https://doi.org/10.65295/SZF.2025.4-4

ZŐDI ZSOLT[1]

„Félelemtörvény”-e az európai mesterségesintelligencia-rendelet?

Is the European Artificial Intelligence Regulation a “Law of Fear”?

Absztrakt

Az EU mesterségesintelligencia-rendeletét (MI-rendelet) mint a „világon első átfogó szabályozását” a mesterséges intelligenciának[2] a sajtó, a jogi szakma és maga az EU bürokrácia is mérföldkőként,[3] egyértelműen pozitív fejleményként ünnepelte. Az eufóriában kissé háttérbe szorultak azok a hangok, amelyek szerint az európai szabályozás inkonzisztens és fragmentált,[4] visszaveti Európa versenyképességét,[5] és túl korai.[6] A dolgozat az utóbbi álláspontón van, ezt az álláspontot erősíti.

Miért kellett ennyire sietni? Amellett, hogy az EU a GDPR relatív sikere, a „Brüsszel-hatás” miatt szereti világelső szabályozónak látni magát, a jelenségnek van egy mélyebb oka is. Cass Sunstein 2005-ös könyve, a *Laws of Fear*[7] azzal érvel, hogy a modern demokráciákban a politikai döntéshozók gyakran túlreagálják az időről időre kitörő pánikot, ez a reakció pedig gyakran vezet „félelemtörvényekhez”. Írásomban amellett érvelek, hogy az MI-rendelet sunsteini értelemben vett félelemtörvény: valóban elsietett és túl szigorú, ráadásul nagyobbrészt egy társadalmi szinten jelentkező, bizonyos szempontból politikai természetű szorongás indukálta. A szorongás talán indokolt, de irracionális erővel zúdul rá a mesterséges intelligenciára. Ugyanakkor mivel még nem vagyunk tisztában a technológia valódi kockázataival és főleg azok méretével, a jogszabály megalkotói értelemszerűen néhány elszigetelt korábbi esetből indultak ki. Ezek az esetek azonban korlátozottan alkalmasak arra, hogy egy átfogó szabályozás kiindulópontjai legyenek, három ok miatt is.

1 Kutatóprofesszor, Nemzeti Közszolgálati Egyetem Információs Társadalom Kutatóintézet.

2 Latham&Watkins, https://www.lw.com/en/insights/eu-ai-act-published-a-new-era-for-ai-regulation-begins

3 Például: EU: Digital Skills and Jobs Platform, https://digital-skills-jobs.europa.eu/en/latest/news/milestone-ai-regulation-eu-secures-worlds-first-ai-act

4 Fortune, https://fortune.com/2024/09/19/mark-zuckerberg-meta-europe-ai-regulation-inconsistent-max-schrems-gdpr/

5 Financial Times, https://www.ft.com/content/9b72a5f4-a6d8-41aa-95b8-c75f0bc92465

6 Geopolitical Intelligence Services AG, https://www.gisreportsonline.com/r/ai-act-eu-regulation-innovation/

7 Sunstein 2006.

Először is azért, mert az ezekben az elszigetelt esetekben manifesztálódó veszélyeknek egyike sem kötődik szükségszerűen a mesterséges intelligenciához mint technológiához. Legtöbbjük más technológiák használatával is bekövetkezett volna, illetve „elkövethető" technológia nélkül is, ami nemcsak azt teszi kétségessé, hogy pusztán az MI szabályozásával célt lehetne érni, hanem azt is, hogy bármilyen technológiaszabályozással lehetne ellenük küzdeni. Másodszor, az esetek által felvetett problémák mögött régebb óta létező társadalmi szorongások, politikai megosztottságok és társadalmi problémák húzódnak meg, amelyek egy technológia betiltásával és korlátozásával nem oldódnának meg. Végül harmadszor, szemben a klasszikus technológiaszabályozással, amely fizikai, egészségi, természetvédelmi, azaz kvantifikálható és könnyen mérhető kockázatokat próbál minimalizálni, az MI-rendeletet egy sor olyan kontextusfüggő, nem kvantifikálható kockázatot igyekszik kezelni, mint a védett kisebbségek diszkriminációja, vagy a demokráciát és a jogállamiságot fenyegető veszélyek.

Kulcsszavak: az EU Mesterséges Intelligencia Rendelete, elővigyázatossági elv, technológiaszabályozás, *ex ante* szabályozás, alapvetőjogi hatásvizsgálat, megfelelés, félelemtörvények

Abstract

The EU Artificial Intelligence Act (AI Act) was hailed as the "world's first comprehensive regulation" of artificial intelligence by the press, the legal profession and the EU bureaucracy itself, as a milestone and a clearly positive development. The euphoria somewhat overshadowed the voices claiming that the European regulation is inconsistent and fragmented, that it undermines Europe's competitiveness and that it is too early to act. The study takes the latter view, and this is a contribution to strengthen this position.

Why was there such a rush? In addition to the EU liking to see itself as the world's leading regulator due to the relative success of its GDPR, known as the „Brussels effect", there is a deeper reason for this phenomenon. Cass Sunstein's 2005 book, Laws of Fear argues that in modern democracies, policymakers often overreact to occasional panics, and that this overreaction often leads to "laws of fear." This paper argues that the AI Act is a law of fear in the Sunsteinian sense: indeed, it is hasty and overly restrictive, driven largely by societal and, in some ways, political anxiety. The anxiety may be justified, but it is being unleashed on AI with irrational force. However, since we do not yet know the true risks of the technology, and especially their magnitude, the drafters of the law naturally started from a few isolated previous cases. However, these cases are of limited use as a starting point for comprehensive regulation, for three reasons.

Firstly, because none of the dangers manifested in these isolated cases are necessarily linked to AI as a technology. Most of them would have occurred using other technologies, or even without the "perpetrated" technology, which makes it doubtful not only that the goal could be achieved by regulating AI alone, but also that any technology regulation could combat them. Secondly, the problems raised by the cases are rooted in long-standing social anxieties, political divisions, and social problems that will not be solved by banning and restricting one kind of technology. Finally, in contrast to classic technology regulation, which tries to minimize physical, health, environmental, i.e. quantifiable and easily mea-

surable risks, the AI Act tries to address a range of context-dependent, non-quantifiable risks, such as discrimination against protected minorities or threats to democracy and the rule of law.

Keywords: EU Artificial Intelligence Act, precautionary principle, regulation of technology, fundamental rights impact assessment, ex ante regulation, compliance, laws of fear

Bevezetés

Cass Sunstein 2004-ben a Cambridge-i Egyetemen John Robert Seeley-ről, a híres brit történészről elnevezett előadás-sorozatában meglepően heves támadást intézett az ekkorra már – főként Európában – általánosan bevettnek mondható „elővigyázatossági elv” (*precautionary principle*) ellen.[8] Sunstein szerint a modern demokráciákra egyre nagyobb nyomás helyeződik, hogy valahogy kezeljék az emberek – néha megalapozott, sokszor azonban megalapozatlan – félelmeit és pánikreakcióit. Azoknak a félelmeknek a kezelésére, amelyek az egészség, a biztonság és a természeti környezet területén jelentkeznek, már az 1960-as évektől kezdődően kialakult, és egyre jobban megerősödött az ún. elővigyázatossági elv mint a politikai cselekvés és a jogalkotás legfőbb igazolása. Az elv lényege, hogy egy adott területen, a három védett értéket veszélyeztető kockázatoknak nem szabad megvárni a bekövetkezését, még ez előtt megfelelő intézkedéseket kell tenni ezek elkerülésére akkor is, ha nincsen egyértelműen megállapítható és bizonyítható kauzális kapcsolat a kockázat és a kár bekövetkezése között. Sunstein szerint azonban nemcsak hogy az egész életünk „kockázatos”, hanem minden kockázatelhárításra irányuló erőfeszítésünk is további kockázatokat generál. Ha valaki fél a repüléstől, és autóval utazik, valójában nagyobb baleseti kockázatnak teszi ki magát. Ez a példa szépen mutatja, hogy a kockázatok nem egyformák, és az, hogy a közvélemény mit kezd el kockázatként kezelni, szinte esetleges. (Sunstein hosszan elemzi a kockázatokkal kapcsolatos percepciókat, de végső soron ezeknek az irracionalitását bizonyítja.) A hétköznapi emberek ugyanis nem nézegetnek statisztikákat, hanem megtörtént esetekre támaszkodnak. Ez az időről időre kitörő pánik egyik forrása, a másik pedig, hogy egy

8 Sunstein 2006.

sor technológia és tudományos módszer nem átlátható egy hétköznapi ember számára, így az kénytelen azoknak a szakembereknek a meglátásaira támaszkodni, akik vagy önös érdekből, vagy csak egyszerűen szakmai csőlátásból szintén hajlamosak bizonyos kockázatokat túlértékelni. A megtörtént kevés számú eset utáni pánikokból azután „hirtelen felindulásból elkövetett" politikai akciók és jogszabályok lesznek, azaz „félelemtörvények". Sunstein egyébként egy, nagyjából a Learned Hand (20. század eleji amerikai legfelsőbb bírósági bíró) felelősségi formulájára emlékeztető megoldást ajánl,[9] amely a potenciális kockázat bekövetkezési valószínűségét és annak súlyosságát egyaránt (szorzatként) figyelembe veszi a jogi szabályozás megalkotásakor. (A Hand-formula a felelősség mértékét határozza meg ez alapján.) Ez egyben azt is jelenti, hogy az elővigyázatossági elvet, mondja Sunstein, „antikatasztrófa elvvé" kellene változtatni, aminek a lényege, hogy nem bizonyított kauzális összefüggések és csekély bekövetkezési valószínűség esetén csak a potenciálisan katasztrofális következményekkel járókra kellene felkészülni (jogot alkotni).

Az itt következőkben azt próbálom bizonyítani, hogy a mesterséges intelligencia jelenlegi kockázatai nem mutatnak egzisztenciális katasztrófa felé. Az MI-rendelet sem számol ilyen következményekkel, ugyanakkor egy elég erős szorongás lengi körül, és ezek a szorongások – és a belőlük kibontakozó kockázatkép – szorosan kötődnek néhány megtörtént esethez. Ezeknek az eseteknek ugyanakkor egyike sincsen összefüggésben az egészség, biztonság, környezet (*health, safety and environment, HSE*) triászával, minden példa visszavezethető néhány emberi jog – egyébként nem túl súlyos – megsértésére, esetleg olyan absztrakt értékek sérelmére, mint a jogállamiság vagy a demokrácia. A rendelet pedig derék félelemtörvényként szisztematikusan túlreagálja ezeket.

Bár a technológiák jogi szabályozása hosszú és színes múltra tekint vissza,[10] a mesterséges intelligencia (MI) szabályozása ezen belül is különleges helyet foglal el. Az európai kultúrának ugyanis nagyon régóta része az az elképzelés, hogy az ember képes lehet olyasmit létrehozni, ami kiszabadulhat az uralma alól, és egzisztenciális fenyegetést jelent-

9 Grossman et al. 2006.

10 Brownsword 2017; Wiener 2004; Marchant et al. 2009.

het számára. Az ipari forradalom táján ez a félelem a gépekre és bizonyos újonnan felfedezett technológiai megoldásokra vetül ki, amely például a mesterségesen alkotott emberszerű lények kontroll alóli kiszabadulásával kapcsolatos szorongásban ölt testet.[11] Ezek miatt a lények miatt, akiket az 1920-as évektől – a névadó Čapek drámájától kezdve[12] – „robotnak” neveznek, az 1920-as évek végén az USA-ban például valóságos „robothisztéria” tört ki.[13] Elég valószínű, hogy ez nagyban inspirálta Isaac Asimovot az 1940-es évek elején,[14] hogy megalkossa a később szédületes karriert befutó „három törvényét” a robotikáról (*three laws of robotics*). A törvények konkrét tartalma talán nem is olyan lényeges (különösen, mert azok erősen kötődnek a novella kontextusához), inkább az üzenetük az: törvényeket kell alkotnunk, hogy megtiltsuk a robotoknak, hogy bántsanak vagy elpusztítsanak bennünket.

A robotapokalipszis miatti szorongás és a robotok jogi eszközökkel történő „féken tartásának” gondolata ezután változatos formákban ugyan, de folyamatosan része marad a populáris kultúrának. Mint az ismert, az 1950-es évek közepén megszületik a „testetlen robotok” gondolata a mesterséges intelligencia formájában.[15] Majd nagyjából az 1970-es évek elején beköszönt az „MI-tél”,[16] amely aztán a Deep Blue MI sakkgép Kaszparov elleni 1996-os győzelme,[17] az IBM Watson Jeopardy műveltségi vetélkedőben elért győzelme[18] és persze bizonyos technológiai újítások hatására ér végleg véget. Ezek közül az újítások és felfedezések közül csak néhány fontosabb: a vektoros jelentésábrázolás,[19] a figyelemblokkok ötlete,[20] valamint a régebb óta ismert, de ekkor újrafelfedezett neurális hálók[21] használata. Az MI jogi szabályozásával kapcsolatos hangok a tél elmúltával természetesen felerősödnek, kez-

11 Shelley 2020.
12 Čapek 1920.
13 Novak 2011.
14 Asimov 1942.
15 Russel–Norvig 2000; Masayuki 2024; Pasquinelli 2023.
16 Kim 2022.
17 Goodrich 2020.
18 Johnson 2011.
19 Mikolov et al. 2012.
20 Vaswani et al. 2017.
21 Hopfield 1982.

detben közvetetten, a Big Data narratíva keretei közt.[22] A mesterséges intelligencia szabályozásának „aranykora" a 2010-es évek közepétől kezdődik, főként az autonóm járművek és az egyre több életterületen használt profilozó és predikciós szoftverek hatására. Előbbiek az MI felelősségi,[23] utóbbiak az etikai,[24] és ezen belül is a diszkrimináló MI[25] vagy elfogult algoritmusok, gépek (*machine bias*)[26] diskurzusában jelennek meg. A 2010-es évek közepén virágzó MI-etika nagyon erősen hatott az EU MI-rendeletének szövegére, de azt természetesen más hírek és társadalmi hatások is alakították.

Fő célom tehát ezzel az írással, hogy meggyőzzem az olvasót arról, hogy az MI-rendelet irreálisan túlbecsüli az MI-ben rejlő kockázatokat, és messze túlterjeszkedik azon az észszerű akciórádiuszon, amelyre egyáltalán lehet szabályozást, főleg technológiaszabályozást létrehozni, és ezért az MI-rendelet túlzás nélkül minősíthető a sunsteini értelemben vett „félelemtörvénynek". Az érvelésem a következő gondolatmenetet követi: először is, a 2. pontban igyekszem azokat a példákat és eseményeket bemutatni, amelyek „inspirálták" a jogalkotót. Ezek az esetek a 2000-es évek végétől még a Big Data narratívába illeszkedtek, majd a 2010-es évek közepétől az MI-etika keretei között születnek. A 3. pontban röviden leírom, hogy ezek az esetek miképpen tükröződnek a rendeletben. Végül a 4. pontban igyekszem az esetekből kiindulva három érvet felsorakoztatni, amelyekben kifejtem, hogy a szabályozás indokolatlanul eltúlzott félelmeken alapszik. Első érvem az, hogy az esetekből kitűnő kockázatok nem MI-specifikusak, nem szükségszerűen kötődnek a mesterséges intelligenciához. Második érvem ezzel szorosan összefüggően az, hogy a botrányok és ügyek mögött rendszerint komoly társadalmi problémák húzódnak meg, amelyeket nem az MI szabályozásával kell felszámolni, és az MI-szabályozás nem is fog semmilyen kézzelfogható hatást gyakorolni rájuk. Végül harmadik érvem az, hogy a technológiaszabályozás a kockázatok egy részének (az alapjogi sérelmeknek) a kezelésére alkalmatlan, mivel azt fizikai rend-

22 Ződi 2018, 2017; Mayer Schönberger–Cukier 2012; Crawford–Schultz 2016; Tene–Polonetsky 2013.

23 EU parlamenti határozat 2017.

24 EU HLEG etikai iránymutatás 2019.

25 Borgesius 2018.

26 Angwin et al. 2016.

szerek kvantifikálható paramétereire találták ki. Az MI-rendelet azzal, hogy a kockázatok közé beemelte az alapjogi kockázatokat, egy nagyon komoly bizonytalansági tényezőt és elég sok felesleges adminisztratív terhet tett a jogszabály részévé.

Ügyek – egy szorongás története

Ahogy azt fentebb jeleztem, az emberi kontroll alól kiszabaduló technológia sok száz éve része az európai populáris kultúrának, de a hollywoodi filmek világán túl a valóságban a 2010-es évek elejéig nem jelentett valódi problémát. Az első ilyen jellegű ügyek a Big Data narratíva keretében jelentek meg. Ennek a narratívának a lényege, hogy az interneten és a szenzorokban spontán termelődő adatokat rejtett összefüggések feltárására, példátlanul pontos profilozásra és predikciókra lehet használni. Ezek közül a leghíresebb az USA-beli ún. Target-ügy.

> A Target-ügy: A Target amerikai áruházlánc 2002 környékén felvett egy matematikus elemzőt, aki hamarosan azt a feladatot kapta, hogy próbálja meg az egyes vevőkről rendelkezésre álló adatok alapján kitalálni, hogy melyikük terhes. A terhesség és a szülés mint életesemény a kereskedők „szent grálja”, mert ilyenkor az egyébként szinte megváltoztathatatlan, sztereotip vásárlási szokások radikálisan képesek átalakulni. Ha ilyenkor, főként a második trimeszter elején egy kereskedelmi lánc jól célzott marketinggel megtalálja a vevőt, akár évekre is képes lehet őt elkötelezni maga mellett. Hiszen ha valaki a pelenkát egy adott boltban vásárolja, nagy valószínűséggel minden mást is ott fog. A Target évek óta gyűjtött minden hozzáférhető adatot a vevőiről az ún. Guest ID felhasználásával. Ezek mellé pedig – az USA-beli jóval lazább adatvédelmi szabályokat kihasználva – még mindenféle más demográfiai információt is megvásárolt a lakóhelytől kezdve a gyermekek számán keresztül egészen a politikai meggyőződésig. Mindezen adatokat a Target saját gyűjtésű vevőinformációival összekapcsolva a Vevői Marketing Elemző részleg előrejelzési módszereket dolgozott ki. A cég statisztikusa a korábbi, ismerten terhes nők vásárlási adatait elemezve a vásárlási szoká-

> sok határozott megváltozásait tárta fel, mint például hogy a terhes nők elkezdenek illatmentes testápolókat, vitaminokat, ásványianyag-étrendkiegészítőket és hatalmas csomag vattákat vásárolni. Bizonyos termékek együttvásárlása a terhesség tényét csaknem 100 százalékos biztonsággal mutatta. Így történt, hogy az egyik vásárlójuknak, egy fiatal hölgynek gratuláló levelet és mindenféle újszülöttekkel kapcsolatos termékkatalógust küldtek. A lány apja felháborodottan tiltakozott a cég központjában egészen addig, amíg a lány be nem vallotta, hogy valóban terhes. Beszédes, hogy az újságíró, aki az esetet igyekezett feldolgozni, és egy ideig együttműködőkre talált, fokozatosan falakba ütközött: informátora egy idő után elzárkózott a kommunikációtól, a Target pedig a végén már egyáltalán nem volt hajlandó vele szóba állni.[27]

Hasonlóan nagy port vert kavart az ún. AOL (America Online) ügye is.

> Az AOL-ügy: 2006-ban az America Online (AOL) nevű internetszolgáltató nyilvánosságra hozta 650 ezer felhasználója háromhavi keresési adatait anonimizált adatbázis formájában, kutatási célokkal. Az adatbázisból igen hamar kiderült, hogy a keresések a lélek tükrei, és igencsak szenzitív információkat közvetíthetnek (például: „hogyan öljük meg a feleségünket" vagy a „depresszió és halál orvosi segítséggel" típusú keresések). Ebből, valamint az AOL egy másik, filmajánló adatbázisának tapasztalataiból hamar kiderült, hogy ezek az álnevesített adatbázisok, összekapcsolva más, nyilvános adatbázisokkal, igen könnyen visszanevesíthetők a Big Data-korszakban. Így az első adatbázisból egyértelműen beazonosíthatóvá vált egy furcsa keresőkérdéseket feltevő felhasználó, egy idős hölgy, aki elborzadva nyilatkozta az eset után, hogy: „nem tudtam, hogy valaki kukucskál a vállam felett".

Nem véletlen, hogy a 2010-es évek elejének szakirodalma hemzseg a magánélet sérelmével, az egyén „önmeghatározását korlátozó" gyakor-

27 Duhigg 2012.

latokkal, valamint a visszanevesítéssel kapcsolatos félelmektől és az ezeket a veszélyeket kezelni igyekvő jogi megoldásoktól.[28]

Válaszként a jogi narratívában történő gondolkodás és a tételes jogi szabályozással kapcsolatos javaslatok a 2010-es évek második felében jelentek meg.

A Big Datával kapcsolatos jogi szabályozási javaslatok mellett azután ugyanekkor elkezd az autonóm járművekkel kapcsolatos szakirodalom is exponenciálisan növekedni, hiszen egy ideig úgy tűnik, hogy elérhető közelségbe kerülnek a teljesen autonóm módon működő autók. Ez előbb az autonóm járművek és a robotok által okozott károk felelősségi kérdéseivel foglalkozó irodalom dömpingjét eredményezi.[29] Több mint érdekes, hogy az MI-rendelet hatálya nem terjed ki az autonóm járművekre.

A csaknem kétszáz évre visszatekintő „MI-szorongás” miatt nem meglepő, hogy más technológiák szabályozási problémáihoz képest nagyon korán és nagyon hangsúlyosan napirendre kerülnek az etikai kérdések, gyakran alapjogi (emberi jogi) köntösben. Az MIT (Massachusetts Institute of Technology, az első számú amerikai műszaki egyetem) például a 2010-es évek közepén egy ekkor már több mint negyven éves filozófiai „geget”, az ún. troliproblémát melegíti fel az autonóm járművek morális kérdéseit illusztrálandó.[30] Egy sor kormányzat, think tank, akadémiai intézet és iparági egyesület tartja fontosnak, hogy etikai kódexeket alkosson. Egy ezeket összegyűjtő weboldalon 160 ilyen dokumentumot találhatunk, és ez messze nem tartalmazza az összeset.[31] Az etikai kérdéseken belül is a legerősebb hangok a Big Data,[32] az algoritmusok,[33] majd a diszkrimináló mesterséges intelligencia lesznek.

Az összes ügy közül, amelynek a lenyomatát tartalmazza az MI-rendelet, talán a legnevezetesebb az ún. COMPAS-ügy.

28 Ohm 2010; Munoz et al. 2016.

29 Pagallo 2011; EU Robot-felelősségi állásfoglalás 2017.

30 Britannica Trolley-problem; Wallach–Allen 2009; MoralMachine, https://www.moralmachine.net/

31 Algorithm Watch, algorithmwatch.org

32 Hirsch 2014.

33 Angwin et al. 2016.

A COMPAS-ügy: A COMPAS-rendszert (Correctional Offender Management Profiling for Alternative Sanctions) a Northpointe amerikai cég fejlesztette ki, és az amerikai bíróságok nagy számban használták a 2010-es évek eleje óta. A rendszer az elkövetőtől felvett adatlap, az elkövető által elkövetett bűncselekmény adatai és a bűnügyi nyilvántartás adatai alapján egy pontszámot adott, amely az illető visszaesési (újabb bűncselekmény-elkövetési) valószínűségét fejezte ki. A pontszámot elsősorban a büntetés kiszabásakor alkalmazták.

Miután egy konkrét ügyben egy elkövető szerette volna tudni, hogy miért kapott a vártnál szigorúbb büntetést („Loomis-ügy"), nyilvánosságra került, hogy a bíró a COMPAS ajánlása alapján döntött. Ennek nyomán a Propublica jogvédő szervezet egy komoly elemzésben mutatta ki, hogy a rendszer a korábbi bűncselekmények torzító adatai alapján súlyosan diszkriminálja a színes bőrű elkövetőket, és preferálja a fehéreket. (Előbbieknek a vártnál rosszabb, utóbbiaknak a vártnál jobb pontszámot ad.) Ezenfelül a tanulmány a rendszer predikciós képességét is kétségbe vonta.[34]

Az ügy a diszkrimináló algoritmusok diskurzusának egyik legfontosabb példája lett, amely azután nemcsak egy sor tudományos tanulmányban tűnt fel, hanem különféle *policy paper*ekben és az EU döntés-előkészítő anyagaiban is, jóllehet mind a gyenge predikciós képességet, mind az egyértelmű és rendszerszintű diszkriminációs hajlandóságot is cáfolták. A rendszerrel kapcsolatban a legnagyobb tévedés az, hogy gépi tanulási elemeket tartalmaz. A rendszer ugyanis statisztikai következtetéseket alkalmaz, de nem tanul.[35]

A másik, sokat emlegetett eset a holland gyermeknevelési támogatási ügy (*Dutch childcare scandal*).

A holland gyermekgondozási támogatási ügy: A holland kormány 2004-ben fogadott el egy törvényt, amely tulajdonképpen a hiányzó megfizethető óvodai és bölcsődei hálózatot volt hivatott gyer-

34 Angwin et al 2016.
35 Kamin et al. 2024.

mekfelügyelettel kapcsolatos díjtámogatással pótolni. A lényege az volt, hogy az állam hozzájárult az óvodai, illetve a gyermekfelügyelőket (bébiszittereket) közvetítő cégek által felszámított díjakhoz, a család jövedelmétől függően 4 és 66 százalék közötti összeggel. A konstrukcióra azonnal lecsapott néhány trükkös vállalkozás, akik lényegében ál-bébiszitterközvetítő cégeket (*childminding agency*) hoztak létre, majd jellemzően a rokonokkal (például a nagyszülővel) kötöttek formális munkaszerződést, amelybe egy fiktív díjat írtak, majd megigényelték az állami támogatást, és osztoztak a konstrukcióban részt vevőkkel. A csalás hamar kitudódott, az álbébiszittereket kiközvetítő cégeket leleplezték, vezetőiket megbüntették, és elkezdték a jogosulatlanul kifizetett pénzek visszaszedését. Ennek kapcsán használtak egy már korábban, egy másik szisztematikus csalás nyomán kifejlesztett kockázatelemző algoritmust, amely – a korábbi csalási eseteket tanítóadatként használva – az idegen hangzású neveket, a bevándorlói háttérrel rendelkező családokat egyértelműen gyanúsként jelölte meg. A családok nemcsak tíz-, sőt százezres nagyságrendű euróösszeg visszafizetéséről kaptak felszólítást, hanem csaknem ezer gyermeket ki is emeltek a családokból. A bíróságra vitt esetekből egyértelműen kiderült, hogy a szoftver számtalan alkalommal jelzett helytelenül, és az eset nagy nyilvánosságot kapott, amelynek következményeként a holland Rutte-kormány lemondásra kényszerült.[36]

Az ügyben azt mindenképp látni kell, hogy bár a szoftver ez esetben csakugyan gépi tanuláson alapult, de a hatóság a holland jog alapján is jogsértő gyakorlatot követett azzal, hogy egy szoftver által megjelölt személyeket azonnal eljárás alá vont.

A jogállamra és a demokratikus folyamatokra gyakorolt hatásával kapcsolatos félelmek esetén általában két ügyet szoktak emlegetni, a 2016-os Cambridge Analytica-ügyet, és a kínai társadalmi pontozási rendszert.

36 Heikkillä 2022.

> A Cambridge Analytica-ügy: A 2010-es évek elején egy Aleksandr Kogan nevű adattudós kidolgozott egy módszert arra, hogyan lehet a közösségi médián keresztül kitölthető kérdőív (és a kérdőív kitöltését kísérő hozzájárulás), valamint az illető profiljában található adatok és aktivitás alapján pszichológiai profilokat felállítani. A „This is Your Digital Life" kérdőívet néhány százezren töltötték ki, és ezzel voltaképp még nem is volt probléma. A gond ott kezdődött, amikor a Facebook (máig tisztázatlan módon) engedélyt adott a kitöltők *ismerősei* adatainak a begyűjtésére és használatára is. Ezzel a Cambridge Analytica cég, amelynek ekkor Kogan már az alkalmazottja volt, nagyjából 50 millió felhasználó adataihoz jutott hozzá, elsöprő többségben azok engedélye nélkül, tehát még a laza USA-beli adatvédelmi sztenderdek alapján is illegálisan. A bomba akkor robbant, amikor Christopher Wylie, aki a Cambridge-i Egyetem egyik kutatójával dolgozott az adatok megszerzésén, felfedte egy angol lapnak, hogy a Cambridge Analytica – amelynek akkoriban Trump fő tanácsadója, Steve Bannon állt az élén – hogyan használta fel a 2014 elején engedély nélkül átvett személyes adatokat egy olyan rendszer felépítésére, amely képes profilt készíteni az egyes szavazókról, és amelyek alapján aztán személyre szabott politikai reklámokkal célozzák meg őket. Wylie így nyilatkozott: a „Facebookot használtuk emberek milliói profiljának begyűjtésére. Modelleket építettünk, hogy kihasználjuk, amit tudtunk róluk, és megcélozzuk belső démonjaikat. Ez volt az alapja az egész cégnek." A célzott hirdetéseket a cég állítólag mind az amerikai elnökválasztás, mind a Brexit 2016-os kampányai során felhasználta.[37]

A másik „ügy", vagy inkább projekt, amelyet gyakorta szoktak a mesterséges intelligencia etikátlan és demokráciát veszélyeztető használatára példaként felhozni, a kínai társadalmi pontozási rendszer.

> A kínai társadalmi pontozási rendszer még az 1980-as évekre vezethető vissza, amikor a kommunista párt igyekezett kitalálni

37 Osborne–Parkinson 2016; Cadwalladr–Graham-Harrison 2016; Wikipedia Cambridge Analyitica scandal.

> valamilyen banki hitelbírálati módszert azoknak a főleg falusi környezetben élő magánszemélyeknek, akiknek semmilyen banki előéletük nem volt, és akikkel kapcsolatban semmilyen dokumentáció nem állt rendelkezésre. 2002-ben a Kínai Kommunista Párt kongresszusán jelentették be a társadalmi pontozási rendszer felállítását, amely még nem tartalmazott részleteket. A rendszert később nemcsak a hitelbírálat megkönnyítésére szánták, hanem a társadalmilag rendkívül gyenge bizalomszint fokozására is, illetve más adminisztratív célok, például a régóta végrehajthatatlan bírósági ítéletek végrehajtásának segítésére. A rendszer még nincsen teljesen kiépítve, de a koncepció szerint négy eleme (vagy „modulja”) lenne, az üzleti, a kormányzati, a társadalmi és a jogi megbízhatósági rendszer, amelyekben pozitív és negatív listákra lehet felkerülni. Fontos, hogy a negatív listákon nem általában „rosszul viselkedő” embereket listáznak, hanem eleve valamilyen jogsértést elkövetőket, jellemzően a bírósági (vagy más) határozatokat ignorálókat. Amíg ők nem teljesítik a kötelezettségeiket, addig bizonyos közszolgáltatásokból (például gyorsvasúti, repülőjegy-vásárlás, ingatlanvásárlás, üdülés stb.) ki vannak zárva. A társadalmi pontozási rendszernek tehát nincsen specifikus szankciója, csak előnyök elvesztésével jár. A kínai társadalom, főként a rendkívül gyakori csalások visszaszorításának ígérete miatt elsöprő többségében támogatja a rendszert, amely egyébként nem is központi, hanem tartományi szinteken működik.[38]

Utolsóként az USA-beli köztéri kamerarendszerekkel kapcsolatos ügyet ismertetem, azzal, hogy a lentebbi ügy az USA több nagyvárosában is előfordult, és a köztéri arcfelismerést ennek nyomán csaknem minden amerikai nagyvárosban betiltották azóta.

> Köztéri arcfelismerő rendszerek az USA-ban: a Williams-ügy. Robert Julian-Borchak Williams fekete férfit a rendőrség teljesen váratlanul elfogta és letartóztatta. Miután egy éjszakát a fogdán töltött, a nyomozók egy ékszerboltban készült felvételt mutattak, ahol egy hozzá hasonló testalkatú fekete férfi órákat és ékszere-

38 Social Credit System, Wikipédia.

ket tulajdonít el. A képekről hamar kiderült, hogy nem őt ábrázolják, így Williams hamar szabadlábra került, de az ügynek nagy visszhangja lett, mivel az is kiderült, hogy a rendőrség Amerika-szerte használ nyilvános helyeken arcfelismerő algoritmusokat, amelyek ugyan (akkor még) a fehér férfiakat egészen nagy pontossággal képesek voltak felismerni, de a feketék esetén rengeteg hamis pozitív találatot adtak. Williams egy ilyen rendszernek lett az áldozata. Detroitban és több amerikai nagyvárosban teljesen leállították, sőt kifejezetten betiltották az arcfelismerő rendszerek használatát.[39]

Az MI-rendelet alapvonalai és logikája

Lássuk, hogyan épültek be ezek az ügyek az MI-rendeletbe. Elöljáróban érdemes felidézni, hogy az MI-rendeletet az európai jogalkotók szerették volna a meglevő szabályozási környezetbe illeszteni, és ehhez az ún. termékmegfelelőségi szabályok logikáját találták irányadónak. Az MI-rendelet így szabályozási megoldásaiban, a megcélzott kötelezetti körben, a rájuk rótt kötelezettségek szerkezetében, a végrehajtásra hivatott szervezeti és eljárási környezetben, a szóhasználatban, a CE-jelzés használatában és még sok egyéb apróság alapján teljesen egyértelműen termékmegfelelőségi szabályozásnak tekinthető. A termékmegfelelőségi szabályok az EU számára az egységes piac megteremtése szempontjából mindig is kulcsfontosságúak voltak. Ennek a szabálycsoportnak az egyik központi jogszabálya, amely a termékek akkreditálásáról és a piacfelügyleti szabályokról szól, a preambulumában így írja le a terület célját:

> „Biztosítani kell, hogy a Közösségen belül az áruk szabad mozgásának előnyeit élvező termékek megfeleljenek az olyan közérdek magas szintű védelme követelményének, mint az egészség és a biztonság általában, a munkahelyi egészség és biztonság, a fogyasztóvédelem és a környezetvédelem, valamint a biztonság, annak biztosítása mellett, hogy a termékek szabad mozgása ne korlátó-

39 Hill 2020.

> zódjék nagyobb mértékben, mint amelyet a közösségi harmonizációs jogszabályok vagy más vonatkozó közösségi szabályok lehetővé tesznek.”[40]

A Bizottság vonatkozó tájékoztató oldalán egymás alatt sorjáznak azok a termékcsoportok, amelyeknek ún. megfelelőségértékelési eljáráson kell átesniük, mielőtt piacra helyezik őket. Ezzel a termékek kontrollja nem ér véget: a teljes életciklus alatt piacfelügyeleti szervezetek folyamatosan nyomon követik a jellemzően fizikai jellemzőkkel körbeírt megfelelőségüket. Az MI-rendelet igyekszik követni ezt a logikát. A MI-rendszereket is megfelelőségértékelési eljárásnak kell alávetni a piacra helyezés előtt, majd a „gyártóknak” a „termék” használatának teljes életciklusa során ún. forgalomba hozatal utáni nyomonkövetési rendszert kell működtetniük, amely azt jelenti, hogy a szolgáltatóknak

> „aktívan és szisztematikusan gyűjtenie, dokumentálnia és elemeznie kell az adott esetben az alkalmazók által szolgáltatott, vagy más forrásokból származó, a nagy kockázatú MI-rendszerek teljes élettartamuk során mutatott teljesítményére vonatkozó releváns adatokat, amelyek lehetővé teszik a szolgáltató számára annak értékelését, hogy az MI-rendszerek folyamatosan megfelelnek-e a III. fejezet 2. szakaszában meghatározott követelményeknek”.[41]

A rendelet az MI-rendszereket alapvetően három kockázati kategóriába sorolja, és ezenfelül a termékmegfelelőségi logiától némileg idegen módon meghatároz néhány tiltott „MI-gyakorlatot” is. (Ezek esetén tehát nem az MI-rendszereket tiltja be, hanem bizonyos MI-rendszerek bizonyos helyzetekben történő alkalmazását, ez is árulkodó, és még vissza is térek rá.) A tiltott gyakorlatok közt találjuk például MI-rendszerek társadalmi pontozásra történő használatát, a valós idejű köztéri arcfelismerést, a tudatalatti manipulációt megvalósító rendszerek használatát, valamint a bűnelkövetők visszaesési kockázatát előrejelző rend-

40 765/2008/EK rendelet.
41 MI-rendelet 72. cikk (2).

szereket. Mint láthattuk, ezek mind korábban megtörtént eseteken alapulnak.

A jogszabály nagy része az ún. magas kockázatú MI-rendszerek előállításának, használatának szabályait rögzíti, és természetesen a megfelelőségértékelési követelményeit és folyamatát, valamint a kapcsolódó intézményrendszert írja le. Árulkodó a magas kockázatú rendszerek listája is, amelybe egy sor olyan rendszer is belekerült, amelyet egyébként magánszervezetek alkalmaznak, például oktatási, munkaügyi vagy banki környezetekben. Az egyik kiemelt nagy kockázatú rendszertípus a banki hitelbírálati osztályozó (*credit scoring*) rendszer.

Ami a jogszabályt különlegessé teszi, hogy szemben az összes többi termékmegfelelőségi jogszabállyal, már a Bizottság által beterjesztett tervezetben is megjelennek az alapvető jogok, szinte mindig az élet, a testi épség, az egészség és a környezet megóvása után beszúrva. Milyen alapvető jogokról van szó? Szinte bármilyenről. A 48. preambulumbekezdés, amely az MI-rendszerek kockázatairól beszél, az egészség és a biztonság után azonnal megemlíti, hogy

> „a[z Alapjogi] Charta által védett alapvető jogokra az MI-rendszer által gyakorolt kedvezőtlen hatás mértéke különösen fontos egy adott MI-rendszer nagy kockázatúként való besorolásakor. E jogok közé tartozik az emberi méltósághoz való jog, a magán- és családi élet tiszteletben tartása, a személyes adatok védelme, a véleménynyilvánítás és a tájékozódás szabadsága, a gyülekezés és az egyesülés szabadsága, a megkülönböztetésmentesség, a fogyasztóvédelem, a munkavállalók jogai, a fogyatékossággal élő személyek jogai, a hatékony jogorvoslathoz és a tisztességes eljáráshoz való jog, a védelemhez való jog és az ártatlanság vélelme, valamint a megfelelő ügyintézéshez való jog."[42]

A jogalkotási folyamatban az alapvető jogokra történő utalások számban is megsokasodtak, sőt, a jogszabály kiegészült egy olyan jogintézménnyel, amelyet ebben a normában alkalmaznak először: az alapjogi hatásvizsgálattal. Ezt a közigazgatási szervezetek, valamint a bankok

42 MI-rendelet 48. preambulumbekezdés.

és a biztosítók kötelesek elvégezni, amennyiben MI-rendszereket alkalmaznak.

Az MI-rendelet egyik kétségkívül leginnovatívabb, egyben legkevésbé operacionalizálható eleme, hogy szeretné megvédeni a jogállamot és a demokratikus folyamatokat is a mesterséges intelligencia negatív hatásaitól. A rendelet 8. preambulumbekezdése így fogalmaz:

> „Ezért a mesterséges intelligencia belső piacon történő fejlesztésének, használatának és elterjedésének elősegítése érdekében szükség van a mesterséges intelligenciára vonatkozó harmonizált szabályokat meghatározó uniós jogi keretre, amely ugyanakkor garantálja az uniós jog által elismert és oltalmazott közérdek – például az egészség és a biztonság –, valamint az alapvető jogok – köztük a demokrácia, a jogállamiság és a környezetvédelem – magas szintű védelmét.”[43]

Amiket nem tartalmaz az MI-rendelet, egy termékmegfelelőségi szabálytól kissé szokatlanul, azok a kvantifikálható paraméterek. Míg az összes többi ilyen jogszabály hosszú, számszerű műszaki paramétereket tartalmazó mellékletekkel operál, addig az MI-rendeletben vannak ugyan mellékletek, de kvantifikált paraméterek szinte egyáltalán nem. Ennek nagyrészt az a magyarázata, hogy az MI egyszerűen nem fizikai termék, amelynek lennének fizikai tulajdonságai. Persze egy szoftvernek ettől még nagyon is lehetnek kvantifikálható paraméterei, de mivel az MI-rendelet „a” mesterséges intelligenciát akarja szabályozni általában, (ún. „horizontális szabályozás”),[44] ilyen paraméterek a rengeteg szabályozott terület miatt nem kerültek bele.

Miért félelemtörvény?

Az itt következő részben amellett érvelek, hogy az MI-rendelet a sunsteini értelemben vett „félelemtörvény” jegyeit mutatja. A fentebbi esetek, amelyek nagyban inspirálták a jogszabály tilalmait és a benne

43 MI-rendelet 8. preambulumbekezdés.

44 Pl. Gilbert 2024.

foglalt kötelezettségeket, messze nem fedik le a veszélyek teljes skáláját, és ezek a tilalmak és kötelezettségek nem fognak bennünket megvédeni, nemhogy a „gyilkos robotok", és az „öntudatra ébredt és az emberiséget fenyegető MI-k" (egyébként egyelőre teljesen irreális) veszélyeitől, de még a diszkrimináló vagy manipuláló MI-től sem. Érvelésem három, egymással összefüggő érvre épül. Az első, hogy a fentebbi esetekben testet öltő veszélyek nemhogy nem MI-specifikusak, hanem olykor még technológia sem kell az előidézésükhöz. A második ezzel összefüggő érvem, hogy a veszélyek valójában mélyen fekvő társadalmi és politikai okokra vezethetők vissza, amelyeket nem fogunk tudni megoldani az MI (túl)szabályozásával. A harmadik pedig, hogy a veszélyek egy része – az alapvető jogokat, a jogállamot és a demokráciát fenyegető veszélyek – olyan természetűek, amelyeket az MI-rendelet szabályozási módszerével, az ún. *ex ante* termékmegfelelőségi szabályozással nem lehet kezelni. Összességében a helyzet az, hogy még nincsen kellő tapasztalatunk az MI mint technológia veszélyeit illetően, amelyet a ChatGPT berobbanása és a rendelet jogalkotási folyamatában okozott hatalmas zavar bizonyított a leginkább.

Az esetekben megnyilvánuló veszélyek nem MI-specifikusak

Ha végigtekintünk az eseteken, azt láthatjuk, hogy vagy nem is MI-rendszer okozta a problémát, a bennünk kialakult helyzetet, vagy ha volt is MI a történetben, nem lényegi elemként. A kínai pontozási rendszer, amely még az 1980-as évekre nyúlik vissza, eredeti formájában értelemszerűen semmilyen MI-elemet nem tartalmazott, és az állampolgárok aktivitásainak regisztrálása, a negatív listán szereplő állampolgárok kizárása bizonyos szolgáltatásokból *semmilyen logikai kapcsolatban nincsen az MI-vel*. A nyilvántartás és az egész rendszer működtetése nem igényel MI-t, sőt még számítógépet sem. Ha a rendszer nagyon sok polgárt figyel meg, a mennyiségek miatt persze valamilyen szoftveres támogatás egy idő után szükségszerű, de az MI-elem nem. Az MI továbbá kétségtelenül *megkönnyíti* az ilyen rendszerek működtetését, hiszen az arcfelismerés csak fejlett gépi érzékelési eszközökkel lehetséges, de az arcfelismerés nem szükségszerű eleme egy ilyen rendszernek. A kínai rendszerrel az a probléma, ami a tiltott MI-gyakorlatok (5. cikk)

definíciójának kulcseleme, hogy a felírt pontszámokat „személyekkel vagy személyek csoportjaival” szemben „olyan szociális kontextusokban” használják, „amelyek nem függenek össze azokkal a kontextusokkal, amelyek között az adatokat eredetileg létrehozták vagy gyűjtötték”, illetve „olyan hátrányos vagy kedvezőtlen” bánásmódot alkalmaznak, „amely indokolatlan vagy aránytalan” az elkövetett magatartáshoz képest.[45] Ha a kínai pontozási rendszernek ez a két jellegzetesség valóban sajátja (szögezzük le, hogy elég homályos tévképzetek élnek a rendszerről Európában), akkor ez valóban nem épp demokratikus és jogállami megoldás, de a kérdés az, hogy az európai jogalkotó miért éppen az MI-rendeletben látta szükségesnek elítélni ezt a gyakorlatot, és deklarálni, hogy Európában soha nem lesz ilyen.

A problémát az jelenti, hogy ezzel egy súlyos inkonzisztencia került bele a rendeletbe. A termékmegfelelési jogszabályok logikája az, hogy a termékek jellemzőit írják le. A termékek használatával kapcsolatos korlátozásokat a termékleírások kell, hogy tartalmazzák, vagy magukat a termékeket kell olyan jellegzetességekkel, funkciókkal stb. ellátni, hogy ne jelentsenek veszélyt a felhasználókra. A tiltott gyakorlatokat belefoglalhatták volna egy *ex post* szabályba (például büntető törvénykönyvbe vagy a polgári jogi felelősségi szabályok közé), de a „mit nem szabad a termékkel csinálni” logika eddig nagyrészt ismeretlen volt a termékmegfelelőségi szabályokra. A járművekkel kapcsolatos jogellenes magatartások tiltását és szankcionálását (a záróvonal-átlépéstől a tömegkatasztrófa okozásáig) sem a járművek termékmegfelelőségi szabályai közt találjuk.

Hasonló gyökerű, de kicsit más karakterű a COMPAS és a Target-ügy lenyomata a jogszabályban. Ezekben az ügyekben nem a rendszerek jelentették a problémát, amelyek ráadásul nem gépi tanuláson alapulnak, hanem ún. statisztikai következtetőrendszerek,[46] inkább *magának a predikciónak a ténye*. Tágabb kontextusban a probléma ezzel az, hogy a jog és az üzlet is gyakorta kényszerül a jövőről szóló előrejelzésekre támaszkodni. Minden rossz érzésünk ellenére egyszerűen objektívebb(nek tűnik) a múltbeli adatok alapján matematikai

45 MI-rendelet 5. cikk (1) c).
46 Kamin 2024.

(statisztikai) módszerekre támaszkodni, mint egyéni intuíciók segítségével megpróbálni előre látni a jövőt.[47]

Az, hogy a veszélyek nem MI-specifikusak, a leglátványosabban az Európai Parlament beismeréssel felérő tanulmánya[48] mutatja, amely mások mellett azt tartja az MI felelősségi irányelv legnagyobb hiányosságának, hogy nem rendezi a csaknem hasonló komplexitással rendelkező *nem* MI-szoftverekkel okozott károk kérdését.

A tanulmány szerzője, Philip Hacker egy évvel korábban írt kritikus hangú cikkében árnyalja ezt a kijelentést, illetve máshogy keretezi. Kifejti, hogy a rendelet az MI definíciójából fakadóan valójában inkább eleve a „komplex szoftverek" által felvetett kockázatokról szól, és nem az MI-ről, hiszen a definícióba beleveszi a szabályalapú szakértői rendszereket és az olyan statisztikai módszereket alkalmazó szoftvereket is, mint például a COMPAS.[49] Nehéz szabadulni attól a gondolattól, hogy az MI ilyen kiterjesztő definíciója mögött épp a COMPAS-tól való irracionális rettegés és az erre alapuló, sokszor teljesen indokolatlan európai erkölcsi felsőbbrendűségi érzés áll.

Az esetek mögött meghúzódó társadalmi problémákat nem lehet technológiaszabályozással megoldani, mert azok jóval mélyebben gyökereznek

Ez az írás nem kíván mély társadalomelméleti fejtegetésekbe bocsátkozni, csak rögzíteni azt a tulajdonképpen triviális meglátást, hogy az ügyek nagy részében manifesztálódó kockázatok és negatív jelenségek, főként a magánéletbe történő túlzó mértékű behatolás, valamint a diszkrimináció, nem az algoritmusok és a mesterséges intelligencia miatt

47 Bár a Northpoint cég utóda, az Equivant honlapja azzal is mentegeti a céget, hogy a „klaszszifikáló" és a „valószínűségbecslő" szoftverek két teljesen külön kategória, és a COMPAS az utóbbi, és nem is szabad másként használni (mármint annak megállapítására, hogy „visszaeső"-e valaki), ez az érvelés azért nem állja meg a helyét, mert a valószínűség kiszámítása után mind az áruházlánc, mind a bíróságok kategorizálták a populáció tagjait, ami komoly következményekkel járt. Másképp: az elvileg 0 és 1 közé eső valószínűségnél határokat húztak, így logikailag klasszifikáló rendszerekké változtak, bármilyen tilalom vagy ajánlás ellenére. Ezt a problémát (ajánlások versus döntések) egyébként az EU Bírósága a Schufa-ügyben igyekezett rendezni, de ennek ismertetése szétfeszítené a dolgozat kereteit.

48 EU Parliament study 2024.

49 Hacker 2023.

keletkezett problémák, és ami fontosabb, nem is oldhatók meg a technológia szabályozásával.

A banki környezetben történő diszkrimináció például az USA-ban nagyjából az 1970-es évek óta a közbeszéd része (*Common Future*), és elég hamar kiderült, hogy a hitelbírálati pontozáshoz (*scoring*) használt szoftverek az adatbázisokban található korábbi ügyek történeti adatait átvéve ezt a diszkriminációt is átveszik.[50] A szoftverek által alkalmazott diszkrimináció ugyanakkor azért veszélyesebb, mint az emberek által alkalmazott, mert az objektivitás *látszatát* hordozza, nagyon hasonlóan a statisztikai predikciós szoftverekhez. A leggyakoribb eset az, amikor ugyan nem adat az etnikum vagy a bőrszín a szoftverben, de a lakcím, az iskolai végzettség és a vagyoni helyzet összességében gyakorta rámutathatnak egy etnikai csoportra, és bizonyos szoftverek ennél jóval több adatot tárolnak vagy gyűjtenek, amelyek még pontosabban jelölnek ki még kisebb védett csoportokat. További probléma a diszkrimináció esetén, hogy a bankok nagyon gyakran nem a hitelkérelem elutasításával diszkriminálják mondjuk a kisebbségeket, hanem az ún. „árdiszkriminációval", azaz egyedi árazással. Míg a diszkriminatív alapon elutasított hitelkérelem (ha bizonyítható egy védett kisebbségi kimondott vagy gyakrabban nem kimondott, de szisztematikus diszpreferálása) egyértelműen jogellenes, addig a kockázatokhoz igazodó egyedi díjszabás (például törlesztőrészlet vagy biztosítási díj) nem az.

Nagyon hasonlónak látszó problémát figyelhetünk meg a holland gyermektámogatási ügy mélyén is. A 2000-es évek közepén, amikor Bulgária csatlakozott az EU-hoz, nagy port vert az ún. bolgár migránsok csalása, amikor egy nagyobb bűnözőcsoport utaztatott bolgár állampolgárokat abból a célból, hogy azok Hollandiában egészségbiztosítási és lakhatási segélyeket vegyenek fel (pár ezer eurós nagyságrendben), majd hagyják el az országot szinte azonnal. Mivel a holland rendszer a bizalmon alapult, így a kérelmezők azonnal megkapták az összeget, az ellenőrzés pedig csak utólag történt meg, így a banda elég komoly összegekkel károsította meg a holland költségvetést. Az ügyből hatalmas botrány kerekedett, és komoly következményei lettek. Egyrészt a kelet-európai bevándorlókkal kapcsolatos indulatok magasra csaptak, másrészt a csalások detektálására (anomáliadetekció) szolgáló szoft-

50 Garcia et al. 2023.

verbe ezek az adatok bekerültek, így a botránynak adatszinten is nyoma maradt. A szoftverben található adatszintű torzítás (ha ez az) kihatott a gyermekgondozási segély ügyére is. Ugyanakkor azt sem szabad szem elől téveszteni, hogy a holland eset egy sor más problémától is terhes. Így nehezen magyarázható, hogy lehetett több mint ezer gyermeket kiemelni a családból annak ellenére, hogy a szülők tiltakoztak, hogy lehetett elkezdeni az összegek végrehajtását kizárólag egy szoftver döntésére alapozva, és így tovább. Ebben az esetben nemcsak, sőt *nem elsősorban* az MI döntése vezetett a súlyosan diszkrimináló eredményekre, hanem a hatóságok nemtörődömsége, túlkapásai és a sorozatos eljárási szabálysértések, amelyeket nyilván az olyan sajtótudósítások is szítottak, amelyekben a bolgárok dicsekszenek azzal, hogy a segélyből házat és autót vásároltak.[51]

A helyzettel kapcsolatban többféle érvelés lehetséges. Az egyik szerint, amelyet én is képviselek, a társadalmi problémákat nem lehet szoftverbeállításokkal vagy az adatgyűjtés és -feldolgozás manipulálásával megoldani. Logikailag az adatbázisok és/vagy a szoftverek diszkriminációmentesítése pozitív beavatkozást igényel, másképpen nagyon nehezen képzelhető el és kezelhető. Itt tehát tulajdonképpen szoftveresen megvalósuló pozitív diszkriminációról beszélünk. Csakhogy míg például a kvótákban megnyilvánuló pozitív diszkrimináció esetén tisztában vagyunk, hogy a „normál" szabályok alapján nem illetné meg egy adott szolgáltatás, egyetemi hely vagy tisztség egy védett kisebbség tagját, és ezt a torzítást tudatosan vállaljuk (most tekintsünk el attól, hogy ez policy szempontból helyes vagy sem), addig az adatokkal és az algoritmusokkal való manipuláció a valóságérzékelésünket torzítja el úgy, hogy nem tudjuk adott esetben végül, hogy mi lenne a „normál állapot", mitől tértünk el. Persze erre lehet azt válaszolni, hogy szülessen két kimenet: egy manipulált (de diszkrimináló) adatkészletekkel és egy másik a kisebbséget preferálókkal, ez nem sokat segít, ha az alapbeállítás a torzított kimenet kötelező elfogadása lesz, ráadásul különösen visszatetsző lehet egy profitorientált magánvállalkozás esetén.

A másik álláspont azt mondja, hogy a társadalmi problémákat igenis lehet súlyosbítani és elmélyíteni, ha az előítéleteket még szoftverkódokba is „beégetjük". Én azért érzem erősebbnek az első érvet, mert

51 Frederik 2020.

azok az emberi jogi jogsértések, amelyekről itt szó van, annyira esetfüggőek és egyediek, hogy valójában csak utólag lehet megállapítani azt, hogy megtörtént-e a jogsértés, és semmi esetre sem előre. Ez átvezet majd bennünket az utolsó, harmadik érvemre.

Végül itt kell beszélni arról, hogy a Cambridge Analytica-ügy úgy került be az MI-rendelet szövegébe, hogy egy sor ponton (szerencsére csak a nem normatív preambulumba) beszúrták a jogállamiság és a demokrácia védelmét is.

> „A mesterséges intelligencia számos hasznos felhasználási módja mellett, azt rendellenesen is fel lehet használni, valamint új és hatékony eszközöket biztosíthat manipulatív, kizsákmányoló és társadalmi ellenőrzési gyakorlatokhoz. Az ilyen gyakorlatok különösen károsak és visszaélésszerűek, és meg kell tiltani őket, mivel ellentétesek az emberi méltóság tiszteletben tartása, a szabadság, az egyenlőség, *a demokrácia és a jogállamiság* uniós értékeivel, továbbá a Chartában rögzített alapvető jogokkal, ideértve a megkülönböztetésmentességhez, az adatvédelemhez és a magánélet tiszteletben tartásához való jogot, valamint a gyermek jogait is.”[52]

A Cambridge Analytica-ügy 2016-ban hatalmas port vert, és kialakult az a szinte egyöntetű vélemény, hogy Trump megválasztását és a Brexitet a Facebookon folyt sötét manipulációk okozták. A kezdeti sommás vélemény egy idő után aztán árnyaltabbá vált. Egyrészt nagyon vitatott, hogy a cég mikrotargetáláson alapuló kampánya mennyire volt hatékony, még a hatalmas erőforrásokat mozgató, az angol adatvédelmi hatósági által lefolytatott nyomozás szerint is ezt „lehetetlen megmondani”,[53] másrészt az MI-rendelet hatálya a közösségi média ajánlószoftvereire nem terjed ki, így ettől az AI Act bizonyosan nem fogja megvédeni az európai polgárokat. Talán ha bizonyítható, hogy a kampány tudatalatti manipuláción alapul (ami valószínűtlen, mert a vonatkozó magyarázat szerint ez inkább vizuális és hanggal való trükközést jelent),[54] akkor lehet tiltott gyakorlatról szó, de ez esetben ez szóba sem

52 MI-rendelet (28) preambulumbekezdés. A szerző kiemelése (a szerk.).

53 ICO 2018, 19.

54 EU Commission Guidelines on prohibited practices, 20.

jöhet, illetve talán más jogszabályok[55] bizonyos rendelkezései korlátozhatják az ilyen jellegű kampányok mozgásterét. De semmi esetre sem az MI-rendelet.

Az emberi jogi kockázatokat nem lehet ex ante termékmegfelelőségi szabályozással kiküszöbölni

Harmadik érvem amellett, hogy az MI-rendelet félelemtörvény, az, hogy mivel a termék (MI-szoftver) előállításának folyamatában nem lehet az alapjogi kockázatokat mindenre kiterjedően látni, ezért kezelni sem. Ez az érv szoros összefüggésben van az előző kettővel, de alapvető különbség, hogy a termékmegfelelőségi szabályozás *ex ante* jellegét hangsúlyozza, és azt, hogy ez éles ellentétben áll az emberi jogok – és különösen azon emberi jogok, amelyek az MI kapcsán kockázatként felmerülnek – legbelső magjával, természetével. Itt elsősorban a diszkriminációmentességről és a magánélethez való jogról, az adatvédelemhez kapcsolódó információs önrendelkezési jogról, ritkábban a szólásszabadságról, esteleg a fair bírósági eljáráshoz fűződő jogról van szó.

A termékmegfelelőségi szabályozás logikája az, hogy előírjuk általánosabb formában és az ezeket az általános előírásokat specifikáló kvantitatív, számszerűsíthető paraméterekkel is, hogy egy adott termék mikor minősül biztonságosnak. Biztonságos egy játék, ha a használat közben a gyermek nem kap mérgezést, mert nem oldódnak ki a játékból káros anyagok. Mivel tudjuk, hogy a gyermekek a szájukba veszik a játékokat, a „káros anyagot" és a „kioldódási értéket" is meg kell határozni, tekintettel a gyermekek érzékenységére is, amelyet a vonatkozó rendelet részletes mellékletek formájában meg is tesz.[56] A mellékletekben azért lehet pontosan meghatározni ezeket az anyagokat és határértékeket, mert tisztában vagyunk vele, hogy egy gyermek számára mi az a mennyiség, amely még nem és már káros az egészségére. Persze itt van egy mozgástér, de a jogalkotó egy egyértelműen számszerűsített,

55 Pl. a DSA 26. cikk (3), illetve a Politikai reklámról szóló rendelet (Politikaireklám-rendelet 18. cikk (1) c).

56 Játékbiztonsági irányelv.

általában nagyon biztonságos értéknél leteszi a garast, meghúz egy határt.

Az emberi jogokkal az a probléma, hogy ilyen számszerűsíthető határokat nem tudunk meghúzni. A probléma gyökere kettős: egyrészt nem tudjuk egzakt módon meghatározni, hogy mikortól diszkriminatív például egy hitelbírálati gyakorlat. Az igaz, hogy vannak teljesen egyértelműen pozitív és negatív esetek, de a kettő közt van egy hatalmas szürke zóna. Épp ebben a szürke zónában tenyésznek a legnehezebb esetek, amelyek olykor az alkotmánybíróságokat is évekig tartó eljárásokba kényszerítik. Másrészt az MI-t „vezérlő” adat nem áll olyan kauzális viszonyban az emberi jogsértéssel, ahogy a veszélyes anyag az egészségkárosodással. Egy adatbázisban, amely múltbeli döntéseket tartalmaz, nem lehet egzakt módon meghúzni egy határt, hogy menynyi, a védett kisebbséget negatívan érintő döntést tartalmazhat, és mekkora mennyiségtől lesz diszkriminatív. A statisztikai alapú predikciós algoritmusok pedig nem is húznak határokat, hanem általában egy 0 és 1 közötti értéket mutatnak.

Konklúzió és epilógus

Miért jelent problémát, hogy az MI-rendelet félelemtörvény? Egyrészt a félelemtörvények felesleges bürokratikus koloncok, rövid távon a költségeket emelik, közép- és hosszú távon pedig Európa versenyképességét csökkentik. A Draghi-jelentés a jogi (túl)szabályozást Európa egyik legnagyobb versenyképesség-csökkentő tényezőjének tartja. Az MI-rendeletből tucatszámra lehetne hozni példákat a teljesen felesleges bürokratikus terhekre, de ezek közül csak egyet emelek ki, a korábban már említett alapjogi, a rendelet szövege szerint „alapvetőjogi” hatásvizsgálat intézményét,[57] amelyre „közjogi”, köszszolgáltatásokat nyújtó és bizonyos magánszervezetek (bankok, biztosítók) kötelezettek. Ez az előzetes „elemzés” arra szolgálna, hogy a közszférában használt MI-rendszerek alapjogokra gyakorolt hatását előre megbecsülje. Amellett, hogy a fentiekből talán elég egyértelmű lett, hogy egy ilyen hatást teljeskörűen biztosan nem, de még nagy vonalakban is csak nagyon

57 MI-rendelet 27. cikk.

durván lehet megbecsülni, testvérintézményéhez, az adatvédelmi hatásvizsgálathoz hasonlóan nagy valószínűséggel csak valamiféle formálisan letudandó, innen-onnan átvett sablonokkal véghezvitt „lepapírozás" lesz belőle. Foglalkozni viszont kell vele, költségek és fáradság, összességében a sokat kárhoztatott „adminisztratív terhek" járnak vele.

De van más probléma is. A 2022 novemberében megjelent a ChatGPT 3.5, az OpenAI cég nagy nyelvi modellre támaszkodó előtanított transzformer csevegőbotja nemcsak a közvéleményt lepte meg, hanem az EU jogalkotóit is. A meglepetés két irányból érkezett. Egyrészt, hogy ezek a modellek nem illettek a szabályozási logikába. A 3. pontban leírtam az MI-rendelet alaplogikáját, a termékmegfelelőségi logikát, és ott láthattunk, hogy ennek lényegi eleme, hogy a termék-előállítás teljes folyamatának az előírásoknak megfelelően kell történnie. Az MI-rendeletben a magas kockázatú MI-ket alapvetően nem a technológia jellege, hanem a felhasználási terület sorolja be kockázati kategóriába, azaz például egy anomáliadetekciós funkciójú MI akkor lesz magas kockázatú, ha az az áramszolgáltató rendszer biztonsági alrendszerének része, és például nem egy magáncégnél jelzi az ügyfelek magatartását. A nagy nyelvi modellekkel emiatt több hatalmas probléma is van. Az egyik triviálisabb az, hogy ezeket nem lehet létrehozni meghatározott szabályok mentén, mert már készen vannak. Ez talán a kisebbik. A nagyobbik probléma az, hogy ezek a rendszerek multifunkciósak: bárhol, ahol a nyelvet használjuk, az orvosi és műszaki dokumentációtól kezdve a jogi szférán keresztül egészen a művészi alkotásokig, mindenhol lehet őket használni. Még egyszerűbben: lehet, hogy teljesen ártatlan célra használjuk, lehet, hogy a jogszabály szerinti „közepes kockázatú" célra, például ügyfélszolgálati robotként, és lehet, hogy egy bírósági ítélet megírásához vagy egy hitelbírálati folyamatban.

Az európai jogalkotók az első problémát egy két elemből álló áthidaló megoldással igyekeztek kezelni: egyfelől elválasztották egymástól a modelleket, és a rájuk épülő ún. *downstream* szoftvereket, utóbbiakat ugyanazzal a kötelezettséghalmazzal megterhelve, mint a nem modellekre épülő MI-k gyártóit és alkalmazóit. Másfelől külön szabályokat alkottak a modellek fejlesztőire, mivel azonban itt az alkalmazási terület kockázataiból nem lehetett kiindulni, egy műszaki jellegű kockázati besorolást alkalmaztak (a modellépítéshez használt számítási teljesítményt használva paraméterként), és a modellek építőire csak

(kétlépcsős) transzparenciakövetelményeket telepítettek. Azaz nekik nem kell az MI-rendelet nagy részét kitevő követelményeket teljesíteniük, csak bizonyos dolgokat átláthatóvá kell tenniük, dokumentálniuk kell. A követelményeknek való megfelelés a downstream szolgáltatók feladata, akik esetén azonban nagyon kétséges, hogy hogyan fognak tudni megfelelni ezeknek a követelményeknek, ha a modellt, amelyre építenek, nem ők fejlesztették, és nem tudják pontosan, mi van benne.

De van egy jóval általánosabb probléma is, amelyet az LLM-ek feltűnése látványosan illusztrál, és ez már közvetlenül összefügg azzal, hogy az MI-rendelet egy félelemtörvény: az LLM-ek miatt a jogszabály még hatályba lépése előtt bizonyult alkalmatlannak a feladata teljesítésére. Az LLM-ekre épülő csevegőrobotok ugyanis teljesen új, addig fel sem merült kockázatokat hoztak felszínre. Így például a szerzői jogilag védett alkotások felhasználását a tanításra, ezzel a szerzői jogosultak jogainak tömeges megsértését vagy az ún. hallucinációs hajlamot, amely minden helyzetben és felhasználási területen jelentkezhet. Így például rágalmazhat embereket, vagy félretájékoztathat fontos ügyekben. Az LLM-ek rosszindulatú felhasználása adathalász e-mailek vagy más megtévesztő szövegek létrehozására, a kép- és videógenerálók felhasználása hamisított képek és videók előállítására csak néhány illusztráció azokra a teljesen új veszélyekre, amelyek legnagyobb részére az MI-rendelet egyszerűen nem volt felkészülve, hiszen nem is tudhatott róla.

Mi a megoldás? Vannak olyan radikális hangok, amelyek szerint csakis egy felelősséget szabályozó egyértelműen *ex post* jellegű szabályozás kellene.[58] Én ezzel a radikális állásponttal nem értek egyet. Viszont két dolgon érdemes lenne elgondolkodni. Az első az emberi jogi kockázatok kiemelése ebből az *ex ante* termékmegfelelőségi szabályozásból. Az emberi jogi kockázatok nagy része a felhasználás során jön elő, és voltaképpen mindegy, hogy egy szervezet mivel sérti meg az emberi jogokat. A jelenlegi szabályozás eszközsemleges: ha egy nagy szervezet megsérti az ügyfele magánélethez való jogát, jelenleg teljesen mindegy, hogy ezt egy papíron vezetett nyilvántartás, egy hagyományos eszközzel készült fénykép, egy MI-vel hamisított videó vagy egy döntés során használt statisztikai predikciós rendszerrel teszi. A jelen-

58 Kretschmer et al. 2023.

legi antidiszkriminációs és adatvédelmi szabályok, ha nem is tökéletes, de legalábbis betartható megoldást nyújtanak. Az MI-rendeletnek pedig csak az élet, a testi épség, az egészség és a környezet kvantifikálható paraméterekkel mérhető értékeit kellene szem előtt tartania. Konzisztensebb és betarthatóbb jogszabályhoz jutnánk.

Irodalom

765/2008/EK rendelet – az Európai Parlament és a Tanács 765/2008/EK rendelete (2008. július 9.) a termékek forgalmazása tekintetében az akkreditálás és piacfelügyelet előírásainak megállapításáról és a 339/93/EGK rendelet hatályon kívül helyezéséről (EGT-vonatkozású szöveg). (letöltve: 2025. 06. 26.)

Algorithmwatch.org AI Ethics Guidelines Global Inventory. *Algorithmwatch.org* https://inventory.algorithmwatch.org/ (letöltve: 2025. 06. 26.)

Angwin, Julia – Larson, Jeff – Mattu, Surya – Kirchner, Lauren 2016: Machine Bias. *Propublica,* május 23. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (letöltve: 2025. 06. 26.)

Asimov, Isaac 1943: Runaround. *Astounding Science Fiction,* március.

Borgesius, Frederik Zuiderveen 2018: Discrimination, Artificial Intelligence and Algorithmic Decision Making. *Council of Europe.* https://www.coe.int/en/web/inclusion-and-antidiscrimination/ai-and-discrimination (letöltve: 2025. 06. 26.)

Brownsword, Roger – Scotford, Eloise – Yeung, Karen (szerk.) 2017: *The Oxford Handbook of Law, Regulation and Technology.* Oxford University Press.

Cadwalladr, Carole – Graham-Harrison, Emma 2018: Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian,* március 17 https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (letöltve: 2025. 06. 26.)

Capek, Karel 1920: *Rossumovi Univerzální Roboti.* Angolul: Rossum's Universal Robots, Prága, Aventinum.

Common Future 2021: Why Credit Scores Are Racist *Medium,* július 26. https://medium.com/commonfuture/why-credit-scores-are-racist-da109fcfb300 (letöltve: 2025. 06. 26.)

Crawford, Kate – Schultz, Jason 2014: Big Data and Due Process, Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review,* 93–128.

Duhigg, Charles 2012: How Companies Learn Your Secrets. *The New York Times Magazine,* február 26. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp (letöltve: 2025. 06. 26.)

Draghi-jelentés: *The future of European competitiveness: Report by Mario Draghi.* https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en (letöltve: 2025. 06. 26.)

DSA: Az Európai Parlament és a Tanács (EU) 2022/2065 rendelete (2022. 10. 19.) a digitális szolgáltatások egységes piacáról és a 2000/31/EK irányelv módosításáról (digitális szolgáltatásokról szóló rendelet).

EU Commission Guidelines on prohibited practices Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act).

EU HLEG etikai iránymutatás 2019: Ethics guidelines for trustworthy AI, EU High Level Expert Group for AI (az EU Bizottság által létrehozott magas szintű független szakértői csoport megbízható mesterséges intelligenciára vonatkozó etikai iránymutatása). https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (letöltve: 2026. 06. 26.)

EU Parlamenti határozat 2017: Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról (2015/2103 [INL]) (EU Robot-felelősségi állásfoglalás). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_HU.html (letöltve: 2025. 06. 26.)

Facebook–Cambridge Analytica data scandal szócikk. *Wikipédia.* https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal (letöltve: 2025. 06. 26.)

Frederik, Jesse 2020: Tienduizenden gedupeerden, maar geen daders: zo ontstond de tragedie achter de toeslagenaffaire. *de Correspondent,* november 13. https://decorrespondent.nl/10628/tienduizenden-gedupeerden-maar-geen-daders-zo-ontstond-de-tragedie-achter-de-toeslagenaffaire/0f71ac81-d117-0d12-23a4-89513b0bb824 (letöltve: 2025. 06. 26.)

Garcia, Ana Cristina Bicharra – Garcia, Marcio Gomes Pinto – Rigobon, Roberto 2023: Algorithmic discrimination in the credit domain: what do we know about it? *AI & Soc.* https://doi.org/10.1007/s00146-023-01676-3

Gilbert, Stephen 2024: The EU passes the AI Act and its implications for digital medicine are unclear. *npj Digit. Med.,* 7, 135. https://doi.org/10.1038/s41746-024-01116-6

Goodrich, Joanna 2021: How IBM's Deep Blue Beat World Champion Chess Player Garry Kasparov. The supercomputer could explore up to 200 million possible chess positions per second with its AI program. *IEEE Spectrum,* január 25. https://spectrum.ieee.org/how-ibms-deep-blue-beat-world-champion-chess-player-garry-kasparov (letöltve: 2025. 06. 26.)

Grossman, Peter Z. – Cearley, Reed W. – Cole, Daniel H. 2006 Uncertainty, insurance and the Learned Hand formula. *Law, Probability and Risk,* 5. évfolyam, 1, 1–18. https://doi.org/10.1093/lpr/mgl012

Hacker, Philipp 2023: The European AI liability directives – Critique of a half-hearted approach and lessons for the future. *Computer Law & Security Review,* 51. évfolyam, 105871. https://doi.org/10.1016/j.clsr.2023.105871

Hacker, Philipp 2024: Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence Complementary impact assessment. *EPRS | European Parliamentary Research Service,* szeptember. https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf (letöltve: 2025. 06. 26.)

Haesik, Kim 2022: *Artificial Inelligence for 6G.* Cham, Springer Nature.

Heikkillä, Melissa 2022: Dutch scandal serves as a warning for Europe over risks of using algorithms. *Politico.eu,* március 29. https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/ (letöltve: 2025. 06. 26.)

Hill, Kashmir 2020: Wrongfully Accused by an Algorithm. *The New York Times,* augusztus 3. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (letöltve: 2025. 06. 26.)

Hirsch, Dennis D. 2014: That's Unfair! Or is it? Big Data, Discrimination and the FTC's Unfairness Authority. *Kentucky Law Journal,* 345–362.

Hopfield, John J. 1982: Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA,* Apr., 79. évfolyam, 8, 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

ICO 2018: Information Commissionar's Office, (UK) Investigation into the use of data analytics in political campaigns. https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf (letöltve: 2025. 06. 26.)

Játékbiztonsági irányelv Az Európai Parlament és a Tanács 2009/48/EK irányelve (2009. június 18.) a játékok biztonságáról (EGT-vonatkozású szöveg).

Johnson, Colin 2011: Johnson IBM's Watson computer beats humans at Jeopard. *EETimes,* április 2. https://www.eetimes.com/ibms-watson-computer-beats-humans-at-jeopardy/ (letöltve: 2025. 06. 26.)

Kamin, Chris 2024: Debunking Misconceptions About the COMPAS Core Instrument: What You Need to Know. *Equivant,* augusztus 12. https://equivant-supervision.com/debunking-misconceptions-about-the-compas-core-instrument-what-you-need-to-know/ (letöltve: 2025. 06. 26.)

Kretschmer, Martin – Kretschmer, Tobias – Peukert, Alexander – Peukert, Christian 2023: The risks of risk-based AI regulation. *Create blog,* november 22. https://www.create.ac.uk/blog/2023/11/22/ai-risks-of-risk/ (letöltve: 2025. 06. 26.)

Loomis-ügy: Loomis kontra Wisconsin, 881 N.W.2d 749 (Wis. 2016).

Marchant, Gary E. – Sylvester, Douglas J. – Abbott, Kenneth W. 2009: What Does the History of Technology Regulation Teach Us about Nano Oversight? *Journal of Law, Medicine & Ethics,* 37. évfolyam, 4, 724–731. https://doi.org/10.1111/j.1748-720X.2009.00443.x

Masayuki, Ida 2024: *A Narrative History of Artificial Intelligence The Perpetual Frontier of Information Technology.* Singapore, Sringer Nature.

Mayer-Schönberger, Viktor – Cuckier, Kenneth 2012: Big Data. Budapest, HVG Könyvek.

Mikolov, Tomas – Chen, Kai – Corrado, Greg – Dean, Jeffrey 2012: Efficient Estimation of Word Representations in Vector Space. *arXiv,* 1301.3781.

MI-rendelet: az Európai Parlament és a Tanács (EU) 2024/1689 rendelete (2024. június 13.) a mesterséges intelligenciára vonatkozó harmonizált szabályok megállapításáról, valamint a 300/2008/EK, a 167/2013/EU, a 168/2013/EU, az (EU) 2018/858, az (EU) 2018/1139 és az (EU) 2019/2144 rendelet, továbbá a 2014/90/EU, az (EU) 2016/797 és az (EU) 2020/1828 irányelv módosításáról (a mesterséges intelligenciáról szóló rendelet) (EGT-vonatkozású szöveg).

Munoz, Cecilia – Smith, Megan – Patil, Dj 2016: *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights, Executive Office of the President (Obama).* Május 12. [obamawhitehouse.archives.gov]

NLF weboldal 2025: Az Új jogszabályi keret (New Legislative Framework) tájékoztató weboldala. https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en (letöltve: 2025. 06. 26.)

Novak, Matt 2011: A Robot has Shot its Master. The 1930s hysteria about machines taking jobs and killing people. *Slate,* november 30. https://www.slate.com/articles/technology/future_tense/2011/11/robot_hysteria_in_the_1930s_slide_show_.html (letöltve: 2025. 06. 26.)

Ohm, Paul 2009: Broken Promises of Privacy: Responding to the Surprising Failure of Anonimization. *UCLA Law Review,* 57 (2009–2010), 1701–1777.

Osborne, Hillary – Parkinson, Hannah J. 2018: Hilary Osborne and Hannah Jane Parkinson Cambridge Analytica scandal: the biggest revelations so far. *The Guardian,* március 22. https://www.theguardian.com/uk-news/2018/mar/22/cambridge-analytica-scandal-the-biggest-revelations-so-far (letöltve: 2025. 06. 26.)

Pagallo, Ugo 2013: The Laws of Robots; Crimes, Contracts and Torts. Cham, Springer.

Pasquinelli, Matteo 2023: *The Eye of The Master, The Social History of Artificial Intelligence*. London – New York, Verso.

Politikai reklám rendelet: Az Európai Parlament és a Tanács (EU) 2024/900 rendelete (2024. március 13.) a politikai reklámtevékenység átláthatóságáról és célzott folytatásáról (EGT-vonatkozású szöveg).

Russel-Norvig, Stuart J. – Norvig, Peter 2000: *Mesterséges intelligencia modern megközelítésben.* Panem, Prentice Hall.

Shelley, Mary 2019: *Frankenstein - Avagy a modern Prométheusz.* Budapest, Móra.

Social Credit System szócikk: Wikipédia. https://en.wikipedia.org/wiki/Social_Credit_System (letöltve: 2025. 06. 26.)

Sunstein, Cass 2006: Laws of Fear; Beyond the Precauitionary Principle, Cambridge University Press, Cambridge, 2006.

Tene, Omer – Polonetzky, Jules 2013: Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. & Intell. Prop., 239, 243–251.

Trolley problem. Enciclopedia Britannica. https://www.britannica.com/topic/trolley-problem (letöltve: 2025. 06. 26.)

Vaswani, Ashish – Shazeer, Noam – Parmar, Niki – Uszkoreit, Jakob – Jones, Llion – Gomez, Aidan N. – Kaiser, Łukasz – Polosukhin, Illia 2017: "Attention is All you Need". Advances in Neural Information Processing Systems. 30. *Curran Associates, Inc.,* arXiv:1706.03762.

Wallach, Wendell – Allen, Colin 2009: *Moral Machines, Teaching Robots Right from Wrong.* Oxford, Oxford University Press.

Wiener, Jonathan B. 2004: The regulation of technology, and the technology of regulation *Technology in Society,* 26, 483–500.

Ződi, Zsolt 2018: *Platformok, robotok és a jog. Új szabályozási kihívások az információs társadalomban.* Budapest, Gondolat.

Ződi, Zsolt 2019: Law and legal science in the age of big data, *Intersections. East European Journal of Society and Politics,* 3. évfolyam, 2, 69–87.