természetes nyelv kombinációjával történő formális közzétételre.⁹² Ahhoz, hogy számítógéppel lehessen őket megvalósítani, programozási nyelven kell kódolni őket. A számítógépes kód nyilvánosságra hozatala akkor lehet megfelelő, ha fennáll a veszélye annak, hogy a számítógépes megvalósítás hibás lehet.⁹³ A számítógépes program azonban általában lényegesen nehezebben olvasható és érthető az emberek számára, mint a matematikai vagy logikai jelölés vagy a természetes nyelv, és ezért a számítógépes kód közzététele a könnyebb értelmezési eszközökkel szemben kevésbé hasznos alternatíva lehet.

Még ha az algoritmus formális összetevői feltárulnak is, az algoritmikus folyamat még mindig nem biztos, hogy kiértékelhető. Az algoritmus érvényességének állítása nem korlátozódik az algoritmus saját szabályainak való megfelelésre. Az állítás a mögöttes adathalmazban lévő tények és eredmények közötti összefüggéseken alapul. Ezt az állítást nem tudjuk megvizsgálni anélkül, hogy valamit ne tudnánk a képzési adatokról. Hogyan választották ki az adatokat, miért választottak ki bizonyos működési szabályokat, míg másokat elvetettek, és milyen lépéseket tettek e választások érvényesítésére?⁹⁴ Az alapul szolgáló adatokhoz való hozzáférés vagy legalább azok leírása segítené megérteni, hogy valójában mennyire erősek az állítólagos összefüggések,

⁹¹ Lásd PASQUALE, *Supra* note 6, at 1-18; lásd még O'NEIL, *Supra* note 6, at 28-31 (szemléltetve, hogy a prediktív algoritmusok alkotói milyen módon rejtik el mechanikájukat); Selbst & Barocas, *Supra* note 85, at 9 (felvázolva a prediktív algoritmusok "Mi-hogyan-miért" magyarázó modelljét, amelynek a prediktív algoritmusok nem felelnek meg).

⁹² Az ilyen közzététel egyik példáját lásd MARIE VANNOSTRAND & GINA KEEBLER, PRETRIAL RISK ASSESSMENT IN THE FEDERAL COURT 48 (2009. április 14.), <http://tiny.cc/r3qrmy> [<http://perma.cc/YKZ5-6FV6>] (matematikai jelöléssel közlésezi a tárgyaláson való meg nem jelenés kockázatának és az előzetes letartóztatásból való szabadlábra helyezéskor a bűncselekmények kockázatának előrejelzésére szolgáló képletet, valamint természetes nyelven az egyes alkalmazott tényezők leírását).

⁹³ Lásd Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1268-69 (2008) (megállapítva, hogy 2004 szeptembere és 2007 áprilisa között a kódírók több mint kilencszáz hibás szabályt építettek be a coloradói állami ellátási rendszerbe, ami több százezer hibás jogosultsági megállapítást és ellátási számítást eredményezett a Medicaid és az élelmiszerjegyek esetében ebben az időszakban); *id.* at 1270 (megállapítva, hogy a kódírók hibásan fordították le a szabályzatot a kaliforniai automatizált állami ellátási rendszerbe, ami túl- és alulfizetéseket és az állami ellátások helytelen megszüntetését okozta).

⁹⁴ A Fifth Circuit számára első ízben felmerült ügyben a houstoni tanárok "az iskolai körzet által használt, magánkézben lévő értékelési algoritmusba épített egyenleteket, számítógépes forráskódokat, döntési szabályokat és

feltételezéseket" kérték. Módosított összefoglaló ítélet, Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist., Civil Action No. H-14-1189, at *17 (S.D. Tex. May 4, 2017).

milyen volt a minta mérete, és egyéb, a statisztikai érvényességet befolyásoló kérdések.

A most jellemzően homályba vesző egyéb információ típusok közé tartozik az algoritmus kifejlesztésének közcélja, az adatok tulajdonjogát és hozzáférését szabályozó szerződési feltételek, valamint a validálási és nyomonkövetési tervek. Néha ezek az információk kifejezetten vagy hallgatólagosan kitérnek az algoritmus által megkövetelt politikai kompromisszumokra is. Mindez fontos annak értékeléséhez, hogy az algoritmus hatékony, tisztességes és egyébként politikailag elfogadható-e.

Elismerjük, hogy még akkor is lehetetlen lenne megérteni egy algoritmikus folyamat eredményeit, ha a fentiekben meghatározott összes információ nyilvánosságra kerülne. Ennek oka, hogy az átláthatóság nem feltétlenül teszi "értelmezhetővé" az algoritmust.⁹⁵ Ha egy algoritmus több száz súlyozatlan bemenetet használ egy összetett döntési fában, amelyben egyetlen bemenet több ponton is megjelenhet, nem feltétlenül tudjuk kitalálni, hogy egy adott esetben mely bemenetek voltak döntőek.⁹⁶ Ez különösen megnehezíti annak megértését, hogy az algoritmus korrelál-e a méltányossági érzékünkkel, és megnehezíti a kormánytisztviselők számára, hogy az algoritmus kimenetét saját helyzetérzékelésük fényében értékeljék, ami azt követeli meg tőlük, hogy vagy figyelmen kívül hagyják ezt a kimenetet, vagy figyelmen kívül hagyják saját ítélőképességüket, és végül talán elveszítik ezt az ítélőképességet.

Végül, az algoritmikus folyamatok dinamikusak lehetnek, szabályaik folyamatosan változnak, hogy az adatok új mintáihoz igazodjanak. Ennek eredményeképpen az 1-es időpontban nyilvánosságra hozott kód és adatkészletek - még ha elszigetelten "értelmezhetőek" is - kevésbé hasonlíthatnak a 2-es időpontban végrehajtott folyamatra. A dinamikus algoritmusok, ahogy Rob Kitchin mondja, "ontogenetikus természetűek", "szerkesztik, felülvizsgálják, törlik és újraindítják őket".⁹⁷ Ez a dinamizmus akadályokat gördít az átláthatóság elé, mivel az algoritmusok

⁹⁵ Természetesen az emberek néha képesek magyarázatot adni nekünk az érvelési folyamataikról. Nincs azonban garanciánk arra, hogy ezek a magyarázatok valóban megfelelnek annak, hogyan jutottak döntéseikre. *Lásd* Zachary C. Lipton, *The Mythos of Model Interpretability* (2016. június 16.) (kiadatlan kézirat) (kézirat 98. o.), <http://arxiv.org/abs/1606.03490v2> [<http://perma.cc/H3E5-TVA2>] (ezt megjegyzi, és egy modell utólagos magyarázhatóságát "utólagos értelmezhetőségként" definiálja).

⁹⁶ Ha egy modell úgy próbálja megjósolni a feltételesen szabadlábra helyezett visszaesését, hogy súlyokat rendel néhány tényezőhöz, például a korábbi erőszakos bűncselekményekhez és az életkorhoz, akkor megérthetjük és megmagyarázhatjuk, hogy mit csinál. Tegyük fel azonban, hogy a modell több mint ezer tényezőt használ a feltételesen szabadlábra helyezett visszaesésének előrejelzésére, amelyek közül néhánynak látszólag nincs intuitív oksági kapcsolata a visszaeséssel (például a magasság). Tegyük fel továbbá, hogy a modell egy összetett döntési fát használ, amely sok tényezőt többszörösen tartalmaz. Nehéz lesz megérteni, hogy az egyes tényezők milyen befolyással bírnak egy adott esetben vagy az esetek egy sorában, vagy

megfogalmazni a modell ok-okozati elméletét (ha van ilyen).

⁹⁷Kitchin, *Thinking Critically*, supra note 6, 18. o., o.

maguk is nehezebben érthetővé válnak. ⁹⁸ Ezzel a fajta dinamizmussal nagyrészt azért nem foglalkozunk, mert az általunk vizsgált helyi és állami kormányzati szereplők még nem használják ezeket a folyamatosan kiigazított előrejelző algoritmusokat.

Ahogy az átláthatóság nem feltétlenül támogatja az értelmezhetőséget, úgy az átláthatóság nem azonos az elszámoltathatósággal. ⁹⁹ Ez csupán egy eszköz. Egy algoritmikus folyamat akkor válik elszámoltathatóvá, ha az érdekeltek az átláthatóság birtokában képesek beavatkozni, hogy változtatásokat érjenek el az algoritmusban, illetve annak használatában vagy végrehajtásában. ¹⁰⁰ A közszférában ez azt jelenti, hogy a kormányzatot a szavazóközönség ténylegesen elszámoltathatóvá teszi az általa alkalmazott algoritmusokért. Ez az elszámoltathatóság nem *tökéletes* átláthatóságot igényel - az algoritmus működési szabályainak, valamint létrehozásának és érvényesítésének teljes körű ismeretét -, hanem az *érdemi* átláthatóság alacsonyabb követelményét - az algoritmus teljesítményének jóváhagyásához vagy elutasításához elegendő ismeretet. A szükséges input biztosításához elegendő lehet a mögöttes számítógépes kódot nem tartalmazó nyilvántartás. Természetesen az elszámoltathatóság a gyakorlatban olyan közoktatást és politikai folyamatokat igényelhet, amelyek meghaladják azt, amivel itt foglalkozhatunk. De az értelmes átláthatóság lesz a szükséges első lépés.

III. A NYÍLT IRATOKRÓL SZÓLÓ TÖRVÉNY PROJEKTJE: AZ ALGORITMUSOK DOKUMENTÁCIÓJÁNAK BESZERZÉSE

A nyílt adatszolgáltatási gyakorlatok valószínűleg a legjobb módja annak, hogy átláthatóvá tegyük a kormányzatban használt algoritmusok egyes aspektusait. ¹⁰¹ Azaz a kormányoknak már a kezdetektől fogva önkéntesen fel kellene tárniuk az algoritmusok releváns struktúráit, logikáját és politikáit. A szövetségi FOIA 2016-os módosításai kodifikálták a kormányzat részéről az átláthatóság "push" módszerének előnyben részesítését, ami csökkenti a nyilvánosság "pull" kormányzati nyilvántartások iránti kérelmeinek terhelését. ¹⁰² ^A kormányok mégsem adtak ki információkat az általuk használt algoritmusokról. Nagy szakadék tátong az algoritmusok fontossága között

⁹⁸ Lásd EXEC. OFFICE OF THE PRESIDENT, *supra* note 24 (megjegyezve, hogy a gépi tanulási módszerek fejlődésével "egyre nehezebbé válhat a gépek által e folyamat során hozott döntések magyarázata vagy elszámoltatása, hacsak nem építenek be olyan mechanizmusokat a tervezésükbe, amelyek biztosítják az elszámoltathatóságot").

⁹⁹ Kroll et al., 26. lábjegyzet, 657-60. o.; lásd még Ananny, 6. lábjegyzet, 109. o.

¹⁰⁰ Lásd Kroll et al., 26. lábjegyzet, 657-60. o.; Selbst & Barocas, 85. lábjegyzet, 15. o.

¹⁰¹ Lásd: JOSHUA TAUBERERER, A NYILVÁNOS KORMÁNYZATI ADATOK ELVEI ÉS PRAKTIKÁJAI 10 (2d ed. 2014).

¹⁰² A FOIA Improvement Act of 2016 módosította a 44 U.S.C. § 3102-t (amely mostantól előírja, hogy az ügynökségeknek "eljárásokat kell létrehozniuk a

nyilvánosság számára általános érdekű vagy a nyilvánosság számára hasznos, nyilvánosságra hozatalra alkalmas nyilvántartások azonosítására, valamint az ilyen nyilvántartások nyilvánosan hozzáférhető elektronikus formátumban történő közzétételére").

az irányítási folyamatok és az algoritmusokhoz való nyilvános hozzáférés.¹⁰³ Az átláthatóság előmozdításának hiányában a nyílt iratok iránti kérelmek jelentik a következő legjobb módot a hiányosságok megszüntetésére és a nyilvánosság azon érdekének érvényesítésére, hogy megértse az algoritmusokat, amelyeket rájuk és polgártársaikra alkalmaznak. Megvizsgáltuk, hogy a kormányok mennyire reagálnak a prediktív algoritmusokra és a kapcsolódó adatelemzésre vonatkozó ilyen információkérésekre. Először a projekttervünket tárgyaljuk, bemutatjuk a nyílt iratokra vonatkozó törvényeket, és elmagyarázzuk, hogyan választottuk ki, hogy mely algoritmusokat és közhivatalokat célozzuk meg, és hogyan fogalmaztuk meg a nyílt iratokra vonatkozó kéréseinket. Ezután az eredményeinket tárgyaljuk.

A. *Projekttervezés és Megvalósítás*

1. Nyílt iratok Törvények

Az algoritmikus folyamatokra vonatkozó információk keresése során támaszkodtunk az információs szabadságról szóló állami törvényekre, amelyeknek mind ugyanaz a központi célja: a kormányzat működésének feltárása az emberek előtt.¹⁰⁴

Az információs szabadságról szóló törvények fő célja az elszámoltatható kormányzat. Az eredeti szövetségi FOIA 1964-es aláírásakor Johnson elnök "mély büszkeségét fejezte ki, hogy az Egyesült Államok olyan nyitott társadalom, amelyben az emberek megismeréshez való jogát nagyra becsülik és őrzik".¹⁰⁵ Azzal, hogy egy évtizeddel később a FOIA módosításai határidőket szabtak az ügynökségek válaszára, a Kongresszus ünnepelte "[a] kormányzatot. . mint a legjobb biztosítékot arra, hogy a kormányzatot a köz érdekében végzik".¹⁰⁶ És amikor a Kongresszus 1996-ban megerősítette, hogy a FOIA központi átláthatósági mandátuma az elektronikus iratokra is vonatkozik, a szenátusi bizottsági jelentés kifejtette, hogy

¹⁰³ Lásd Nicholas Diakopoulos, *Algoritmikus elszámoltathatósági jelentés: On the Investigation of Black Boxes*, TOW CTR. FOR DIGITAL JOURNALISM 2 (2014), <http://academiccommons.columbia.edu/catalog/ac:2ngf1vhhn4> [<http://perma.cc/7N2Q-7XHA>] ("Amit közönségként általában hiányolunk, az a tisztánlátás arról, hogy az algoritmusok hogyan gyakorolják felettünk a hatalmukat.").

¹⁰⁴ Lásd *pl.*, New York Freedom of Information Law, N.Y. PUB. OFF. L. § 86(4) (McKinney 2003) ("Az emberek joga, hogy megismerjék a kormányzati döntéshozatal folyamatát, és betekintheessenek a döntésekhez vezető dokumentumokba és statisztikákba, alapvető társadalmunkban. Az ilyen információkhoz való hozzáférést nem szabad megghiúsítani azzal, hogy a titoktartás vagy a bizalmas kezelés köpenyében burkolják azokat. A jogalkotó ezért kijelenti, hogy a kormányzat a nyilvánosság ügye, és hogy a nyilvánosságnak egyénileg és kollektíven, valamint a szabad sajtó által képviselve hozzáférést kell biztosítani a kormányzat irataihoz.").

¹⁰⁵ Lyndon B. Johnson elnök nyilatkozata a Pub. L. 89-487 on July 4, 1966, in ATTORNEY GENERAL'S MEMORANDUM ON THE PUBLIC INFORMATION SECTION OF THE ADMINISTRATIVE PROCEDURE ACT (1967),

<http://www.justice.gov/oip/67agmemo.htm> [<http://perma.cc/XD4G-FLNR>].

¹⁰⁶ S. REP. NO. 93-854, 1. o. (1974).

a kormányzati átláthatóság "összhangban van demokratikus kormányzati formákkal, mivel elősegíti a polgárok azon érdekét, hogy tudják, mit tesz a kormányuk. "¹⁰⁷ A bíróságok következetesen úgy ítélték meg, hogy a FOIA "a teljes ügynökségi nyilvánosság általános filozófiáját "¹⁰⁸ testesíti meg.

Ugyanezen átláthatósági elvek alapján mind az ötven állam és a Columbia körzet nyílt iratokra vonatkozó jogszabályai biztosítják az egyének számára a kormányzati iratokhoz való hozzáférés jogát, különböző mentességek mellett. Ezek közé tartoznak az egyéni magánélet, a bűnügyi nyomozati anyagok és az ügynökségi tanácskozási folyamatok védelmét szolgáló mentességek. ¹⁰⁹ Szinte valamennyi törvény mentesít az üzleti titkok alól is, amelyeket az alábbiakban tárgyalunk.

A FOIA csak az "ügynökségi iratokra" vonatkozik - ez a fogalom a jogszabályban nincs meghatározva. ¹¹⁰ A Legfelsőbb Bíróság értelmezése szerint az "ügynökségi iratok" közé tartozik minden olyan irat, amelyet egy ügynökség 1) létrehoz vagy megszerez, és

2) a FOIA-kérelem benyújtásának időpontjában az ellenőrzésük alatt áll. ¹¹¹ Bár az állami törvények jellemzően a "kormányzati iratok" kifejezést használják, a lefedettség hasonló.¹¹²

A FOIA hatálya kiterjed a digitális nyilvántartásokra, beleértve a szoftvereket és adatbázisokat is. ¹¹³ Egyes állami törvények kifejezetten tartalmazzák a szoftvereket, mint nyilvános iratokat. ¹¹⁴ New Jersey nyílt iratokra vonatkozó törvénye szerint például a "kormányzati irat" magában foglal minden "adatfeldolgozott vagy képfeldolgozott dokumentumot" és "elektronikusan tárolt vagy tárolt információt", ha azt egy állami tisztviselő vagy alkalmazott "a saját vagy a saját munkája során" készítette, tartotta fenn vagy kapta meg

¹⁰⁷ S. REP. NO. 104-272, 5. o. (1996). *Lásd általában* MICHAEL SCHUDSON, RISE OF THE RIGHT TO KNOW (2015) (a FOIA történetének krónikája).

¹⁰⁸ Dept of the Air Force v. Rose, 425 U.S. 352, 360 (1976).

¹⁰⁹ *Lásd pl.*, 5 U.S.C. §§ 552(b)(1)-(9) (2012). A FOIA tekintetében a Legfelsőbb Bíróság "ismételten kijelentette, hogy ezeket a nyilvánosságra hozatal alóli mentességeket szűken kell értelmezni, oly módon, hogy a lehető legnagyobb hozzáférést biztosítsák". Vaughn v. Rosen, 484 F.2d 820, 823 (D.C. Cir. 1973).

¹¹⁰ 5 U.S.C. § 552(a)(4)(B).

¹¹¹ *Lásd* Forsham kontra Harris, 445 U.S. 169, 182 (1980).

¹¹² *Lásd pl.*, ARIZ. REV. STAT. § 41-151.18; CAL. GOV'T CODE § 6252.

¹¹³ *Lásd például az* 5 U.S.C.C. §(f)(2)(A) pontot (amely kimondja, hogy "'rekord" és bármely más, ebben a szakaszban az információra hivatkozva használt kifejezés magában foglal minden olyan információt, amely az e szakasz követelményei alá tartozó ügynökségi rekordnak minősül, ha azt egy ügynökség bármilyen formátumban, beleértve az elektronikus formátumot is, tárolja"). Egyes állami nyílt nyilvántartások kizárnak bizonyos típusú szoftvereket. *Lásd pl.* CAL GOV'T CODE §§ 6254.9(a)-(b) (kivéve az "állami vagy helyi ügynökség által kifejlesztett számítógépes szoftvereket, beleértve a számítógépes térképezési rendszereket, számítógépes programokat és számítógépes grafikai rendszereket").

¹¹⁴ *Lásd általában* Cristina Abello, Reporters Committee for Freedom of the Press, *Access to Electronic Communications*, REPORTERS COMMITTEE FOR FREEDOM PRESS (2009), <http://www.rcfp.org/rcfp/orders/docs/ELECCOMM.pdf> [<http://perma.cc/QBM4-UCZ8>]; Andrea G. Nadel, Annotation, *What Are "Records" of Agency Which Must Be Made Available under State Freedom of*

Information Act, 27 A.L.R.4th 680 (Supp. 2014); Marjorie A. Shields, Annotation, *Disclosure of Electronic Data under State Public Records and Freedom of Information Acts*, 54 A.L.R.6th 653 (Supp. 2014).

hivatalos ügyek. ¹¹⁵ Más állami törvények kifejezetten *kizárják* a szoftvereket a nyilvános nyilvántartások köréből. ¹¹⁶ Mások nem foglalkoztak a kérdéssel. ¹¹⁷ Az algoritmikus átláthatóság szempontjából ritkán lesz a legfontosabb kérdés, hogy a nyilvánosság egy tagja jogosan ragaszkodhat-e a szoftverek bemutatásához az állami nyílt iratokról szóló törvény alapján, mivel az érdemi átláthatóság más típusú nyilvántartásokon keresztül is megvalósítható.

A közzététel legsúlyosabb akadálya a nyilvántartás tulajdonjoga. A legtöbb nyílt iratokra vonatkozó törvény csak a kormányzati iratokra vonatkozik. Amennyiben a magánvállalkozók kizárólagos ellenőrzést gyakorolnak a nyilvántartások felett, ezek a nyilvántartások kívül eshetnek az átláthatósági törvények hatókörén. A FOIA előírja, hogy ha a nyilvántartásokat "egy ügynökség számára egy kormányzati szerződés alapján működő szervezet iratkezelési céllal vezeti", akkor ezek a nyilvántartások "ügynökségi nyilvántartások" maradnak, amelyek a FOIA nyilvánosságra hozatalának hatálya alá tartoznak. Ez vonatkozik azokra a helyzetekre, amikor egy ügynökség magánszolgáltatóval köt szerződést a nyilvántartások, például a rendőrségi kamerafelvételek megőrzésére. ¹¹⁸ Ezek a nyilvántartások akkor is ügynökségi nyilvántartások, ha magán szervereken tárolódnak. Ha azonban egy magánfél *saját* célra készít nyilvántartásokat, és azokat soha nem helyezi letétbe egy ügynökségnél, az ilyen nyilvántartások valószínűleg nem tartoznak a FOIA és az állami nyílt iratokról szóló törvények hatálya alá. ¹¹⁹ Algoritmusok esetében ezek közé tartozhatnak a képzési adatok és az algoritmus létrehozásának és validálásának dokumentációja. Amint azt alább tárgyaljuk, az ezekhez a nyilvántartásokhoz való nyilvános hozzáférés attól függ, hogy a kormányhivatalok mennyire ragaszkodnak az adatok tulajdonjogához és/vagy a nyilvántartások birtoklásához.

2. Algoritmusok, ügynökségek és a kérelmek megfogalmazása.

-
- ¹¹⁵ N.J. STAT. ANN. § 47:1A-1.1 (West 2015); *cf.* FLA. STAT. § 119.011(12) (2017) (a "közhiteles nyilvántartás" fogalmának meghatározása, amely magában foglalja az "adatfeldolgozó szoftvereket"); 5 ILL. COMP. STAT. 140/2(c) (2016) (a "nyilvántartások" fogalmának meghatározása, amely magában foglalja az "elektronikus adatfeldolgozási nyilvántartásokat"); N.Y. PUB. OFF. L. § 86 (4) (McKinney 2003) (a "rekord" meghatározása: "bármely információ, amelyet egy ügynökség vagy az állami törvényhozás tart, tart, iktat, készít vagy reprodukál, vagy amelyet egy ügynökség vagy az állami törvényhozás vagy az állami törvényhozás számára bármilyen fizikai formában tárolnak, tárolnak, iktatnak, előállítanak vagy reprodukálnak").
- ¹¹⁶ *Lásd pl.*, ALASKA STAT. § 40.25.220(3) (2010) (amely előírja, hogy a "közhiteles iratok" fogalommeghatározása nem foglalja magában a "védett szoftverprogramokat").
- ¹¹⁷ *Lásd pl.* ALA. CODE § 36-12-40 (2004) (amely előírja, hogy "[e]gy állampolgárnak joga van betekinteni és másolatot készíteni az állam bármely nyilvános írásművébe, kivéve, ha a törvény kifejezetten másként rendelkezik", anélkül, hogy foglalkozna az "írásmű" fogalmával).

¹¹⁸ *Lásd* Alexandra Mateescu, et al., Police Body-Worn Cameras 9 (Data & Soc'y

ResearchInstitute, WorkingPaper, Feb.2,2015), <http://www.datasociety.net/pubs/dcr/PoliceBodyWornCameras.pdf> [<http://perma.cc/TB9A-K63G>] (a rendőrségi testkamerás felvételek harmadik fél felhőszervereken történő rendőrségi tárolásáról).

¹¹⁹ ^A magánszervezetek birtokában lévő nyilvántartások állami kezelése összetett és változatos. Az áttekintést lásd Alexa Capeloto, *Transparency on Trial: A legal review of public information access in face of privatization*, 13 CONN. PUB. INT'L L.J. 19, 27 (2013).

Bár a nyílt iratokkal kapcsolatos projektünk "empirikus" volt abban az értelemben, hogy kéréseket küldtünk ki a világba, hogy lássuk, hogyan reagálnak a kormányok, nem volt, és nem is lehetett volna mennyiségi vagy statisztikai jellegű. Nincs központi nyilvántartás a kormányok által használt algoritmusokról, és az algoritmusok természetesen nem láthatóak úgy, mint mondjuk a felhőkarcolók vagy a hidak. Így nem tudjuk, hány algoritmus van jelenleg használatban, ki fejlesztette ki őket, vagy mely kormányok használják őket. Ezen ismeretek nélkül nem tudunk olyan módszert kidolgozni az algoritmusok használatának mintavételezésére, amely lehetővé tenné, hogy általánosíthassunk az eredményeinkből.

Amit mi tettünk, az sokkal kevésbé hivatalos. Felmértük a nyilvános információkat, hogy azonosítsuk a prediktív algoritmusok önkormányzati használatát. Ezután kiválasztottunk hat programot, hogy különböző témakörök (rendészet, büntető igazságszolgáltatás, gyermekjólét és oktatás), különböző fejlesztők (alapítványok, magánvállalatok és kormányzati szervek) és különböző joghatóságok (negyvenkét különböző ügynökség huszonhárom államban) keverékét kapjuk. A hat program a következő: Közbiztonsági értékelés; Eckerd Rapid Safety Feedback; Allegheny Family Screening Tool; PredPol; HunchLab; és New York City Value-Added Measures. Nyílt iratok iránti kérelmet fogalmaztunk meg, amely eleve nagyon általános és átfogó volt, és megpróbált minden olyan feljegyzésre kiterjedni, amely a szóban forgó algoritmushoz kapcsolódik.¹²⁰ Amikor egy ügynökség azt válaszolta, hogy a kérés terjedelme késedelmet okoz, módosított kérést küldtünk.¹²¹ Az idő és az erőforrások szűkössége miatt nem támadtuk meg a végleges elutasításokat vagy a válaszadás tartós elmaradását a bíróságon, és nem fizettünk jelentős díjakat sem, amikor egy ügynökség követelte, hogy válaszoljanak a kérésünkre.¹²² Az általunk tapasztalt akadályok jelentősen korlátozzák az algoritmusokkal kapcsolatos információkhoz való nyilvános hozzáférést, még akkor is, ha némelyikük több idővel és pénzzel leküzdhető lenne. Néhány esetben közvetlen kommunikációt is folytathattunk az algoritmusfejlesztőkkel, hogy megpróbáljunk további információkhoz jutni.

¹²⁰ A szabványos kezdeti kérésünk nyelvezetének egyik példáját lásd a HunchLab használatával kapcsolatos, a Lincoln, Nebraska állambeli rendőrséghez intézett kérésünkben, Michael Morisy levele a Lincoln Police Dep't-nek (2016. november 15.), <http://www.muckrock.com/foi/lincoln-4033/lincoln-police-department-hunchlab-documents-30110/> [<http://perma.cc/Z8LY-RT32>].

¹²¹ Lásd például a Cocoa (Florida) rendőrkapitánysággal folytatott levelezésünket: Michael Morisy levele a Cocoa Police Dep't-nek (2016. november 16.), <http://www.muckrock.com/foi/cocoa-10491/cocoa-police-department-predpol-documents-30104/> [<http://perma.cc/3KEA-HJF3>].

¹²² *Lásd pl.*, Letter from Okla. Dep't of Human Servs. to Muck Rock [sic] (2016. november 3.), <http://www.muckrock.com/foi/oklahoma-248/oklahoma-department-of-human-services-eckerd-rapid-safety-feedback-28151/#file->

108045 [<http://perma.cc/Q88A-8JVQ>] (2.472 dollár kifizetését kérte, mielőtt megkezdenék a munkálatokat a kérésünkre való válaszadással kapcsolatban).

B. Eredmények

Általában véve nagy eltéréseket találtunk abban, hogy a joghatóságok válaszoltak-e a megkereséseinkre; hogy hivatkoztak-e a nyílt iratok alóli mentességre; és ha nem, akkor milyen információkat adtak meg. A joghatóságok közül azonban csak egy, Allegheny megye volt képes mind az általa használt tényleges előrejelző algoritmusokat (beleértve atényezők teljes listáját és az egyes tényezők súlyát), mind pedig a kidolgozásuk módjára vonatkozó lényeges részleteket megadni. Egyes fejlesztők is közlékenyebbek voltak, mint mások. Az Arnold Foundation, a Public Safety Assessment fejlesztője viszonylag egyszerű algoritmusait nyilvánosságra hozta, de a fejlesztési folyamatról szinte semmit sem közölt, míg a HunchLabot fejlesztő Azavea, Inc. sokkal többet közölt. Ezek az eredmények azt sugallják, hogy az átláthatóság a joghatóságok és szállítók döntése - egy olyan döntés, amelynek kevésbé van köze a megváltoztathatatlan üzleti titkokhoz vagy a titoktartási aggályokhoz, mint inkább a nyilvánosságra hozatal kultúrájához. A továbbiakban részletezzük a hat algoritmus mindegyikére vonatkozó válaszokat.

1. Közbiztonsági értékelés-előzetes tárgyalás Felszabadítás

A közbiztonsági értékelés (PSA) a Laura és John Arnold Alapítvány által kifejlesztett, a tárgyalás előtti kockázatértékelési eszköz, amelynek célja, hogy segítse a bírakat annak eldöntésében, hogy a vádlottat a tárgyalás előtt őrizetbe vegyék vagy szabadon engedjék.¹²³ E sorok írásakor harmincnyolc joghatóságban használják, beleértve Arizona, New Jersey és Kentucky államok egészét.¹²⁴ A PSA három különböző kockázatértékelési algoritmust tartalmaz, amelyek célja annak a kockázatnak a felmérése, hogy a szabadlábra helyezett vádlott nem jelenik meg a tárgyaláson; bűncselekményt követ el a szabadlábra helyezés alatt; vagy erőszakos bűncselekményt követ el a szabadlábra helyezés alatt.

A három algoritmus úgy működik, hogy a vádlott büntetett előéletére vonatkozó kilenc tény alapján pontokat oszt ki; egyes tényeket csak egy vagy két algoritmushoz használnak, míg másokat mindháromhoz. A meg nem jelenés és a bűncselekmény elkövetésének értékeléséhez a nyers pontszámokat egy hatpontos skálára alakítják át, ahol az egy a legalacsonyabb kockázatot, a hat pedig a legmagasabb kockázatot jelenti. Az erőszakos bűncselekmények elkövetésének értékelésénél a nyers pontszámot egy bináris igen/nem válaszra alakítják át; az elkövetett bűncselekmény vagy valószínűleg erőszakos, vagy valószínűleg nem erőszakos.

¹²⁵

¹²³ *Közbiztonsági értékelés*, 89. lábjegyzet.

¹²⁴ *Lásd id.*

- ¹²⁵ Mindhárom algoritmus leírása, beleértve a tényezőket, a nyers pontkiosztást és a nyers pontszámok végső eredményre történő átváltását, a *Közbiztonsági értékelés* címen érhető el: *Kockázati tényezők és képlet*, LAURA & JOHN ARNOLD FOUND. (2013-

Néhány más algoritmussal ellentétben a PSA viszonylag egyszerű - számítógép nélkül is megvalósítható a különböző tényezőkre adott pontok összeszámlálásával, majd egy átváltási képlet alkalmazásával a végső kockázatértékeléshez. A PSA algoritmusok, sok más algoritmussal ellentétben, teljes mértékben nyilvánosak. Az Arnold Alapítvány azonban nem hozta nyilvánosságra, hogyan hozta létre az algoritmusokat, vagy hogy végzett-e a bevezetés előtt vagy után validációs teszteket, és ha igen, milyen eredménnyel. Azt sem tette közzé, hogy mennyiségi vagy százalékos értelemben mit jelent az "alacsony kockázat" és a "magas kockázat": egy az tízhez vagy egy az ötszázhoz az esélye annak, hogy egy "alacsony kockázatú" vádlott nem jelenik meg? Annak az esélye, hogy egy "magas kockázatú" vádlott nem jelenik meg, kétszerese vagy ötvenszerese az alacsony kockázatú vádlotténak?

Annak megállapítására, hogy a PSA-t alkalmazó bíróságok rendelkeznek-e válaszokkal ezekre vagy hasonló kérdésekre, a PSA-programmal kapcsolatos nyílt iratokra vonatkozó kéréseket küldtünk tizenhat különböző bíróságnak. Nagyszámú kérelmet ~~küldtünk~~ a legtöbbet az általunk kiválasztott algoritmusok közül - részben azért, mert tudtuk, hogy számos nyílt iratokra vonatkozó törvény mentesíti a bíróságokat a legtöbb közzététel alól. Az öt bíróság közül, amely válaszolt, és rendelkezésre bocsátott néhány dokumentumot¹²⁶, négy - a Mesa Municipal Court, a Pima és Navajo County Court rendszer Arizonában, valamint a San Francisco Superior Court rendszer Kaliforniában - azt állította, hogy nem tudnak információt szolgáltatni a PSA-ról, mivel ezek az információk az Arnold Alapítvány tulajdonában és ellenőrzése alatt állnak.¹²⁷ Három

16), <http://www.arnoldfoundation.org/wp-content/uploads/PSA-Risk-Factors-and-Formula.pdf> [<http://perma.cc/GNF5-89PV>].

¹²⁶ Négy bíróság nem válaszolt, egy bíróság pedig elismerte a megkeresésünk kézhezvételét, de nem válaszolt tovább. Négy bíróság azt válaszolta, hogy nem rendelkezik releváns dokumentumokkal, két bíróság pedig elutasította a kérésünket, arra a következtetésre jutva, hogy a vonatkozó, nyílt iratokra vonatkozó törvények nem kötelezik őket az általunk kért anyagok rendelkezésre bocsátására. A New Jersey-i Legfelsőbb Bíróság például azzal utasította el kérésünket, hogy szabályai szerint nem kell közzétenni "az előzetes döntéshozatali szolgáltatási programmal kapcsolatos iratokat", valamint "az igazságszolgáltatás, a bíró vagy a bírósági személyzet tagja által hivatalos feladatai ellátása során készített vagy annak használatára szolgáló bármilyen formában vezetett feljegyzéseket, memorandumokat vagy egyéb munkadokumentumokat". Michelle M. Smith, a Legfelsőbb Bíróság jegyzőjének levele Michael Morisy-nak (2016. dec. 22.), <http://www.muckrock.com/foi/new-jersey-229/new-jersey-superior-court-public-safety-assessment-court-28835/#file-114392> [<http://perma.cc/M9ZZ-JU3G>]. Az Allegheny megyei bíróság szintén elutasította a kérelmünket, azzal az indokkal, hogy a pennsylvaniai törvény a bíróságokra csak a pénzügyi nyilvántartások tekintetében vonatkozik, és a mi kérelmünk nem pénzügyi nyilvántartásokra vonatkozott. E-mail from Christopher H. Connors, Chief Deputy Court Adm'r, to Michael Morisy (Oct. 6, 2016), <http://www.muckrock.com/foi/allegheny-county-306/allegheny-county-public-safety-assessment-court-28150/> [<http://perma.cc/3BYS-4RDM>].

¹²⁷ *Lásd:* Ann E. Donlan, Commc'ns Dir., Superior Court of Cal., Cty. of

San Francisco, to Michael Morisy (Dec. 16, 2016), <http://www.muckrock.com/foi/san-francisco-city-and-county-3061/san-francisco-public-safety-assessment-court-30096/#file-113829> [<http://perma.cc/A48W-RN5A>]; Letter from Ronald G. Overholt, Court Adm'r,

Ez a négy (Pima megye, Navajo megye és San Francisco) elküldte nekünk az Arnold Alapítvánnyal kötött egyetértési megállapodásuk másolatát, amely azonos megfogalmazást tartalmazott, amely megtiltotta a bíróságoknak, hogy a PSA-programmal kapcsolatos információkat nyilvánosságra hozzák.¹²⁸

Az egyetlen bírósági rendszer, amelytől az egyetértési nyilatkozaton kívül más dokumentumot is kaptunk a PSA-ról, a floridai hetedik bírósági kerület bírósága volt, a Volusia megye - az egyik általa kiszolgált megye, Volusia County Pretrial Services Programja - nevében. Ennek oka az lehet, hogy a floridai törvények megkövetelik a magánfelektől, hogy kifejezetten megjelöljék az üzleti titkokat, vagy lemondjanak a bizalmas kezeléstről - ez a törvény azon jellemzője, amelyet az Arnold Alapítvány és a hetedik bírósági kerület közötti szándéknyilatkozat is tükröz, amelyet szintén megkaptunk.¹²⁹

A Hetedik Igazságügyi Körzet által készített dokumentumok további érdekes információkkal szolgálnak. Az egyik dokumentum például a kockázati pontszámok szerint közli azon vádlottak tényleges százalékos arányát, akik nem jelennek meg, illetve akik új bűncselekményt vagy új erőszakos bűncselekményt követnek el. Az Arnold által az algoritmusok létrehozásához használt eredeti képzési adatokban a következő százalékos arányok szerepelnek a kockázati pontszám szerint: 1 (10%), 2 (15%), 3 (20%), 4 (31%), 5 (35%) és 6 (40%).¹³⁰ Így a legmagasabb kockázati pontszámot úgy határozták meg, hogy a

Ariz. Superior Court, Pima Cty., to Michael Morisy (Jan 18, 2017), <http://www.muckrock.com/foi/pima-county-183/pima-county-superior-court-public-safety-assessment-30130/#file-116730> [<http://perma.cc/EER2-PS5L>]; E-mail from Marla Randall, Court Adm'r, Navajo Cty. Courts, to Michael Morisy (Dec. 20, 2016), <http://www.muckrock.com/foi/navajo-county-9526/navajo-county-superior-court-public-safety-assessment-30129/> [<http://perma.cc/YXB9-2454>]; E-mail Paul Thomas, Court Adm'r, Mesa

Mun.Court,toMichaelMorisy(Nov.17,2016),<http://www.muckrock.com/foi/mesa-4736/mesa-municipal-court-public-safety-assessment-30126/> [<http://perma.cc/V6P5-ZAP5>].

¹²⁸ *Lásd pl.*, Memorandum of Understanding Between the Laura and John Arnold Foundation and the Superior Court of California, County of San Francisco, *supra* note 42 ("A Bíróság beleegyezik abba, hogy az illetékes bíróság által hozott bírósági határozat hiányában tartózkodik az eszközzel kapcsolatos információk, beleértve az eszköz fejlesztésével, működtetésével és bemutatásával kapcsolatos információkat, harmadik felek számára történő nyilvánosságra hozatalától az Alapítvány előzetes írásbeli jóváhagyása nélkül.")

¹²⁹ Memorandum of Understanding Between the Laura and John Arnold Foundation and the Seventh Judicial Circuit of the State of Florida 3 (2015), http://cdn.muckrock.com/foia_files/2016/12/15/Memorandum_of_Understanding.pdf [<http://perma.cc/4QAE-NUCP>].

¹³⁰ *Lásd* Zach Dal Pra, LJAF közbiztonsági értékelés-PSA, JUST. SYS. PARTNERS 31, http://cdn.muckrock.com/foia_files/2016/12/15/Volusia_Stakeholder_Training_10162015.pdf [<http://perma.cc/BH7A-PNPC>] (diakép). Ez a prezentáció, amelyet a Hetedik Igazságügyi Körzet bocsátott rendelkezésünkre, csak egy rövid összefoglalót tartalmaz a tanulmányról, nagyon kevés részletességgel. Mivel nem tartalmaz hivatkozást semmilyen közzétett forrásra, feltételezzük, hogy a tanulmányt maga az Alapítvány készítette, és nem került publikálásra.

a megjelenés elmaradásának kockázata négyszerese a legalacsonyabb kockázati pontszámnak. Miután azonban a PSA-algoritmust elkezdtek használni, az Arnold Alapítvány úgy találta, hogy az eredmények szűkebb sávját generálja: 1 (12%), 2 (16%), 3 (18%), 4 (23%), 5 (27%) és 6 (30%).¹³¹ A

a hatos pontszám kisebb kockázatot jelentett a megjelenés elmaradására, mint a négyes pontszám a képzési adatokban. Sajnos az egyetlen validálási vizsgálati eredmény három összefoglaló táblázat. Ezért nem tudhatjuk például, hogy az egyes kockázati kategóriákba tartozó vádlottak hány százalékát vették őrizetbe ahelyett, hogy a tárgyalás előtt szabadon engedték volna, és így nem szerepeltek a validációs tanulmányban.

A Hetedik Igazságügyi Körzet által készített két dokumentum egy másik Arnold Alapítványi kezdeményezésről is szolgáltat némi információt, amelyet maga az Alapítvány nem hozott széles körben nyilvánosságra.¹³² Az alapítvány azt ajánlja, hogy a bíróságok használjanak egy "döntéshozatali keretrendszert", amely a vádlott PSA-kockázati pontszámát és a folyamatban lévő vádakot veszi alapul, és kimenetként konkrét ajánlásokat generál a tárgyalás előtti kezelésre vonatkozóan, az óvadék nélküli szabadon bocsátástól az őrizetbe vételig.¹³³ A döntéshozatali keretrendszer egy második algoritmus, amely konkrét ajánlásokat generál a kezelésre vonatkozóan (a kockázati pontszámok helyett). Az alapítvány szerint a döntéshozatali kereteket az egyes joghatóságok képviselői hozzák létre az egyes joghatóságok számára egy olyan vállalkozóval együttműködve, amely a PSA-programnak az adott bírósági rendszerekben történő végrehajtására szakosodott.¹³⁴ A hetedik bírósági kerület dokumentumai azonban nem adnak tájékoztatást arról, hogy az adott bíróság döntéshozatali keretrendszerét hogyan hozták létre, vagy hogy azt tesztelték-e.

¹³¹ *Id.* 46. o. (ezeket a számokat a PSA alkalmazásának nyomon követése alapján állapította meg 100 000 esetben Kentuckyban és három meg nem nevezett Kentucky-n kívüli városban).

¹³² *Lásd Dal Pra, Supra* 130. lábjegyzet, 49-54. dia (a döntéshozatali keretrendszer leírása); *Zach Dal Pra, Volusia County Case Review*, JUST. SYS. PARTNERS 3(2016

),
http://cdn.muckrock.com/foia_files/2016/12/15/Volusia_County_Case_Review_Redacted.pdf [<http://perma.cc/A4PA-2EQR>] (megjegyezve, hogy "[a] tárgyaló személyzet a döntéshozatali keretrendszert követi, amely útmutatóul szolgál a vádlott kockázati szintjén alapuló ajánlások megtételéhez, a jelen vád természetének megfelelő kiigazításokkal").

¹³³ *Lásd Dal Pra*, 130. lábjegyzet, 52. dia (amely elmagyarázza, hogyan használják a döntéshozatali keretrendszert "az előzetes ajánlás szerinti szabadlábba helyezési típus és a megfelelő feltételrendszer meghatározására").

¹³⁴ *Lásd Leila Walsh*, a Laura & John Arnold Found. kommunikációs alelnöke és Robert Brauneis, a George Washington Univ. jogászprofesszorának e-mailje 5

noldFoundationEMail.pdf [

[http://perma.cc/7CU3-](http://perma.cc/7CU3-4JQP)

4JQP]("BeforeimplementingthePSA,localstakeholders-such as representatives from the courts, law enforcement, district attorney's office, and public defender's office-work together on a decision-making framework (DMF) for their jurisdiction.").

Végül közvetlenül az Arnold Alapítványhoz fordultunk, és egy sor e-mail és telefonbeszélgetés során kifejezetten olyan technikai jelentéseket, validációs tanulmányokat és egyéb dokumentumokat kértünk, amelyek az Alapítvány birtokában lehetnek, és amelyek részletesebb tájékoztatást nyújtanak a PSA-algoritmusok létrehozásáról és teszteléséről. Az Alapítvány egy rövid, háromoldalas nyilatkozatot küldött, amely nagyrészt az Alapítvány honlapján már elérhető szövegből állt.¹³⁵

Az Alapítvány honlapjáról, a Hetedik Igazságügyi Körzet által rendelkezésre bocsátott dokumentumokból és az Alapítvány által számunkra készített nyilatkozatból tudjuk, hogy az Alapítvány a PSA-algoritmusokat mintegy 750 000 ügy adatainak elemzésével hozta létre. Semmit sem tudunk arról, hogy hogyan elemezte ezeket az adatokat, milyen alternatívákat próbált ki, vagy hogy ezek az alternatívák hogyan viszonyultak a végül elfogadott PSA-algoritmusokhoz.

Konkrétan megkérdeztük, hogy az Arnold Alapítvány miért ragaszkodott olyan egyetértési megállapodásokhoz, amelyek megtiltják a bíróságoknak, hogy bármilyen információt nyilvánosságra hozzanak a PSA-ról. Az alapítvány válaszolt:

Az algoritmus kiadása előtt a korai bevezető joghatóságokkal kötött titoktartási megállapodások miatt a PSA használata korlátozott volt, amíg az eredmények mérésére szolgáló helyi adatinfrastruktúrát fejlesztettük ki, megvártuk és tanulmányoztuk a bevezetés utáni előzetes vizsgálati eredményeket, és további kutatásokat kezdeményeztünk. Ezek a titoktartási megállapodások azt is megakadályozták, hogy a profitorientált vállalatok a PSA elemeit a joghatóságok számára értékesítendő, nem megfelelő kockázati eszközök kifejlesztésére használják fel.¹³⁶

Amennyire azonban meg tudjuk állapítani, a titoktartási rendelkezések nem korlátozódnak a "korai elfogadást végző joghatóságokra", és a rendelkezések mindegyike kimondja, hogy a titoktartást örökre megkövetelik.

2. Eckerd Rapid Safety Feedback-Child Welfare értékelések

Az Eckerd Rapid Safety Feedback (RSF) egy olyan kockázatértékelési eljárás, amelynek célja a gyermekjóléti esetek azonosítása, amelyekben nagy a valószínűsége a gyermek súlyos sérülésének vagy halálának.¹³⁷ Az RSF-et az Eckerd Kids (Eckerd), egy nonprofit család- és gyermekjóléti szolgálat fejlesztette ki.

¹³⁵ *Id.*

¹³⁶ *Id.* 5.

¹³⁷ *Összefoglaló és sokszorosítási információk*, ECKERD RAPID SAFETY FEEDBACK,
<http://static.eckerd.org/wp-content/uploads/Eckerd-Rapid-Safety-Feedback-Final.pdf> [<http://perma.cc/6H3D-PNBR>].

szervezet, valamint a Mindshare Technology, egy profitorientált szoftvercég. A gyermekjóléti esetek nagy csoportjának áttekintése révén, beleértve azokat is, amelyekben gyermekek sérültek meg vagy haltak meg, az Eckerd azonosította a gyermekek sérüléséhez vagy halálához hozzájáruló legnagyobb kockázati tényezőket, nevezetesen "a három év alatti gyermek, az otthonban élő szerető, a kábítószerrel való visszaélés és a családon belüli erőszakos előzmények, valamint a korábban nevelőszülőknél elhelyezett szülő".¹³⁸ Az Eckerd a Mindshare Technology-val együttműködve olyan szoftvert hozott létre, amely elemzi a meglévő gyermekjóléti jelentési rendszerek adatait, és beavatkozás céljából megjelöli az ilyen tényezőkkel rendelkező, magas kockázatú eseteket.¹³⁹

Nyílt iratokat tartalmazó kéréseket küldtünk az Eckerd RSF algoritmusának használatára vonatkozó információkért öt állami gyermekjóléti ügynökségnek, amelyek az Eckerd jelentése szerint az RSF-rendszert használják: Alaszka, Connecticut, Illinois, Maine és Oklahoma. Alaszkából, Connecticutból és Illinoisból számos dokumentumot kaptunk. Oklahoma azt válaszolta, hogy körülbelül 2500 dollárra lenne szüksége ahhoz, hogy válaszoljon a kérésünkre, ami nyilvánvalóan magában foglalja az Eckerdnek küldött gyermekjóléti ügyek adatainak a személyazonosító adatok eltávolításával történő átadásának költségeit.¹⁴⁰ Maine tudomásul vette a kérésünket, de a mai napig nem nyújtott be semmilyen dokumentumot.¹⁴¹

Az Alaszkai Egészségügyi és Szociális Szolgáltatások Minisztériuma számos dokumentumot küldött nekünk, köztük a Gyermekjóléti Szolgáltatások Hivatala ("OCS") és az Eckerd közötti egyetértési nyilatkozatot, amely az Eckerd által a gyermekjóléti ügyek RSF-értékelésének az OCS számára történő nyújtásáról szól. A megállapodásból egyértelműen kiderül, hogy az Eckerd megtartja az ellenőrzést a szoftver felett, amely feldolgozza az OCS gyermekjóléti ügyeivel kapcsolatos információkat és kockázatértékeléseket készít. A gyermekjóléti ügyekkel kapcsolatos információkat továbbítják az Eckerdnek vagy a Mindshare-nek, és az ezekről az ügyekről készített kockázatértékeléseket egy weboldalon teszik elérhetővé.

¹³⁸ *Id.; lásd még* Bryan Lindert, *Eckerd Rapid Safety Feedback: Bringing Business Intelligence to Child Welfare*, 2014 POL'Y & PRAC. 25, <http://static.eckerd.org/wp-content/uploads/Eckerd.pdf> [<http://perma.cc/SKU8-UJQH>]. Az Eckerd azt is megállapította, hogy a gyermekek sérülésének vagy halálának megelőzése érdekében megtehető legkritikusabb lépések a minőségi biztonsági tervezéshez, a minőségi felügyeleti felülvizsgálatokhoz, valamint a házi látogatások mennyiségéhez és gyakoriságához kapcsolódnak. *Id.* Nem világos, hogy Eckerd ezeket az elemzéseket gépi tanulással, emberi értékeléssel vagy a kettő valamilyen kombinációjával végezte-e.

¹³⁹ *Lásd: Összefoglaló és replikációs információk*, 137. lábjegyzet.

¹⁴⁰ *Lásd* számla, nyílt iratbetekintési kérelem, OKLA. DEP'T HUMAN SERVICES (2016. november 3.), http://cdn.muckrock.com/foia_files/2016/11/09/11-

3- 16_MR28151_FEE_2472.pdf [<http://perma.cc/L78B-DRBK>] (2472 dollár
kifizetését kérte a nyílt iratok iránti kérelem megválaszolásához).

¹⁴¹ *Lásd:* Kevin C. Wells, főtanácsos levele, Me· Dep't of Health and Human Servs.
to Michael Morisy, MUCKROCK (2016. november 21.),
http://cdn.muckrock.com/foia_files/2016/12/06/1249s5dayletterEckerdRapid_SafetySoftware11.21.16.pdf [<http://perma.cc/34RS-TZLX>].

az Eckerd által fenntartott, az OCS személyzetének hozzáférése.¹⁴²

Az állami szerv, az OCS nem fér hozzá a kockázatértékeléseket generáló algoritmushoz, és nem fér hozzá ahhoz a folyamathoz sem, amelynek során az algoritmus létrejön és kiigazításra kerül. Sőt, amennyiben az OCS bármit is megtud az algoritmusról, vállalja, hogy azt nem hozza nyilvánosságra. Az Eckerd-Alaska OCS egyetértési megállapodás előírja, hogy az összes "Eckerd IP", beleértve az Eckerd által fenntartott weboldalt és a hozzá kapcsolódó szoftvert, az Eckerd által készített jelentéseket, valamint az összes kapcsolódó találmányt, eljárást, fejlesztést és algoritmust "bizalmas információként" kell kezelni, amelyet az OCS vállalja, hogy nem hoz nyilvánosságra.¹⁴³

A connecticuti gyermek- és családjogi minisztérium számos dokumentumot nyújtott be az Eckerd RSF-ről, köztük egy brosúrát, egy adatlapot, egy diabemutatót és egy folyamatábrát, amelyek megerősítik, hogy az állami szerv a gyermekjóléti ügyekről információkat szolgáltat az Eckerdnek, amely ezt követően feldolgozza ezeket az információkat, és kockázatértékeléseket készít, amelyeket az ügynökség megtekinthet.¹⁴⁴

Az Illinois-i Gyermek- és Családjogi Minisztérium az Eckerddel kötött szerződéseit a 2016-os és 2017-es pénzügyi évre vonatkozóan nyújtotta be. Ezekből kiderül, hogy Illinois állam a becslések szerint milyen összegeket fog fizetni az Eckerdnek a szolgáltatásaiért - 107 000 dollárt a 2016-os pénzügyi évben és 171 000 dollárt a 2017-es pénzügyi évben.¹⁴⁵ A szerződések tartalmazznak olyan, a jelek szerint szokásos állami szerződéskötési rendelkezéseket is, amelyek kedvezőbbek a közzététel és a köztulajdon szempontjából, mint az Alaszkai Memorandum.

¹⁴² Az Eckerd Youth Alternatives, Inc. és az Alaszkai Egészségügyi és Szociális Szolgáltatások Minisztériumának 2015. február 20-i egyetértési nyilatkozata 2,

http://cdn.muckrock.com/foia_files/2017/03/02/MOU_Eckerd_and_OCS_February_2015_signed.docx.pdf [<http://perma.cc/58KW-FXKP>] (amely szerint az Eckerd "[h]ost, karbantartja és támogatja a portált azzal a céllal, hogy az ügynökség számára 24 órás technikai támogatást és hozzáférést biztosítson a portálhoz és az általa generált jelentésekhez"); *id.* 1. pont (a "portál" meghatározása: "olyan weboldal és kapcsolódó technológia, amely [a gyermekjóléti esetekre vonatkozó elektronikus információk] olvasására, automatizált elemzésre és jelentések készítésére szolgál, amelyek az Eckerd gyors biztonsági visszajelzésének megvalósítására és támogatására használhatók").

¹⁴³ *Lásd id.* 1., 4., 5. pont.

¹⁴⁴ Például a "CT Eckerd Rapid Safety Feedback Process Flow" című diagram a Mindshare-hez rendeli a "Mindshare Tool/Predictive Analysis Generates List for Review" (Mindshare eszköz/prediktív elemzés létrehozza a listát felülvizsgálatra) lépést. *CT Eckerd Rapid Safety Feedback Process Flow*, ECKERD, http://cdn.muckrock.com/foia_files/2016/11/13/CT_ERSF_Process_Flowchart9-12.pdf [<http://perma.cc/4B3Z-ZVPG>].

¹⁴⁵ *Lásd*: Contract # 5445089016 between Eckerd Youth Alternatives, Inc. and State of Illinois, Department of Children and Family Services 7 (Sept. 15,

2015), <http://www.robertbrauneis.net/algorithms/ILERSFFY16.pdf>
[<http://perma.cc/D5LY-44C5>] [a továbbiakban: Illinois FY16 Contract];
Contract # 5445089027 between Eckerd Youth Alternatives, Inc. and State of
Illinois, Department of Children and Family Services 7 (July 1, 2016) ,
<http://www.robertbrauneis.net/algorithms/ILERSFFY17.pdf>
[<http://perma.cc/M7QS-XLH7>] [a továbbiakban: Illinois FY17 Contract].

Megértés. Ezek kimondják, hogy "minden olyan információ, amelyet a szövetségi törvény, az állami törvény vagy az alkalmazandó FOIA-mentesség nem tilt vagy nem mentesít a nyilvánosságra hozatal alól, nyilvános."¹⁴⁶ Azt is előírják, hogy az állam tulajdonában van minden, a szerződések alapján előállított információ, beleértve a szellemi tulajdonjogokat és a szerződésekből származó termékeket.¹⁴⁷ Az illinois-i szerződésekben van néhány olyan megfogalmazás, amely arra utal, hogy az Eckerdnek az illinois-i adatok elemzésén alapuló új előrejelző algoritmust kell készítenie; az "1. fázis: A modell kifejlesztése" egyik intézkedése "[a]z előrejelző modell kifejlesztése, amelyet a súlyos sérülés vagy halálozás legnagyobb valószínűségével járó bejövő vizsgálatok azonosítására használnak."¹⁴⁸ Nem világos azonban, hogy ez valójában teljesen új adatelemzést vagy egy meglévő algoritmus valamilyen kiigazítását jelenti.

3. Allegheny Family Screening Tool-Child Welfare Értékelések

Az Eckerd RSF-hez hasonlóan az Allegheny Family Screening Toolt (AFST) is a gyermekjóléti esetek osztályozásának megkönnyítésére fejlesztették ki. Az AFST-t az Aucklandi Műszaki Egyetem Adatelemzési Központja (az "Aucklandi Konzorcium") által vezetett konzorcium fejlesztette ki az Allegheny Megyei Humánszolgáltatási Főosztállyal együttműködve. Az Allegheny DHS pályázati felhívást tett közzé az Allegheny megye adatbázisainak hasznosítására irányuló projektekre, és az Auckland Konzorcium sikeres pályázatot nyújtott be. Míg az Eckerd RSF-et nyilvánvalóan folyamatosan használják a gyermekjóléti rendszeren belüli esetek nyomon követésére, az AFST-t a gyermekbántalmazás bejelentésekor alkalmazzák. Segít annak meghatározásában, hogy a bejelentés indokolja-e a hivatalos vizsgálatot. Az Allegheny FST-t jelenleg csak Allegheny megyében alkalmazzák.

Miután az AFST-vel kapcsolatban nyílt iratokra vonatkozó kérelmet nyújtottunk be Allegheny megyéhez, a megye tisztviselői felvették velünk a kapcsolatot, átadták az Auckland Konzorcium által az algoritmus fejlesztéséről készített jelentést,¹⁴⁹ és jelezték, hogy szívesen beszélnek velünk az algoritmusról és annak fejlesztéséről. A jelentés sok tekintetben a legátfogóbb, amit az algoritmus fejlesztéséről láttunk. Részletesen ismerteti a döntések nagy részét, amelyeket a

¹⁴⁶ Lásd Illinois FY16 Contract, *Supra* 145. lábjegyzet, 11. pont; Illinois FY17 Contract, *Supra* 145. megjegyzés, 11. pont.

¹⁴⁷ Lásd Illinois FY16 Contract, *Supra* 145. lábjegyzet, 11. pont; Illinois FY17 Contract, *Supra* 145. lábjegyzet, 11. pont.

¹⁴⁸ 145. megjegyzés, 11. pont.

¹⁴⁹ Illinois FY16 szerződés, 145. lábjegyzet, 6. pont.

¹⁴⁹ *Lásd* Vaithianathan et al., 72. lábjegyzet.

a fejlesztési folyamatot, a döntések indoklását, valamint a felhasznált adatokat és módszereket.

A fejlesztők végül két algoritmust hoztak létre: az egyik annak valószínűségét jelzi előre, hogy egy hivatalosan ki nem vizsgált állítást két éven belül egy másik, ugyanezt a gyermeket érintő állítás követ, a másik pedig annak valószínűségét, hogy egy hivatalosan kivizsgált állítás két éven belül a gyermek nevelőszülőkhöz kerül.¹⁵⁰ Az algoritmusok képzési adatait a megye 2008-ban létrehozott integrált adatkezelési rendszeréből nyerték; a fejlesztők úgy döntöttek, hogy az adatállományban szereplő minden egyes bántalmazási vád esetében az adott vádat megelőző tizenhét hónagra, valamint a vádat követő két évre vonatkozó adatokra van szükségük.¹⁵¹ Az adatkészlet több mint 800 változót tartalmazott.¹⁵² A fejlesztők a nemlineáris regressziót használták fő elemzési módszerként, nagyrészt azért, mert ez ugyanolyan jó eredményeket hozott, mint más módszerek, és előnye, hogy értelmezhető.¹⁵³ Más szóval, alkalmas volt az átláthatósági és elszámoltathatósági célok elérésére. A fejlesztők mind belső validációs vizsgálatokat végeztek - a képzési adatok egy fenntartott részének felhasználásával -, mind külső validációs vizsgálatokat, a kórházi felvételek és a "kritikus események" (súlyos sérülés vagy haláleset) nyilvántartásának felhasználásával.¹⁵⁴ Az algoritmust nem használták elég régóta ahhoz, hogy a bevezetést követő vizsgálatokat végezhesse.

A jelentés egy függelékben 112 változót tesz közzé, amelyek végső soron

Az Allegheny DHS kérésére az Allegheny DHS rendelkezésére állnak a felhasznált adatok - 71 a nevelőszülőknél történő elhelyezést előrejelző modellhez, és 59 a megismételt vádakat előrejelző modellhez -, valamint az egyes változókhoz rendelt súlyok.¹⁵⁵ Az algoritmusok eredményét két kockázati pontszámként mutatják be - az egyik az ismételt vádakra vagy "ismételt beutalásra", a másik pedig a nevelőszülőknél történő elhelyezésre vonatkozik - egy 1-től 20-ig terjedő skálán, ahol minden egyes szám az összes figyelembe vett gyermek 5%-ának megfelelő sávot képvisel.¹⁵⁶ Így a "10" pontszám azt jelenti, hogy a gyermek újbóli beutalásának vagy elhelyezésének kockázata az összes gyermek 50-55%-a között van; a "15" pontszám a 75-80%-os tartományban van. A fejlesztők úgy döntöttek, hogy létrehozhatnak egy olyan küszöbértéket is, amely vélhetően kötelező vizsgálatot eredményez, azzal a lehetőséggel, hogy a felügyelő lemondhat erről az eredményről; a jelentés nem hozza nyilvánosságra a küszöbértéket.¹⁵⁷

¹⁵⁰ *Lásd id.* 10. pont

¹⁵¹ *Lásd id.* 11. pont

¹⁵² *Lásd id.* 12. pont

¹⁵³ *Lásd id.* 13-14. o.

¹⁵⁴ *Lásd id.* 15-17., 19-23. pont.

¹⁵⁵ *Lásd id.* 37-43. o. Ilyen kérelmet nyújtottunk be, és megkaptuk a súlyokat.

¹⁵⁶ *Lásd id.* 27. pont

¹⁵⁷ *Lásd id. 28. pont*

Allegheny megye végül úgy döntött, hogy nem használja a gyermek vagy a gondviselők faji hovatartozását, mint változót, mivel az nem javította lényegesen az előrejelző erőt, és egyébként is problémás volt.¹⁵⁸ A jelentés hosszasan tárgyalja a hamis negatív és hamis pozitív eredmények veszélyeit, de azt nem mondja meg, hogy ezeket végül egyformán vagy egyenlőtlenül súlyozták-e.¹⁵⁹ Bár az Auckland Konzorcium megtartotta a szerzői jogokat. az algoritmus végrehajtásához használt kódban, az Allegheny megyével kötött szerződés feljogosítja arra, hogy további fizetés nélkül engedélyezze más joghatóságok számára a kód használatát, és a megyei tisztviselők jelezték, hogy érdekelték ebben. Így, bár ez a projekt nem teljesen nyílt forráskódú projekt, közelebb áll hozzá, mint az általunk vizsgált másik öt algoritmus bármelyike.

4. PredPol-Predictive Rendőrség

A PredPol egy olyan szoftver, amely megjósolja, hogy hol és mikor várhatóak különböző típusú bűncselekmények, és így segíti a rendőrséget a járőrözés megtervezésében, hogy megakadályozza ezeket a bűncselekményeket. Eredetileg a Los Angeles-i Kaliforniai Egyetem és a Santa Clara Egyetem matematikusai és viselkedéskutatói fejlesztették ki a Los Angeles-i és a Santa Cruz-i rendőrség bűnügyi elemzőivel és rendőreivel együttműködve,¹⁶⁰ de ma már a PredPol Inc. nevű profitorientált vállalat kezeli. A PredPol alkotói megállapították, hogy a bűncselekmények előrejelzéséhez szükséges három legfontosabb információ, azaz "adatpont" a bűncselekmény típusa, a bűncselekmény helyszíne, valamint a bűncselekmény időpontja.¹⁶¹ A PredPol a múltbeli bűncselekményekről szóló adatokat egy algoritmusba táplálja, amely megjósolja, hogy hol és mikor fognak új bűncselekményeket elkövetni.¹⁶² Egy forrás szerint a PredPol "közismert arról, hogy algoritmusát "szigorúan őrzött" titokként kezeli"¹⁶³.

Tizenegy rendőrkapitányságnak, köztük az alabamai Oxford, az arkansasi Little Rock, a kaliforniai Los Angeles, Modesto, Orange County és Santa Cruz, a floridai Cocoa, a georgiai Atlanta, a marylandi Hagerstown, a pennsylvaniai Reading és a washingtoni Tacoma rendőrkapitányságának küldtünk a PredPolra vonatkozó iratokat. A tizenegy rendőrkapitányság közül nyolc vagy nem válaszolt, vagy tudomásul vette a megkeresésünket, de nem válaszolt.

¹⁵⁸ *Lásd id.* 15., 30. pont.

¹⁵⁹ *Lásd id.* 10. pont.

¹⁶⁰ *PredPol Is Predictive Policing*, PREDPOL, <http://www.predpol.com/about/> [<http://perma.cc/Y2R8-C2XM>].

¹⁶¹ *Id.*

¹⁶² *Id.*

¹⁶³ Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. ONLINE 101, 119 (2017) (idézi Ali Winston,

Arizona Bill Would Fund Predictive Policing Technology, REVEAL (2015.
március 25.)
,<http://www.revealnews.org/article/arizona-bill-would-fund-predictivepolicing-technology/> [<http://perma.cc/V4BV-YMZP>]).

dokumentumokat, több időt kértek a válaszadásra, de még nem válaszoltak, vagy azt válaszolták, hogy nem rendelkeznek releváns dokumentumokkal. A három hivatal, amely dokumentumokat bocsátott rendelkezésre, Tacoma, Cocoa és Santa Cruz volt.

A washingtoni Tacoma városa volt az egyik legnyilvánosabb azon joghatóságok közül, amelyeknek bármilyen algoritmusra vonatkozó iratkérést küldtünk. Kétszáz e-mail üzenetet küldött a Tacoma Rendőrség és a PredPol munkatársai közötti levelezésből a PredPol bevezetésével kapcsolatos legkülönbözőbb kérdésekről. Tíz prezentációt is készített a PredPol és a prediktív rendőri munka működéséről. Ezek a dokumentumok igen hasznosak lennének azok számára, akiket érdekel, hogyan néznek ki a PredPol-jelentések, milyen adatokat használ a PredPol-algoritmus bemenetként, és így tovább. Egyik dokumentumból sem derül ki azonban, hogy a PredPol milyen algoritmust használ a korábbi bűnözési adatokból történő előrejelzések létrehozására, és az sem, hogy a PredPol milyen eljárást alkalmazott az algoritmus létrehozásához.

Cocoa városa számos dokumentumot küldött nekünk, amelyek mind a PredPol-tól való szolgáltatásvásárlással kapcsolatosak. Talán a legbeszédesebb az a háttérdokumentum volt, amelyet a városi tanács tagjai rendelkezésére bocsátottak, amikor a PredPol szolgáltatásainak megvásárlása a tanács napirendjén szerepelt. Ez a dokumentum nem tartalmaz semmilyen részletet a PredPolról, de kimondja, hogy "[a] város ügyvédje azt tanácsolta, hogy a felügyeleti technikákat, eljárásokat vagy személyzetet feltáró információk a Florida Statutes 119.071(2)(d) s. alapján mentesülnek a nyilvános betekintés alól. "¹⁶⁴ Valószínű, hogy a város erre a tanácsra támaszkodott, amikor elutasította, hogy magáról a PredPolról dokumentumokat adjon át, bár nagyon valószínű, hogy a város elrejtette a felügyeleti technikákat, miközben átláthatóbbá tette az algoritmus értékeit és végrehajtását.

A kaliforniai Santa Cruz városa több képernyőképet is küldött a PredPol szoftverről. Az egyik képernyő arra kéri a felhasználót, hogy adja meg a közelmúltban elkövetett bűncselekmények helyére (szélességi és hosszúsági fokban), időpontjára és típusára (jármű vagy lakóhely) vonatkozó adatokat, és azt állítja, hogy a következő huszonnégy óras időszakban elkövetett bűncselekmények helyére vonatkozó előrejelzések megjelennek egy térképen. A másik képernyő a város térképe, amelyen színes területek jelölik a bűncselekmények valószínűsíthető előfordulási helyeit. Ezek a képernyőképek információt nyújtanak az adatbevitel típusáról és a kimenet formátumáról, de ennél többet nemigen. Úgy tűnik, hogy a Santa Cruz által használt PredPol verzió kevésbé kifinomult, mint a Tacoma által használt.

5. HunchLab-Predictive Rendőrség

- ¹⁶⁴Jogszabályok részletei (szöveggel), 15-361. sz. ügyirat, Kakaó városa 1 (2015. július 30.), [http://cdn.muckrock.com/foia_files/2017/01/13/15-361_City_Council_Agenda_Item 8-25-15.pdf](http://cdn.muckrock.com/foia_files/2017/01/13/15-361_City_Council_Agenda_Item%208-25-15.pdf) [<http://perma.cc/K9JZ-ZAR9>].

A PredPolhoz hasonlóan a HunchLab is egy olyan szoftver, amely megjósolja, hogy hol és mikor fognak bűncselekmények előfordulni, és egy kartográfiai kimenettel jelzi azokat a területeket, ahol bizonyos időszakokban nagyobb a bűncselekmények bizonyos típusainak kockázata. A HunchLabot az Azavea, Inc. fejleszti és tartja fenn, amely egy nyereségorientált, tanúsított B társaság.¹⁶⁵ A HunchLab számos bemeneti adatot használ a bűncselekmények kockázatának előrejelzéséhez, és lehetővé teszi az egyes rendőrkapitányságok számára, hogy a kiválasztott bűncselekmények tekintetében prioritásokat állítsanak fel.¹⁶⁶

A HunchLabra vonatkozó nyílt iratokat négy rendőrkapitányságnak küldtük el, köztük a floridai Miami, a Missouri állambeli St. Louis megye, a nebraskai Lincoln és a pennsylvaniai Philadelphia rendőrkapitányságainak.

A Miami végül azt válaszolta, hogy nem rendelkezik válaszadó dokumentumokkal.¹⁶⁷ St. Louis megye 400 dollár kifizetését kérte, mielőtt bármilyen dokumentumot előállítana, és megismételte, hogy 400 dollár kifizetése nélkül nem fog eljárni, amikor megkérdeztük, hogy szűkíthetjük-e a díj csökkentésére irányuló kérelmünket.¹⁶⁸ Philadelphia városa benyújtotta a HunchLab szolgáltatásra vonatkozó megrendelést,¹⁶⁹ de egyébként elutasította a kérésünket azzal az indokkal, hogy nem kértünk konkrét dokumentumokat.¹⁷⁰ A nebraskai Lincoln városa több dokumentumot is benyújtott, köztük egy kézikönyvet, amely bemutatja a HunchLabot a személyzetnek, valamint a város közbiztonsági igazgatójának a HunchLabról szóló blogbejegyzését. Talán a leghasznosabb, hogy Lincoln a HunchLab számára egy mintát adott a bemeneti adatokból, amelyeket úgy azonosított, mint amelyek egy harmincnapos gördülő görgetett

¹⁶⁵ A B Corporations olyan profitorientált vállalatok, amelyeket a B Lab, egy nonprofit tanúsító szervezet tanúsít, hogy megfelelnek bizonyos társadalmi és környezeti teljesítményre, elszámoltathatóságra és átláthatóságra vonatkozó előírásoknak. *Lásd: Mik a B vállalatok?*, B CORP, <http://www.bcorporation.net/what-are-b-corps> [<http://perma.cc/Q5FH-8KCK>].

¹⁶⁶ A HunchLab meghatározza, hogy mely adatok a leghasznosabbak az egyes bűncselekmények előrejelzéséhez. Bizonyos esetekben a földrajz - a korábbi bűncselekmények helyszínei vagy bizonyos tereptárgyak - a legfontosabb tényező. Más esetekben az idő - a hét napja, az év hónapja - élvez elsőbbséget." Maurice Chammah, *Policing the Future*, MARSHALLPROJECT (Feb.3, 2016), <http://www.themarshallproject.org/2016/02/03/policing-the-future#.vVL53xF4m> [<http://perma.cc/YKC9-CDU6>] (beszámoló a HunchLab rendszer St. Louis-i telepítéséről).

¹⁶⁷ *Lásd MiamiPDHunchLabDocuments*, MUCKROCK, <http://www.muckrock.com/foi/miami-103/miami-pd-hunchlab-documents-30109/> [<http://perma.cc/9WPG-YH3F>] (levelezés megjelenítése).

¹⁶⁸ *Lásd St. Louis CountyPDHunchLabDocuments*, MUCKROCK, <http://www.muckrock.com/foi/st-louis-county-8838/st-louis-county-pd-hunchlab-documents-30113/> [<http://perma.cc/P68R-8M4F>] (a levélváltás megjelenítése).

- ¹⁶⁹ *Vásárlási megbízás*, CITY OF PHILADELPHIA PROCUREMENT DEP'T. (2015) http://cdn.muckrock.com/foia_files/2017/02/06/POXX16106457_-_Redacted.pdf [<http://perma.cc/FQ6W-TEDN>].
- ¹⁷⁰ *Lásd* Robert Kieffer, a városi jogtanácsos asszisztensének Michael Morisy-nak küldött e-maile (Febr. 6 2017), http://cdn.muckrock.com/foia_files/2017/02/06/Final_Response_-_Morisy_CP_2016-1710.pdf [<http://perma.cc/46LH-M3A2>].

a rendőrségi incidensjelentések ablaka.¹⁷¹ Ebben az időszakban Lincoln 3057 rendőrségi eseményről szóló jelentést rögzített; minden egyes ilyen jelentés tartalmazza a bejelentett bűncselekmény utcai címét, szélességi és hosszúsági fokát, a bűncselekmény típusát, valamint a bejelentés és a bűncselekmény dátumát és időpontját.¹⁷²

Jeremy Heffner, a HunchLab projektmenedzsere és az Azavea vezető adatelemzője megkeresett minket, miután értesült a Lincoln, Nebraska államban benyújtott nyílt iratok iránti kérelmünkről. E-mail váltást és telefonbeszélgetést folytattunk vele, és végül elkészített és elküldött nekünk egy dokumentumtervezetet "A Citizen's Guide to HunchLab" címmel, amely a HunchLab algoritmusairól, valamint azok létrehozásáról és validálásáról nyújt tájékoztatást. A dokumentumból úgy tűnik, hogy a HunchLab algoritmusai kevésbé értelmezhetőek, mint sok más algoritmus. Az algoritmusokat egy "gradiens növelő döntési fa" technikával építik fel, amelyben egymást követő döntési fákat próbálnak ki és tesztelnek; a fejlesztők nem csak a bejelentett bűncselekményekre és azok helyére és idejére vonatkozó adatokat, hanem az ismert elkövetők helyére, a bűncselekmények ismert és valószínű célpontjainak helyére, az időjárásra, a napi, heti és szezonális ciklusokra, a társadalmi-gazdasági mutatókra és így tovább vonatkozó adatokat is beépítik.¹⁷³ A rendőr nem tudhatja, hogy az algoritmus döntéshozatala hogyan viszonyul a saját tudásához és ítélőképességéhez. A HunchLab algoritmus az általunk vizsgált algoritmusok közül is a legdinamikusabb. A HunchLab minden egyes ügyfél esetében néhány hetente új "modellezési futtatást" végez a modell újrakalibrálása érdekében, és minden egyes ilyen modellezési futtatás új előrejelző algoritmust hoz létre.¹⁷⁴

A HunchLab nyíltan megvitatja a bemeneti adatok esetleges elfogultságának kérdését és az ezzel kapcsolatos ítéleteit is. A torzítás egyik típusa a "bejelentési torzítás" - egyes közösségek a bűncselekmények nagyobb százalékát jelenthetik, mint mások. A HunchLab azon az állásponton van, hogy a rendőrségi tevékenységbe helyénvaló beépíteni az ilyen előítéletesség nagy részét. Azt állítja: "Úgy véljük, hogy a rendőrségi tevékenységnek tükröznie kell azt, amit a közösség problémaként jelent.

elfogultságok a rendőrséggel szembeni bizalmatlanságból fakadnak, akkor úgy véljük, hogy helyénvaló az adatokban hagyni az elfogultságot."¹⁷⁵ Megjegyzi, hogy ez nem biztos, hogy így van, ha a bejelentés elmulasztása félelem vagy szégyen miatt történik, de nem világos, hogyan lehet ezt orvosolni. A HunchLab a "végrehajtási torzításról" is nyilatkozik - annak lehetőségéről, hogy a rendőrség végül több letartóztatást hajt végre, és több végrehajtást végez.

¹⁷¹ *Lásd:* Tonya Peters, Police Legal Advisor, Lincoln Police Dep't, MichaelMorisy(Nov.302016)e-mailje

onse.pdf [<http://perma.cc/2KU2-DP3Z>].

¹⁷² PoliceIncidentReports
,http://cdn.muckrock.com/foia_files/2016/11/30/Archive.zip
[<http://perma.cc/7D2N-EL2Q>].

¹⁷³ Lásd: HUNCHLAB, *A Citizen's Guide to HunchLab*, 77. lábjegyzet, 5-6. o.

¹⁷⁴ *Id.* 19. o.

¹⁷⁵ *Id.* 2. pont.

tevékenység egyes közösségekben, mint más közösségekben, még akkor is, ha a bűnözés szintje azonos. Meggyőződése, hogy ez az elfogultság kevésbé érvényesül a súlyos bűncselekmények, például az emberölések, rablások vagy támadások esetében; az egyéb kábítószerrel kapcsolatos és kellemetlen bűncselekmények esetében pedig azt állítja, hogy inkább a közösség szolgáltatási igényeit - panaszokat - tükröző adatokat próbálja felhasználni, mint a rendőrség bűnüldözési tevékenységét tükröző adatokat.¹⁷⁶

A HunchLab program három további érdekes funkcióval rendelkezik. Először is, az algoritmus lehetővé teszi, hogy az egyes közösségek súlyokat határozzanak meg az egyes bűncselekménytípusok relatív súlyosságára - mennyivel fontosabb egy gyilkosságot megállítani, mint egy betörést? Lehetővé teszi továbbá a járőrözés hatékonyságának személyre szabott súlyozását is - a belső bűncselekményeket kevésbé valószínű, hogy a fokozott rendőri jelenlét visszatartja.¹⁷⁷ Másodsor, a HunchLab azt javasolja, hogy az algoritmus tartalmazzon véletlenszerűséget annak biztosítása érdekében, hogy a rendőrök ne minden nap ugyanazokat az útvonalakat járják be, a munka monotóniája ellen és a területen való állandó rendőri jelenlét negatív mellékhatásainak csökkentése érdekében.¹⁷⁸ Harmadszor, a HunchLab mostanra kiterjesztette hatókörét a járőrözési taktikákra is, ajánlva bizonyos típusú rendőri tevékenységeket a járőrözési területeken, például autós járőrözést, gyalogos járőrözést, autómegállítást és hasonlókat, és figyelemmel kíséri az alkalmazott taktikák hatékonyságát az idő múlásával.¹⁷⁹

6. New York City és New York állam hozzáadott érték modellek-tanárértékelés

New York City és New York állam azon joghatóságok közé tartozik, amelyek a tanárok értékelésére a hozzáadottérték-modellt ("VAM") alkalmazzák.¹⁸⁰ Általában a hozzáadott érték modell algoritmusai összehasonlítják a tanulók tesztpontszámait egy adott év elején és végén, hogy mérjék a tanulók fejlődését. Ezeket az eredményeket ezután kiigazítják, hogy megpróbálják figyelembe venni a tanárok hatékonyságán kívüli olyan tényezőket, mint például a társadalmi-gazdasági helyzet, amelyek felelősek lehetnek a tanulók fejlődéséért vagy annak hiányáért. A diákok korrigált eredményei

¹⁷⁶ *Lásd id.* 26.

pont.¹⁷⁷ *Id.* 9-10.

o.¹⁷⁸ *Id.* 10-11. o.

¹⁷⁹ *Id.* 11-14. o.

¹⁸⁰ A New York-i Legfelsőbb Bíróság úgy ítélte meg, hogy a New York-i növekedési mérések a panaszos tanár tekintetében önkényesek és önkényesek voltak. *Matter of Lederman v. King*, 47 N.Y.S.3d 838 (N.Y. Sup. Ct. 2016). *Lásd általában* Valerie Strauss, *Judge Calls Evaluation of N.Y. Teacher 'Arbitrary' and 'Capricious' in Case Against New U.S. Secretary of Education*, WASH. POST (2016) (beszámoló a tanárnő jogtalan elbocsátás miatt a város ellen

indított peres eljárásáról)

,<http://www.washingtonpost.com/news/answer-sheet/wp/2016/05/10/judge-calls-evaluation-of-n-y-teacher-arbitrary-and-capricious-in-case-against-new-york-secretary-of-education/>
[<http://perma.cc/5CH4-RYQE>].

amelyeket egy adott tanár tanít, azután az adott tanár hatékonyságának értékeléséhez használják fel.

Nyílt iratok iránti kérelmet nyújtottunk be New York városához és New York államhoz is a VAM-programjukkal kapcsolatos dokumentumokért.¹⁸¹ A mai napig New York városa öt levelet küldött nekünk, amelyben értesít minket arról, hogy több időre van szüksége az iratok benyújtásához, de nem küldött nekünk egyetlen iratot sem.¹⁸² New York állam oktatási minisztériuma számos dokumentumot nyújtott be, többek között a szállítójával, az American Institutes for Research-rel kötött eredeti szerződést a New York-i VAM program megvalósítására; a szerződés két megújítását; öt közzétett cikket különböző szerzőktől, amelyek általában a hozzáadott érték modellek érvényességét értékelik, és amelyek egyike sem a New York-i VAM megvalósítására összpontosít; valamint a VAM algoritmus mintakimeneteit - ötven diák és ötven tanár kimeneteit, nevük és egyéb azonosítóik eltávolításával.

A mintakimenetek némi információt nyújtanak a VAM algoritmus által előállított eredmények formátumáról, és bepillantást engednek az algoritmus működésébe, mivel a bemenetek egy részét - például a diákok tesztpontszámait - és a kimeneteket is tartalmazzák. Ötven mintakimenet azonban túlságosan kevés ahhoz, hogy az algoritmust visszafejteni lehessen, és az Oktatási Minisztérium és az American Institutes for Research közötti szerződés úgy rendelkezik, hogy "a szerződés teljesítésének időpontjában a szerződő fél tulajdonát képező módszerek vagy intézkedések" "védett információk", amelyeket az Oktatási Minisztérium "kizárólag [saját] oktatási céljaira" használhat fel.¹⁸³ Így az algoritmus vagy algoritmusok nem nyilvánosak, és a folyamat, amelynek során létrehozták őket, nem került nyilvánosságra.

C. Következtetés

A prediktív algoritmusok megismerésére irányuló erőfeszítéseink a nyílt iratok iránti kérelmek révén sok tekintetben frusztrálóak voltak. Sok kormány nem válaszolt, és azok közül is sokan, akik válaszoltak.

¹⁸¹ Nem a miénk volt az első kísérlet. Cathy O'Neil is megpróbálta megszerezni New York City VAM-nyilvánartását, de sikertelenül. *Lásd* Cathy O'Neill, *An Attempt To FOIL Request the Source Code of the Value-Added Model*, MATHBABE (2014), <http://mathbabe.org/2014/03/07/an-attempt-to-foil-request-the-source-code-of-the-value-added-model/> [<http://perma.cc/HNB3-LC5X>].

¹⁸² *Lásd: New York állam vagy New York város hozzáadottérték-intézkedései a tanárok számára*, MUCKROCK (2016. november 9.), <http://www.muckrock.com/foi/new-york-city-17/new-york-állam-vagy-new-york-város-érték-többlet-intézkedései-a-tanárok-számára-29739/> [<http://perma.cc/MN4R-U3U5>] (a levelezés megjelenítése).

¹⁸³ *Lásd a C010834. számú szerződést New York állam népe és az Am. Insts. for Research* (2011. szeptember 19.), app. D, http://robertbrauneis.net/algorithms/AIR_CONTRACT_2011_Redacted_FOIL

azt állították, hogy vagy általában mentesülnek a nyílt iratokról szóló törvények alól (mint például a bíróságok), vagy nem tudnak eleget tenni a kéréseknek, mert ígéretet tettek az információk bizalmas kezelésére. Míg számos joghatóság rendelkezésre bocsátotta a beszállítókkal kötött szerződéseit, így megtudhattunk valamit a szerződési feltételekről, az algoritmusok fejlesztéséről nagyon keveset tudtunk meg, valószínűleg azért, mert a kormányok soha nem rendelkeztek olyan iratokkal, amelyek tartalmazzák ezeket az információkat. Allegheny megye, amely szerződést kötött egy egyetemi kutatókból álló konzorciummal egy előrejelző algoritmus kifejlesztésére a semmiből, volt a legnagyobb kivétel, mivel megrendelte és birtokában volt olyan jelentéseknek, amelyek részletesen bemutatták az algoritmus fejlesztését, és nyilvánosságra hozták magát az algoritmust.

IV.

AZ ÁTLÁTHATÓSÁG LEGFŐBB AKADÁLYAI

Miután részleteztük erőfeszítéseinket, amelyek célja, hogy a prediktív algoritmusok állami és helyi kormányzati alkalmazására vonatkozó nyílt iratok iránti kérelmek révén hasznos információkhoz jussunk, most rátérünk az akadályokra, amelyekkel talákoztunk. Ezek közül a legfontosabbak a fontos feljegyzések létrehozásának vagy a kormányok való átadásának elmulasztása, valamint az üzleti titoktartásra vonatkozó állítások voltak. Kitérünk továbbá a nyílt iratokra vonatkozó törvényekben szereplő bűnüldözési és tanácskozási eljárásra vonatkozó mentességekre is, amelyeket valószínűleg túlzásba visznek, mert a kormányok félnek az algoritmikus átláthatóságtól.

A. *A dokumentáció hiánya*

A kormányok nem hozhatnak nyilvánosságra több információt, mint amennyivel rendelkeznek. A legtöbb nyílt iratokra vonatkozó törvény csak a már létező "iratok" vagy "információk" megszerzésére jogosítja fel a kérelmezőket. Az ügynökségek általában nem kötelesek új iratokat létrehozni, ha nyílt iratigényléssel szembesülnek.¹⁸⁴ Kutatásunk azt mutatta, hogy a kormányok egyszerűen nem rendelkeztek sok, az algoritmusok létrehozására és alkalmazására vonatkozó irattal, vagy azért, mert ezeket az iratokat soha nem készítették el, vagy azért, mert a vállalkozók készítették el, és soha nem bocsátották a kormányzati ügyfelek rendelkezésére. Ezek közé tartoznak a modelltervezési döntésekre, az adatok kiválasztására, a tényezők súlyozására és a validálási tervekhez kapcsolódó feljegyzésekre. Még alapvetőbb szinten a legtöbb kormányzat nem rendelkezett semmilyen feljegyzéssel arról, hogy a modelleknek milyen problémákat kellett volna megoldaniuk, és hogy mik voltak a siker mérőszámai.

¹⁸⁴ Lásd pl. , *A Pocket Guide to the California Public Records Act, FIRST AMEND.*

PROJECTSOCIETY OF PROFESSIONAL JOURNALISTS,
<http://www.thefirstamendment.org/publicrecordsact.pdf>
[<http://perma.cc/K3KP-Y8QT>] ("A PRA csak a már létező nyilvántartásokra
terjed ki, és egy ügynökséget nem lehet arra kötelezni, hogy készítsen
nyilvántartást, listát vagy összeállítást.").

A nagy adatmennyiségű alkalmazásokban a legfontosabb döntések közül sok a modell tervezésének "nagykereskedelmi" szintjén születik, nem pedig a konkrét esetre való alkalmazás "kiskereskedelmi" szintjén. Az analóg világban a nem törvényi szinten meghozott nagykereskedelmi politikai döntéseket valószínűleg közigazgatási szabályalkotás útján hozzák meg. Bejelentik a javasolt politikát, lehetőséget adnak a javaslattal kapcsolatos észrevételek megtételére, végül pedig közzéteszik a végleges politikát, az elfogadásának okait és a végrehajtás módját. Ezek a normák és törvények nem vonatkoznak az algoritmikus politika kialakítására. A nagyadat-előrejelző modelleket gyakran anélkül építik és használják, hogy a kulcsfontosságú szakpolitikai döntéseket valaha is megfogalmaznák, indokolnák vagy rögzítenék. A legjobb esetben nyilvános ajánlattételi felhívásokat tesznek közzé magánszolgáltatók számára, hogy előrejelző algoritmusokat szállítsanak a kormányzatnak.¹⁸⁵ Jellemzőbb, hogy egyszerűen csak egy formális megállapodás születik egy magánszolgáltatóval, amely nem fogalmazza meg az algoritmusba ágyazott politikai döntéseket.

B. Agresszív üzleti titok és titoktartási igények

Még ha a kormányok rendelkeznek is kulcsfontosságú magyarázó feljegyzésekkel, megtagadhatják azok közzétételét, tekintettel a magáneladók azon állításaira, hogy ezek az információk bizalmasak. A szabadalmaztatott algoritmusok tulajdonosai gyakran titoktartási megállapodásokat követelnek meg a közhivatalok ügyfeleitől¹⁸⁶, és üzleti titokvédelemre hivatkoznak az algoritmus és a kapcsolódó fejlesztési és telepítési folyamatok tekintetében.¹⁸⁷ A kormányok ezután ezeket az állításokat arra használják fel, hogy az eladói anyagokat mentesítsék a nyilvánosságra hozatal alól, gyakran olyan módon, amely sérti a nyílt iratokra vonatkozó törvények viszonylag szűk körű üzleti titokra vonatkozó mentességeit.

¹⁸⁵ *Lásd pl.*, *Request for Proposal to Design and Implement Decision Support Tools and Predictive Analytics in Human Services*, ALLEGHENY COUNTY DEPT' HUM. SERV. (2014), <http://www.county.allegheny.pa.us/Human-Services/Resources/Doing-Business/Solicitations/2014/Decision-Support-Tools-and-Predictive-Analytics-in-Human-Services-RFP.aspx> [http://perma.cc/UHS3-8VW3].

¹⁸⁶ *Lásd Joh*, *Supra* 163. lábjegyzet, 7. o. (a rendőrkapitányságok és a Harris Corporation közötti, a Stingray rendőrségi megfigyelési technológia használatára vonatkozó titoktartási megállapodások tárgyalása).

¹⁸⁷ *Lásd pl.*, *Amended Summary Judgment Op.*, *Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist.*, Civil Action No. H-14-1189, at *1 (S.D. Tex. May 4, 2017) ("[A tanárértékelési] pontszámokat összetett algoritmusok generálják, "kifinomult szoftvereket és számítások zámos rétegét alkalmazva számítások [Azeladó]. ezeket az algoritmusokat és szoftvereket üzleti titokként kezeli, és nem hajlandó azokat sem [a kerületnek], sem maguknak a tanároknak kiadni."). *Lásd általában* Kitchin, *Thinking Critically*, *supra* note 6, 20. o. ("[I]t often a

company's algorithms that provide it with a competitive advantage and they are unwilling to expose their intellectual property even with non-disclosure agreements in place").

Az algoritmikus átláthatóság útjában álló üzleti titkok különösen problematikusak a büntető igazságszolgáltatásban, ahol az egyéni szabadság a tét. Az újságíróknak nem sikerült információt szerezniük a NorthPointe COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) elítélési algoritmusáról a nyílt iratokra vonatkozó kérelmek révén, mert állítólag üzleti titokvédelem alatt áll.¹⁸⁸ Az algoritmussal kapcsolatos peres eljárásban a Wisconsin Legfelsőbb Bíróság a COMPAS használatát egy vádlott tisztességes eljárásra való hivatkozással szemben jóváhagyta, de elismerte az átláthatóság problémáját, és előírta, hogy az ítélethozatali jelentésekben tájékoztatni kell a bírakat arról, hogy "az [algoritmus] tulajdonjogi jellegére hivatkozva megakadályozták a tényezők mérlegelésével vagy a kockázati pontszámok meghatározásával kapcsolatos információk közzétételét"¹⁸⁹.

A fentiekben említettek szerint találoztunk olyan joghatóságokkal, amelyek üzleti titokra és titoktartásra hivatkozva nem tudtak többet elárulni előrejelző modelljeikről. Ez igaz volt például a Mesa Municipal Court, az arizonai Pima és Navajo megyei bírósági rendszerre, valamint a kaliforniai San Francisco Superior Court rendszerére, amelyek az Arnold Foundation PSA-Court programját használták.¹⁹⁰ Igaz volt ez Alaszka is, amely az Eckerd Rapid Safety Feedback gyermekekre vonatkozó kockázatértékelést használta.¹⁹¹ Arizona,¹⁹² Kalifornia,¹⁹³ és Alaszka¹⁹⁴ nyílt iratokra vonatkozó törvényei¹⁹⁴ mind mentesítik az üzleti titkokat és a bizalmas információkat, és egyik sem jogosítja fel a nyilvánosságot olyan iratokhoz való hozzáférésre, amelyekkel a kormány nem rendelkezik. Jelentős további vizsgálatot és esetleg peres eljárást igényelne annak megállapítása, hogy a kormányhivatalok jogszerűen jártak-e el. Azt azonban kijelenthetjük, hogy az ügynökségek rendelkeznek hatáskörrel. Több nyilvántartást is nyilvánosságra hozhattak volna egyszerűen a titoktartás körének csökkentésével és az algoritmusok magyarázatához szükséges nyilvántartások kormányzati birtoklásának biztosításával.¹⁹⁵

A nyílt iratokra vonatkozó kérelmekre adott válaszként a bizalmas kezelésre vonatkozó túlzó állítások gyakoriak a területen. Például annak kutatása során, hogy a kaliforniai rendőrkapitányságok hogyan használják a

¹⁸⁸ Lásd Angwin et al., 7. lábjegyzet; lásd még Diakopolous, 9. lábjegyzet (a nyilvánosságra hozott gátló üzleti titoktartási akadályok tárgyalása).

¹⁸⁹ State v. Loomis, 881 N.W.2d 749, 763-64 (Wis. 2016). Azt is nyilvánosságra kellett hozni, hogy nem készültek validációs tanulmányok, és az eszközöket folyamatosan figyelemmel kell kísérni és újra kell normálni. *Id.* 769. o.

¹⁹⁰ Lásd a 129-130. lábjegyzethez fűzött szöveget.

¹⁹¹ Lásd Illinois FY16 szerződés, 145. lábjegyzet; Illinois FY17 szerződés, 145. lábjegyzet.

¹⁹² Arizona Public Records Law, ARIZ. REV. STAT. ANN. § 39-121 (2017).

¹⁹³ California Public Records Act, CAL. GOVT. CODE §§ 6250-6276.48 (West 2017).

¹⁹⁴ Alaska Public Records Disclosures, ALASKA STAT. §§ 40.25.100-350 (West 2017).

¹⁹⁵ Amint azt fentebb említettük, Florida hetedik bírósági kerülete tett egy lépést a

helyes irányba, *lásd a* 129. lábjegyzethez tartozó *fenti* szöveget, de sokkal többet is tehetett volna.

A Forbes riportere, Matt Drange több mint egy tucatnyi állami információszabadságról szóló kérvényt nyújtott be a Shotspotter által generált lövésekről szóló jelentésekre vonatkozóan, hogy a Shotspotter technológia segítségével reagáljanak a joghatóságuk területén leadott lövésekre.¹⁹⁶ Annak ellenére, hogy a kérelmek nem a mögöttes érzékelő technológiára irányultak, a joghatóságok kezdetben azt jelentették, hogy a Shotspotterrel kötött titoktartási megállapodások miatt nem hozhatják nyilvánosságra az adatokat.¹⁹⁷ A kockázatkerülő önkormányzatok úgy gondolták, hogy nem oszthatják meg a joghatóságuk területén észlelt lövésekkel kapcsolatos információkat, még akkor sem, ha az adatok nem minősülnek üzleti titoknak vagy bizalmasnak.

Az üzleti titok védelmére vonatkozó kormányzati állítások az eladók nevében néha indokoltak lehetnek. A kormányzati ügynökök az üzleti titkok felfedéséért és/vagy a titoktartási megállapodások megsértéséért rendes felelősséggel tartoznak, kivéve, ha valamilyen mentesség védi őket.¹⁹⁸ A legtöbb állam elfogadta az egységes üzleti titokról szóló törvényt,¹⁹⁹ amely védelmet nyújt az üzleti titok "eltulajdonítása" ellen, amelyet úgy határoznak meg, mint "egy másik személy üzleti titkának kifejezett vagy hallgatólagos beleegyezés nélküli felfedése vagy felhasználása olyan személy által, aki ... a felfedés vagy felhasználás időpontjában tudta vagy okkal tudhatta, hogy az üzleti titokról való tudása

¹⁹⁶ Matt Drange, *Milliókat költünk erre a high-tech rendszerre, amelyet a fegyveres erők csökkentésére terveztek. Van-e hatása?*, FORBES (2016. november 17.), <http://www.forbes.com/sites/mattdrange/2016/11/17/shotspotter-struggles-to-prove-impact-as-silicon-valley-answer-to-gun-violence/#27e6920c9dbf> [<http://perma.cc/BNX9-ZK8R>].

¹⁹⁷ A vállalat 2015 júliusában országos feljegyzést küldött ki az ügyfeleknek, amelyben sürgette a városokat, hogy adjanak ki általános elutasító válaszokat a nyilvántartási kérelmekre, vagy adjanak ki erősen szerkesztett információkat, "olyan formában, amely nem károsítja az SST üzletét, és lehetővé teszi az ügyfél számára, hogy a közérdekűség szempontjából válaszoljon". Customer

SuccessTrainingBulletin,SST(2015),<http://www.documentcloud.org/documents/3221020-ShotSpotter-nationwide-memo-July-2015.html> [<http://perma.cc/F2DB-8AN9>]; lásd még Jason Tashea, *Should the Public Have Access to Data Police Acquire Through Private Companies?*, ABAJ.(2016.dec.1

),http://www.abajournal.com/magazine/article/public_access_police_data_privat_e_company [<http://perma.cc/B3L2-DYMK>] (beszámoló az önkormányzati kellemetlenségekről a Shotspotter által a joghatóságon belül leadott lövésekről gyűjtött adatok tulajdonjogára vonatkozó állításokkal kapcsolatban).

¹⁹⁸ *Lásd pl.*, WASH. REV. CODE § 42.56.060 (2017) ("Egyetlen közhivatal, köztisztviselő, közalkalmazott vagy letétkezelő sem felelős, és nem áll fenn keresetindítási jogalap sem a közokirat kiadásán alapuló veszteségért vagy kárért, ha a közhivatal, köztisztviselő, közalkalmazott vagy letétkezelő jóhiszeműen járt el, amikor megpróbált megfelelni e fejezet rendelkezéseinek."); *accord* Levine v. City of Bothell, 2015 WL 2567095 at *3 (W.D. Wash. 2012) (elismeri, hogy "egy közhivatal és annak alkalmazottai mentesülnek a felelősség alól a közhiteles iratok kiadása esetén, ha jóhiszeműen jártak el, amikor megpróbáltak megfelelni" a washingtoni nyílt iratokról szóló törvénynek); *vö.* Peter S. Menell, *Tailoring a Public Policy*

Exception to Trade Secret Protection, 105 CAL. L. REV. 1, 30-31 (2017) (a magánfeleknek az üzleti titkok közérdekű nyilvánosságra hozatalára vonatkozó kiváltságait tárgyalja).

¹⁹⁹ A törvényt elfogadó joghatóságok táblázatát lásd: UNIF. TRADE SECRETS ACT (UNIF. LAW COMM'N 1985) (West 2016).

a titkot ... olyan körülmények között szerezték meg, amelyek alapján kötelesek voltak titokban tartani vagy korlátozni annak használatát. ²⁰⁰ A kormányok személyek, és ezért potenciálisan felelősek. ²⁰¹ Így az eladók a titoktartás felé vonzódnak azért, hogy védelmet érvényesítsenek, vagy megkövetelik, hogy a kormánytisztviselők titoktartási megállapodásokat írjanak alá. Ezt a vonzerőt erősíti a kormányzati szerv saját titoktartási érdeke, az alább tárgyalt okok miatt.

A nyílt iratokról szóló törvények nem gyakorolnak akkora ellenhatást az átláthatóságra, mint amekkorát gyakorolhatnának, mivel általában mentesítik az üzleti titkokat. ²⁰² A FOIA 4. mentessége számos szoros párhuzamot mutat az állami nyílt iratokról szóló törvény mentességeivel. A mentesség kizárja a nyilvánosságra hozatal alól "az üzleti titkokat és a személytől kapott kereskedelmi vagy pénzügyi információkat, [amelyek] kiváltságosak vagy bizalmasak".²⁰³ A mentesség tehát két széles kategóriát fed le: (1) üzleti titkok; és (2) olyan információk, amelyek a) kereskedelmi vagy pénzügyi jellegűek, b) egy személytől származnak, és c) kiváltságosak vagy bizalmasak.

A vizsgálat során ezek az üzleti titok alóli mentességek szűkebbek, mint ahogyan azt a vállalatok állítják. Túlságosan nagyvonalú ügynökség

²⁰⁰ *Id.* § (1)(2)(ii).

²⁰¹ *Id.* § (1)(3). Ezen túlmenően a szövetségi törvény kifejezetten tiltja az üzleti titkok felfedését. 18 U.S.C. § 1905 (büntetőjogi felelősségre vonás minden olyan amerikai kormányzati alkalmazottat, aki hivatalos feladatai ellátása során "közvetesen, nyilvánosságra hoz, nyilvánosságra hoz vagy nyilvánosságra hoz bármilyen módon vagy mértékben, a törvény által nem engedélyezett módon vagy mértékben bármilyen információt ... amely információ üzleti titokra, eljárásokra, műveletekre, munkastílusra vagy készülékekre, vagy bármely személy személyazonosságára, bizalmas statisztikai adataira, jövedelmének, nyereségének, veszteségének vagy kiadásainak összegére vagy forrására vonatkozik vagy azokkal kapcsolatos").

²⁰² *Lásd pl.*, CONN. GEN. STAT. § 1-210(b)(5) (2017) (amely mentesíti a nyilvánosságra hozatal alól az "üzleti titkot", amelyet a következőképpen határoznak meg: A) "információ, beleértve a képleteket, mintákat, összeállításokat, programokat, eszközöket, módszereket, technikákat, eljárásokat, rajzokat, költségadatokat vagy ügyféllistákat, amelyek i) tényleges vagy potenciális független gazdasági értéket képviselnek, mivel nem közismertek és megfelelő eszközökkel nem szerezhetőek meg könnyen más személyek számára, akik gazdasági értéket szerezhetnek a közzétételükből vagy felhasználásukból, és ii) a körülményekhez képest ésszerű erőfeszítések tárgyát képezik a titoktartás fenntartása érdekében; és (B) Bizalmasan adott kereskedelmi vagy pénzügyi információk, amelyeket nem ír elő jogszabály."); DEL. CODE ANN. 29 § 10002(g)(2) (2017) (nem tekinti nyilvánosnak "[t]árutitkokat és egy személytől kapott kereskedelmi vagy pénzügyi információkat, amelyek kiváltságosak vagy bizalmas jellegűek"); Pennsylvania Right-to-Know Law, 65 PA. STAT. AND CONS. STAT ANN. § 67.708 (b)(11) (mentesíti a nyilvánosságra hozatal alól "[olyan] feljegyzéseket, amelyek üzleti titkot vagy bizalmas tulajdonosi információt képeznek vagy tárnak fel"). Más államokban a bíróságok általánosabb mentességek alapján ismerik el az üzleti titkokat. *Lásd pl.*, Phx. Newspapers, Inc. v. Keegan, 35 P.3d 105, 112 (Ariz. Ct. App. 2001) (elismeri, hogy az üzleti titkokat "a nyilvánosságra hozatal alóli titoktartási kivétel védi" az arizonai nyílt iratokról szóló törvényben). *Lásd általában* Linda B. Samuels, *Protecting Confidential Business Information Supplied to State Governments: Exempting Trade Secrets from State Open Records Laws*, 27 AM. BUS. L.J. 467, 468-69 (1989) (az

állami üzleti titkok nyílt iratokra vonatkozó törvények alóli mentességeinek
tárgyalásával); *Open Government Guide*

,REPORTERSCOMMITTEEFORFREEDOMPRESS ,

<http://www.rcfp.org/ogg/index.php> [<http://perma.cc/2QNF-V83V>] (mind az
ötven állam nyílt iratokra vonatkozó törvényének linkje).

²⁰³ 5 U.S.C. § 552(b)(4) (2016).

a védintézkedéseket megtámadás esetén megsemmisítették. A D.C. Circuit - a FOIA esetjogának vezető forrása - az "üzleti titok" kifejezést az egységes üzleti titokról szóló törvény²⁰⁴ (és a 2016. évi szövetségi Defend Trade Secrets Act) értelmében szűkebb értelemben értelmezte.²⁰⁵ A kormány a 4. mentesség alapján csak olyan "titkos, kereskedelmi szempontból értékes terv, formula, eljárás vagy eszköz esetében tarthat vissza iratokat, amelyet [kereskedelmi árukkal kapcsolatban] használnak, és amelyről elmondható, hogy vagy innováció vagy jelentős erőfeszítés végterméke."²⁰⁶ A kérdéses információnak "közvetlen kapcsolatban kell állnia a termelési folyamattal", nem pedig pusztán "járulékos üzleti titoktartás"²⁰⁷ Más szóval, az eltitkolt információnak a kereskedelmi termék szempontjából központi szerepet kell játszania, és nem csupán mellékes mellékterméknek. E korlátozásra tekintettel nem minden olyan algoritmikus folyamat, amelyet egy eladó üzleti titoknak tekinthet a kereskedelmi szférában, számíthat üzleti titoknak a nyílt iratok alóli mentesség szempontjából.

A 4. mentesség második pontja lehetővé teszi a titoktartást bizonyos típusú pénzügyi vagy kereskedelmi információk esetében.²⁰⁸ A mentességnek ez a része is korlátozott. Az információnak "kiváltságosnak vagy bizalmasnak" kell lennie.²⁰⁹ A D.C. Circuit úgy ítélte meg, hogy az információ forrásának tett puszta titoktartási ígéret nem elegendő. A kormányzatnak inkább azt kell bizonyítania a kötelező nyilvántartások tekintetében, hogy a nyilvánosságra hozatal valószínűleg (1) "károsítaná a kormányzat

²⁰⁴ UNIF. TRADE SECRETS ACT § 1.4 (az üzleti titok meghatározása: "információ, beleértve a képletet, mintát, mintát, összeállítást, programot, eszközt, módszert, technikát vagy eljárást, amely: (i) tényleges vagy potenciális, független gazdasági értéket képvisel, mivel nem közismert, és megfelelő eszközökkel nem könnyen megismerhető más személyek számára, akik gazdasági értéket szerezhetnek a közzétételéből vagy felhasználásából, és (ii) a körülményekhez képest ésszerű erőfeszítések tárgyát képezi a titoktartás fenntartása érdekében").

²⁰⁵ 18 U.S.C. § 1839(3) (2016) (beleértve "a pénzügyi, üzleti, tudományos, műszaki, gazdasági vagy mérnöki információk minden formáját és típusát, beleértve a mintákat, terveket, összeállításokat, programeszközöket, képleteket, terveket, prototípusokat, módszereket, technikákat, eljárásokat, eljárásokat, programokat vagy kódokat, akár tárgyi, akár immateriális, és akár fizikailag, elektronikusan, grafikusán, fényképen vagy írásban tárolt, összeállított vagy megörökített formában vagy módon").

²⁰⁶ Pub. Citizen Health Research Grp. v. U.S. Food & Drug Admin., 704 F.2d 1280, 1288-89 n.25. (D.C. Cir. 1983) ("A Restatement megközelítése a vétkességre és a hűtlen kezelésre helyezett hangsúlyával nem alkalmas arra, hogy megfelelő egyensúlyt teremtsen a szabályozott iparágak és a közvélemény egymással versengő érdekei között."); lásd még Anderson v. Dep't of Health & Human Servs., 907 F.2d 936, 944 (10th Cir. 1990) (ugyanazt a meghatározást elfogadva).

²⁰⁷ Pub. Citizen, 704 F.2d at 1287-88.

²⁰⁸ Guide to the Freedom Of Information Act, U.S. DEPT JUSTICE (2004. május), <http://www.justice.gov/oip/foia-guide-2004-edition-exemption-4> [<http://perma.cc/LC5X-EJPK>] (amely szerint a nyilvántartások akkor minősülnek kereskedelmi jellegűnek, ha a benyújtónak "kereskedelmi érdeke fűződik hozzájuk").

²⁰⁹ HARRY A. HAMMITT ET AL., LITIGATION UNDER THE FEDERAL OPEN

GOVERNMENT LAW 119 (25th ed. 2010) (A "bizalmas" kifejezés "a kulcsfogalom a 4. mentességi joggyakorlatban").

a szükséges információk jövőbeni beszerzésének képességét", vagy (2) "jelentős kárt okoz az információforrás versenyhelyzetében".²¹⁰ Az önkéntesen nyilvánosságra hozott nyilvántartások tekintetében a kereskedelmi és pénzügyi információk akkor "bizalmasak", ha a forrás ezeket az információkat szokás szerint nem adná ki a nyilvánosságnak.²¹¹ Összefoglalva, a FOIA üzleti titok alóli mentessége csak a "céges" üzleti titkok egy szűk körére és a pénzügyi vagy kereskedelmi információk egy részhalmazára vonatkozik. A nyilvánosságra hozatal véelme megmarad. Tekintettel a kormányt terhelő teherre, egyes államok megkövetelik, hogy az állami szervek értesítsék a magánszervezeteket az üzleti titokra vagy bizalmas információkra vonatkozó kérésekről, és kérik a magánfél védekezését a kijelölésükkel kapcsolatban.²¹²

Az üzleti titokra vonatkozó mentességek értelmezésekor a kormánytisztviselőknek szem előtt kell tartaniuk a kivételek célját. A FOIA 4. mentesítésének célja, hogy megőrizze a kormányzat azon képességét, hogy információkat gyűjtsön a szabályozott szervezetektől,²¹³ vagy egy másik megfogalmazás szerint "arra ösztönözze az egyéneket, hogy bizonyos típusú bizalmas információkat adjanak át a kormányzatnak".²¹⁴ Hasonlóképpen, az állami nyílt iratokra vonatkozó törvények az üzleti titkokat és a bizalmas információkat a közcélok előmozdítása érdekében védik.²¹⁵ Mivel a nyílt iratokra vonatkozó törvények a nyitottság véelmét írják elő, és mivel az államok követték a FOIA bíróságait az üzleti titkok és bizalmas anyagok szűken értelmezésében, az üzleti titkok nyílt iratokra vonatkozó mentessége szűkebb, mint azt a magáneladók szeretnék. Ez különösen igaz akkor, ha a kormányzat ügyfélként és nem szabályozó hatóságként jár el, mert a titoktartás nem segíti elő a kormányzat szabályozási hatáskörét. A jelenleg New Yorkban folyamatban lévő peres ügyek azt a kérdést vetik fel, hogy milyen mértékben kell tiszteletben tartani az üzleti titokra vonatkozó igényeket, amikor a kormányzat cselekszik.

²¹⁰ Nat'l Parks & Conservation Ass'n v. Morton, 498 FD 765, 770 (D.C. Cir. 1974).

²¹¹ Critical Mass Energy Project v. Nuclear Regulatory Comm'n, 975 FD 871, 872-74 (D.C. Cir. 1992).

²¹² *Lásd pl.* Pennsylvania Right-to-Know Law, 65 PA. STAT. AND CONS. STAT. ANN. 67.707. § (előírja, hogy az állami ügynökségnek öt munkanapon belül értesítenie kell a vállalatot az üzleti titok vagy bizalmas információ felfedésére irányuló kérelemről. A vállalatnak ezután öt munkanap áll rendelkezésére, hogy közölje az állami ügynökséggel a vállalat álláspontját az információk nyilvánosságra hozatalával kapcsolatban. A vállalat értesítését követő tíz napon belül az állami ügynökségnek döntenie kell az információ kiadásáról vagy visszatartásáról).

²¹³ Az Egyesült Államok Igazságügyi Minisztériuma szerint a 4. mentesség "védelmet nyújt azoknak a benyújtóknak, akiknek kereskedelmi vagy pénzügyi információkat kell szolgáltatniuk a kormánynak, megvédve őket a nyilvánosságra hozatalból eredő versenyhátrányoktól". *Guide to the Freedom Of Information Act, supra* note 208; *lásd még id.* n. 2. pont.

²¹⁴ Soucie v. David, 448 F.2d 1067, 1078 (D.C. Cir. 1971).

²¹⁵ *Lásd pl.* Verizon N.Y., Inc. kontra N.Y. State Pub. Serv. Comm'n, 23 N.Y.S.3d

446 (N.Y. App. Div. 2016) ("[T]he policy behind Public Officers Law § 87(2)(d) (§ 87(2)(d)) is simply to protect businesses from the adverse consequences of disclosure of confidential commercial information, so to promote the State's economic development efforts and attract business to New York.").

vállalati minőségében. Az üzleti titok védelmére hivatkozva ²¹⁶ New York City elutasította a Brennan Center for Justice információszabadságról szóló törvény alapján benyújtott kérelmét a Palantir Gotham néven ismert ítélezési algoritmussal kapcsolatos feljegyzésekre vonatkozóan. ²¹⁷

Az általunk benyújtott valamennyi kérelmet a vállalkozási jogkörökben eljáró joghatóságokhoz nyújtottuk be. Biztosak lehetünk abban, hogy az algoritmusokkal kapcsolatos összes anyagra vonatkozó üzleti titok állítása túlságosan széleskörű volt. Még ha feltételezzük is, hogy a forráskód és a modell bizonyos részletei üzleti titoknak vagy bizalmas információnak minősülnek, képzési anyagokat, meglévő és tervezett validációs tanulmányokat, valamint az algoritmusban tükröződő célokra és tervezési döntésekre vonatkozó egyéb dokumentációt kértünk. Nehéz elképzelni, hogy ezeknek az anyagoknak a nagy része, vagy egyáltalán bármelyike megfelelné a mentességnek.

Szinte bizonyos, hogy az algoritmusok üzleti titokként való védelme néha ösztönzi a vállalatokat, hogy nyilvános alkalmazásokhoz készítsenek előrejelző modelleket. ²¹⁸ Ugyanakkor az állítólagosan az üzleti titokjog által védett információk alapvető közfunkciók középpontjában állhatnak, és régóta vizsgálható politikai döntéseknek minősülhetnek. Ahogy David Levine írja, "az üzleti titoktartás és az átlátható és elszámoltatható demokratikus kormányzat közötti konfliktus végső soron a kormányzás elméletének és értékeinek összeütközése." ²¹⁹ Ezt a konfliktust enyhíthetik a bíróságok és a jogalkotók az üzleti titok alóli mentesség hatályának a nyílt iratokra vonatkozó törvényekre való korlátozásával, valamint azzal, hogy a kormányzati szervek ragaszkodnak az átláthatósághoz, amikor algoritmusokra szerződnek.

²¹⁶ N.Y. PUB. OFF. L. § 87(2)(d) (mentesíti a nyilvánosságra hozatal alól azokat a nyilvántartásokat, amelyek "üzleti titok, vagy amelyeket egy kereskedelmi vállalkozás nyújtott be egy ügynökségnek, vagy amelyek egy kereskedelmi vállalkozástól kapott információból származnak, és amelyek nyilvánosságra hozatala jelentős kárt okozna az érintett vállalkozás versenyhelyzetében").

²¹⁷ Memorandum of Law in Support of Verified Petition, Brennan Center for Justice at New York University School of Law v. New York City Police Department, Case No. 0160541/2016 (N.Y. Sup. Ct. 2016) (Doc. No. 8), <http://www.brennancenter.org/sites/default/files/8%20-%20Memorandum%20of%20Law%20in%20Support%20of%20Verified%20Petition.pdf> [<http://perma.cc/E5E6-Q7HB>].

²¹⁸ Kroll et al., *supra* note 26, 15-17. o.; lásd még David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 180-81 (2007) (a szavazógépekkel összefüggésben példát hozva arra, hogy a forráskódok közzétételét előíró állami törvények hogyan tarthatják vissza a vállalatokat attól, hogy az állammal közszolgáltatásokra szerződést kössenek).

²¹⁹ Lásd Levine, *Supra* 218. lábjegyzet, 157. o.; lásd még Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885, 918-19 (2006) (megfigyelve egy "alapvető konfliktust a kormányzati ügynökségekre vonatkozó törvények és az ügynökségek egyre nagyobb mértékben magáncégekre való támaszkodása között", és megjegyezve, hogy az állami bíróságok és törvényhozók "nem tudtak konszenzust vagy egyértelműséget kialakítani a

nyílt kormányzati törvények tekintetében", hogy kezeljék ezt a konfliktust).

**C. Egyéb kormányzati érdekek és nyílt iratok
törvény mentességek**

Még ha a kormányhivatalok elegendő nyilvántartást hoztak létre vagy szereztek is be, és biztosították, hogy ezek a nyilvántartások nem tartoznak az üzleti titoktartás hatálya alá, más okuk is lehet arra, hogy ellenálljanak az algoritmikus átláthatóságnak: játék vagy kijátszás; a tanácskozási őszinteségének elvesztése; és indokolatlan nyilvános viták.

A kormánytisztviselők aggódhatnak amiatt, hogy a nyilvánosan közzétett algoritmusokat kijátsszák vagy megkerülik, ami az előrejelzéseket kevésbé megbízhatóvá teszi, és megghiúsítja azok célját.²²⁰ Ha egy büntetőjogi vádlott tudja, hogy az általa tett kijelentések magasabb visszaesési kockázati pontszámot eredményeznek, akkor hazudhat.²²¹ Ha egy terrorista tudja, hogy a neveket hogyan helyezik el a terrorista átvilágítási adatbázisban, és hogyan egyeztetik össze a vízumkérelmekben szereplő nevekkkel, megpróbálhatja elkerülni az ilyen elhelyezést és egyeztetést.²²²

Ezek az aggodalmak érthetőek, de nem mentés arra, hogy nem reagálnak a nyílt iratok iránti kérelmekre. A nyílt iratokról szóló törvények foglalkoznak a bűnüldözési vizsgálatok és nyomozási technikák összefüggésében történő lehetséges játékokkal.²²³ A FOIA 7(E) mentessége kifejezetten megkérdezi, hogy a nyomozati iratok nyilvánosságra hozatala

²²⁰ A gépi tanulással kapcsolatos szakirodalomban a szerencsejáték-probléma általánosabban "ellenséges tanulásként" ismert - a modellek fejlesztésének problémája, amikor már a kezdetektől fogva várható, hogy az ellenfelek megpróbálják majd legyőzni őket. *Lásd pl.* Daniel Lowd & Christopher Meek, *Adversarial Learning*, in PROCEEDINGS OF THE ELEVENTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (KDD) 641 (Robert Grossman et al. eds., 2005); Pavel Liskov & Richard Lippmann, *Machine Learning in Adversarial Environments*, 81 MACHINE LEARNING 115 (2010).

²²¹ Például a COMPAS, a bűnügyi vádlottak visszaesésének valószínűségét felmérő eszköz részben arra alapozza előrejelzéseit, hogy a vádlott egyetért-e vagy sem az olyan kijelentésekkel, mint például "Az éhes embernek joga van lopni" vagy "Ki lehet beszélni magunkat a problémából". *Lásd* Brittney Via et al., *Exploring How to Measure Criminogenic Needs: Five Instruments and No Real Answers*, in HANDBOOK ON RISK AND NEED ASSESSMENT: THEORY AND PRACTICE (Faye S. Taxman szerk., 2016).

²²² *Lásd* JEROME P. BJELOPERA, BART ELIAS & AARON SISKIN, CONG. RESEARCH SERV., R44678, THE TERRORIST SCREENING DATABASE AND PREVENTING TERRORIST TRAVEL 12 (2016) (dokumentálva a névkereső algoritmusok használatát a vízumkérelmezők szűrése során). A végrehajtás és a megelőzés közötti homályos határvonal vizsgálatához *lásd* Coglianesse & Lehr, *Supra* 58. lábjegyzet, 1210. o. (megjegyezve, hogy egy algoritmikus szabályalkotási folyamat modellezhető a szabályozott szervezetek megfelelési döntéseit, amely esetben az utólagos végrehajtási algoritmusokhoz hasonló lenne, és jogosan mentesülhetne a közzététel alól).

²²³ *Lásd pl.*, 5 U.S.C. § 552(b)(7)(E) (mentesíti a "bűnüldözési célokra összeállított nyilvántartásokat vagy információkat", amelyek nyilvánosságra hozatala "ésszerűen elvárhatóan a törvény megkerülésének kockázatával járhat"); 5 ILL. COMP. STAT. 140/7(d)(v) (mentesíti azokat a bűnüldözési nyilvántartásokat, amelyek "az általánosan használtaktól eltérő, egyedi vagy speciális nyomozási technikákat tárnak fel"); MICH. COMP. LAWS

§15.243(1)(b)(v) (mentesítve az olyan nyilvántartások alól, amelyek "[a]z igazságügyi nyomozati technikákat vagy eljárásokat nyilvánosságra hoznák").

technikák "a törvény megkerülésének kockázatát jelentenek".²²⁴ A "nyomozás" azonban alapvetően a már elkövetett bűncselekmények elkövetőinek azonosítására és a bizonyítékok összegyűjtésére vonatkozik. Kivételes esetekben egyes bíróságok hajlandóak voltak a "nyomozás" fogalmát kiterjeszteni bizonyos megelőző intézkedésekre,²²⁵ és az egyik nyílt iratokra vonatkozó kérelmünkből kiderült, hogy az egyik joghatóság mentesítette magát a rendőrségi megfigyelési technikákkal kapcsolatos adatok szolgáltatása alól, ami vitathatatlanul a megelőzés egyik rokona.²²⁶ A PredPolhoz és a HunchLabhoz hasonló prediktív rendőrségi programok azonban, amelyek inkább az elrettentésre, mint a nyomozásra összpontosítanak, a mentesség periferiáján vagy azon túl helyezkednek el. A bűnelkövetők kockázatértékelése a visszaesés és a megjelenés elmulasztása tekintetében még kevésbé tűnik a "nyomozáshoz" kötődőnek. Ezen túlmenően a nyílt iratokra vonatkozó törvények alól nem mentesülnek a nem büntető igazságszolgáltatási játékokkal kapcsolatos egyéb aggályok. Az olyan gyermekjóléti programok, mint az Eckerd Rapid Safety Feedback és az Allegheny Family Screening Tool erőfeszítései nem elsősorban a bűnüldözéshez kapcsolódnak.²²⁷

Az ügynökségek a manipulációval szemben viszonylag védett algoritmusok elfogadásával tudnak a legjobban megbirkózni a játékkal kapcsolatos aggályokkal. Az Arnold Alapítvány például azt állítja, hogy

²²⁴ 5 U.S.C. § 552(b)(7)(E).

²²⁵ *Lásd Coastal Delivery Corp. v. U.S. Customs Serv.*, 272 F. Supp. 2d 958 (C.D. Cal. 2003) (megállapítva, hogy a vámszolgálat a 7. E. mentesség alapján visszatarthatja a különböző tengeri kikötőkbe érkező áruk vizsgálatának számáról szóló nyilvántartásokat, mivel azok segíthetik az áruk illegális behozatalát azáltal, hogy tájékoztatják az importőröket arról, hogy hol és mikor kevesebb vizsgálatra kerül sor);

U.S. News & World Report v. Dept. of the Treasury, 1986 U.S. Dist LEXIS 27634 (D.D.C.) (megállapítva, hogy az elnök limuzinjainak építésével kapcsolatos részletek visszatarthatók a 7(E) mentesség alapján, és elfogadva a "nyomozás" tág értelmezését, amely magában foglalja az elnöknek okozott lehetséges károk megelőzését). *Lásd azonban a Living Rivers, Inc. kontra U.S. Bureau of Reclamation*, 272 F. Supp. 2d 1313, 1320-22 (D. Utah 2003) (amely szerint a gátak átszakadása esetén elöntött területekről készült térképek nem tarthatók vissza a 7(E) mentesség alapján, mivel a térképek nem fedtek fel vizsgálati gyakorlatot).

²²⁶ A floridai Cocoa városa küldött nekünk egy dokumentumot, amelyben megjegyzi, hogy a PredPolról szóló részleteket nem adnak meg nyilvános dokumentumban, mivel "a megfigyelési technikákat, eljárásokat vagy személyzetet feltáró információk" a floridai nyílt iratokról szóló törvény értelmében mentesülnek a nyilvánosságra hozatal alól. *Lásd: Legislation Details (With Text)*, File # 15- 361, *supra* note 164; FLA. STAT. § 119.071(2)(d) (2017) ("Minden olyan információ, amely megfigyelési technikákat, eljárásokat vagy személyzetet tár fel, mentesül a s. 119.07(1) és a s. 24(a), Art. I. cikke szerinti állami alkotmány."). Még ha a rendőrségi személyzet meghatározott területeken, meghatározott időpontokban történő bevetésének rendszere megfigyelési technikának vagy eljárásnak is minősül, a megfigyelésre vonatkozó külön mentesség nem gyakori a nyílt iratokról szóló törvényekben.

²²⁷ Amint fentebb említettük, *lásd a* 16. oldalon, a kormánytisztviselők aggódhatnak az algoritmusok nyilvánosságra hozatalának véletlenszerű, káros viselkedési hatásai miatt is, mint például a szükséges mentális

egészségügyi kezelés elkerülése azok által, akik megtudják, hogy az ilyen kezelésben való részesülés a gyermekjóléti kockázatértékelésnél tényező. A szerencsejátékhoz hasonlóan ez is jogos aggodalom lehet az átláthatósággal szemben; nincs olyan mentesség a nyílt iratokra vonatkozóan, amely ezzel foglalkozna.

A PSA-Court, amely kizárólag a vádlott előéletére vonatkozó objektív, ellenőrizhető tényekre támaszkodik, ugyanolyan pontos kockázatértékelést készít, mint a vádlottak szubjektív nyilatkozataira támaszkodó algoritmusok.²²⁸ Az Azavea véletlenszerűséget vezetett be a HunchLab prediktív rendőri algoritmusába, ami többek között megghiúsítaná a járőrözési tervek levezetésére irányuló erőfeszítéseket még egy nyilvánosságra hozott algoritmusból is.²²⁹

A tisztviselőknek az is aggályos lehet, hogy nem akarják felfedni a prediktív algoritmusokkal kapcsolatos bizonytalan gondolkodásukat. Mind a FOIA, mind számos állami nyílt iratokról szóló törvény tartalmaz egy mentességet a végrehajtó hatalmi ágon belüli tanácskozási folyamat védelmére.²³⁰ Egyetlen nyílt iratigénylésünket sem utasították el a végrehajtó hatalmi ági tanácskozási folyamatra vonatkozó mentesség alapján, így az ilyen mentesség alkalmazása az algoritmikus folyamatokra továbbra is spekulatív marad. A mérlegelési folyamatra vonatkozó kiváltság feltételezi, hogy az ügynökségek már bejelentettek egy szabályt és kifejtették annak indoklását. A tanácskozási folyamatra vonatkozó mentesség lényege, hogy "megvédjen a kérdések összezavarásától és a nyilvánosság félrevezetésétől az olyan dokumentumok terjesztése által, amelyek olyan okokat és indokokat sugallnak egy intézkedéshez, amelyek valójában nem voltak az ügynökség intézkedésének végső okai."²³¹ Ha a kormányzat soha nem magyarázza el egy algoritmus "szabályait" vagy azt, hogy miért fogadták el azt, akkor nincs olyan mérvadó nyilatkozat, amely megvédhetné az eltévedt tanácskozástól. Valójában az algoritmus kidolgozása során keletkezett feljegyzések jelentenek az egyetlen ablakot az algoritmikus folyamathoz kötődő szabályokra és ésszerűségekre.

Az igazságszolgáltatási ágazat gyakran mentesül a nyílt iratokra vonatkozó törvények alól.²³² Számos, a nyílt iratokról szóló törvényre vonatkozó kérelmünket azzal az indokkal utasították el, hogy a bíróságok nem tartoznak megfelelően a kérelem tárgykörébe. Nem állíthatjuk, hogy ez minden esetben helytelen volt, de ez

²²⁸ *Lásd: Developing a National Model for Pretrial Risk Assessment*, LAURA & JOHN ARNOLDFOUND (Nov.2013)

,http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf

[<http://perma.cc/KTY8-VCP9>] (megjegyezve, hogy más kockázatértékelési eszközök "olyan adatokra támaszkodtak, amelyeket csak a vádlottal folytatott kihallgatások révén [lehetett] gyűjteni", a PSA-Court pedig kizárólag olyan adatokat használ, amelyek "a vádlott büntetett előéletéből származnak").

²²⁹ *Lásd A Citizen's Guide to HunchLab*, 77. lábjegyzet, 10-11. o.

²³⁰ *Lásd pl.*, 5 U.S.C. § 552(b)(5) (mentesítve az "ügynökségek közötti vagy ügynökségen belüli memorandumokat vagy leveleket, amelyek a törvény szerint nem állnának az ügynökséggel perben álló más fél rendelkezésére"); 5 ILL. COMP. STAT. 140/7(f) (mentesítve "[p]reliminary draft, notes, recommendations, memoranda and other records in which are expressed opinions, or policies or actions are formulated"); N.Y. PUB. OFF. L. 87(2)(g) (mentesítve a legtöbb "ügynökségek közötti vagy ügynökségen belüli anyagot").

²³¹ *Coastal States Gas Corp. v. Dep't of Energy*, 617 F.2d 854, 866 (D.C. Cir. 1980).

²³² *Lásd pl.*, 5 U.S.C. § 552(f)(1) (az "ügynökség" meghatározása a bíróságok kizárásával); 65 PA. STAT. & CONS. STAT. ANN. § 67.304 (West 2017) (amely csak a "[j]udicial agencies" számára írja elő a pénzügyi nyilvántartásokhoz való hozzáférést).

kellene lennie. Egy bírósági rendszer algoritmusának kidolgozása és elfogadása kevéssé hasonlít az egyedi ügyekben hozott (általában egyébként is nyilvános magyarázatokkal megvilágított) bírói döntésekhez. Sokkal inkább hasonlít egy olyan bizonyítási szabály kidolgozásához és elfogadásához, amelyet az esetek nagy számára alkalmaznak. A bírósági szabályalkotás, akár csak a közigazgatási szabályalkotás, jellemzően nyilvánosan történik. A szövetségi törvények előírják, hogy a Legfelsőbb Bíróságon kívüli szövetségi bíróságok által kihirdetett szabályokat "csak megfelelő nyilvános értesítés és észrevételezési lehetőség biztosítása után írhatják elő"²³³, és a Legfelsőbb Bíróság is alkalmazza az értesítés és észrevételezés alapú szabályalkotást a Bírósági Konferencia által kiadott eljárások alapján.²³⁴ Az állami bíróságok hasonló nyilvános eljárásokkal rendelkeznek.²³⁵ Az algoritmus kidolgozásának és elfogadásának folyamatára vonatkozó, a nyílt iratokra vonatkozó megbízás hiányában a bíróságoknak fontolóra kell venniük a szabályok elfogadásához és módosításához alkalmazott eljáráshoz hasonló nyilvános eljárás valamilyen formáját.

Végezetül a kormányok aggódhatnak amiatt, hogy egyes választópolgároknak kényelmetlen az algoritmusok alkalmazása, diszkriminációt vagy tisztességtelenséget fognak észlelni ott, ahol nincs ilyen, vagy indokolatlanul vitatják az algoritmikus ajánlásokat. A kormányok a torzításokon, tudománytalan következtetéseken vagy tévedéseken alapuló, szerintük indokolatlan viták elkerülése érdekében inkább nem hozzák nyilvánosságra az algoritmikus modelleket. Nem ismerünk olyan mentességet a nyílt iratokról szóló törvény alól, amely megakadályozná a vitás kérdések nyilvánosságra hozatalát, és bár a kormánytisztviselők joggal tarthatnak a torzításoktól és a tudománytalan következtetésektől, a viták elkerülhetetlenek a demokratikus folyamatban. Gyakran a középpontjában áll.

V. FIXES

Hogyan segíthetik elő a kormányok az átláthatóságot a prediktív algoritmusok használata során? A törvényhozás valószínűleg nem vonja vissza az üzleti titkok és más bizalmas információk védelmét.²³⁶ Még ha ez meg is történne, az üzleti titoktartalom megszüntetése

²³³ 28 U.S.C. § 2071(b).

²³⁴ *Lásd* ADMIN. AZ A M E R I K A I BÍRÓSÁGOK HIVATALA, A BIZOTTSÁGOK ELJÁRÁSAI ELJÁRÁSI ÉS GYAKORLATI SZABÁLYZAT 440 .20.40 . §

<http://www.uscourts.gov/rules-policies/about-rulemaking-process/laws-and-procedures-governing-work-rends-committees-0> [<http://perma.cc/MEL6-GJ3A>].

²³⁵ *Lásd pl.*, ILL. S. CT. R. 3(a)(1) (2017) (amely olyan elemeket tartalmazó szabályalkotási folyamatot ír elő, mint "minden ... javasolt szabály és javasolt módosítás nyilvános nyilvántartása" és "a nyilvánosság, a bírói kar és az ügyvédi kamara észrevételeinek és javaslatainak lehetősége").

²³⁶ Azzal az érveléssel kapcsolatban, hogy az üzleti titoktartás nem használható fel

arra, hogy egy prediktív algoritmussal kapcsolatos információkat visszatartsanak egy büntetőjogi vádlottól, lásd Rebecca Wexler, *Life, Liberty and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. (megjelenés előtt 2018).

a titokvédelem önmagában nem oldaná meg a nem megfelelő dokumentáció és a magánjellegű iratanyag birtoklásának problémáját. Sokkal eredményesebb lenne, ha a kormányok szerződéskötési jogkörükkel ragaszkodnának a megfelelő nyilvántartások létrehozásához, rendelkezésre bocsátásához és közzétételéhez.²³⁷ Először a rendelkezésre bocsátási és közzétételi követelményeket vizsgáljuk meg, majd rátérünk a nyilvántartások létrehozásával kapcsolatos legjobb gyakorlatokra.

**A. Anyilvántartások rendelkezésre bocsátását
megkövetelő és azok közzétételét
engedélyező szerződéses nyelvezet**

Az állami szervek és a vállalkozók közötti megállapodások, amelyeket a nyílt iratok iránti kérelmek révén szereztünk be, azt mutatják, hogy a kormányok nem egyformán engednek, és nem is kell, hogy engedjenek a vállalkozók titoktartási és adattulajdonlási kívánságainak.

Például úgy tűnik, hogy amikor az Arnold Alapítvány a PSA-programjához egy szabványos egyetértési nyilatkozatot készített, az a titoktartásra vonatkozó erős, széleskörű megfogalmazást tartalmazott. Azok a bíróságok, amelyek nem kértek változtatást ezen a nyelven, ígéretet tettek arra, hogy a PSA-val kapcsolatos minden információjukat bizalmasan kezelik.²³⁸ A floridai hetedik bírósági kerület azonban nyilvánvalóan olyan nyelvet kért, amely lényegesen szűkebb titoktartási kötelezettséget írt elő. Az Arnold Alapítványra hárította az üzleti titkok megjelölésének, a nem védett anyagok szerkesztésének és a megjelölt másolatok kormányának való átadásának terhét.²³⁹ Ez a megközelítés - amely a vállalkozóra hárítja azt a terhet, hogy a dokumentumban meghatározott részeket üzleti titokként azonosítsa és jelölje meg - nagyban hozzájárul az üzleti titok túlzott igénybevételének elkerülése érdekében, és arra kényszeríti a vállalkozót, hogy pontosan mérlegelje, hogy az adott információ nyilvánosságra hozatala miért és hogyan ássa alá versenyhelyzetét.²⁴⁰ Az ilyen megfogalmazás összhangban van a nyílt iratokról szóló törvényekben az üzleti titok alóli mentességek megfelelően szűk értelmezésével.²⁴¹

²³⁷ Vö. Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 589-90 (1998) (a közbeszerzési normák politikai célok elérése érdekében történő alkalmazása mellett érvelve).

²³⁸ *Lásd a* 127-128. lábjegyzeteket.

²³⁹ *Lásd a* 129. lábjegyzetet.

²⁴⁰ Hasonlóképpen, New York állam oktatási minisztériumának az American Institutes of Research kutatóintézetével a hozzáadott érték mérésére irányuló projektre kötött szerződése előírja, hogy "a szerződő félnek egyértelműen azonosítania kell ... a védett információkat [a szerződés ... végrehajtásának időpontjában a szerződő fél tulajdonát képező módszerekkel vagy intézkedésekkel kapcsolatban], és ... engedélyt kell adnia ... a NYSED számára, hogy az ilyen védett információkat kizárólag a NYSED oktatási céljaira használja tovább a szerződés megszűnésétől számított tíz éven

keresztül". *Lásd a* C010834. sz. szerződést New York állam népe és az American Institutes for Research között, 183. lábjegyzet.

- ²⁴¹ *Lásd a* 204-217. lábjegyzethez tartozó szöveget (a nyílt iratokról szóló törvényekben az üzleti titokra vonatkozó mentességek megfelelően szűk értelmezésének megvitatása).

Fontos felismerni, hogy a sokkal szűkebb titoktartási nyilatkozatra vonatkozó igény nem okozta azt, hogy az Arnold Alapítvány megtagadta a szerződéskötést a hetedik bírósági kerülettel. Az alapítvány elfogadta a kevésbé kedvező nyelvezetet, annak ellenére, hogy a PSA-t ingyenesen biztosítja, és a Hetedik Igazságügyi Körzetnek nem volt olyan alkupozíciója, hogy visszatartsa a fizetést. A nonprofit szervezeteknek és alapítványoknak ugyanúgy szükségük van ügyfelekre, mint a profitorientált vállalatoknak - meg kell mutatniuk adományozóiknak, hogy olyan szolgáltatásokat nyújtanak, amelyek változást hoznak, és hatással vannak a kormányok működésére. Ezért a kormányoknak meg kell érteniük, hogy akkor is van befolyásuk, ha nem fizetnek a szolgáltatásokért.²⁴²

Ha a kormányok fizetnek a szolgáltatásokért, akkor további befolyással rendelkeznek a titoktartási és tulajdonjogi kérdések tekintetében. Így például az Illinois állam által az Eckerd Kidsnek a Rapid Safety Feedback szolgáltatásért fizetett szerződésben nyilvánvalóan a közbeszerzési szerződések szokásos, az állam számára kedvező közzétételi és tulajdonjogi rendelkezéseket tartalmazó nyelvezetét használták. A nyilvánosságra hozatal tekintetében a szerződés úgy rendelkezik, hogy az alapértelmezett feltételezés szerint az Eckerd által szolgáltatott minden információ nyilvános²⁴³ - bár még tovább is mehetne, ahogyan azt a hetedik bírói körzet és az Arnold Alapítvány közötti megállapodás tette, és a vállalkozóra háríthatná azt a terhet, hogy konkrét, megjelölt üzleti titoktartásra vonatkozó állításokat tegyen, különben elveszíti a nyilvánosságra hozatal elleni kifogásolási jogát. Ami a tulajdonjogot illeti, a szerződés előírja, hogy Illinois tulajdonában van minden, a szerződés alapján előállított termék, beleértve az e termékekhez fűződő valamennyi szellemi tulajdonjogot is.²⁴⁴ Ezzel szemben, amikor az alaszka Egészségügyi és Szociális Szolgáltatások Minisztériuma aláírt egy egyetértési nyilatkozatot, amelynek értelmében az Eckerd Kids beleegyezett abba, hogy ellenszolgáltatás nélkül nyújt RSF-szolgáltatásokat, Alaszka megígérte, hogy az Eckerd minden alkotását és termékét bizalmas információként kezeli, és beleegyezett abba, hogy az Eckerd tulajdonában van minden, ami a Rapid Safety Feedback programmal kapcsolatos, beleértve az összes szoftvert és a szoftver által előállított összes jelentést.²⁴⁵

Az a vállalkozó, aki kifejlesztett egy algoritmust, amelynek célja a több joghatóság módosítás nélkül nem akarja átruházni a forráskód tulajdonjogát, amely végrehajtja a

²⁴² Bizonyos esetekben a kormánytisztviselők azért üdvözölhetik a titoktartási nyilatkozatot, mert el akarják kerülni cselekedeteik nyilvános ellenőrzését. Nehezebb lehet egy olyan kormányhivatallal szemben, amely titoktartást ígér egy vállalkozónak, hogy ezzel indokolja saját döntéshozatali folyamatának titokban tartását, de megfelelő esetben jogi lépéseket lehet tenni, amelyek megtámadják az ilyen intézkedést, mivel az ellentétes az ügynökség nyílt kormányzati kötelezettségeivel.

²⁴³ Lásd Illinois FY16 szerződés, 145. lábjegyzet, 11. pont.

²⁴⁴ *Lásd id.* 11-12. o.

²⁴⁵ *Lásd* a 2015. február 20-i egyetértési nyilatkozatot az Eckerd Youth Alternatives, Inc. és az Alaszka Egészségügyi és Szociális Szolgáltatások Minisztériuma között, 142. lábjegyzet.

algoritmus egy joghatósághoz. Ha azonban a vállalkozó egyedi algoritmust biztosít egy joghatóság számára, akkor helyénvaló lehet, hogy ez a joghatóság ragaszkodjon a tulajdonjoghoz, vagy legalábbis a saját és más joghatóságok általi használatra vonatkozó licenchez. Így például Allegheny megye szerződése az Auckland konzorciummal nem kizárólagos licencet biztosít az állam és a szövetségi kormányzat számára a szerződés alapján előállított szoftver használatára, valamint arra, hogy mások számára is engedélyezze annak használatát, és biztosítja a megye számára a jogot a szerződés alapján előállított minden olyan dolog használatára és terjesztésére, amely szellemi tulajdonjogok által védett.²⁴⁶ A kormányzati szerveknek minden esetben tulajdonjogot kell érvényesíteniük azon jelentések felett, amelyek az adott joghatóság által szolgáltatott adatok alapján értékelik az adott joghatóságban fennálló kockázatokat. Az illinois-i szerződés ilyen állítást tesz,²⁴⁷ míg az alaszki megállapodás az összes jelentés tulajdonjogát az Eckerdre ruházza át.²⁴⁸

Még a tulajdonjogot és a nyilvánosságra hozatalt előíró, nagyon kedvező megfogalmazás sem hatékony azonban, ha nem készült dokumentáció, vagy ha azt soha nem bocsátották a kormányzati ügyfél rendelkezésére. Az Arnold Alapítvánnyal kötött megállapodásban foglalt közzétételi rendelkezések miatt a hetedik bírósági kerület olyan információkat tudott átadni nekünk a PSA kockázati skálákról - az előzetes letartóztatás előtt szabadon bocsátott, meg nem jelent személyek százalékos aránya a kockázati pontszámok szerint -, amelyeket sem más bíróság, sem maga az Arnold Alapítvány nem bocsátott volna rendelkezésünkre. Mégis csak azért tudta megadni ezeket az információkat, mert azok történetesen szerepeltek az Arnold Alapítvány egyik munkatársa által a bíróságnak tartott diavetítésen, így a nyilvánosságra hozatali politika meghatározása teljes mértékben az Arnold Alapítványra maradt. A felelős kormányoknak meg kellene hozniuk ezeket a döntéseket, és a közzétételi rendelkezéseket összekötni azzal a követeléssel, hogy a nyilvántartásokat a kormánynak mutassák be, és hozzák létre, ha azok még nem léteznek.

B. Nyilvántartások létrehozása a elszámoltathatóság érdekében

A kormányoknak tudatosan olyan nyilvántartásokat kell létrehozniuk - vagy megkövetelniük a beszállítóiktól -, amelyek elősegítik az algoritmikus folyamatok nyilvános megértését. Úgy tűnik, hogy az Európai Unió (2018-ban hatályba lépő) általános adatvédelmi rendelete is ezt tervezi, amely előírja, hogy a

²⁴⁶ Lásd AUT Enterprises Ltd. Contract 9-1-14 to 6-30-15, 37-39., 45-46. pont (2014) <http://robertbrauneis.net/algorithms/AlleghenyAUTContract.pdf> [<http://perma.cc/5HWR-CR8U>].

²⁴⁷ Lásd Illinois FY16 szerződés, 145. l. b. §.

²⁴⁸ Lásd az Eckerd Youth Alternatives, Inc. és az Alaszki Egészségügyi és

Szociális Szolgáltatások Minisztériuma között 2015. február 20-án kelt egyetértési nyilatkozatot, 142. lábjegyzet.

az algoritmus működését a nyilvánosság számára érthetővé kell tenni.²⁴⁹

Ideális esetben az érintett érdekelt felek összeállítanának egy sor bevált gyakorlatot a prediktív algoritmusok létrehozásának és végrehajtásának dokumentálására. Egy ilyen bevált gyakorlatokat tartalmazó dokumentum számos meglévő modellre támaszkodhatna. Az átláthatósági és elszámoltathatósági kezdeményezés például kiadott egy útmutatót a kormányzati átláthatóság, elszámoltathatóság és polgári szerepvállalás legjobb gyakorlatairól.²⁵⁰ A National Federation of Municipal Analysts (Önkormányzati Elemzők Országos Szövetsége) az önkormányzati adósságok kibocsátásával kapcsolatos legjobb közzétételi gyakorlatok sorozatát tette közzé.²⁵¹ Az Online Trust Alliance számos bevált gyakorlatot tartalmazó dokumentumot adott ki, köztük az Internet of Things Trust Framework 2.5-t, amely az adatvédelem és a biztonság alapelveinek gyűjteménye, és amely a csatlakoztatott otthoni és viselhető technológiákra összpontosít.²⁵² Talán a leglényegesebb, bár nagyon magas absztrakciós szinten, az Association for Computing Machinery amerikai közpolitikai tanácsa hét "Algoritmikus átláthatóság és elszámoltathatóság elvei" című dokumentumot állított össze²⁵³.

Bár itt nem remélhetjük, hogy olyan bevált gyakorlatokról szóló nyilatkozatot tudunk adni, amelyet a több érdekelt fél részvételével zajló tartós tanácskozás eredményezne, kutatásunk alapján nyolc kategóriában határozzuk meg a kívánatos dokumentációt: az algoritmikus modell általános előrejelzési célja és alkalmazása; releváns, rendelkezésre álló és gyűjthető adatok; az adatok megfontolt kizárása;

²⁴⁹ Az Európai Unió az adatvédelmi irányelv részeként az állampolgároknak is jogot adott arra, hogy magyarázatot kapjanak az olyan (nyilvános és magán) algoritmikus döntésekről, amelyek "jelentősen érintik" az egyéneket. Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (általános adatvédelmi rendelet) 2016 HL (L 119/1) 71; vö. Sandra Wachter et al, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76 (2016) (azzal érvelve, hogy az irányelv "jelenlegi formájában nem a magyarázathoz való jogot, hanem inkább egy korlátozott "tájékoztatáshoz való jogot" valósít meg" az automatizált döntéshozatalról); Goodman & Flaxman, *supra* 52. lábjegyzet, 6. o. (a fejlesztői titoktartást, a nyilvánosság technikai analfabetizmusát és az algoritmikus tervezést a magyarázat akadályaként azonosítva).

²⁵⁰ *Kormány megnyitása: A Guide to Best Practice in Transparency, Accountability and Civic Engagement Across the Public Sector*, TRANSPARENCY & ACCOUNTABILITY INITIATIVE

(2011), <http://www.transparency->

[initiative.org/archive/wp-content/uploads/2011/09/15-Open-government11.pdf](http://www.transparency-initiative.org/archive/wp-content/uploads/2011/09/15-Open-government11.pdf) [<http://perma.cc/V244-AJME>].

²⁵¹ *Közzétételi irányelvek*, NAT'L FED'N MUN. ANALYSTS, <http://www.nfma.org/disclosure-guidelines> [<http://perma.cc/DC9F-WDAV>].

²⁵² IoT Security & Privacy Trust Framework v. 2.5, ONLINE TRUST ALLIANCE (2017), <http://otalliance.actonsoftware.com/acton/attachment/6361/f-008d/1/-/.-/-/IoT%20Trust%20Framework.pdf> [<http://perma.cc/LB23-EM4N>].

²⁵³ *Nyilatkozat az algoritmikus átláthatóságról és elszámoltathatóságról*, ASS'N FOR COMPUTINGMACH . U.SPUB .POLICYCOUNCIL

,(2017.január12),

http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf [<http://perma.cc/5TKE-LFHT>].

konkrét előrejelzési kritériumok; alkalmazott elemzési technikák; főbb szakpolitikai döntések; validálási tanulmányok és ellenőrzések eredményei; valamint az előrejelző algoritmus és az algoritmus kimeneteinek magyarázata.

1. Általános előrejelző cél és Alkalmazás

A kormányoktól elvárható, hogy megfogalmazzák a prediktív algoritmus használatával kapcsolatos céljaikat. Ez fontos viszonyítási alapot biztosít, amelyhez képest a konkrét kritériumokat mérni lehet, és az algoritmikus előrejelzések által megalapozott döntések jobb megértéséhez vezethet. A cél nem mindig magától értetődő. Például egy olyan algoritmus, mint a PredPol vagy a HunchLab legáltalánosabb célja az, hogy megjósolja, hol és mikor fognak bűncselekmények történni. Egy helyi rendőrséget azonban valójában az is érdekelheti, hogy döntéseket hozzon arról, hogy a korlátozott számú járőrök hol tudnak a leghatékonyabban elrettenteni a bűncselekményektől, elismerve, hogy a beltérben elkövetett bűncselekményeket nehéz elrettenteni járőrozéssel. Ezért az osztály pontosabban, szűkebben fogalmazná meg célját, mint annak előrejelzése, hogy a rendőri járőrök jelenléte hol és mikor tartaná vissza a bűncselekményeket.

Az általános előrejelzési cél megfogalmazásának részeként egy kormányzatnak érdemes lehet egy lépéssel hátrébb lépnie, és megfogalmazni a problémát, amelyet kezelni próbál. Például egy kormányzatot, amely segítséget kér annak megjósolásához, hogy mely fogvatartottak követnek el a legnagyobb valószínűséggel bűncselekményeket, ha feltételes szabadlábra helyezik őket, többféle szempont is motiválhatja. Lehet, hogy a börtönök létszámát szeretné csökkenteni a túlszűfolttság miatt; vagy a feltételesen szabadlábra helyezettéket számát szeretné csökkenteni, akik újabb bűncselekményeket követnek el; vagy a feltételes szabadlábra helyezésről szóló döntési gyakorlatának igazságosságával kapcsolatos kihívásokkal kell szembenéznie. E helyzetek mindegyike valószínűleg különböző érzékenységet igényel a prediktív algoritmusok létrehozásakor. A prediktív algoritmusok a kormányzati döntések támogatására is alkalmazhatók különböző helyzetekben. Például, míg az Allegheny Family Screening Tool és az Eckerd Rapid Safety Feedback mindkettő gyermekjóléti értékelést nyújt, az előbbit úgy tervezték, hogy a gyermekjóléti forródrótra érkező hívás pillanatában, azonnali szűrőeszközként alkalmazzák; az utóbbit nyilvánvalóan az ügynökség által jelenleg kezelt összes gyermekjóléti ügy időszakos felülvizsgálatára használják. Ideális esetben el kellene gondolkodni azon a konkrét döntéshozatali folyamaton, amelyhez az algoritmust tervezik, hogy ez-e az algoritmikus előrejelzés legjobb alkalmazása az adott ügynökség működésében, és hogy az algoritmus kialakítása megfelelő-e az adott alkalmazáshoz.

2. Adatok: .

Elektronikusan elérhető a következő címen:

<https://ssrn.com/abstract=3012499>

Ha a prediktív célt szem előtt tartjuk, a következő lépés az, hogy megvizsgáljuk, milyen adatok lehetnek relevánsak a predikció elkészítéséhez. Mind az algoritmus értékelése, mind a mérlegelés indukálása szempontjából hasznos, ha dokumentáljuk, hogy kezdetben milyen adatokat tarthatunk elképzelhetőnek a kérdéses eredmény előrejelzése szempontjából. Például, az adattudósok, akik esetleg a vádlott korábbi letartóztatási előéletére és foglalkoztatási előéletére vonatkozó adatokkal számoltak, figyelembe vették-e a vádlott edzésprogramjára és oktatási hátterére vonatkozó adatokat is? Ha nem, miért nem? A legtöbb előrejelző algoritmust olyan adatokon képzik ki, amelyeket már más célból gyűjtöttek. Az adattudósok tehát meglévő adatforrások után fognak kutatni, és fontos lesz dokumentálni, hogy hol keresték és mit találtak.

3. Adatok Kizárás

A rendelkezésre álló adatok végül kizárhatók abból az adathalmazból, amelyet az algoritmus képzésére használnak, és amelyet végül bemeneti adatként használnak egy adott témára vonatkozó előrejelzés létrehozásához. Az adatok kizárásának legalább öt oka van: minőségi aggályok, manipulációra való hajlam, időbeli és térbeli korlátok, a relevancia hiánya, valamint a relevancia hiányától eltérő politikai megfontolások. Mindezek dokumentálása fontos az algoritmus képzési adatainak és bemeneti adatainak megértéséhez.

a. *Adatminőség.* Az adattudósok aggódhatnak amiatt, hogy az adatkészletek vagy bizonyos adatmezők túl sok pontatlanságot tartalmaznak, nem következetesen lettek meghatározva az adatgyűjtés során, vagy különböző módon sérültek. Előfordulhat például, hogy a címeket kézzel írt eredetiből kézzel írták át, és a tesztek érvénytelennek bizonyultak.²⁵⁴ Vagy előfordulhat, hogy bizonyos ideig kétféle adatot vittek be egy mezőbe. Az ilyen problémák dokumentálása, valamint az arra vonatkozó döntések, hogy az adatokat - még a hibáikkal együtt is - meg kell-e tartani, vagy ki kell-e zárni, fontosak lehetnek az előállított algoritmus minőségének értékeléséhez.

b. *Manipuláció és játék.* A prediktív algoritmusok készítői dönthetnek úgy is, hogy kizárnak bizonyos típusú adatokat, mert azok manipulálhatók vagy "játszadoznak", és így aláássák a képzési adatok pontosságát vagy a kész algoritmus bemeneti adatainak pontosságát. Például, mint fentebb említettük, az Arnold Alapítvány úgy döntött, hogy létrehoz egy olyan előzetes szabadlábra helyezési algoritmust, amely nem igényel bemeneti adatot az alábbiakból

²⁵⁴ Vö. Julia Andre, Luis Ceferino & Thomas Trinelle, *Prediction Algorithm for Crime Recidivism*, MACH. LEARNING PROJECT, STAN. U. 1 (2015), http://cs229.stanford.edu/proj2015/250_report.pdf [<http://perma.cc/YWP2-5CTX>] (figyelmeztetve arra, hogy "a nyilvánosan elérhető adatkészletek [a szabadult elítéltek visszaeséséről] az előírások miatt régiek, ami azt jelenti, hogy gyakran a kézzel tárolt adatok számszerű átírása").

a vádlottal folytatott kihallgatás során gyűjtött tények.²⁵⁵ Ezt a kizárást részben az az aggodalom indokolta, hogy a kihallgatás során gyűjtött információk, amikor a vádlott tudja, hogy a válaszok meghatározhatják a tárgyalás előtti szabadlábra helyezést, manipulálhatók.

c. *Időbeli és helybeli korlátozások.* Az adatokat szükségszerűen olyan alanyokról gyűjtik, akik különböző időpontokban és helyeken cselekszenek. Ha minden más dolog egyenlő, minél nagyobb a képzési adathalmaz, annál jobb. De nem biztos, hogy minden más dolog egyenlő. A tíz évvel ezelőtti visszaesés kockázata ma más lehet az azonos profilú elítéltek esetében, a gazdaság, a rendelkezésre álló szociális szolgáltatások és sok más tényező miatt. Ha a különböző évekből származó adatacsoportok jelentősen eltérő korrelációt mutatnak, dönteni lehet a régebbi adatok elavultként való kizárásáról. Másrészt, ha az a cél, hogy megjósoljuk, hogy egy feltételese szabadlábra helyezett személy elkövet-e bűncselekményt a következő öt évben, akkor a képzési adathalmazból ki kell zárni az öt évnél nem régebben feltételese szabadlábra helyezett foglyok adatait, mivel az újabb feltételese szabadlábra helyezettek nem rendelkeznek kellően hosszú előélettel. Bizonyos esetekben tehát előfordulhat, hogy egyes adatokat ki kell zárni, mivel túl régiek, más adatokat pedig túl újaknak kell tekinteni.²⁵⁶

A HunchLab esetében a Lincoln Rendőrkapitányság felfedte, hogy a HunchLab által egy adott napon előállított eredmény az előző harminc nap rendőrségi jelentéseinek alapul.²⁵⁷ A harmincnapos ablak kiválasztása nyilvánvalóan magában foglalja az egymással versengő tényezők egyensúlyát. Az elmúlt hónapra való korlátozás viszonylag friss adatokat biztosít, és lehetővé teszi a heti és havi aktivitási ciklusok vizsgálatát. Ugyanakkor nem teszi lehetővé a szezonális ciklusok vizsgálatát, és a viszonylag ritka bűncselekménytípusokra vonatkozó adatok nagyon vékonyak lehetnek.

Az algoritmusok fejlesztőinek a képzési és bemeneti adatok földrajzi hatókörét is meg kell határozniuk. Az eltérő társadalmi és gazdasági feltételek, és ami talán még ellentmondásosabb, az eltérő etnikai összetétel, jövedelmi profil vagy más tényezők miatt az egyik területről - esetleg egy városi területről - származó vádlottak egy csoportja, akik egyébként hasonlóak egy másik területről - esetleg egy vidéki területről - származó vádlottak csoportjához, eltérő kockázatot jelenthet az eljárás előtti szökés szempontjából.

Tudjuk, hogy az Arnold közbiztonsági értékelő algoritmusát olyan adatokon képezték ki, amelyeket három, a közbiztonságra vonatkozó

²⁵⁵ *Developing a National Model for Pretrial Risk Assessment (A bíróság előtti kockázateértékelés nemzeti modelljének kidolgozása)*, uo. 228. lábjegyzet, 3. o., 3. pont.

²⁵⁶ Az adatok időbeli és térbeli korlátozásának megválasztásáról lásd Andreas M. Olligschlaeger, *Crime Forecasting on a Shoestring Budget*, CRIME MAPPING

& ANALYSIS NEWS 8, 9-10 (2015 tavasz), http://crimemapping.info/wp-content/uploads/2015/03/CrimeMappingNews_Issue23.pdf[<http://perma.cc/DF5F-FBFF>].

²⁵⁷ *Lásd* Tonya Peters e-mailje, 171. lábjegyzet.

száz különböző joghatóságban országszerte.²⁵⁸ Nem tudjuk, hogy az Arnold Alapítvány tesztelte-e, hogy az adatállomány különböző államokból vagy régiókból származó részhalmazai ugyanazokat a prediktív korrelációkat mutatják-e, mint az adatállomány egésze. Ha a különböző régiókból származó adatok lényegesen eltérő prediktív korrelációkat mutatnak, döntés születhet az adatállomány földrajzi korlátozásáról. Függetlenül attól, hogy az adathalmazt idő és hely szerint korlátozzák-e vagy sem, a legjobb gyakorlat lehet az idő és hely szerinti különbségek vizsgálata és az eredmények dokumentálása.

d. Relevancia. Egyes adatelemek kizárhatók, mert úgy tűnik, hogy nem állnak kellő összefüggésben a megjósolni kívánt eredménnyel. Hasznos lenne dokumentálni ezt a kizárást, valamint azt a prediktív értékküszöböt, amely alá a kizárt adatok estek.

e. A relevancián kívüli politikai okok. Talán a legjelentősebb és legellentmondásosabb az, hogy bizonyos adatokat - a potenciális előrejelző értékük ellenére - különböző politikai okokból kizárnak. Az Arnold Foundation például algoritmusának előnyeként hirdeti, hogy nem vesz figyelembe olyan kérdéseket, mint "faj, nem, jövedelem, iskolai végzettség, lakcím, drogfogyasztási előzmények, családi állapot, családi állapot, családi állapot, nemzeti származás, foglalkoztatás, [vagy] vallás".²⁵⁹ Az olyan megváltoztathatatlan jellemzők, mint a faj és a nem, alkotmányosan problematikusak; a lakcím sok esetben szoros korrelációt mutathat a fajjal. Az olyan jellemzők kizárására vonatkozó döntés, mint az iskolai végzettség és a kábítószer-használati előzmények, ha ezekről kiderül, hogy jelentős előrejelző értékkel bírnak, feltehetően ellentmondásosabb lenne, és dokumentálni kellene.

4. Speciális prediktív kritériumok

Fentebb már említettük, hogy hasznos lehet megfogalmazni egy általános előrejelzési célt, amelyre egy algoritmusfejlesztési projekt törekszik. Ha azonban egyszer már döntöttek arról, hogy milyen képzési adatokat használjanak, valószínűleg kiderül, hogy a tényleges előrejelzéseket az eredeti előrejelzési céltől némileg eltérően kell leírni. Ezért az előrejelzéshez használt kritériumok kiválasztását dokumentálni kell, különösen akkor, ha azok eltérnek a nyilvánvalótól.

Például egy algoritmus általános előrejelzési célja lehet az, hogy megjósolja, hol és mikor fog bűncselekmény bekövetkezni, de az egyetlen rendelkezésre álló képzési és bemeneti adat a legvalószínűbb bűncselekmények, amelyeket

²⁵⁸ *Developing a National Model for Pretrial Risk Assessment (Az eljárás előtti kockázateértékelés nemzeti modelljének kidolgozása)*, uo. 228. lábjegyzet, 3. o., 3. pont.

²⁵⁹ *A közbiztonsági értékelés (PSA)*, LAURA & JOHN ARNOLD FOUND., <http://www.arnoldfoundation.org/wp-content/uploads/PSA-Infographic.pdf>

[<http://perma.cc/PLR5-2SZ6>].

jelentettek, és amelyeket viszonylag hamar az előfordulásuk után jelentettek. Így az algoritmus végül nem azt fogja pontosabban megjósolni, hogy hol fognak bűncselekmények történni, hanem azt, hogy hol fognak *bejelentett* bűncselekmények történni. Ez nem csak azért aggasztó, mert sok bűncselekményt nem jelentenek be,²⁶⁰ hanem azért is, mert a különböző városrészekben eltérő arányban jelentenek be bűncselekményeket.²⁶¹ Egy tanulmány például azt találta, hogy a hátrányos helyzetű városrészekben kisebb valószínűséggel jelentenek be egyszerű testi sértést.²⁶² Egy másik tanulmány azt találta, hogy a bűncselekményeket különösen az erősen bevándorló lakónegyedekben jelentették be alul.²⁶³ Egy harmadik azt találta, hogy a bűncselekmények bejelentése az áldozat életkorával együtt nő, így az idősebb lakosokkal rendelkező városrészek valószínűleg nagyobb arányban jelentik be a bűncselekményeket.²⁶⁴ Így a bejelentett bűncselekményeken alapuló algoritmus a rendőrséget a hátrányos helyzetű, bevándorló és fiatal áldozatoktól távolíthatja el, akik vitathatatlanul a legkiszolgáltatottabbak közé tartoznak. Ezek a problémák nem korlátozódnak a prediktív rendőri tevékenységre. Allegheny megyét például leginkább annak előrejelzése érdekelte, hogy a bejelentett gyermekbántalmazás mikor vezethet súlyos sérüléshez vagy halálhoz, de úgy döntött, hogy nem tud olyan algoritmust készíteni, amely ezt közvetlenül meg tudná tenni, mivel azok az esetek, amelyekben valóban súlyos sérülés vagy haláleset történt, (szerencsére) túl kevés adatpontot szolgáltatottak. Ezért úgy döntöttek, hogy ehelyett a nevelőszülőknél történő elhelyezést és a bántalmazásról szóló további jelentéseket használják a konkrét előrejelző kritériumként, az Auckland Konzorcium jelentésében részletesen kifejtett okok miatt.²⁶⁵ Hasonlóképpen, a COMPAS visszaesési algoritmus a bűnismétléssel kapcsolatos letartóztatási adatokon alapul, nem pedig az elítéléseken;²⁶⁶ bár az Arnold Alapítvány nem hozta nyilvánosságra a következő adatokat

²⁶⁰ Lásd Lynn Langton et al., *Victimizations Not Reported to Police, 2006-2010*, U.S. DEPT'JUST .., BUREAUJUST. STAT (2012aug.), <http://www.bjs.gov/content/pub/pdf/vnrrp0610.pdf> [<http://perma.cc/ASK6-TJ82>].

²⁶¹ A bejelentett bűnözés és a valós bűnözési arányok általános eltéréséről lásd David Robinson & Logan Koepke, *Stuck in a Pattern: Early Evidence on Predictive Policing and Civil Rights*, UPTURN5 (2016), www.teamupturn.org/static/reports/2016/stuck-in-a-pattern/files/Upturn_-_Stuck_In_a_Pattern_v.1.01.pdf [<http://perma.cc/C5SL-MF4Q>].

²⁶² Eric P. Baumer, *Neighborhood Disadvantage and Police Notification by Victims of Violence*, 40 CRIMINOLOGY 579, 597 (2002).

²⁶³ Carmen M. Gutierrez & David S. Kirk, *Silence Speaks: The Relationship Between Immigration and the Underreporting of Crime*, 63 CRIME & DELINQUENCY 928, 946 (2015).

²⁶⁴ Lásd Stacey J. Bosick et al., *Reporting Violence to the Police: Predictors Through the Life Course*, 40 J. CRIM. JUST. 441 (2012). Az Azavea, Inc., a HunchLab létrehozója, csodálatra méltó módon részletesen tárgyalja a bejelentett bűncselekmények képzési adatokként való kiválasztását, a

választás okait, valamint az általa preferált bűncselekmény-bejelentések típusát. *Lásd:* A Citizen's Guide to HunchLab, 77. lábjegyzet, 25-26. o.

²⁶⁵ *Lásd* Vaithianathan et al., 72. lábjegyzet.

²⁶⁶ *Lásd* COMPAS Risk & Need Assessment System, NORTHPOINTE 2 (2012), http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf [<http://perma.cc/6TL7-GCA4>] (az általános képzési módszer leírása).

részleteket a PSA képzési adatairól, szinte biztosan letartóztatásokat is használ, nem pedig ítéleteket. Fontos megérteni, hogy ez a kettő hogyan térhet el egymástól. Abe Gong arra kér minket, hogy fontoljuk meg: "Mi van akkor, ha a rendőrök nagyobb valószínűséggel üldöznek, kutatnak és tartóztatnak le fekete gyanúsítottakat, mint fehér gyanúsítottakat? Mi van, ha a bűnüldöző szervek aránytalanul nagy mennyiségű erőt alkalmaznak, vagy agresszívebb rendőri taktikákat alkalmaznak a fekete negyedekben?"²⁶⁷ A kisebbségi közösség tagjainak letartóztatásai mesterségesen magasra torzulnak, ezért más előrejelző kritériumokat kellett választani; ezeket a választásokat nyilvánosságra kell hozni.

5. Elemzési és fejlesztési technikák Használt módszerek

Viszonylag kevés elemzési technikát használnak az előrejelzés tárgyainak jellemzői vagy jellemzői közötti összefüggések feltárására. A legnépszerűbbek közé tartoznak a regressziós technikák (lineáris, logisztikus és polinomiális), a véletlen erdők, a neurális hálózatok és a támogató vektor gépek.²⁶⁸ Hasznos dokumentálni, hogy mely technikákat próbálták ki, melyeket választották és miért. A lineáris regresszió például akkor lehet megfelelő, ha valószínűnek tartják, hogy egy vagy több bemenet és a kimenet között valóban lineáris kapcsolat áll fenn - például egy vádlott életkora és annak valószínűsége között, hogy a vádlott bűncselekményt követ el, ha a tárgyalás előtt szabadul. Előfordulhat, hogy egy nem lineáris előrejelző modell (például azért, mert bizonyos életkori határértékeket használ) statisztikailag hasonlóan szignifikáns eredményeket produkál, amelyek statisztikailag ugyanolyan szignifikánsak.

Vannak olyan standard algoritmusfejlesztési technikák is, mint például az adathalmaz véletlenszerű felosztása részhalmazokra, amelyeket az algoritmus képzésére használnak, majd egy vagy több szakaszban történő tesztelése ("validálás").²⁶⁹ E fejlesztési technikák dokumentálása valószínűleg szintén a legjobb gyakorlat.

6. Fő politikai irányvonalak Választási lehetőségek

Az algoritmus fejlesztése során hozott különböző típusú politikai döntésekről már beszéltünk. Az egyik az egyébként releváns adatok különböző okokból történő kizárásáról szóló döntés. Egy másik a hamis negatív és hamis negatív eredmények súlyozására vonatkozó döntés.

A visszaesés kockázati skála algoritmus, amely arra vonatkozó adatokon alapul, hogy a vádlottakat letartóztatták-e a felvételi értékelést követő két éven belül).

²⁶⁷ Abe Gong, *Ethics of Powerful Algorithms (2 of 4)*, MEDIUM (2016. július 12.), <http://medium.com/@AbeGong/ethics-for-powerful-algorithms-2-of-3-5bf750ce4c54> [<http://perma.cc/VVR9-PF9G>].

²⁶⁸ *Lásd pl.*, SHAI SHALEV-SCHWARTZ & SHAI BEN-DAVID, UNDERSTANDING MACHINE LEARNING: FROM THEORY TO ALGORITHMS 89-240 (2014).

²⁶⁹ *Lásd pl.* YASER S. ABU-MOSTAFA, MALIK MAGDON-ISMAIL & HSUAN-TIEN LIN,

Elektronikusan elérhető a következő címen:
<https://ssrn.com/abstract=3012499>

pozitívumok egyformán vagy különbözőképpen. Ezeket a döntéseket dokumentálni kell, valamint azt is, hogy miért döntöttek úgy, ahogyan döntöttek.

7. Validálási tanulmányok, ellenőrzések, naplózás és nem átlátható elszámoltathatóság

A bevezetés előtti validálás a prediktív algoritmusok kezdeti fejlesztésének szokásos lépése. Az algoritmus üzembe helyezése után azonban további, a megvalósítás utáni validálási vizsgálatokat lehet végezni az algoritmus előrejelző erejére és az algoritmus által a valós körülmények között esetleg előidézett kimeneti torzításokra vonatkozóan. Ki lehetne dolgozni a legjobb gyakorlatokat arra vonatkozóan, hogy mikor és hogyan kell ilyen vizsgálatokat végezni, és mikor célszerű ragaszkodni ahhoz, hogy a vizsgálatokat független szervezet végezze el. Az állami ügyfelek megkövetelhetnék, hogy az ilyen tanulmányokat az ő ügyeiken végezzék el és adják át nekik.

Az auditok a validációs vizsgálatok alternatívájaként vagy kiegészítéseként szolgálhatnak. Amennyiben üzleti titok, biztonsági vagy adatvédelmi okokból nem kerül sor az optimális közzétételre, fontos lehet, hogy az algoritmusfejlesztést egy harmadik fél bizalmasan auditálja.²⁷⁰ A nyilvános ügyfelek ragaszkodhatnak az auditáláshoz, ha egy algoritmus bizonyos célokat nem ér el, vagy ha az ügyfelek bizonyítékot találnak arra, hogy a fejlesztési folyamat hibás volt. Szintén helyénvaló lenne megkövetelni a fejlesztőtől, hogy a fent leírt dokumentáció számos vagy valamennyi kategóriáját tartalmazó naplót vezessen, még akkor is, ha a teljes naplót általában nem hoznák nyilvánosságra, arra az esetre, ha szükségessé válna az ellenőrzés.²⁷¹ Az állami szervezeteknek szerződést kellene kötniük az algoritmusok végrehajtásának auditálására is, amit a Hetedik Igazságügyi Körzet a PSA-algoritmus végrehajtásához kapott (az Arnold Alapítvány egyik alvállalkozója végezte el).²⁷² Az állami ügyfeleknek tudniuk kell, és a nyilvánosság előtt nyilvánosságra kell tudniuk hozni, hogy helyesen adják-e be az adatokat és helyesen értelmezik-e az eredményeket.

8. Algoritmus és kimenet Magyarázatok

Gyakran fontos lesz, hogy közérthető nyelven magyarázzuk azokat az összefüggéseket, amelyeken egy algoritmus alapul.

²⁷⁰ Az algoritmusok auditálásáról lásd Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*, ANN. MEETING INT'L COMM. ASS'N (2014) <http://tiny.cc/61wrmy> [<http://perma.cc/F29K-BALH>].

²⁷¹ Lásd az *algoritmikus átláthatóságról és elszámoltathatóságról* szóló nyilatkozatot, 253. l. ábjegyzet, 2. pont ("Ellenőrizhetőség: A modelleket, algoritmusokat, adatokat és döntéseket rögzíteni kell, hogy azok ellenőrizhetők legyenek olyan esetekben, amikor gyanúja.").

Elektronikusan elérhető a következő címen:

<https://ssrn.com/abstract=3012499>

²⁷² Lásd Dal Pra, *Volusia County Case Review*, 132. lábjegyzet.

alapján, valamint az általános útvonalról, amelyen a jóslathoz eljut, legyen az egy képlet, amely súlyozza a tényezőket, egy döntési fa vagy más útvonal.²⁷³ Ez lehetővé teszi mind a nyilvános elszámoltathatóságot, mind pedig azt, hogy az algoritmus felhasználói megítélhessék az algoritmus eredményét. Ha az algoritmus annyira bonyolult, hogy egyszerű magyarázat nem tűnik lehetségesnek, akkor valószínűleg ezt is nyilvánosságra kell hozni, hogy azok, akik az algoritmus előrejelző kimenetét használják, megértsék, hogy az egy fekete doboz, amely nem kapcsolódik semmilyen jól megfogalmazható magyarázathoz vagy oksági elmélethez. Ha egy értelmezhető algoritmus ugyanolyan jól teljesít, mint egy nem értelmezhető algoritmus, a kormányoknak a kormányzati kapacitás és a nyilvános átláthatóság érdekében az értelmezhető algoritmust kellene előnyben részesíteniük. Ha a kormányzati ügynökök (vagy az általuk megbízott személyek) megértik az algoritmust, jobban fel lesznek szerelve ahhoz, hogy elfogadják vagy felülbírálják annak ítéletét.²⁷⁴

Gyakran fontos lesz az algoritmus kimenetének magyarázata is. Ez különösen igaz akkor, ha az algoritmus nem kalibrált skálát állít elő, mint például a PSA egytől hatig terjedő kockázati skálái. Az algoritmus korai bevezetése során végzett validációs vizsgálat során tízből majdnem kilenc vádlott, aki a tárgyalás előtti szökés kockázatának legalacsonyabb pontszámát kapta, valóban megjelent a tárgyaláson; azok közül, akik a *legmagasabb* kockázati pontszámot kapták, tízből hét jelent meg. Ha az előzetes letartóztatási szolgálatok tisztviselői és a bírák nem ismerik ezeket a százalékos arányokat, akkor feltételezhetik, hogy a legalacsonyabb és a legmagasabb kockázati pontszámok közötti különbség nagyobb, mint amekkora valójában, vagy eltérő feltételezéseik lehetnek arról, hogy egy "alacsony kockázatú" vádlott milyen alacsony, illetve egy "magas kockázatú" vádlott milyen magas kockázatot jelent.²⁷⁵

VI. KÖVETKEZTETÉS

A nyílt forráskódú kódot használó vagy más módon a prediktív elemzést futtató kódot közlétező közintézményeknek mindig lesz értéke. A kódhoz való hozzáférés azonban általában nem szükséges az érdemi átláthatóság eléréséhez, és néha még csak nem is szükséges.

²⁷³ *Lásd id.* ("Magyarázat: Az algoritmikus döntéshozatalt alkalmazó rendszereket és intézményeket arra ösztönzik, hogy magyarázatot készítsenek mind az algoritmus által követett eljárásokról, mind a meghozott konkrét döntésekről. Ez különösen fontos közpolitikai kontextusokban."), Diakopoulos, *supra* note 6, 411. o. (javasolja, hogy az algoritmusokra vonatkozó átláthatósági politika tartalmazza "a hasonlósági vagy osztályozási algoritmusok által használt definíciókat, operacionalizálásokat vagy küszöbértékeket").

²⁷⁴ Az értelmezhető algoritmusok fejlesztéséről lásd Jiaming Zeng, Berk Ustun & Cynthia Rudin, *Interpretable Classification Models for Recidivism Prediction*, J. ROYAL STAT. SOC'Y: SERIES A (STATISTICS IN SOC'Y) (2016), <http://arxiv.org/abs/1503.07810> [<http://perma.cc/MZ87-LNF9>].

- ²⁷⁵ Annak értékelését, hogy a vizsgálati alanyok csoportosítása a kockázati különbségeket tükröző módon történik-e, "kalibrálásnak" nevezik. *Lásd pl.* , Nicholas Serrano, *Calibration Strategies to Validate Predictive Models: Is New Always Better?* , 38 INTENSIVE CARE MED. 1246 (2012).

segítség. A közintézményeknek inkább arra kellene összpontosítaniuk, hogy az algoritmikus folyamatok tervezését, beszerzését és végrehajtását átgondoltabb és átláthatóbb módon végezzék. A közintézmények szerződéseinek meg kellene követelniük a szállítóktól, hogy olyan nyilvántartásokat készítsenek és szolgáltatassanak, amelyek magyarázatot adnak a legfontosabb szakpolitikai döntésekre és érvényesítési erőfeszítésekre, anélkül, hogy feltétlenül nyilvánosságra hoznák a pontos képleteket vagy algoritmusokat. Ezeket a nyilvántartásokat ezután közzé lehet tenni, és támogatni lehet a nyílt politikai vitákat anélkül, hogy ez hátrányosan befolyásolná a vállalkozók versenypozícióját. Amennyiben a prediktív algoritmus-projektekben feloldhatatlan üzleti titkok maradnak fenn, a nyílt iratok iránti kérelmekre válaszoló kormányzati iratmegőrzőknek szűken kell értelmezniük ezeket az igényeket. A bíróságoknak ugyanígy kellene eljárniuk, és kötelezniük kellene a vállalkozókat, hogy adják ki azokat a feljegyzéseket (akár szerkesztett formában is), amelyek nem gyengítik versenyhelyzetüket. Ez lehetővé teszi az érdemi átláthatóságot, és ezáltal a kormányzat elszámoltathatóságát ezen algoritmusok használata során.