Ságvári Bence

Diszkrimináció, átláthatóság és ellenőrizhetőség

Bevezetés az algoritmusetikába

Egyre fontosabb szerepet játszanak az algoritmusok abban, hogy milyen információk jutnak el közvetlenül hozzánk, kikkel lépünk kapcsolatba a közösségi oldalakon, milyen hírek kerülnek a „hot", „trending" vagy éppen „most discussed" kategóriákba az egyes online szolgáltatóknál. Mégis, aligha tévedünk nagyot azzal a kijelentéssel, mely szerint az emberek (vagy hívjuk őket átlagos felhasználóknak) többsége egyáltalán nincs tisztában ezzel a ténynyel. De aki tisztában van vele, vélhetően még az sem tudja, hogy valójában ezek miként is működnek, milyen adatokat gyűjtenek róla, és azokat pontosan ki és milyen módon használja fel. Nem tudhatjuk, hogy egy ártalmatlannak tűnő kattintásból (mondjuk egy Facebook-csoporthoz való csatlakozás) milyen következtetést von le egy algoritmus a háttérben, és ezt az információt ki és hogyan fogja a személyünkkel kapcsolatban felhasználni a jövőben.

Az adatalapú algoritmusok elsődleges gazdasági, „megoldásközpontú" ígérete a hatékonyság, a költségcsökkentés, a gyorsaság, a pontosság, a korábban nem elérhető információk hozzáférhetősége és elemezhetősége, a testreszabhatóság, a különböző (köztük bűnüldözési és bűnmegelőzési) rendszerek működésének adatalapú és valós idejű racionalizálása. Ezt a felsorolást bizonyára még hosszasan lehetne folytatni, annyi azonban bizonyos, hogy ennek a digitális adatkornak az építőelemei ma már szinte mindenhol körülvesznek bennünket. A mobiltelefonok cella- és GPS-helyadatai, a közösségi médián keresztül megosztott üzenetek és képek, az online és offline vásárlások részletes adatai, az internetes keresések mind egy-egy apró építőelemet jelentenek az életmódunkról és személyiségünkről egyre pontosabb képet alkotó adatbázisok számára. A „kik vagyunk", „mit csinálunk" és „mi érdekel bennünket" kérdésekre közvetlen és közvetett válaszokat adó adatokat ma már szolgáltatók tömegei gyűjtik rólunk. Ennek egy része a felhasználók tudtával és beleegyezésével zajlik, sok esetben azonban a formális beleegyezés megtörténik ugyan, ám az emberek fejében aligha tudatosul, ki és milyen formában gyűjti, elemzi, vagy éppen értékesíti tovább a róluk szóló információkat.

A tanulmány az OTKA K112713 számú („Egy online közösségi hálózat életciklusa: big data elemzés") kutatási projekt keretében készült.

A mindennapok szintjén ezeknek az adatbázisoknak az „etetése" kétségkívül hozzájárul a folyamatos fejlődéshez, az egyéni igények mind pontosabb kiszolgálásához. Ugyanakkor az is nyilvánvaló, hogy ennek a folyamatnak számtalan negatív következménye is lehet. A napról napra univerzálisabbá váló adatalapú szolgáltatások világában, ahol az emberek függősége folyamatosan nő ezektől a rendszerektől, egyre aktuálisabb kérdéssé válik az adatok, illetve az ezeket feldolgozó algoritmusok (társadalmi) igazságossága, és ezeknek az egyénre, illetve a társadalom egészére gyakorolt hatása.

Írásom célja betekintést nyújtani azokba a folyamatokba, ahogyan a Big Data és algoritmusalapú rendszerek automatizált döntései meglévő diszkriminációs folyamatokat erősíthetnek fel, illetve, ahogyan az objektívnek gondolt modellek a társadalmi egyenlőség és igazságosság elveit kikezdve igazságtalan és káros rendszerek alapjául szolgálhatnak. És ehhez kapcsolódik az a kérdés is, vajon miként bizonyosodhatunk meg arról, hogy egy algoritmus rosszul működik. Milyen gyakorlati lehetőségek vannak az algoritmusok transzparenciájának és külső kontrolljának megteremtésére?

A tanulmány ezeket a témákat az alábbi szerkezetben tekinti át. Az első részben röviden definiálom a három kulcsfogalmat (Big Data, algoritmus és társadalmi diszkrimináció). Ezt követően néhány konkrét példán keresztül bemutatom a diszkrimináció, a modellalapú társadalmi igazságtalanság megjelenésének lehetséges formáit. A harmadik rész az algoritmusok átláthatóságának és auditálásának nehézségeiről szól. Végezetül a tanulmány konklúziójában néhány olyan szempontra hívom fel a figyelmet, amelyek hozzájárulhatnak a negatív hatások tompításához, a probléma felhasználói és szolgáltatói szinten való felismeréséhez.

Big Data, algoritmus és diszkrimináció

Big Data

A Big Data[1] mint kifejezés közel két évtizedes múltra tekint vissza, azonban a köztudatba csupán néhány éve került be, elsősorban az analitikai megoldásokat szállító nagy IT-vállalatok marketingtevékenységének köszönhetően (Gandomi és Haider 2015). A Big Data szociológiai szempontból való értelmezésének is napról napra növekvő irodalma van (Borgman 2015; Csepeli 2015; Dessewffy és Láng 2015; McFarland, Lewis és Goldberg 2015; Mutzel 2015; Székely 2015), azonban továbbra sem igazán létezik a fogalomnak egységes meghatározása. Történtek persze kísérletek olyan átfogó definíció megalkotására, amely figyelembe veszi az eddig megszületett eltérő fókuszú (pl. üzleti vs. akadémiai) megközelítéseket, ezek eredményei azonban túlságosan összetettek és így rutinszerűen aligha használhatók (Puschmann és Burgess 2014). De Mauro és szerzőtársai 2014-ben például több mint 1500 tanulmány alapján négy olyan kulcsterületet azonosítottak, amelyek a különböző Big Data meghatározások döntő többségében megtalálhatók voltak. Ezek (1) az *információ jellege;* (2) a *begyűjtéshez és tároláshoz alkalmazott technológia;* (3) az *elemzéshez használt módszerek;* (4) továbbá a

1 Az angol nyelvű irodalomban a Big Data kifejezés írásmódja nagy kezdőbetűkkel terjedt el, elsősorban azért, mert így lehetett fogalomként megkülönböztetni az egyszerű „nagy adat" szókapcsolattól. Bár magyar nyelvű szövegekben ez a megkülönböztetés nem lenne szükséges, az egységesség jegyében mégis így terjedt el és én is így használom tanulmányomban.

mindezek segítségével elérhető társadalmi, gazdasági hatások voltak (De Mauro, Greco és Grimaldi 2015). Jelen írás szempontjából elsősorban az utolsó két területnek van jelentősége. A Big Data rendszerekben található információknak nem csak a digitális formátum a közös jellemzője, hanem ezeknek a strukturált, félig strukturált vagy strukturálatlan adatként való hozzáférhetősége is (*datafication*), hiszen csak így válhat belőlük valódi nyersanyag. (Az adatokhoz való hozzáférés korlátaiból fakadó újfajta hatalmi aszimmetriák kialakulását is itt érdemes megemlíteni, azonban ezzel a tanulmányban részletesen nem foglalkozom.) A technológia felől közelítve az adatok tárolásához is újfajta megoldásokra van szükség, ahol egyszerre kell, hogy szempont legyen a gyors hozzáférhetőség, illetve a hatékony fizikai tárolás, továbbá – a harmadik szempontból kiindulva – a Big Data típusú elemzések újfajta elemzési technikákat is igényelnek. Természetesen itt nem arról van szó, hogy a „nagy adatok" elemzéséhez minden esetben merőben új statisztikai eljárásokra lenne szükség. (Nagyon sok strukturált adattal dolgozó elemzés nem lép túl a klasszikus statisztikai módszerek alkalmazásán.) Azonban a félig strukturált és strukturálatlan információk (pl. hálózati adatok, képek, hangok, videók, szövegek) rendszerezése, összekapcsolása és értelmezése valóban új módszereket igényel. Az elemzésekhez használt eszközök és alkalmazások (pl. adatvizualizáció, dashboardok), a real-time elemzési technikák dinamikus fejlődéséhez pedig aligha férhet kétség.

Végezetül – a negyedik kulcsterületből kiindulva – a Big Data jelenségéhez közelíthetünk az eredmények és hatások felől is. Danah boyd és Kate Crawford meghatározásában például kulcsszerepet kap az a „mitológiai" elem is, melynek lényege az a széles körű meggyőződés, hogy a nagy adatbázisok és az újfajta elemzési technikák egy „magasabb szintű értelmet" képviselnek: olyan ismereteket, melyek korábban nem voltak (nem lehettek) birtokunkban. Ezt az új tudást pedig az „igazságosság, az objektivitás és a pontosság aurája lengi körül" (boyd és Crawford 2012). Ez az egyértelműen pozitív várakozás nyilvánvalóan árnyalásra szorul. Jelen írás is ehhez kíván hozzájárulni.

Láthatjuk tehát, hogy – túllépve a technológiai fókuszon, illetve a ma már közhelyszerűen használt *„Big Data* = öt *V betűvel kezdődő angol szó"*[2] definíciós univerzumon – az adatvezérelt világ legfontosabb kérdései a potenciális pozitív és negatív társadalmi hatások felől ragadhatók meg.

Algoritmus

Az algoritmust, mint fogalmat néhány évvel ezelőttig lényegében csak a matematika és a számítástechnika világában jártas emberek használták. Napjainkban azonban a digitalizáció, és ezen belül is a minket körülvevő adatvezérelt világ egyik legfontosabb kulcsszavaként tekinthetünk rá, szűk szakmai közegéből kilépve pedig ma már egyre inkább a közbeszéd része. Ennek következménye, hogy egyre többen vannak azok is, akik tisztában vannak azzal, hogy életük bizonyos területeit mind gyakrabban szervezik és alakítják különböző algoritmusok. A széles körű ismertség azonban az eredeti jelentést is elhomályosította és részben átalakította. Ezért az algoritmusok egyre inkább úgy jelennek meg, mint az életünk részévé vált misztikus – tehát az emberek többsége számára ismeretlen és átláthatatlan – „dolgok".

2 Eredetileg a Big Data körülírására a (1) Volume (mennyiség), (2) Variety (változatosság), (3) Velocity (sebesség), (4) Veracity (hitelesség/valódiság) kifejezéseket használták. Ez később kiegészült a (5) Value (érték) fogalmával.

A világ bármely pontján élő tetszőleges felhasználónak az algoritmus szó hallatán nagy valószínűséggel a Google Page Ranking vagy a saját személyre szabott Facebook feed-je jut eszébe, mint a két leginkább kézenfekvő és legismertebb példa. Az algoritmusok ugyanakkor szinte mindenhol jelen vannak, ahol a digitális világ által generált nagy mennyiségű adatban mintázatokat kell keresni, csoportokba rendezni vagy rangsorolni felhasználókat, vagy éppen előre jelezni viselkedéseket és ezek alapján teljesen automatizált döntéseket hozni, vagy azokat „igazi" emberek számára valamilyen módon előkészíteni. A többnyire emberi beavatkozás nélküli, hihetetlen mennyiségű és gyorsaságú automatizált döntést hozó rendszerekre jó példa lehet a programozott tőzsdei kereskedés. Ezeket az eljárásokat még mindig elsősorban az üzleti életben hasznosítják, de egyre több olyan állami feladat van, amelyben közvetve vagy közvetlenül megjelennek döntéseket előkészítő, egy-egy alrendszer működését optimalizálni kívánó, állampolgárokat különböző szempontok szerint osztályozó, rangsoroló algoritmusok (Schneier 2015). Ilyen algoritmusok szűrik ki azokat az egyéneket is, akiket származásuk, bőrszínük, nevük, vallásuk, utazási mintázataik és más adataik alapján külön biztonsági ellenőrzésnek vetnek alá, vagy végső esetben nem engednek felszállni egy repülőjáratra. Végezetül, a technológia társadalmakat átalakító „mumusaként" az algoritmus fogalma több olyan technológiával is összeforrt a közbeszédben, amelyek alapvetően formáltak és formálnak majd át teljes gazdasági szektorokat és az ezekhez kapcsolódó foglalkozásokat (pl. önvezető autók) (Dourish 2016).

Miközben a fenti példák a gyakorlat felől közelítik meg az algoritmus jelentését, maga a fogalom ennél egyszerűbben is meghatározható. A számítástechnika világában az algoritmus egy számítási folyamat absztrakt, formalizált leírása. Önmagában ez ugyanaz a jelenség, amelynek segítségével a mindennapi életünk során felmerült feladatokat, problémákat véges számú, logikailag egymásra épülő lépések sorozatán keresztül megoldjuk (pl. meghatározott alapanyagokból főzünk, eljutunk A pontból B-be, stb.). Miközben ezeket a tevékenységeket végezzük, természetesen a legritkább esetben gondolunk arra, hogy most éppen egy algoritmust hajtunk végre.

A számítástechnikában a program és az algoritmus közötti határvonal elsőre korántsem egyértelmű. Az algoritmus egyszerre kevesebb is és több is, mint egy program. Kevesebb abban az értelemben, hogy egy program egyszerre tartalmazhat algoritmusalapú és nem algoritmusalapú elemeket. Egy algoritmus absztrakt lépései operatív szinten a programban jelennek meg és válnak végrehajthatóvá. Itt tehát az algoritmus a program egyik komponense. Egy program viszont csupán az adott implementációs környezet (programnyelv) jelentette korlátok között képes működni, miközben az algoritmus alapvetően absztrakt mivoltából fakadóan ilyen értelemben nem kötött. Ennyiben tehát többet és mást jelent a programnál. (De természetesen az algoritmust magát le kell programozni az adott környezetben.)

Összefoglalóan tehát azt mondhatjuk, hogy a digitális világ algoritmusai – miközben végtelen számú feladat autonóm megoldását segíthetik elő – minden esetben szerves részei egy adott programnak. Mint a későbbiekben látni fogjuk, ez a tény a külső ellenőrzés lehetőségének szempontjából döntő jelentőségű lesz.

Diszkrimináció

A Big Data és az algoritmusok mellett a tanulmány harmadik kulcskifejezése a diszkrimináció. Ezt elsősorban szociológiai, nem pedig jogi értelemben használom. A kettő természe-

tesen nem választható el egymástól, azonban a diszkrimináció szociológiai meghatározása több olyan „szoft" jellegzetességet is magába foglal, amelyek önmagukban még nem feltétlenül merítik ki a hátrányos megkülönböztetés jogi kategóriáit. A tankönyvi definíció szerint a diszkrimináció az egyenlő bánásmód elvével ellentétes módon, (esély)egyenlőtlenséget, valamilyen gazdasági vagy társadalmi igazságtalanságot hoz létre különböző társadalmi csoportok között. A folyamat lényege a hatalmi aszimmetria, amely bizonyos tulajdonságok (életkor, nem, származás, bőrszín, szexuális beállítottság, társadalmi státusz, technológiához való hozzáférés stb.) alapján különbséget tesz emberek egyes csoportjai között, miközben figyelmen kívül hagyja az egyéni érdemeket és képességeket, azaz valamely csoport-hovatartozás alapján befolyásolja az egyén esélyeinek alakulását.

Szociológiai értelemben a diszkrimináció kétféleképpen valósulhat meg. Közvetlen diszkriminációról akkor beszélünk, amikor – többnyire szándékosan – valakit vagy valakiket egy adott tulajdonság alapján másoknál kedvezőtlenebbül kezelnek, és így jön létre a hátrányos megkülönböztetés. Ezzel szemben a közvetett diszkrimináció nem egy új egyenlőtlenség létrehozataláról szól, hanem a már meglévő káros gyakorlatok további alkalmazásáról és a hatás felnagyításáról. Ilyen például az, amikor a közvetlen diszkrimináció során megvalósuló, valaki számára méltánytalan eljárások további döntések alapjául szolgálnak, és öngerjesztő folyamatként még inkább intézményesülnek. Abban az esetben is közvetlen diszkriminációról beszélünk, amikor egy látszólag semleges, elfogulatlan, azaz nem hátrányos megkülönböztetés valamilyen létező diszkrimináció „proxyjaként" hoz létre egyenlőtlen társadalmi helyzeteket, megsértve ezzel az egyenlő bánásmód követelményét. Fontos kiemelni, hogy a munkaerőpiacon, oktatásban és más fontos alrendszerekben megvalósuló diszkrimináció kutatása hosszú múltra és jól bevált módszertani megoldásokra tekint vissza.

Jogi szempontból közelítve a diszkrimináció tilalma egy speciális, háromszereplős viszony, olyan különleges jog, amely a szabadságjogokkal ellentétben nem a cselekvés állami beavatkozástól mentes, privát területét jelöli ki, hanem – az igazságos bánásmódot célozva az összehasonlítható helyzetek között – az állami védelmet garantálja a hátrányos megkülönböztetéssel szemben.

Az egyenlőség érvényesülésének zálogai az egyes államok alkotmányai, illetve a nemzetközi dokumentumok diszkriminációtilalmat megfogalmazó szabályai, amelyek gondolata először a 18. században, az amerikai Függetlenségi nyilatkozatban jelent meg, azóta pedig az egyenlőség eszméje mind tágabb teret hódít, és olyan univerzális elvvé vált, amely minden fejlett ország jogrendszerének alappillérét jelenti.

Az egyenlő bánásmód általános elvének emellett számtalan speciális megjelenési formája is van, így például Magyarországon az egyenlő bánásmódról és az esélyegyenlőség előmozdításáról szóló 2003. évi CXXV. törvény,[3] de hasonló törvényi szabályozások vannak hatályban más országokban is. Az Egyesült Államokban a klasszikus tiltott klasszifikáció (nem, életkor, bőrszín, származás, családi állapot stb.) mellett külön törvény tiltja többek között a hitelezés[4], lakáskiadás és -értékesítés[5] során megvalósuló diszkriminációt. Mindezen elveknek a betartatására, a diszkrimináció feltárására és szankcionálására különböző állami szervek hivatottak.

3 https://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=A0300125.TV.
4 Equal Credit Opportunity Act.
5 Fair Housing Act.

A fogalmak rövid áttekintése után joggal merülhet fel a kérdés, vajon milyen módon kapcsolódik a társadalmi diszkrimináció a Big Data és az algoritmusok világához? A diszkrimináció és a hátterében meghúzódó előítéletek a társadalmak legalapvetőbb működési mechanizmusai közé tartoznak. Nagyon is emberi és szubjektív, érzelmi vagy éppen irracionális viselkedésekről beszélünk, amelyeket alapvetően a rendelkezésre álló ismeretek korlátozottsága vagy figyelmen kívül hagyása működtet. Ezzel szemben az adatok és az algoritmusok a technológia oldaláról szemlélve *per definitionem* ennek szöges ellentétei: gépies kiszámíthatóság, racionalitás és objektivitás. Ez azonban csak akkor igaz, ha figyelmen kívül hagyjuk a természetesen itt is meglévő emberi tényezőt. Az adatokat emberek gyűjtik, tárolják és dolgozzák fel, az algoritmusok lépéseit emberek tervezik és programozzák, még akkor is, ha egyre inkább teret kapnak a gépi tanulás (*machine learning*) és a mély tanulás (*deep learning*) elvein alapuló algoritmusok is. Ez utóbbi esetében már elmosódni látszódnak a különbségek ember és gép felelőssége között, hiszen bár magának a gépi tanulásnak a kezdeti ismeretanyagát és magát a célt a mérnökök határozzák meg, és a folyamatot is emberi beavatkozás indítja el, az eredmény viszont egy olyan folyamat során jön létre, amelybe kívülről nem lehet belelátni, és az algoritmus által hozott döntésekhez kapcsolódó magyarázatok az emberek számára értelmezhetetlenek. A célok és a rendelkezésre álló adatnyersanyag tekintetében az emberi tényező tehát minden esetben döntő jelentőségű, de egyre inkább előtérbe kerül annak a kérdése, hogy miként tudnak az algoritmusok a készítőik, illetve a működésük által érintett emberek számára folyamatos visszacsatolást adni az egyes döntések részletes ok-okozati hátteréről és azokról az etikai dilemmákról, amelyekkel a döntéseik során szembesülnek.

Az Obama-kormányzat Big Datával foglalkozó munkacsoportja (*Big Data Working Group*) már 2014-ben és 2015-ben is közzétett egy-egy jelentést, amelyek az adatok kormányzati és üzleti célú felhasználásának lehetőségeit és társadalmi kockázatait vették számításba (The White House, 2014, 2015). A 2014-ben megjelent összegzés egyik fontos üzenete volt, hogy felhívta a figyelmet a társadalmi diszkrimináció automatizált döntésekben való kódolt megjelenésére, illetve ezeknek a rendszereknek a nem transzparens és a kívülállók számára csak nagyon korlátozottan kontrollálható működésére.

A lehetséges társadalmi következmények nagyságrendjét tekintve tehát egy viszonylag új jelenségről van szó. A probléma és a mögöttes mechanizmusok ismertek, egyre több az ezzel kapcsolatos hír, és egyre többet tudunk arról is, milyen szándékos vagy nem szándékos lépéseken keresztül jöhetnek létre igazságtalan és diszkriminatív döntések vagy az ezekre épülő komplex rendszerek.

Kétféle módon jelenhet meg a diszkrimináció az algoritmusalapú rendszerekben. Az első esetben már maguk az algoritmus számára bemenő információt (input) szolgáltató adatok is megkérdőjelezhető minőségűek. A társadalmi igazságosság, az egyenlő bánásmód szempontjából pedig az eredmény az alkalmazott algoritmus tökéletességétől és részrehajlásától függetlenül problémás, hiszen már a kezdetektől fogva jelen van a „rossz" adat a rendszerben.

Második esetben nem az adattal, hanem az azt feldolgozó algoritmus működésével van a probléma. Ennek lényege, hogy a matematikai-statisztikai alapon működő döntések vagy az egyre inkább az emberi agy működéséhez hasonló neurális hálózatok elvén nyugvó mesterséges intelligencia (*Artificial Intelligence* – AI), az eredeti szándéktól akár teljesen füg-

getlenül is, a bennük megjelenő „kulturális kód" következményeként a meglévő társadalmi igazságtalanságokat erősítik fel vagy hoznak létre újfajta diszkriminatív helyzeteket.

Nézzünk most meg mindkét alaptípust részletesebben.

Amikor nem az algoritmus a hibás…

Ahhoz, hogy egy algoritmusalapú rendszer működni tudjon, adatokra van szüksége. A hagyományos kutatások világában jól bevált receptek vannak annak eldöntésére, hogy mitől *jó* vagy *rossz* egy adat. Lényegében ugyanezek a szabályok érvényesek az algoritmusok világában is. Az adatok gyűjtésénél, előkészítésénél, tisztításánál, aggregálásánál annak meghatározásakor, hogy mit tartunk fontosnak és megőrzendőnek, bármilyen elemzésről is van szó, számtalan olyan döntést kell meghoznunk, amelyeknek később nem várt következményei lehetnek. Rosszul megválasztott, hibás, vagy a valóságot csak részben lefedő adatokból jó eséllyel rossz eredmények fognak születni. Ez természetesen a hagyományos kutatások világában is problémát jelent. Egy olyan rendszer esetében azonban, ahol az adatok alapján a másodperc tört része alatt születnek emberek életére hatással levő döntések oly módon, hogy az algoritmus belső folyamatait egy adott személyre vonatkozóan sem az érintettek, de sok esetben maguk a készítők sem tudják kontrollálni, már súlyosabb és nehezebben korrigálható következményekkel járhat. Például egy modell, amelynek működése elméletileg a teljes populációra lehet hatással, ám mégsem reprezentatív adatokból dolgozik, akarva-akaratlanul bizonyos társadalmi csoportoknak kedvez, illetve másokat hátrányosan érint.

Természetesen nem csak az adatbázisokat kezelők döntésein múlik a bemenő adat minősége és torzítatlansága. Bizonyos rendszerek esetében azt létrehozhatják maguk a felhasználók is. Ilyenkor nem történik más, mint az, hogy a társadalomban meglévő előítéletek és igazságtalanságok jelentik a bemeneti adatokat, az algoritmus eredményei pedig ezek következményeit felerősítik és matematikai-statisztikai alapokon tovább intézményesítik.

Ennek alátámasztására az elmúlt években számtalan olyan kutatási eredmény látott napvilágot, amelyek az online (piac)terekben megvalósuló diszkriminációt és társadalmi igazságtalanságokat próbálták feltárni empirikus módszerekkel. Doleac és Stein például az egyik legnagyobb amerikai apróhirdetésekre szakosodott weboldalon (craigslist.com) vizsgálták, hogy mekkora az érdeklődések, illetve konkrét vásárlási ajánlatok számában az eltérés akkor, ha a készüléket bemutató fényképen egy szemmel láthatóan fehér, illetve fekete bőrű személy keze látható. Az eredmények azt mutatták, hogy a felhasználók által feketének gondolt eladók szignifikánsan kevesebb ajánlatot kaptak, és azokat is alacsonyabb árakon (Doleac és Stein 2013). Egy ehhez nagyon hasonló kutatás ugyanerre az eredményre jutott 2015-ben: az Ebay-en árult, baseballjátékosokat ábrázoló képek esetében derült az ki, hogy a sötét bőrű eladók képei átlagosan 20%-kal kevesebb pénzért keltek el (Ayres, Banaji és Jolls 2015). Ugyanezek a diszkriminatív társadalmi reflexek a világ mára legnagyobb online szállásközvetítőjének, az Airbnb-nek a rendszerében is kimutathatók voltak. Az afroamerikai hangzású névvel rendelkező potenciális jelentkezőket 16%-kal kisebb valószínűséggel fogadták el bérlőnek a tipikusan fehér hangzású nevű emberekhez képest (Edelman 2016). Egy másik kutatás pedig a bérbeadói oldallal kapcsolatban tárta fel, hogy a Kalifornia államban lévő Oakland városában az ázsiai származásúak átlagosan 20%-kal kevesebbet kerestek a hasonló ingatlanokat kínáló fehér lakossággal összehasonlítva (Wang D. 2015). Végezetül

Hannák és szerzőtársai a munkaerő-közvetítésre szakosodott TaskRabbit és Fiverr esetében találtak empirikus bizonyítékot a nem-, bőrszín- és származásalapú diszkriminatív eltérésekre (Hannák et al. 2017).

Ezek a(z amerikai) kutatási eredmények akár a hétköznapi tapasztalatok, akár a társadalomtudományok szemüvegén keresztül szemlélve nem túl meglepők. Az online közvetítő szolgáltatások itt nem tettek mást, mint teret engedtek az amúgy is létező sztereotípiákból táplálkozó diszkriminációnak. Tehát nem elsősorban az algoritmus, hanem a felhasználóknak a szolgáltatással való interakciója, azaz a bemenő adat volt előítéletes. Erre a hatásra pedig még maga az algoritmus is ráerősíthet azáltal, ha például bizonyos sztereotipizált tulajdonságai alapján lepontozza, hátrébb sorolja az adott felhasználót a rendszerben, aki így további hátrányba kerül a többiekhez képest. Erre lehet jó példa az Uber saját sofőrjeit értékelő rendszere. Ennek lényege, hogy azok a sofőrök, akiknek az aggregált értékelési pontszáma egy bizonyos szint alá kerül, alacsonyabb jövedelemben részesülnek, vagy csak egyszerűen a rendszer számukra nem kínál fel olyan lehetőségeket, amelyek bizonyos minőségi paraméterek (pl. minimális arányú visszautasított fuvar, adott idő alatt meghatározott mennyiségű fuvar lebonyolítása, illetve kiváló értékelés az utasoktól) elérése esetén kiemelt bérezésben részesítik a sofőrt. Fontos az is, hogy a tartósan alacsony pontszám az Uber rendszeréből való kizárást (azaz a sofőr elbocsátását) eredményezi (Rosenblat 2016). A diszkrimináció folyamata itt úgy működik, hogy – egyrészt – valamilyen etnikai kisebbségi csoportba tartozó sofőr nagyobb valószínűséggel kap rosszabb értékelést a (nem kisebbségi) utastól; másrészt pedig a sofőrök sem szívesen vesznek fel ilyen utast, illetve inkább részesítik előnyben a gazdagabb és „fehérebb" városrészeket (Rogers 2015). Ezek a torzítások természetesen a hagyományos taxizásban is megfigyelhetők, és ehhez nagyon hasonló mechanizmusokat a klasszikus diszkriminációkutatások korábban részletesen feltártak (Riach és Rich 2002; Sik és Simonovits 2011). A kérdés itt inkább az, vajon az Uber algoritmusa ezekre a folyamatokra „ráerősít-e" azáltal, hogy figyelmen kívül hagyja ezeket a szempontokat; vagy éppen ellenkezőleg, készítői megtanítják-e a diszkriminációs mechanizmusok felismerésére és a lehetőségekhez mért aktív semlegesítésére. Az ilyen jellegű korrekció ugyanakkor pozitív diszkriminációnak tekintendő, amely újabb módszertani, jogi és etikai kérdések sorát veti fel, arról nem is beszélve, hogy ez a profit maximalizálásának alapvető elvével is ellentétes lehet. A tapasztalatok, illetve a minőségi kutatási eredmények hiányában pedig ma még alig tudjuk, miként lehetne az algoritmusainkat megtanítani arra, hogy egy olyan bonyolult kérdésben, mint az esélyeknek a társadalmi igazságosság jegyében való kiegyenlítése, sikeresek legyenek.

Amikor az algoritmus a hibás…

A fentiekben néhány olyan példát mutattam be, ahol a diszkrimináció forrása elsősorban nem maga az algoritmus volt. Következzen most egy rövid áttekintés azokról a jelenségekről és problémákról, ahol kifejezetten az alkalmazott matematikai-statisztikai modell, az adatokon alapuló emberi és gépi döntések hoznak létre előítéletes, vagy csak egyszerűen rossz döntéseket.

Ideális esetben az adat- és algoritmusalapú döntéseken nyugvó rendszerek egyik fontos ígérete, hogy kiküszöbölik az emberi tényezőt, azaz esetünkben a társadalmi igazságosság és az egyenlő bánásmód szempontjából helytelen és elfogult döntéseket. Ez a feltételezés azt jelenti, hogy az algoritmusok alapvetően értéksemlegesek. Ez azonban csak akkor igaz, ha

az ilyen rendszereknek mind a megtervezése, mind pedig a működtetése értéksemlegesen történik, azaz esetünkben folyamatosan és teljes mértékben érvényesül az egyenlő esélyek és a társadalmi igazságosság elve (angolul: *equal opportunity by design*). Az algoritmusokat viszont – mint arról már szó esett – emberek hozzák létre, és ezeknek a folyamatoknak, akár tetszik, akár nem, szerves része, hogy a készítőiknek szubjektív, értékalapú döntéseket kell meghozniuk. Ugyanannak a problémának a megoldására szolgáló algoritmust két különböző személy vagy csoport nagy valószínűséggel kétféle módon készít el. Minél összetettebb a probléma, annál több olyan döntési pont van, ahol az értékalapú döntések miatt más és más lehet a végeredmény. Általánosságban pedig kijelenthetjük, hogy az algoritmusok készítői, a programozók, és az őket foglalkoztató szervezetek erkölcsi szempontból (is) felelősek az általuk létrehozott modellekért.

De mik lehetnek az értékalapú döntések egy algoritmus esetében? A digitális világ bináris elemi alkotóelemei az „IGAZ" (1) vagy „HAMIS" (0) értékek. Ebből következik, hogy az algoritmusokat is végső soron – a legelemibb döntési szinteken – egyszerű igaz-hamis kérdések sorozatára kell lebontanunk, még akkor is, ha tudjuk, hogy természetesen a végeredmény nem feltétlenül csak kétféle (IGEN/NEM), hanem ennél jóval több kategória is lehet. Például a gyakorlatban ez azt jelenti, hogy összetett társadalmi jelenségekkel kapcsolatban is számos olyan értékalapú döntést kell meghozni, ahol a *jó-rossz, sok-kevés, nagyon-kicsit* stb. ellentétpárok egy kontinuumon helyezkednek el, tehát nincsenek egyértelműen kijelölhető, objektív határvonalak. Ez pedig elvezet bennünket a döntéselmélet klasszikus kérdéséhez: az elsőfajú (*false positive)* és a másodfajú (*false negative*) hiba minimalizálása közötti egyensúlyozáshoz. Az ilyen helyzetekben muszáj valamilyen értékalapú döntést meghozni, vagy más szavakkal „valamelyik ujjunkba muszáj beleharapnunk". Például egy diagnosztikában használt képfelismerő algoritmusnak döntést kell hoznia arról, hogy egy röntgenfelvételen látható folt vajon annyira sötét-e (vagy világos), hogy ez alapján IGAZ (pozitív) vagy HAMIS (negatív) jelzést kell továbbküldenie egy meghatározott diagnózissal kapcsolatban. Egy bankban működő másik algoritmusnak pedig arról kell döntenie, hogy a hitelkérelmet benyújtó ügyfél több évre visszamenőleg összesített élelmiszer-kiadásai alapján „IGAZ" vagy „HAMIS" jelzést küldjön azzal kapcsolatban, hogy várhatóan visszafizeti-e a felvett hitelét.

Az első algoritmus esetében, amennyiben túl szigorúak ezek az ember által meghatározott küszöbértékek (elsőfajú hiba), olyanokat is betegként fog azonosítani az algoritmus, akik valójában nem azok, illetve a második esetben a bank olyan embereket is hitelképesnek fog tartani, akik nem tudják majd visszafizetni a felvett hiteleiket. Ellentétes esetben viszont – ha túl „megengedő" az algoritmus – beteg embereket fog egészségesként diagnosztizálni, vagy amúgy jó adósoktól fogja megtagadni a hitelfelvétel lehetőségét. A bűnüldözési és bűnmegelőzési szervek, amelyek számára igencsak vonzó a gyanúalapú, nyomozást igénylő esetkezelés kiváltása algoritmikus módszerekkel, hasonló dilemmával szembesülnek. Ezeknek az intézményeknek az alapvető működési logikájából fakad, hogy az elsőfajú hiba inkább megengedett, mint a másodfajú hiba – legyen szó akár a hagyományos, akár az algoritmusalapú felderítésről.[6] A valóságban természetesen soha nem ilyen egyszerű egy algoritmus

6 Ennek talán legismertebb korai esete Hasan Elahié, akit neve és más adatai alapján, a mainál természetesen sokkal kisebb adatállomány elemzésével, tévesen terroristagyanús személynek kategorizáltak. Reakcióként azóta az élete minden pillanatát dokumentálja és ezzel bombázza a nyomozó szerveket, és voltaképpen egész életét egy művészeti projektnek fogja fel. Thousand Little Brothers című, mintegy 32 ezer önmegfigyelő fotót tartalmazó nagyméretű alkotása (2014) Budapesten is látható volt a Watching You, Watching Me kiállításon 2015 őszén.

működése, azonban ezek a példák is jól jelzik, hogy az sem egyértelmű, vajon egy-egy rendszer működését milyen eredményekre optimalizáljuk. Inkább a téves vagy az elmaradt riasztás az elkerülendő? Az egészségügyben mindenképpen az az elkerülendő, hogy pozitív esetek maradjanak felderítetlenül, akár annak kárára is, hogy negatív esetek pozitívként kerülnek diagnosztizálásra. Egy bank esetében azonban éppen fordított a helyzet, hiszen itt a konzervatív stratégia lényege, ha a rosszul fizető adósok számát minimalizáljuk, miközben potenciálisan jó adósokat is szükségképpen elveszítünk. (A pénzbőség és a piaci verseny közepette ezt a szabályt felejtette el alkalmazni sok bank a 2000-es években.) Annak eldöntése, vajon mikor melyik stratégia a helyes és hol húzzuk meg a határvonalakat, folyamatosan értékalapú döntéseket igényel az algoritmusok készítői részéről. Ennyiben tehát máris sikerült árnyalnunk az algoritmusok objektivitásával és értéksemlegességével kapcsolatos gondolatokat. Az igazi problémát ugyanakkor nem a szubjektív értékítéletek megléte, hanem ezek transzparenciája, ellenőrizhetősége és esetleges korrigálhatósága jelenti.

A társadalmi szempontból igazságtalan, diszkriminatív algoritmusok problémája, az ezzel kapcsolatos tudományos és policyjellegű tanulmányok mind ez idáig elsősorban az USA-ban jelentek meg (Upturn 2014). (Részben emiatt is érezheti úgy az olvasó, hogy az itt leírtak túlságosan távoliak és „nagyon amerikaiak".) Ez nem véletlen, hiszen az informatikai ipar globális zászlóshajó-vállalatai (Google, Apple, Microsoft, Facebook, Amazon, Netflix stb.) mellett az *on-demand* és a *sharing-economy,* a hagyományos kereskedelem (Wal-Mart, Target stb.), illetve az elmúlt közel másfél évtizedben a terrorizmus elleni harc jegyében kiépített kormányzati és titkosszolgálati rendszerek működtetése mind-mind az adatok begyűjtéséről és az algoritmusalapú döntéshozatali rendszerek bevezetéséről szólt.

Robert Pasquale 2015-ben megjelent könyvében napjaink társadalmát egy fekete dobozhoz hasonlítja. (Pasquale, 2015) A fogalom itt kettős értelemmel bír. Egyrészt, életünk egyre több pillanatának információmorzsáit figyelik és tárolják el különböző vállalatok, illetve kormányzatok. Másrészt viszont aligha vagyunk tisztában azzal, hogy ezek az információk valójában hová is kerülnek, hogyan használják fel őket, illetve milyen közvetlen vagy közvetett következményekkel járhatnak.

Ennek a jelenségnek már a szélesebb célközönségnek szánt populárisabb műfajban is van irodalma, egyik friss példája Cathy O'Neil *Weapons of Math Destruction* című könyve (O'Neil 2016). A szerző az akadémiai pályáját feladva éveken át dolgozott egy amerikai hedge fundnál, illetve számos start-up vállalkozásnál. Adattudósként (*data scientist*) az volt a feladata, hogy olyan matematikai modelleket, algoritmusokat hozzon létre, amelyek a befektetéseken elérhető profit maximalizálását, illetve emberek vásárlásainak vagy kattintásainak az előrejelzését szolgálták. Munkája során döbbent rá arra, hogy micsoda hatalom összpontosul a rendszereket létrehozó szakemberek (matematikusok, statisztikusok, informatikusok) kezében, és ezek egyéni és társadalmi szinten, akarva-akaratlanul, milyen kártékonyak tudnak lenni. Cathy O'Neil ezeknek a kártékony, algoritmus(modell)-alapú rendszereknek az összefoglaló jellemzésére a tömegpusztító fegyverek bevett angol kifejezésére rímelve bevezette a *Weapons of Math Destruction (WMD)* fogalmát.

Fontos, hogy természetesen nem minden algoritmus kártékony. O'Neil három olyan tulajdonságot nevez meg, amelyek teljesülése esetén egy modell ebbe a kategóriába sorolandó. Az első a *láthatatlanság,* amelynek lényege, hogy az érintettek vagy egyáltalán nem is tudnak a háttérben „zakatoló", kategorizáló algoritmus létezéséről, vagy pedig csak egyszerűen nincsen információjuk arról, hogy ezek milyen bemenő adatok alapján milyen típusú dön-

téseket hoznak. A második tulajdonság a *méret (skálázhatóság),* ami esetünkben arra utal, hogy a modell képes legyen „nagyra nőni", azaz emberek tömegeinek életére hatással lenni. Az itt tárgyalt algoritmusok lényege, hogy egyszerre nagyon sok ember életét befolyásolhatják. Végezetül a harmadik szempont maga a *kártékonyság,* az, hogy emberek életét negatívan legyen képes befolyásolni a modell. Ez a három tulajdonság nem csak igen-nem alapon jelenhet meg egy-egy algoritmusban: mind az átláthatatlanságnak, mind a kiterjedtségnek, mind pedig a kártékonyságnak fokozatai vannak. Sőt természetesen univerzális kártékonyságról sincsen szó: mindig vannak olyanok, akik az algoritmus döntéseinek haszonélvezői. Ahol viszont vannak győztesek (kedvezményezettek), ott lennie kell veszteseknek is. Nem mindegy azonban, hogy valaki okkal, vagy igazságtalanul, azaz ok nélkül kerül a hátrányosan megkülönböztetett csoportba. Ez a nagyon is gyakorlati következmény pedig visszavezet minket az első- és másodfajú hibák absztrakt világába.

O'Neil könyve a befektetések, a hitelezés, a biztosítási szektor, a felsőoktatás, az online hirdetési piac, a bűnmegelőzés, az igazságszolgáltatás, a munkaerő-kiválasztás, a munkaszervezés és a munkavállalók értékelése, illetve a közösségi média közvéleményt formáló szerepén keresztül mutatja be az algoritmusok sokszor igazságtalan és torz működését. Minden említett szektor esetében konkrét (amerikai) példákon keresztül szemlélteti azokat a folyamatokat, ahogyan a hatékonynak és igazságosnak gondolt vagy éppen a biztonságunkra vigyázni hivatott modellek akarva-akaratlanul egyéneket stigmatizálnak, társadalmi csoportokat diszkriminálnak vagy éppen használnak ki.

Nagyjából ugyanezen szektorok működését vette górcső alá a Barack Obama elnöksége alatt 2016 májusában kiadott hivatalos kormányzati jelentés is (Muñoz, Smith és Patil 2016). Ez a dokumentum a potenciális diszkrimináció egyik lehetséges okaként tekint arra a tényre, hogy jelenleg az Egyesült Államok állampolgárainak valamivel több mint egytizede nem rendelkezik hitelpontszámmal (*credit score*), mivel nincsen róluk elegendő adat, amiből a hitelpontszámokat kalkuláló algoritmus dolgozni tudna. Ezen túl pedig minden ötödik emberről valamilyen hibás információ is megtalálható az adatbázisokban. Ez amiatt fontos, mert az amerikai pénzügyi rendszer negatívan különbözteti meg a hitelpontszámmal nem rendelkezőket, illetve azokat, akiknek alacsony a pontszáma (ennek következményeként pedig nem, vagy csak nagyon drágán kaphatnak hitelt).

Egy másik gyakran idézett, rendszerszintű társadalmi igazságtalanságokat magában hordozó rendszer az online álláskereső szolgáltatások világa. Az elmúlt években egyre elterjedtebb gyakorlattá vált, hogy meghatározott képzettséggel és gyakorlattal rendelkező munkavállalókat kereső vállalatok az álláskeresők adatbázisaiban automatizált algoritmusok alapján előszűrik, „leválogatják" a potenciális jelölteket. De ehhez hasonló algoritmusok arra a célra is rendelkezésre állnak, ha az egy-egy állásra jelentkező több száz jelölt közül kell kezelhető számúra szűkíteni a beérkező életrajzok számát. Ezeknek a kiválasztást támogató eljárásoknak a segítségével kétségkívül időt és pénzt lehet megtakarítani, illetve egy jó algoritmus arra is alkalmas, hogy semlegesítse azokat az ismert mechanizmusokat, amelyek során tudattalanul is a hozzánk (kiválasztókhoz) hasonló jelölteket részesítjük előnyben, illetve valamilyen csoporthoz való tartozás (nem, életkor, bőrszín stb.) alapján diszkriminálunk. Az algoritmusok tehát beteljesíthetik az objektivitás és igazságosság ígéretét, ugyanakkor ennek a gyakorlatban való megvalósulása számos ponton – a legjobb szándék ellenére is – „félrecsúszhat". Nézzük meg ezt egy egyszerű példán keresztül: ezekben az esetekben az algoritmus tipikus végeredménye egy – a hitelpontszámhoz hasonló – jelölt pontszám. Az ezt

kiszámoló modell figyelembe veheti például, hogy az adott személy mikor volt utoljára alkalmazásban. Azok, akik valamilyen külső ok miatt (pl. gazdasági visszaesés, dekonjunktúra) munkanélküliek voltak, értelemszerűen rosszabb esélyekkel indulnak azokkal szemben, akik jelenleg is dolgoznak – feltéve, hogy a modell mérlegelés, azaz mindenféle háttérismeret nélkül lepontozza azokat, akik jelenleg nem dolgoznak. Sőt az is elképzelhető, hogy a modell beállításaiból fakadóan a „vonal alá" kerülnek, tehát be sem jutnak az „előszűrtek" csoportjába. Más algoritmusok azt próbálják megbecsülni, vajon egy állásra jelentkező személy a különböző egyéni jellemzői alapján mekkora valószínűséggel fog hosszabb távon is kitartani. Egy amerikai tanácsadó cég erre a célra kifejlesztett modellje historikus adatokon például azt találta, hogy az ügyfélszolgálati dolgozók körében legerősebb indikátor erre vonatkozóan a munkahelytől való távolság.[7] Más szavakkal: minél könnyebben és gyorsabban tud valaki bejutni a munkahelyére, annál nagyobb valószínűséggel fog hosszabb ideig ott dolgozni. Nem kell ahhoz túlságosan fejlett algoritmikus gondolkodással rendelkezni, hogy oksági kapcsolatot bizonyító érveket találjunk ennek az egyszerű korrelációnak a magyarázatára. Elfogadva az összefüggést, logikus és statisztikai értelemben racionális lépés lehet, ha a bizonyos távolságon túlról érkező jelentkezőket már eleve visszautasítja a modell. Így viszont nem ad esélyt azoknak, akik ugyan messzebb laknak, ugyanakkor ők ebben a modellben az „outlier" szerepét töltenék be, mivel a távolság az ő esetükben nem befolyásolná a hosszú távú elkötelezettségüket. (De ezt a gyakorlatban nem fogják tudni bebizonyítani, hiszen nem is kerülnek a humánerőforrás-osztály hús-vér dolgozóinak látókörébe…) Ha pedig ezt azzal az összefüggéssel is kiegészítjük, hogy a messzebb lakók között felülreprezentáltak a valamilyen védett tulajdonsággal rendelkező, kisebbségi csoport tagjai, azaz a távolság például a bőrszín proxyváltozójának szerepét tölti be (hiszen ők azok, akik elsősorban a külvárosi területeken élnek), már vissza is jutottunk a klasszikus munkaerőpiaci diszkrimináció tankönyvi esetéhez.

A diszkriminatív algoritmusok egy harmadik tipikus példája nem annyira a diszkrimináció klasszikus, eddig tárgyalt jelenségeihez, hanem az információkhoz való szabad hozzáférés korlátozásához, illetve az algoritmizált online médiatartalmakhoz kapcsolódik (Diakopoulos és Koliska 2016). Ez leegyszerűsítve nem más, mint a szűrőbuborékok, a személyre szabott ajánlórendszerek világa, amelyek leggyakrabban számunkra relevánsnak gondolt keresési eredményeket, híreket, termékeket, zenéket, filmeket kínálnak (Pariser 2011; Sweeney 2013). Ezeknek az algoritmusoknak a célja az, hogy azokat az információkat, amelyek közel állnak a saját valóságunkhoz, azokat a termékeket, amelyekre valahol, valamikor rákerestünk, vagy azokat a műveket, amelyeket a hozzánk hasonló emberek néznek és hallgatnak, nagyobb valószínűséggel kattintsuk, vásároljuk és fogyasszuk. Ennek a profilozáson alapuló új világnak a diszkriminációs kockázataira részletesen itt most nem térek ki. Fontos, hogy a napról napra szofisztikáltabb adatgyűjtési és adatfeldolgozási módszerek egyre pontosabban és hatásosabban képesek elérni és megszólítani különböző célcsoportokat. Ennek kockázata pedig leginkább az emberek gondolati és marketing„silókba" való betagozódása, amely számos diszkriminatív, közösségeket és társadalmi csoportokat egyszerre homogenizáló és polarizáló, végső soron pedig a demokratikus alapelveket erodáló folyamat felerősítéséhez járulhat hozzá.

7 „Robot Recruiters: Big Data and Hiring." *The Economist,* 2013. április 6. http://www.economist.com/news/business/21575820-how-software-helps-firms-hire-workers-more-efficientlyrobot-recruiters, letöltve: 2017. január 20.

Az algoritmus mint „fekete doboz”

Az előzőekben bemutatott néhány potenciálisan igazságtalan és diszkriminatív algoritmus példáján keresztül csupán a probléma alapvető jellegzetességeit szerettem volna röviden áttekinteni. Talán ezekből is látható volt, hogy a diszkriminatív funkció egy algoritmus esetében korántsem magától értetődő. Erről azonban csak úgy lehet megbizonyosodni, ha tisztában vagyunk azzal, mi is történik az algoritmus „motorházteteje” alatt. Ez pedig egy még fontosabb kérdéshez, az algoritmusok transzparenciájának és külső kontrolljának biztosításához vezet el bennünket, ami egy nagyon összetett és egyszerű megoldásokkal ma még aligha kezelhető probléma. Ennek oka részben az, hogy sok esetben az sem definiálható és még kevésbé mérhető, vajon mi a jól és a rosszul működő algoritmus közötti különbség. A gyorsaság és az optimalizálás, mint eredmény látható, de az, hogy ez etikai és jogi szempontból mennyire legitim, már jóval kevésbé. Ha elfogadjuk azt az állítást, mely szerint a jövőben egyre több területen jutnak növekvő szerephez a különböző algoritmusok, akkor azt is beláthatjuk, hogy az az állapot hosszú távon nem tartható fenn, hogy ezek működése továbbra is ennyire misztikus és átláthatatlan maradjon. Valamilyen módon tehát fel kell nyitni az embereket osztályozó és rangsoroló algoritmusoknak a „fekete dobozát” (Citron és Pasquale 2014; Rouvroy 2016). Európában a jogi keretek ehhez többé-kevésbé adottak, azonban ezek gyakorlati megvalósulása egy ilyen dinamikusan fejlődő, változó területen kihívásokkal teli (Goodman és Flaxman 2016). A személyes adatok felhasználhatóságát szabályozó európai adatvédelmi jog korábban is előírta, és a 2018-tól kötelezően alkalmazandó új egységes EU-rendelet (General Data Protection Regulation, GDPR) továbbra is előírja az automatikus döntéshozatal tilalmát, vagyis azt, hogy az érintett személyekre joghatással vagy más jelentős hatással járó ügyekben a döntés kizárólag automatizált adatkezelésen alapuljon. A hatályos magyar szabályozás szerint pedig az ilyen esetekben az érintettet – kérelmére – tájékoztatni kell az alkalmazott módszerről és annak lényegéről. A teljesen automatizált adatfeldolgozással történő döntéshozatal (ide tartozik például a profilalkotás) szűk határok közé lesz szorítva, és viszonylag sok lehetősége lesz a (tudatos) felhasználónak arra, hogy bizonyos adatainak a felhasználását letiltsa, illetve, hogy megfelelő információkat kapjon. A rendelet 2018-tól történő bevezetése és az ennek való megfelelés azonban korántsem lesz egyszerű. Ennek részleges alátámasztására az algoritmusok alábbi három jellegzetességét érdemes kiemelni.

Az algoritmus mint üzleti titok, a szerzői jog hatálya alá tartozó szellemi tulajdon és a biztonságos internet alapfeltétele

A legnagyobb internetes vállalatok hétpecsétes titokként őrzik a szolgáltatásaik lényegét adó vagy az alaptevékenységüket kiegészítő különböző algoritmusok kódjait. A Google tulajdonképpen a weboldalakat fontosságuk szerint rangsoroló keresőalgoritmusból indulva vált a világ legnagyobb internetes vállalatává.[8] A Facebook lényegét ma már az egyéni hírfolyamokat fizetett és nem fizetett tartalmak alapján létrehozó algoritmus jelenti.[9] Az Amazon vagy a

8 https://en.wikipedia.org/wiki/List_of_largest_Internet_companies.

9 http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html.

Spotify a sok millió felhasználója által generált adatokból dolgozó ajánló algoritmusa nélkül aligha tudott volna piacvezetővé válni. Ez a néhány példa csupán a jéghegy csúcsa, hiszen ezeken a mindenki által ismert vállalatokon kívül még sok ezer olyan piaci és kormányzati rendszer működik világszerte, amelyek lényegi eleme szintén valamilyen adatokat értelmező és azok alapján döntéseket hozó algoritmus. Ezek pedig a legtöbb esetben a szerzői jogok által védettek, konkrét tartalmuk üzleti titkot képez, így mások által nem megismerhető. (Nem számítva a kisebbségben lévő open-source megoldásokat, illetve azokat a vállalatokat, amelyek valamilyen mértékben nyilvánossá tették a saját algoritmusuk belső működését.) Az algoritmusokhoz való hozzáférés korlátozása a tulajdonosok elemi érdeke, hiszen azok ismeretében, a működésüket kijátszva, súlyos visszaélésekre vagy üzleti károk okozására van mód. Gondoljunk csak bele, hogy mi történne akkor, ha egy az egyben ismertté válna a Google kereső algoritmusa, vagy azok a kódok, amelyek a bankkártya-visszaélések felderítésére hivatott algoritmusokat rejtik. Az algoritmusok fejlesztői, illetve az azok ismeretéből anyagi vagy valamilyen más hasznot remélők közötti macska-egér harc régóta tart, és aligha fog egyhamar véget érni. A keresőoptimalizálás mára már szinte önálló iparággá nőtte ki magát, amely kis túlzással nem akar mást elérni, mint visszafejteni a Google kereső algoritmusát. A Google pedig (szintén kis túlzással) nem csinál mást, mint ezt folyamatosan megakadályozza. Az, hogy bizonyos algoritmusokkal kapcsolatos konkrét információk ne kerüljenek illetéktelen kezekbe, tulajdonképpen a megbízható és biztonságos internet egyik alapfeltételének tekinthető. Ugyanakkor hosszú távon nem tűnik fenntarthatónak az az állapot sem, hogy ezek tartalmához a tulajdonoson kívül senki más ne férhessen hozzá.
Az algoritmusokkal kapcsolatos titkolózás persze egy másfajta önvédelmi mechanizmusból is eredeztethető. A külvilág által nem megismerhető algoritmusok egyben a vállalatok és az állam között is újabb aszimmetriákat hoznak létre a magánvállalatok javára, amennyiben az állam ellenőrző-szabályozó szerepének egyértelmű gyengülését vesszük alapul.

Az algoritmus mint komplex, magas szakmai tudást igénylő szellemi alkotás

A világ vezető internetes vállalatainál, illetve a kiemelt kormányzati szervezeteknél az algoritmusokat magasan kvalifikált programozók, matematikusok, statisztikusok és fizikusok írják. A legjobb szakemberekért ádáz verseny folyik, ami nem véletlen, hiszen a világ ugyan „data scientist" lázban ég, azonban még mindig kevesen vannak azok, akik ennek a dinamikusan fejlődő területnek valóban „A" kategóriás szakemberei. Egy ma használatos komolyabb algoritmus pedig már régen túl van azon a komplexitáson, hogy egyetlen ember átlássa a folyamatokat. 2015-ben a Google összes szolgáltatása kb. 2 milliárd sornyi kód alapján működött, a vállalatnál dolgozó 25 ezer mérnök pedig ezek közül hetente 15 millió sort módosított 250 ezer különböző fájlban.[10] Maguk az algoritmusok ennek persze csak egy kisebb részét teszik ki, de a kereső algoritmus kódjába is legalább naponta egyszer belenyúlnak.[11] A Google természetesen extrém példa, azonban az algoritmus és az algoritmust rejtő kód komplexitása, illetve annak folyamatos változása egy külső szereplő számára igazi kihívássá teszi a kód visszafejtését, illetve az algoritmus működésének megértését. Bárki, aki „olvasni" akarja az algoritmust, egy „szakmai súlycsoportban" kell, hogy legyen annak lét-

10 https://www.wired.com/2015/09/google-2-billion-lines-codeand-one-place/.
11 https://www.cnet.com/news/the-human-process-behind-googles-algorithm/.

rehozójával. Ez a tény ismételten az állami szabályozás és ellenőrzés számára jelent ma még aligha megugorható akadályt.

„Az algoritmus, amit önmagában nem is lehet értelmezni"

Az életünket leginkább befolyásoló algoritmusok nagy része olyan gépi tanuláson (*machine learning*) alapuló eljárás, amelyekhez nagy mennyiségű adatra van szükség. Az algoritmusok ezekben az adatokban keresnek mintázatokat, amelyek tehát emberi beavatkozás nélkül, tisztán statisztikai/valószínűségi alapon jönnek létre. Ez sok esetben, leegyszerűsítve, csupán annyit jelent, hogy oksági kapcsolat helyett csak korrelációk határozzák meg a modell működését. Ez azért lényeges, mert az ellenőrzés hagyományos top-down megközelítése, ahol magából az algoritmusból próbálunk kiindulni és ez alapján megérteni a teljes folyamatot, ebben az esetben nem működik. A fordított logika szerint először az algoritmus által létrehozott eredményt kell megvizsgálnunk, majd eldöntenünk, hogy az megfelelő-e vagy sem. Ha az utóbbi eset áll fenn, akkor kezdődhet a lehetséges okok visszafejtése (Dourish 2016). Ebben az esetben viszont további nehézség, hogy a gépi tanuláson alapuló, valós időben működő algoritmusok lényegében minden időpillanatban változnak, hiszen más lesz a bejövő adat, illetve a modell is folyamatosan módosul. Ahogy kétszer nem lehet belelépni ugyanabba a folyóba, úgy egy adott helyzetet („adatállapotot") is nagyon nehéz reprodukálni egy ilyen rendszerben, ami megint csak az algoritmusok hatékony ellenőrizhetőségének szab gátat.

Hogyan lehet ellenőrizni az algoritmusokat? Az algoritmusaudit lehetősége

Csupán az előzőekben bemutatott három szempont alapján is látható, hogy az algoritmusok ellenőrzése sem technológiailag, sem pedig jogilag nem olyan egyszerű folyamat, mint mondjuk – némiképpen sarkítva – egy üzemi konyha vagy egy pénzügyi szolgáltató ellenőrző hatóságok általi ellenőrzése és auditálása. A napi menüből mintaként félretett és vizsgálatra átadott ételminták utólagos ellenőrzése bármilyen utólagos probléma (pl. szalmonellafertőzés) esetén nem tűnik túl bonyolult feladatnak, hiszen a minta adott, a vizsgálati módszerek és a jogi háttér pedig ismert. Egy bank belső rendszereinek az auditálására szintén rutinszerűen, rendszeres időközönként sor kerül. A diszkriminációra fókuszáló tudományos auditkutatások lebonyolítására szintén bejáratott módszertanok állnak rendelkezésre.

A különböző vállalatok, állami szervek által használt Big Data alapú algoritmusok ugyanakkor nagyon más természetűek. A világ most próbálja kitalálni, hogy ezek valamilyen formában történő tudományos és/vagy hivatalos auditálására szükség van-e, és ha igen, az milyen formában lenne megvalósítható. Sandvig és munkatársai például az algoritmusok auditálásának öt lehetséges módját gyűjtötték össze (Sandvig, Hamilton, Karahalios és Langbort 2014).

Véleményük szerint az első, leginkább kézenfekvő megoldásnak a közvetlen kódaudit tűnne, azaz lehetőséget kéne biztosítani arra, hogy az arra hivatottak az algoritmust tartalmazó kód másolatához hozzáférve vizsgálhassák annak igazságtalan, diszkriminatív működését. Az elv ugyan egyszerűnek tűnik, de a gyakorlatban ez nagyon sok esetben megvalósíthatatlan. Mint arról már volt szó, ezek az algoritmusok magas hozzáadott értékű szellemi tulajdont, azaz komoly üzleti értéket képviselnek. A kód nyilvánosságra kerülése lényegében egyenlő az adott vállalat ellehetetlenülésével: vagy „csak" azért, mert elveszí-

ti a versenytársakkal szembeni előnyét, vagy pedig azért, mert a kód birtokában az adott rendszer könnyedén kijátszható, a működése visszafejthető és „meghekkelhető". Kivételek természetesen akadhatnak, de az algoritmusok teljes körű, rendszerszintű nyilvánosságra hozatala lényegében elképzelhetetlen. Ezt feloldandó, elsősorban Frank Pasquale nevéhez fűződik az a gondolat, hogy olyan „kódlerakatot" kellene létrehozni az algoritmusok számára, ahol ellenőrzött körülmények között tárolnák azokat, lehetőséget biztosítva bizonyos fokú auditálásra és külső kontrollra, ugyanakkor tiszteletben tartva az üzleti érdekeket és a szerzői jogokat is (Pasquale 2010, 2015). Mint azt láttuk, ennek a lépésnek nem csak bonyolult jogi és szervezeti, hanem technikai akadályai is vannak. Az algoritmusok többnyire olyan komplex és folyamatosan változó rendszerként működnek, amelyek egyszerű „copy-paste" formában aligha mozgathatók, továbbá a valós adatok nélkül nem is igazán bírhatók szóra – már ami az esetleges diszkriminatív működések vagy más rendszerszintű társadalmi problémák tettenérését illeti.

A második ellenőrzési módszer a klasszikus noninvazív felhasználói audit. A kifejezés eredeti, orvosi jelentéséből kiindulva ez a vizsgálati módszer arra törekszik, hogy minél kevesebb kellemetlenséggel, „külsérelmi nyomok" nélkül tárja fel az algoritmus működésének belső jellegzetességeit. Ez röviden azt jelenti, hogy a felhasználók tudtával és beleegyezésével, közvetlenül tőlük gyűjtünk információkat azzal kapcsolatban, hogy egy adott algoritmus mit csinált (és mit nem csinált) az esetükben. Természetesen ez a módszer is számos buktatót rejt magában, nehéz ugyanis egy racionálisan működő gépbe programozott, nem feltétlenül racionális rendszer működését sokszor irracionálisan viselkedő, szubjektív személyes tapasztalatokról beszámoló emberek válaszain keresztül megérteni és elemezni.

A harmadik módszer az automatizált adatgyűjtésen (*scraping*) alapszik, és elsősorban akkor használható, ha a keresett információk szabadon elérhetők a weben keresztül. Ez persze csak az esetek nagyon kis részében működik. A Google kereső algoritmusa tesztelhető ezen a módon, több olyan kutatási eredmény is napvilágot látott már, amely a keresőmotor előítéletességét, rasszizmusát támasztotta alá.[12] (Természetesen nem maga az algoritmus volt rasszista, hanem az emberek korábbi előítéletes keresései alapján ezt „tanulta meg") (Angwin 2014; Sweeney 2013).

A következő auditálási eszköz a piac- és közvélemény-kutatásban évtizedek óta alkalmazott klasszikus *mystery shopping* technika online, digitális változata. Ez az eljárás leginkább azokhoz a klasszikus diszkriminációkutatási módszerekhez hasonlítható, amikor erre betanított emberek (színészek) előre megtervezett módon tesztelik egy adott munkaadó, kereskedő stb. viselkedését. Az online térben ez például nem valódi felhasználói profilok létrehozásán keresztül valósulhat meg. Ez a vizsgálat az esetek jelentős részében jogi kérdéseket vet fel, hiszen a legtöbb szolgáltató szabályzatában tiltja nem létező személyekhez kötődő felhasználói fiókok létrehozását. Ez természetesen nem jelenti azt, hogy különböző indíttatásból (az egyszerű teszteléstől a tömeges dezinformáció és propaganda terjesztésén keresztül a pénzügyi csalásokig minden ide tartozik) ezzel a lehetőséggel ne élnének sok millióan a világban. Mégis, egy alapvetően az igazságtalan vagy éppen törvénybe ütköző gyakorlatok feltárására hivatott bizonytalan kimenetelű tevékenység alapja nem lehet maga is

12 Carole Cadwalladr: Google is not "just" a platform. It frames, shapes and distorts how we see the world. *The Guardian,* 2016. december 11. (https://www.theguardian.com/commentisfree/2016/dec/11/google-frames-shapes-and-distorts-how-we-see-world).

egy egyértelmű szabályszegés. A jogi és etikai aggályok mellett pedig az sem mellékes, hogy nagyobb mennyiségű nem valódi felhasználói profil létrehozásával maga a külső auditáló is végső soron ahhoz járul hozzá, hogy az algoritmus rosszul működjön, azáltal, hogy „hamis adatokkal eteti a rendszert".

Végezetül az előző auditálási mód valódi felhasználók közreműködésével is megvalósítható. Ők lehetnek önkéntesek, vagy akár valamilyen formában fizetett tesztelők is. Annyi azonban bizonyos, hogy mindkét esetben komoly szervezésre és adminisztrációra van szükség, még akkor is, ha egy ilyen jellegű munkát hatékonyabbá tehet az Amazon Mechanical Turk vagy más ehhez hasonló mikro-munkaközvetítő rendszerek igénybevétele.

Következtetések

A tanulmány célja az volt, hogy ráirányítsa a figyelmet az algoritmusok igazságtalan és átláthatatlan működésének néhány fontos kérdésére. A szöveg jellegéből és terjedelméből fakadóan csupán „megkapargatni" tudta ennek az összetett jelenségnek a felszínét. A magyar olvasó talán joggal érezheti úgy, hogy Magyarországon és Európában ezek a folyamatok még korántsem annyira előrehaladottak, mint a hatékonyság keresésében és általában az élet kvantifikálásában mindig is élenjáró Egyesült Államokban, továbbá az adatvédelem európai jogi keretei még mindig jóval szigorúbbak az amerikai gyakorlatnál. A digitális világ azonban egyre kevésbé ismer határokat, még akkor is, ha a „pontozásalapú társadalom" nálunk még embrionálisabb állapotban van csak. Várhatóan a magyar piac mérete és közepes gazdasági fejlettségünk miatt a hazai felhasználók életét csak fáziskéséssel, és talán kicsit kevésbé intenzív formában fogják meghatározni a különböző algoritmusok az elkövetkezendő években. Ez azonban egyáltalán nem jelenti azt, hogy a társadalmi igazságosság, az információs sokszínűség és a társadalmi polarizáció szempontjából ez nálunk ne jelentene problémát. Jelenleg nagyon keveset tudunk arról, hogy az emberek (fogyasztóként és állampolgárokként egyaránt) mit tudnak és mit gondolnak az őket körülvevő digitális világnak ezekről a kérdéseiről. Pozitív fejlemény, hogy az elmúlt néhány évben megnőtt azoknak a híreknek és jó értelemben vett felvilágosító cikkeknek a száma, amelyeken keresztül jobban bekerültek a köztudatba ezek a problémák. Várható ugyanakkor, hogy ez a jelenleg nagyon is aszimmetrikus információs viszony az algoritmusok és az ezek tárgyait jelentő emberek között rövid távon jelentősen nem fog változni. Nagyon sok vita, konfliktus és felesleges erőfeszítés várható a közeljövőben, amelyeknek célja az lesz, hogy a jelenlegi állapotból elmozduljunk egy átláthatóbb és ellenőrizhetőbb világ felé. Ennek bekövetkezte azonban korántsem magától értetődő, mivel egy olyan mozgó célpontról van szó, amely néhány év alatt is hatalmas változáson és fejlődésen képes keresztülmenni, így a jelen meglátásai és javaslatai gyorsan meghaladottá válhatnak. Annak jogáról, hogy megismerjük, és ha kell, megkérdőjelezzük a rólunk döntéseket hozó algoritmusok működésének igazságosságát, nem lenne szabad lemondanunk.

Ebben a helyzetben a társadalomtudományok képviselői a természettudósokkal és az IT-specialistákkal együttműködve egy új lehetőséget is kaptak arra, hogy megpróbáljanak beleszólni ebbe a fejlődésbe. Nem máshogy, mint kidolgozva ennek a rendkívül összetett problémának a hatékony kutatási módszertanát, majd pedig a segítségével felhívva a figyelmet azokra az ismert összefüggésekre, törvényszerűségekre és várható következményekre, amelyek egy igazságosabb vagy még inkább igazságtalan világ kialakulásához vezethetnek el.

Irodalom

Angwin, Julia (2014): *Dragnet Nation. A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance*. New York: Times Books, Henry Holt and Company.

Ayres, Ian, Mahzarin Banaji és Christine Jolls (2015): Race Effects on eBay. *Rand Journal of Economics* 46(4): 891–917.

Borgman, Christine L. (2015): *Big Data, Little Data, No Data. Scholarship in the Networked World*. Cambridge, MA: MIT Press.

boyd, danah és Kate Crawford (2012): Critical Questions for Big Data. *Information, Communication & Society* 15(5): 662–679.

Citron, Danielle Keats és Frank A. Pasquale (2014): The Scored Society. Due Process for Automated Predictions. *Washington Law Review (*89): 1–33.

Csepeli György (2015): A szociológia és a Big Data. *Replika* (92–93): 171–176.

De Mauro, Andrea, Marko Greco és Michele Grimaldi (2015): What is Big Data? A Consensual Definition and a Review of Key Research Topics. *1644* (1): 97–104.

Dessewffy Tibor és Láng László (2015): Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon. *Replika* (92–93): 157–170.

Diakopoulos, Nicholas és Michael Koliska (2016): Algorithmic Transparency in the News Media. *Digital Journalism* (megjelenés előtt).

Doleac, Jennifer L. és Luke C. D. Stein (2013): The Visible Hand. Race and Online Market Outcomes. *Economic Journal* 123(572): F469–F492.

Dourish, Paul (2016): Algorithms and Their Others. Algorithmic Culture in Context. *Big Data & Society* 3(2).

Edelman, Benjamin (2016): *Responses to Airbnb's Report on Discrimination*. Interneten: http://www.benedelman.org/news/091916-1.html.

Gandomi, Amir és Murtaza Haider (2015): Beyond the Hype. Big Data Concepts, Methods, and Analytics. *International Journal of Information Management* 35(2): 137–144.

Goodman, Bryce és Seth Flaxman (2016): *European Union Regulations on Algorithmic Decision-making and a „Right to Explanation"*. (Az ICML Workshop on Human Interpretability in Machine Learning (WHI 2016) konferencián elhangzott előadás, New York, NY.) Interneten: https://arxiv.org/pdf/1606.08813v3.pdf.

Hannák Anikó, Claudia Wagner, David Garciaz, Alan Mislove, Markus Strohmaier és Christo Wilson (2017): *Bias in Online Freelance Marketplaces. Evidence from TaskRabbit and Fiverr*. (Az ACM 20. Computer-Supported Cooperative Work and Social Computing [CSCW 2017] című konfenciáján elhangzott előadás, Portland, Oregon, USA.) Interneten: http://datworkshop.org/papers/dat16-final22.pdf.

McFarland, Daniel, Kevin Lewis és Amir Goldberg (2015): Sociology in the Era of Big Data. The Ascent of Forensic Social Science. *The American Sociologist* 47(1): 12–35.

Muñoz, Cecilia, Megan Smith és Patil DJ (2016): *Big Data. A Report on Algorithmic Systems, Opportunity, and Civil Rights*. Interneten: https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

Mützel, Sophie (2015): Facing Big Data. Making Sociology Relevant. *Big Data & Society* 2(2).

O'Neil, Cathy (2016): *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pariser, Eli (2011): *The Filter Bubble. What the Internet is Hiding From You*. New York: Penguin Press.

Pasquale, Frank A. (2010): Restoring Transparency to Automated Authority. *Journal on Telecommunications and High Technology Law* 9(1): 235–256.

Pasquale, Frank A. (2015): *The Black Box Society : The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.

Puschmann, Cornelius és Jean Burgess (2014): Metaphors of Big Data. *International Journal of Communication* 8: 1690–1709. Interneten: http://ijoc.org/index.php/ijoc/article/view/2169/1162.

Riach, Peter A. és Judith Rich (2002): Field Experiments of Discrimination in the Market Place. *The Economic Journal* 112(483): F480–F518.

Rogers, Brishen (2015): The Social Costs of Uber. *The University of Chicago Law Review* 82(1): 85–102. Interneten: http://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1037&context=uclrev_online.

Rosenblat, Alex, Karen EC Levy, Solon Barocas és Tim Hwang (2016): Discriminating Tastes. Customer Ratings as Vehicles for Bias. *Intelligence & Autonomy* (október 19.). Interneten: https://datasociety.net/pubs/ia/Discriminating_Tastes_Customer_Ratings_as_Vehicles_for_Bias.pdf.

Rouvroy, Antoinette (2016): „*Of Data and Men*". *Fundamental Rights and Freedoms in a World of Big Data.* (Council of Europe, Directorate General of Human Rights and Rule of Law, T-PD-BUR(2015)09REV.) Interneten: http://www.academia.edu/16699608/_OF_DATA_AND_MEN_FUNDAMENTAL_RIGHTS_AND_FREEDOMS_IN_A_WORLD_OF_BIG_DATA.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios és Cedric Langbort (2014): *Auditing Algorithms. Research Methods for Detecting Discrimination on Internet Platforms*. (Az International Communication Association 64. éves konferenciáján tartott előadás, Seattle, WA, USA.)

Schneier, Bruce (2015): *Data and Goliath. The Hidden Battles to Collect your Data and Control your World*. New York, NY: W. W. Norton & Company.

Sik Endre és Simonovits Bori (2011): Adalékok a diszkriminációtesztelés kutatási problémáinak megismeréséhez. In *A diszkrimináció mérése*. Sik Endre és Simonovits Bori (szerk.). Budapest: ELTE TÁTK, 177–207.

Sweeney, Latanya (2013): Discrimination in Online Ad Delivery. *Queue* 11(3): 10.

Székely Iván (2015): Az adatmentes zónák szükségessége és esélye. *Replika* (92–93): 209–225.

The White House (2014): *Big Data. Seizing Opportunities, Preserving Values*. Washington, DC: The White House.

The White House (2015): *Big Data. Seizing Opportunities, Preserving Values. Interim Progress Report*. Washington DC: The White House.

Upturn (2014): *Civil Rights, Big Data, and Our Algorithmic Future*. Interneten: https://bigdata.fairness.io/wp-content/uploads/2015/04/2015-04-20-Civil-Rights-Big-Data-and-Our-Algorithmic-Future-v1.2.pdf.

Wang, David, Stephen Xi és John Gilheany (2015): The Model Minority? Not on Airbnb.com. A Hedonic Pricing Model to Quantify Racial Bias against Asian Americans. *Technology Science* (szeptember 1.). Interneten: http://techscience.org/a/2015090104/.