Okoseszközök – Okos jog?

A mesterséges intelligencia szabályozási kérdései

G. KARÁCSONY GERGELY

LUDOVIKA
EGYETEMI KIADÓ

Okoseszközök – Okos jog?
A mesterséges intelligencia szabályozási kérdései

Vákát oldal

G. Karácsony Gergely

OKOSESZKÖZÖK – OKOS JOG?

A mesterséges intelligencia szabályozási kérdései

Ludovika Egyetemi Kiadó ❖ Budapest, 2020

A mű a KÖFOP-2.1.2.-VEKOP-15-2016-00001 „A jó kormányzást megalapozó közszolgálat-fejlesztés” című projekt keretében elnyert, „Államszervezési és kormányzási kihívások Magyarországon az Alaptörvény hatálybalépését követően” című, 2017/96/PTE-ÁJK azonosítójú NKE Államtudományi Kutatóműhely program keretében valósult meg.

Szakmai lektor
Keserű Barna Arnold

A kézirat lezárásának időpontja: 2018. október 10.

Tartalom

Köszönetnyilvánítás

Köszönettel tartozom a feleségemnek, akinek a türelme és a támogatása nélkül ez a könyv nem születhetett volna meg. Köszönöm a családom minden tagjának, hogy mellettem álltak.

Köszönöm oktató és kutató kollégáimnak a Széchenyi István Egyetem jogi karán a szakmai segítséget és az inspiráló szellemi közeget, amely mindig motivált a kutatásban. Köszönöm a könyvem megjelenését és az ahhoz vezető kutatási munkát támogató kutatási projekt vezetésében és menedzselésében közreműködő minden kolléga lelkiismeretes és professzionális munkáját.

Győr, 2018 októbere

G. Karácsony Gergely

Vákát oldal

Ajánlom ezt a könyvet Simon fiamnak, aki a mesterséges intelligencia izgalmas és kihívásokkal teli új korszakában nő fel.

Vákát oldal

I. rész

A mesterséges intelligencia működése és annak jogi kérdései

Vákát oldal

Bevezető gondolatok

A mesterséges intelligenciáról (MI) szóló diskurzus egyáltalán nem új keletű. A robotok és különösen a mesterséges intelligenciát alkalmazó eszközök fejlődésének útja hullámzó tendenciát mutat: az 1950-es évektől kezdve hol felerősödtek, hol elhalkultak a fejlesztéssel és ezzel párhuzamosan a szabályozással és az erkölcsi kérdésekkel foglalkozó hangok. A kutatások sokszor azért torpantak meg vagy futottak zsákutcába, mert a kor műszaki fejlettsége még nem tette lehetővé a kutatók és fejlesztők által kitűzött célok elérését. Ez sokszor kiábrándultságot, csalódást okozott a nagy várakozásokat támasztó kutatóknak, és gyakran évekre, évtizedekre megtorpanásra késztette az ilyen irányú vizsgálódásokat.

Napjainkban a mesterséges intelligenciához kapcsolódó kutatások ismét lendületet kaptak. Ehhez jelentősen hozzájárult az elérhető hardverek fejlettsége és az általuk nyújtott számítási kapacitás nagysága, amely olyan mértékű lehet, ahol már a mesterséges intelligencia működni képes. Ezzel együtt a szoftveres oldalon elért fejlettség is bizonyította, hogy a specifikus feladatokra alkalmazható mesterséges intelligencia működőképes.

Az elmúlt néhány év technológiai hírei között egyre gyakrabban találkozunk sikeres mesterséges intelligencia alapú megoldásokról szóló beszámolókkal. A kutatóintézetek és egyetemek falait elhagyva a kereskedelmi fejlesztések is nagy lökést adtak az MI-iparnak. A mesterséges intelligencián alapuló algoritmusokat napi rendszerességgel alkalmazzuk, legyen az az okostelefonunk digitális asszisztense, a közösségi oldalak tartalomfolyamát személyre szabó gépi agy vagy az online piactereken egyedi ajánlatokat és további vásárlási lehetőségeket kínáló adatelemző szolgáltatás.

Hazánkban még nem elterjedt, azonban az angol nyelvterületen már nem számít lehetetlennek a félig vagy teljesen robotizált online ügyfélszolgálat sem. Annak ellenére, hogy egyelőre hivatalosan még nem sikerült egyetlen gépi agynak sem átmenni a Turing-teszten,[1] a laikus felhasználó sokszor

[1] Alan Turing 1950-es cikkében teszi fel a kérdést, hogy „tudnak-e a gépek gondolkodni". A gondolkodó gépek vizsgálatára olyan tesztet javasol, ahol egy emberi alany két másik partnerrel vált – kizárólagosan írásban – üzeneteket. A két partner közül az egyik ember, a másik gép. A teszt akkor tekinthető sikeresnek, ha a párbeszéd után a kérdező nem tudja megmondani, hogy melyikük melyik (Turing 1950, 433–460.).

már nem tudja teljes bizonyossággal megállapítani, hogy chatrobottal kommunikált-e vagy sem.

Ebben a társadalmi és technológiai közegben egyre gyakrabban merülnek fel a szabályozás kérdései. A mesterséges intelligencia fejlődése elért arra a szintre, hogy sajátos jogi problémák jelentkeztek a működésével, a fejlesztésével és a mindennapi életünkre gyakorolt hatásával kapcsolatban. Az okosotthonunk személyi asszisztense személyes adatokat gyűjt és továbbít rólunk, az automatizált hitelképesség-kiszámító eszköz a kisebbségekhez tartozókat tanult diszkriminációval magától lejjebb pontozza, és túl vagyunk már az önvezető autók által okozott első közlekedési baleseteken is.

Sokan fordulnak hát a bíróságokhoz és hatóságokhoz hatékony jogvédelmet keresve az őket ért sérelemért. A jogalkalmazók pedig zavartan néznek körül, mert erre nincsenek felkészülve. Az általuk ismert jogi dogmatika nem ad egyértelmű iránymutatást a nem emberi aktorok tetteinek megítéléséről, a jogrendszer jelenlegi normatív kereteit pedig az intelligens eszközök kezdik szétfeszíteni.

Az új problématípusok új szóhasználatot és fogalomkészletet igényelnek, vagy pedig a jogrendszer most is létező kategóriáiba kell őket erőszakkal betuszkolni. A nehéz helyzetbe került jogalkalmazók a jogalkotóra néznek ebben a helyzetben, és tőle kérnek iránymutatást. A jogalkotó pedig nem kevésbé van kutyaszorítóban: érzékeli a szabályozási igényt, és maga is látja a technológia rohamos fejlődését, amelyhez hasonlót a – fennálló jogi és társadalmi kereteket igencsak átalakító – ipari forradalom sem produkált. Ugyanakkor nincs kész válasza a jogalkotástan alapkérdéseire, hogy mit, hogyan, mikor és kinek kellene szabályoznia. Olyan területre értünk ugyanis, ahol a jogalkotó olyan, forradalmian új technológiáról kell, hogy nyilatkozzon, amelynek a működési elvét, műszaki jellegzetességeit, belső logikáját és a külvilágra gyakorolt lehetséges hatásait csak a témában nagyon képzett szakemberek értik, sok következményt pedig még ők sem látnak előre. Közben pedig ott ül az egyik vállán a kisördög, amelyik azt súgja a fülébe, hogy ezek a technológiák veszélyt jelentenek az egyes emberre és az egész emberiségre, és korlátozni kell őket; míg a másik vállán a kisangyal azt suttogja, hogy a mesterséges intelligencia soha nem látott távlatokat nyit meg a tudományban, egyúttal könnyebbé és biztonságosabbá teszi az emberek mindennapjait, ezért támogatni és ösztönözni kell a fejlesztéseket. Ezek a hangok a napi sajtóban, a gazdasági teljesítményről szóló hírekben és a *science fiction* irodalomban egyaránt megjelennek, és befészkelik magukat a gondolatainkba.

Ezzel a munkával a jogalkotó, a jogalkalmazó és a jogtudomány művelői számára szeretnék segítséget nyújtani a mesterséges intelligenciáról való gondolkodásban. A könyv első részében felvillantom a mesterséges intelligencia fejlődésének útját, fogalmát és legfontosabb technikai részleteit. Ezt kifejezetten a számítástechnika és a matematika területén járatlan olvasó számára is érthetően próbálom megfogalmazni. Mindazonáltal lényegesnek tartom, hogy kitérjek ezekre a kérdésekre is, mivel e nélkül nem érthetjük meg igazán, hogy a mesterséges intelligencia mit miért csinál, és milyen jogi hatásokkal járhat a működése.

Ezután néhány csomópont köré csoportosítva bemutatom azokat a közjogi természetű problémákat, amelyeket a mesterséges intelligencia gyakorlati alkalmazása okozhat. Ez a felsorolás korántsem kimerítő, a célja az, hogy az olvasó összekapcsolja a mindennapi életben látható mesterséges intelligenciát a közjog klasszikus kérdéseivel.

A könyv második részében kifejezetten azokat a kérdéseket veszem számba, amelyek a mesterséges intelligencia szabályozása terén merülnek fel, és amelyek a jogalkotó részéről alapos végiggondolást igényelnek. Óva intem a jogalkotót és a jogalkalmazókat attól, hogy csak azért, mert azt látják, a technológia gyorsvonata éppen az orruk előtt száguld el, megpróbálják hirtelen kinyújtani a kezüket, és felkapaszkodni rá, mert így súlyos balesetet szenvedhetnek.

Meggyőződésem, hogy a jogalkotó akkor jár el helyesen, ha minél több információt gyűjt a szabályozandó technológia működéséről, társadalomra gyakorolt hatásairól és a felmerülő jogi problémákról. Ezután alaposan mérlegeli, hogy kinek és hol kell beavatkoznia ebbe a helyzetbe, kell-e új jogdogmatikai megközelítés, vagy pedig a jogrendszer meglévő struktúrája megfelelő keretet nyújt az új technológiák kezeléséhez. Ehhez a jogalkotó, a jogalkalmazó, a technológia fejlesztői és a jogtudomány képviselői közötti párbeszéd és együtt gondolkodás szükséges. Munkámmal ehhez szeretnék hozzájárulni, bízva a jogalkotó bölcsességében, és várakozással tekintve a jövőbe.

Vákát oldal

1. A mesterséges intelligenciáról

Napjainkra a technológiai fejlődés elért abba a szakaszba, ahol már nem számít fikciónak a természetes emberi nyelven feltett kérdéseket megértő és azokra adekvát választ adó, mesterséges intelligencia alapú alkalmazás. Az elérhető hardveres eszközök és a piaci alapon fejlesztett mesterséges intelligencia képességei korábban elképzelhetetlen számítási kapacitást és minden eddiginél hatékonyabb működést tettek lehetővé, elsődlegesen a specializált felhasználási célú alkalmazások terén.

A mesterséges intelligencián alapuló megoldások széles körű elterjedése a filozófiai és etikai kérdések boncolgatásán túlmenően számos jogi kérdést is felvetettek. A jogalkotónak hamarosan választ kell tudnia adni arra a kérdésre, hogy a jelenlegi jogi környezet és a hatályos jogszabályok elégségesek-e a mesterséges intelligencia működésével kapcsolatos kérdések mindenki számára megfelelő kezelésére, vagy pedig új jogalkotásra van szükség, és létre kell hozni a mesterséges intelligencia működésének társadalmi kereteit megteremtő jogszabályi környezetet.

1.1. A mesterséges intelligencia fogalma

1.1.1. Fogalomalkotási nehézségek

Egy tudományos mű elején általában definiáljuk a legfontosabb fogalmakat, különösen a kutatás tárgyát illetően. A mesterséges intelligenciáról szóló munka esetében azonban nehéz helyzetben vagyunk, ugyanis a mesterséges intelligenciának nincs általánosan elfogadott tudományos fogalma. Ez természetesen nem a tudományos közösség hanyagsága miatt van így, hanem a kutatás tárgya olyan nehezen megragadható, hogy számos, egymással konkuráló definíció létezik rá párhuzamosan. Különösen megnehezíti a fogalomalkotást, hogy műszaki vagy informatikai kutatómunka során létrejött entitásra kell egy olyan, a humán tudományok területére tartozó fogalmat adaptálni, mint az intelligencia. A *humán* intelligenciának is számos fogalmát ismerjük, és általánosan tartja magát az a nézet, hogy az intelligencia

elsődlegesen emberi tulajdonság, amelyre ezáltal oly büszkék vagyunk, hogy önmagunkat is a *homo sapiens,* vagyis értelemmel bíró ember névvel illetjük.

A mesterséges intelligencia fogalmának megalkotásakor elsőként azzal a kérdéssel szembesül a kutató, hogy az emberi intellektus melyik szeletét ragadja ki és használja fel a definícióhoz. Azt mindenesetre szem előtt kell tartanunk, hogy a mesterséges intelligencia fogalmát egy viszonylag specifikus területre kell fókuszálnunk, és nem nevezhetünk bármilyen olyan mesterséges entitást (számítógépes programot vagy robotot) így, amely bizonyos feladatokat az embernél hatékonyabban vagy gyorsabban tud végrehajtani. A fogalomalkotáshoz ezért meg kell találnunk azt a döntő elemet, amely elválasztja a hatékony munkavégzést az intelligenciától.

1.1.2. Definíciós kísérlet

A mesterséges intelligenciának nevezett jelenség kifejlődésére számos tudományterület jeles képviselői hatottak megtermékenyítőleg az ókortól napjainkig. Ezen a széles bázison állva megpróbálkozhatunk a fogalmi körülhatárolással – legalábbis a jelen kutatás számára. Előzetesen le kell szögezni, hogy mesterséges intelligenciáról olyan, mesterségesen létrehozott entitások esetében beszélhetünk, amelyek önálló feladatmegoldásra képesek, vagyis ahol nincs szükség emberi beavatkozásra a tevékenységük során.

A témával foglalkozó szerzők nézeteit két dimenzió mentén csoportosította a mesterséges intelligenciáról szóló alapműnek számító könyvében Russell és Norvig:[2] egyfelől vizsgálhatjuk a mesterséges entitás gondolkodási folyamatait, illetve logikáját, vagy pedig annak viselkedését, cselekvéseit. A másik tengely mentén pedig elkülöníthetjük azokat a nézeteket, amelyek a gépi agy működését az emberi viselkedéshez mérik, illetve azokat, amelyek a teljesítményt a racionális (az adott információk alapján meghozott legjobb döntés) viselkedésbázisán vizsgálják. Így négy fő csoportot alakítottak ki a definíciókból, ezek szerint a mesterséges intelligencia az, amely képes: emberként gondolkodni, emberként cselekedni, racionálisan gondolkodni, illetve racionálisan cselekedni. Az ember módjára való gondolkodás mint ismérv a mesterséges agy belső folyamatait helyezi előtérbe, és azt veti össze az emberi gondolkodással, amelynek megismerése önmagában is azt igényli, hogy az emberi elme működésébe belelássunk. A kognitív tudomány interdiszciplináris területén folyó

[2] RUSSELL–NORVIG 2016, 2.

vizsgálódások foglalkoznak ezzel a kérdéssel. Ez a diszciplína kölcsönösen megtermékenyítőleg hatott egymásra a mesterséges intelligencia kutatásával, azonban teljes párhuzamot nem vonhatunk. Munkánk szempontjából sem tartjuk ezt a fajta definíciót célravezetőnek, mivel az általunk vizsgálandó mesterséges intelligencia teljességét nem feltétlenül tudja lefedni.

Az emberként cselekvő mesterséges intelligencia esetében az úgynevezett Turing-tesztnek való megfelelést helyezzük előtérbe. Alan Turing tesztje[3] közel hetven év után is releváns maradt, az annak való megfelelés valóban fejlett mesterséges intelligenciát igényel, és a mesterségesintelligencia-kutatások szinte valamennyi területét érinti (a szövegértéstől a gépi tanulásig). Mindazonáltal a mesterséges intelligenciával foglalkozó fejlesztők és kutatók nem helyeztek nagy hangsúlyt arra, hogy rendszereik megfeleljenek a teszt követelményeinek, sokkal inkább az alapul szolgáló működést próbálták tökéletesíteni.

A mesterséges intelligencia fejlesztésének modern irányzatai nem abba az irányba haladnak, hogy az emberi gondolkodást vagy viselkedést kíséreljék meg reprodukálni, sokkal inkább specifikus feladatok ellátására fejlesztenek eszközöket. A racionális gondolkodásra való képesség inkább a formális logika szabályainak való megfelelést igazolja. Azonban a mesterséges intelligencia tipikus működési területein a legritkábban találkozunk olyan feladatokkal, amelyeket le lehet írni a formális logikának megfelelő egyszerű kijelentésekkel, és amelyek megoldhatók a logika szabályai szerint. A modern fejlesztési irányoknak leginkább megfelelő definíció a racionálisan cselekvő mesterséges intelligenciáé. A mesterséges intelligenciát alkalmazó eszközök közös jellemzője általában, hogy valamilyen cél elérése érdekében jöttek létre, és ezért alkalmazzák az intelligens megoldásokat. A racionálisan cselekvő gép (a cselekvés latin megfelelőjéből nevezhetjük ágensnek is) a rendelkezésre álló információk alapján optimálisnak mondható megoldást választ, és így éri el a lehető legjobb eredményt. A racionális cselekvéshez a gépnek szinte minden olyan tulajdonságra szüksége van, ami a Turing-teszt teljesítéséhez kellhet, az emberi nyelv értelmezésétől az információk önálló feldolgozásán át a tanulásra való képességig. Ne felejtsük azonban el, hogy a gép racionális cselekvése két irányból is korlátozott lehet a gyakorlatban: egyfelől nem mindig áll rendelkezésre elegendő információ[4] a legjobb

[3] TURING 1950, 433–460. A teszt ismertetését lásd fent (4. számú lábjegyzet).

[4] Az információhiányos helyzetekben való döntés mechanizmusa önmagában is számos jogi kérdést vethet fel, amelyekkel a későbbiekben foglalkozunk.

megoldás kiválasztásához, másfelől pedig a túl sok faktor figyelembevétele is túlkomplikálhatja és ezáltal nemkívánatos módon lelassíthatja a programot.

Mindezeket szem előtt tartva mégis a racionálisan cselekvő ágens tűnik a modern mesterséges intelligenciát leíró legpontosabb definíciónak. Ezt a meghatározást amiatt éri a legtöbb kritika, hogy egy nehezen megfogható fogalmat (intelligencia) egy másik, hasonlóan nehezen körülhatárolható és sok esetben a szándék tételezésével metafizikai kérdéseket is felvető (lehet-e egy gépnek szándéka bármire is) fogalommal, a cél fogalmával kívánja magyarázni.[5] Mindazonáltal a jogi problémák feltárása szempontjából talán mégis ez a meghatározás hordozza az egyik legfontosabb információt, a cselekvés optimalizálására és a lehető legjobb eredmény elérésére való törekvést. Sőt, a cselekvést magát is a definíció részének tekinthetjük, a jogi problémák körülhatárolása szempontjából a mesterséges intelligenciát alkalmazó ágensek kerülnek a vizsgálódás fókuszpontjába, azt feltételezve, hogy a nem cselekvő (a külvilág számára érzékelhető változást elő nem idéző) mesterséges intelligencia nem vet fel érdemi szabályozási kérdéseket. Cselekvés alatt most az aktusok legszélesebb skáláját értjük, az egyszerű írásbeli kommunikációtól a tárgyak manipulálásáig vagy a gépjárművek vezetéséig. Összességében tehát – ha definíciót nem is alkotunk – kitűzhetünk néhány sarokpontot a mesterséges intelligencia jelen könyvben való szerepeltetéséhez szükséges meghatározására. A következőkben tehát mesterséges intelligencián olyan gépi ágenseket értünk, amelyek egy – előre meghatározott vagy működés közben felmerült – cél elérése érdekében

- képesek cselekvő tevékenységet kifejteni, amelyhez emberi beavatkozás nem szükséges;
- tevékenységük során a rendelkezésre álló információk alapján az elérhető (vagy várható) legjobb eredmény elérése érdekében racionálisan cselekszenek;
- képesek a környezetükből valamilyen módon információkat nyerni és azt feldolgozni;
- képesek kommunikálni a környezetükkel;
- képesek saját működésüket önállóan szabályozni és irányítani, működésüket szükség esetén emberi beavatkozás nélkül megváltoztatni;
- ezek a mesterséges intelligenciát alkalmazó ágensek egyaránt lehetnek kizárólag szoftveres (térbeli megjelenéssel nem bíró) programok vagy pedig testtel is rendelkező gépek.

[5] SCHERER 2015, 361.

Amint láthatjuk, a fenti meghatározás rendkívül tág kereteket szab a vizsgálatunk tárgyának. Ez a megoldás azt a megközelítést tükrözi, amely az egész művet jellemzi majd: nemcsak a jelenben létező technológiákra és problémákra koncentrálunk, hanem a lehető legnagyobb vizsgálati kört választjuk ki, és megpróbálunk a jövőben is releváns kérdéseket és válaszokat megfogalmazni.

1.2. A mesterséges intelligencia kutatásának története

Ahogy az emberi intelligencia rendkívül összetett jelenség, úgy a mesterséges intelligencia is felettébb sokrétű, így számos más tudományágban megtaláljuk a gyökereit és egyes előképeit. Ezek rövid felvillantásával az a célunk, hogy megmutassuk az olvasónak, a mesterséges intelligencia koncepciója egyáltalán nem új keletű, és – bár az alkalmas eszközt az adta az emberiség kezébe – nem a számítástechnika megjelenéséhez köthető jelenség. A tudománytörténeti áttekintésből látjuk majd azt is, hogy a mesterséges intelligencia kutatása, módszertanának fejlesztése nem korlátozódik a reáltudományok terrénumára, hanem a bölcsészettudománytól a biológiai, pszichológiai, közgazdaságtani és természetesen informatikai és matematikai területek eredményeiből szőtték azt a rendkívül színes szövetet, amelyet ma mesterséges intelligenciának nevezhetünk. A mesterséges intelligencia fejlesztésében ma egyaránt helyet kapnak nyelvészek, statisztikusok és orvostudományi kutatók is. Nem mondhatjuk tehát azt, hogy a jogtudomány ne vonhatná vizsgálódási körébe ezt a területet, hiszen a mesterséges intelligenciát alkalmazó megoldások fejlesztése mellett annak társadalomra gyakorolt hatásáról, különösen pedig a jogalkotóval szemben támasztott szabályozási igényéről nem lehet megfeledkezni. Az alábbi rövid áttekintés álljon itt azért, hogy az olvasó lássa azt a fejlődési ívet, amely elvezetett a mesterséges intelligencia problémájának tudományos vizsgálatához, kutatásához, megalkotásához és a napjainkban látható fejlettséghez. A tudománytörténeti bevezetőben igyekszünk megmagyarázni néhány olyan társtudományi alapfogalmat is, amelyek a mesterséges intelligencia működésének későbbi bemutatása során felmerülnek, és amelyek ismerete a felvázolt problémák megértéséhez nélkülözhetetlen.

1.2.1. A mesterséges intelligencia kutatásának kezdetei

A mesterséges intelligencia fejlődésének kezdeteit az 1940-es évek vége és az 1950-es évek eleje jelentette, ez az időszak a mesterséges intelligenciához kapcsolódó alapok lerakásáról szólt. Az egyik kiinduló impulzust az emberi idegrendszer neuronjainak tanulmányozása adta, ennek nyomán jelent meg a neurális hálózatok működésének matematikai modellje és a neuronok közötti kapcsolat tanulás általi erősödését leíró Hebb-féle tanuláselmélet[6] is. A mesterséges intelligenciáról szóló első, teljesnek mondható koncepciót 1950-ben Alan Turing fogalmazta meg.[7] Itt vezette be a Turing-teszt, a gépi tanulás, a genetikus algoritmusok és a megerősítéses tanulás fogalmat is, amelyek a mai napig meghatározzák a területet.

Az évtized kulcsfontosságú eseménye a mesterséges intelligencia kezdeti időszakának egyik legmeghatározóbb személyisége, John McCarthy által 1956-ban a Dartmouth College-ben összehívott munkatalálkozó volt, ahol a témában a következő évtizedekben jelentős eredményeket elérő szakemberek megismerhették egymást. A találkozón a résztvevők elfogadták a McCarthy által javasolt megnevezést, és az új területet mesterséges intelligenciának *(Artificial Intelligence, AI)* nevezték el. Ettől az időszaktól datálódik a mesterséges intelligencia kutatásának önállósodása is, amelynek során kikristályosodott, hogy nem tekinthető önmagában a matematika egyik ágának, vagy – hasonlósága ellenére – miért nem maradt az irányításelmélet, az operációkutatás vagy a döntéselmélet keretein belül. Ezekre a válasz a mesterséges intelligencia kutatásának célkitűzései és módszerei között keresendő. Az előbb említett területek közül tisztán csak az MI tekinthető a számítógépes tudományok egy ágának. Emellett a mesterséges intelligencia kutatása során bonyolult, változó környezetben autonóm módon működő gépek építése a cél, illetve az olyan emberi képességek megjelenítése vagy másolása, mint a kreativitás, az önfejlesztés és a nyelv használata. Ezekkel a kérdésekkel semmilyen más terület nem foglalkozott.[8]

Az évtized végére McCarthy kidolgozta a mesterséges intelligencia programozására használt legfontosabb programnyelvet, a *Lispet,* emellett leírt[9] egy olyan hipotetikus programot, amelyet az első teljes MI-rendszernek tekinthetünk.

---

[6] Donald O. Hebb 1949-ben írta le elméletét, amely szerint a kölcsönható idegsejtek közötti kapcsolat erősödik a tanulás hatására.

[7] TURING 1950.

[8] RUSSELL–NORVIG 2016, 18.

[9] MCCARTHY 1959.

Az *Advice Takernek* nevezett program is tudást használt fel egy probléma megoldásának megtalálásához, ehhez pedig a világra vonatkozó általános tudással kellett rendelkeznie. A program a tudásreprezentáció és a következtetés leglényegesebb elveit testesítette meg, miszerint hasznos, ha rendelkezünk a világot és az ágens cselekvéseinek eredményét leíró explicit és formális reprezentációval, és képesek vagyunk ezt a reprezentációt deduktív módon manipulálni.[10]

A korai évtizedek nem várt sikereket hoztak, a kezdetleges számítógépekkel is olyan eredményeket tudtak elérni, ami felülmúlta a várakozásokat, így sorra pipálták ki az olyan feladatokat, amelyre a „gép biztosan nem lesz képes". A területtel foglalkozó szakemberek a sikereken felbuzdulva egyre ambiciózusabb ígéreteket tettek, amelyek teljesítése akkor még meghaladta az elérhető eszközök teljesítményét. Ezért aztán sokszor vezetett csalódáshoz egy-egy nagy ívű vállalkozás kudarca, ami pedig kiváltotta a – következőkben bemutatott – hanyatlást a kutatások terén.

1.2.2. Lelkesedés és hanyatlás

A technológiai haladás nagyobb lépcsőfokai időről időre a műszaki és társadalomtudományi diskurzus homlokterébe tolják a mesterséges intelligencia kérdéskörét, azonban egy-egy felívelő szakaszt leggyakrabban csalódás, kiábrándultság, ennek következményeként a kutatási források csökkentése és az ezzel járó hosszabb-rövidebb hanyatlás, valamint a tudományos diskurzus elhalkulása követ.

A leghosszabb hullámvölgyek az 1980-as évek környékén voltak tapasztalhatók, ezek a mesterséges intelligencia telének *(AI winter)* is aposztrofált időszakok a fejlesztésben remélt ugrás elmaradása miatti csalódásnak, az emiatt felerősödő kritikus hangoknak és ezt követően a fejlesztésre fordítható források elapadásának voltak nagyrészt köszönhetők.[11]

Az 1974–1980 közötti hanyatló időszak Európában az úgynevezett Lighthill-jelentés[12] megsemmisítő kritikai észrevételeinek hatására alakult ki, míg az Egyesült Államokban ezzel egyidejűleg néhány nagy érdeklődéssel várt projekt kudarca után a DARPA mesterséges intelligenciával kapcsolatos alapkutatásokra szánt keretét csökkentették szinte nullára.

---

[10] Russell–Norvig 2016.

[11] Thierer – Castillo O'Sullivan – Russell 2017, 7.

[12] Lighthill 1973.

Az 1987–1993 közötti „tél” ismét az USA-ban folyó mesterségesintelligencia-kutatásokra szánt források radikális csökkentésének következménye volt.[13] Ezt követően a mesterséges intelligenciát kutató programok – részben a fogalomhoz társított kudarcos felhangok miatt – más név alatt folytak: gépi tanulásnak, kognitív rendszereknek, intelligens rendszereknek vagy tudásalapú rendszereknek nevezték ezeket, részben utalva arra, hogy nem univerzális mesterséges intelligenciát kívánnak kifejleszteni, hanem fókuszáltan egy-egy részterület problémáit kívánják megoldani.

1.2.3. A mesterséges intelligencia kutatása napjainkban

Az 1970-es években egy olyan előrelépés jelent meg a mesterséges intelligencia működésében, amely máig meghatározza az alkalmazási területeit. A kezdeti problémamegoldó módszerek nem specifikus, a konkrét problématerületre specializált algoritmust használtak, hanem többféle probléma megoldása során is alkalmazható, általános módszert alkalmaztak. Ennélfogva ezen metódusok teljesítménye kicsi volt, ezért nevezték őket gyenge módszereknek *(weak methods)* is. A gyenge módszerek a problémák nagyobb vagy nehezebben megoldható példányaira már nem voltak felskálázhatók. Ennek alternatívájaként megjelentek a tudásalapú módszerek, amelyek a problémamegoldáshoz nagyszámú speciális rendeltetésű szabályt használnak fel, az adott problémakörre specifikus tudást véve alapul. Az első ilyen rendszerek az orvostudomány területén jöttek létre.[14] Fontos eredményeket a tudásreprezentáció területén értek el, valamint a jövőbeli fejlesztések alapjául szolgált az itt alkalmazott szabályalapú megközelítés. Ezenkívül szintén fontos hozadéka volt ezeknek a megoldásoknak (különösen a MYCIN-rendszernek) a bizonyossági faktor hozzárendelése mesterséges intelligencia által megalkotott következtetésekhez, amely az orvosi ismeretek és diagnózis felállításának bizonytalanságát is tükrözte.

---

[13] MCCORDUCK 2004, 426–431.

[14] Ilyen volt a DENDRAL-program (1969) és az 1970-es évek elején a Stanford Egyetemen bakteriális fertőzések diagnosztizálására kifejlesztett MYCIN-rendszer is. Ez utóbbi körülbelül 600 szabályból állt, a diagnózist végző orvosnak igen/nem válasszal megválaszolható kérdések sorát tette fel, és az ezekre adott válaszokból állította fel a szóba jöhető fertőzést okozó baktériumok valószínűségi sorrendjét. A rendszer képes volt kezelni az abból fakadó bizonytalanságot is, hogy két különböző kérdésre adott válasz egy bizonyos baktérium érintettségét eltérő valószínűséggel mutatta ki.

A jövőben a mesterséges intelligencia működésének számos területén, különösen heurisztikus módszerekben köszön vissza a bizonytalan helyzetekben való döntéshozatal során a valószínűségszámítási módszerek ilyen alkalmazása. A szakértői rendszerek az 1980-as években hamar elterjedtek az üzleti szektorban, az évtized végére szinte minden nagyvállalat alkalmazott valamilyen mesterséges intelligencia alapú szakértői rendszert, amely kézzelfogható költségmegtakarítást eredményezett és az MI-ipar felfutását és dollármilliárdos forgalmú iparággá válását hozta magával.

Napjainkra a mesterségesintelligencia-kutatás a tudományos módszertant és megközelítést is a magáévá tette, egy-egy tézis igazolhatóságát szigorúbb ellenőrzés és statisztikai vizsgálat előzi meg, nagy hangsúly került az eredmények reprodukálhatóságára is. A kutatásban az intuíció helyett a meglévő tételekre alapozott következtetések, valamint az elméletek valós problémák feldolgozása útján való igazolása vált bevett gyakorlattá.[15] A neurális hálók ismét az érdeklődés középpontjába kerültek, a mesterséges intelligencián alapuló megoldások egyikeként a megfelelő statisztikai, alakfelismerési és gépi tanulási technikákkal lehet összehasonlítani, és az adott alkalmazáshoz meg lehet választani az adott probléma sikeres megoldására leginkább alkalmasat. Újra bekerült a mesterségesintelligencia-kutatások körforgásába a valószínűség- és döntéselméleti megközelítés. A Bayes-hálók *(Bayesian networks*[16]*)* alkalmazásával olyan megközelítést alakítottak ki, amely máig uralja a bizonytalan következtetésekre és szakértői rendszerekre alapozott kutatásokat. Ezzel összefüggésben került elő a normatív szakértői rendszer elképzelése, amely a döntéselméleti törvényeknek megfelelően racionálisan cselekszik, és nem az emberi szakértők gondolkodását utánozza.

A mesterséges intelligencia alapú rendszerek fejlődésének újabb lökést adott és irányt szabott az internet elterjedése is. A folyamatosan adatokat fogadó, valós környezetbe ágyazott MI-ágensek mára teljesen megszokottá váltak: ilyen megoldások működtetik a keresőmotorokat, az ajánlórendszereket és az önműködő tartalomszerkesztő rendszereket is.

Russell és Norvig munkája[17] nyomán teret nyert a mesterséges intelligenciával foglalkozó szakirodalomban az ágensalapú megközelítés. Ez vezetett

[15] Russell–Norvig 2016.

[16] A Bayes-háló valószínűségi változók eloszlását reprezentáló modell. Segítségével lehetővé válik egy eredményváltozó valószínűsíthető értékére való következtetés a szülőváltozók lehetséges értékének és a függőségi struktúráknak az ismeretében.

[17] Russell–Norvig 2016.

el annak felismeréséhez, hogy a mesterségesintelligencia-kutatás egyes részterületeit (érzékelés, mozgás, tudásreprezentáció, nyelvfelismerés, kommunikáció stb.) az eddiginél jobban össze kell kapcsolni, ennek érdekében azokat részben át is kell szervezni. Ezzel együtt pedig a kezdeti elkülönülést követően ismét közelebb került a mesterségesintelligencia-kutatás a társtudományokhoz, így a gazdaságtanhoz vagy az irányításelmélethez.

1.3. A mesterséges intelligencia működési elve

1.3.1. Az ágens fogalma

Ágensnek nevezünk valamit, ami képes érzékelni a környezetét az érzékelőin (szenzorok) keresztül, és amely hatást tud gyakorolni a környezetére valamilyen módon. Ha az emberre vetítjük ezeket a fogalmakat, akkor a szenzoroknak a szem, a fül, az orr vagy a tapintást szolgáló bőrfelület nevezhető, míg az ágens a keze, lába, hangszálai stb. útján tud hatást kifejteni a környezetére. Egy robot ágens szenzorai lehetnek kamerák, lézeres vagy radaros letapogató eszközök, míg különböző motorokkal mozgatott részegységeivel gyakorol hatást a környezetére. Egy szoftveres ágens a billentyűzet gombjainak lenyomását érzékeli, fájlok tartalmából vagy a hálózaton érkező információkból tájékozódik a környezetéről, míg tartalmak képernyőn való megjelenítésével, fájlok létrehozásával vagy a hálózaton való adatküldéssel kommunikál. Az érzékelés fogalmát használjuk az ágens érzékelő bemeneteinek leírására egy tetszőleges pillanatban. Egy ágens érzékelési sorozata az ágens érzékeléseinek teljes története, vagyis minden, amit az ágens valaha érzékelt. Egy ágens cselekvése egy adott pillanatban függhet bármitől, amit az ágens érzékel, vagy akár mindentől, amit valaha érzékelt, azonban semmiképpen nem kapcsolódhat olyasvalamihez, amit az ágens nem érzékelt. Ha meg tudjuk határozni, hogy egy ágens hogyan viselkedik minden elképzelhető érzékelési sorozat esetében, akkor tökéletesen sikerült leírnunk az ágenst.

A matematika nyelvén megfogalmazva azt mondhatjuk, hogy az ágens viselkedését az ágensfüggvény írja le, ami az adott érzékelési sorozatot egy cselekvésre képezi le. Ha táblázatos formában szeretnénk megjeleníteni az ágensfüggvényt, ahol a táblázat minden sorának egyik oszlopában található érzékelési sorozathoz a másik oszlopban egy cselekvés kapcsolódik, akkor egy rendkívül nagy – mondhatjuk azt, hogy végtelen – táblázatot kapunk ered-

ményül. Ez a táblázat az ágens cselekvéseinek külső megjelenítése, a belső működést az ágens számára az úgynevezett ágensprogram határozza meg.[18]

Ahhoz, hogy a fent említett táblázat helyes kitöltését, vagyis az ágensprogram helyes felépítését meg tudjuk tenni, meg kell határoznunk, hogy mit tekintünk helyes viselkedésnek az ágens részéről, vagyis mikor viselkedik valóban „okosan" a cselekvő eszközünk. Ehhez pedig azt kell definiálni, hogy mit tekintünk „helyes viselkedésnek". Legegyszerűbben azt mondhatjuk, hogy helyes viselkedés az, ami az ágensünket a legsikeresebbé teszi. Ezzel persze bevezettük a sikeresség fogalmát, ami szintén értelmezést kíván. Ahhoz, hogy meghatározzuk az ágens cselekvésének sikerességét, a sikeresség mérési módját kell kitalálnunk. Az ágens sikerességének kritériumát egy mérhető adat, a teljesítménymérték *(performance measure)* testesíti meg.[19] Ha egy ágenst elhelyezünk a környezetében, akkor az érzékeli a környezetét, és az érzékelések sorozatára cselekvések sorozatával válaszol. A cselekvések sorozatának hatására a környezet állapotok sorozatán halad végig. Ha a környezetben végbemenő változások sorozata számunkra kívánatos volt, akkor az ágens jól teljesített.

Egyetérthetünk abban, hogy nincs olyan univerzális, rögzített mérték, amely minden ágens számára megfelelő teljesítménymérték volna. Ezért valamilyen objektív teljesítménymértéket fogunk használni, amit általában a tervező előzetesen az ágens számára meghatározott. A teljesítménymérték meghatározásakor körültekintően kell eljárnunk, mert az általunk definiált teljesítménymérték nagyban befolyásolni tudja az ágens jövőbeli működését, és akár nem kívánt irányba is terelheti azt. Russell és Norvig alapműnek számító leírásában erre egy robotporszívót hoz példaként. Ha azt az elsőre megfelelőnek tűnő teljesítménymértéket határozzuk meg számára, hogy a munkaideje alatt minél több port takarítson fel, akkor egy racionális ágens azzal fogja maximalizálni az elért teljesítményét, ha a felszívott port újra meg újra kiönti a padlóra, és utána ismételten feltakarítja.[20] Ebből a példából okulva látjuk, hogy a teljesítménymértéket jobb úgy definiálni, hogy a környezetben elérni kívánt állapotot határozzuk meg, és nem az ágens által végrehajtandó műveletekhez kötjük a sikerességet.

Racionális ágensnek azt nevezzük, amely minden lehetséges érzékelési sorozat esetén helyesen cselekszik, vagyis az ágensfüggvény helyesen van definiálva. Az, hogy egy adott pillanatban mi a racionális cselekvés, négy

[18] Russell–Norvig 2016, 33–36.
[19] Goodfellow–Bengio–Courville 2016, 97.
[20] Russell–Norvig 2016, 36–37.

tényezőn múlik: a siker fokát mérő teljesítménymértéken, az ágensnek a környezetéről való eddigi tudásán, az ágens eddig a pillanatig tartó érzékelési sorozatán és azon, hogy az ágens milyen cselekvéseket képes végrehajtani. Az ideális racionális ágens tehát az, amely minden egyes észlelési sorozathoz a benne található tények és a beépített tudása alapján minden elvárható dolgot megtesz a teljesítménymérték maximalizálásáért. A racionalitás mindazonáltal csak annyit jelent, hogy az ágens az addig a pillanatig felépített érzékelési sorozata alapján választja meg a cselekvését annak érdekében, hogy az elvárt teljesítményt maximalizálja. Egy racionális ágens nem mindentudó, vagyis nem tudja a cselekvései valódi kimenetelét, és nem képes a tényleges teljesítmény maximalizálására sem, csakis az elvárt teljesítményére. Az elvárt és a tényleges teljesítmény maximalizálása között kompromisszumot kell kötnünk, mert máskülönben nem tudunk működő ágenst alkotni, hiszen nem várható el tőle, hogy mindent tudjon, és mindennek a tényleges kimenetét is meg tudja jósolni. A racionális viselkedésnek azonban feltétlenül része az a folyamatos törekvés az ágens részéről, hogy minél több hasznos információt szerezzen a környezetéről, mert ezzel a saját viselkedését is tudja javítani. A fenti elvárás nemcsak azt követeli meg a racionális ágenstől, hogy információt gyűjtsön a környezetéről, hanem azt is, hogy amennyit csak lehet, tanuljon a megfigyeléseiből.

A tanuló rendszerek tervezésének utolsó fontos tényezője az *a priori* információ rendelkezésre állásának kérdése. Az MI, a számítástudomány és a pszichológia területén a tanulással foglalkozó kutatás nagy része azzal az esettel foglalkozott, amikor kezdetben az ágensnek semmi információja nincs arról, amit megpróbál megtanulni. Bár ez egy fontos speciális eset, de egyáltalán nem ez az általános. A legtöbb emberi tanulás is egy jó adag kiinduló háttértudásra épül. Néhány pszichológus és nyelvész szerint még az újszülötteknek is van ismeretük a világról.

Bármi is legyen az igazság ezt illetően, kétségtelen, hogy a kiinduló tudás rengeteget segíthet a tanulásban. Egy ágenst nem küldhetünk ki mindenfajta előzetes tudás nélkül a világba úgy, hogy mindent magának kell megtanulnia, mert ez egyrészt nem hatékony, másrészt azt feltételezi az ágenstől, hogy a kezdeti időszakban teljesen véletlenszerű cselekvéseket végezzen (ide-oda menjen, különböző beavatkozásokat tegyen a környezetében), amelyből felmérheti a környezetét és egyes cselekvései kívánt és nem kívánt hatásait. Ezért általában a programozás során megadunk valamilyen előzetes információt az ágens környezetéről és cselekvései hatásáról a környezetben. Ezután ahogy az ágens tapasztalatot szerez, ez az előzetes tudás

módosulhat és átértékelődhet. A sikeres ágensek három különböző periódusra bontják az ágensfüggvény kiszámításának feladatát:

- amikor az ágenst tervezik, a számítás egy részét a tervezők végzik;
- amikor megfontolja a következő cselekvését, az ágens további számításokat végez;
- amikor tanul a tapasztalataiból, még további számításokat végez annak eldöntésére, hogyan módosítsa a viselkedését.

Addig, amíg az ágens a tervezői által beépített tudásra épít, és nem saját megfigyeléseire és az abból tanultakra, nem nevezhetjük autonómnak. Ahhoz, hogy racionális ágensről beszélhessünk, annak autonómnak kell lennie, mindent, amit csak megtanulhat, meg kell tanulnia ahhoz, hogy a hiányos vagy hibás előzetes tudását kompenzálja. Miután elegendő tapasztalatot szerzett a környezetéről, a racionális ágens viselkedése gyakorlatilag függetlenné válhat a beépített előzetes tudásától. Ily módon a tanulás alkalmazásával olyan ágens tervezhető, amely sokféle környezetben is sikeres lesz.

1.3.2. Az ágens felépítése és környezetéhez való viszonya

Az ágenseket eddig viselkedésük leírásán keresztül vizsgáltuk – azon cselekvés alapján, amelyet egy adott észlelési sorozat hatására végrehajtanak. Az ágens belső felépítésének megismerése azonban szintén fontos részét képezi az ágensekről való ismereteknek. A mesterséges intelligencia feladata az ágensprogram megtervezése: egy függvényé, amely megvalósítja az észlelések és a cselekvések közötti leképezést. Ez a program pedig valamiféle fizikai érzékelőkkel és beavatkozókkal ellátott eszközön fog futni – ezt *architektúrának* nevezzük. Az ágens két fő komponense tehát az architektúra (vagyis az ágens „teste") és az ágensprogram (ami az ágens agyának vagy gondolkodásának feleltethető meg).

Nyilvánvalóan az ágensprogramnak olyannak kell lennie, ami megfelelő az alkalmazott architektúra számára. Ha a program mozgásra utasítja majd az ágenst, akkor célszerű valamilyen helyváltoztatásra képes (például kerekekkel felszerelt) architektúrát választani, különben nem fog az elvárásoknak megfelelően működni. Ez az architektúra lehet egyszerű vagy összetett, de általánosságban a működési elv az, hogy az architektúra a szenzoroktól (kamera, LiDAR, billentyűzet stb.) érkező észleléseket elérhetővé teszi a program számára, futtatja a programot, és a cselekvéseit a létrejöttük pilla-

natában a beavatkozók (karok, kerekek, kijelzők) felé továbbítja. A tanulásra is képes ágensek esetében az ágensprogram tovább oszlik a tanuló és a végrehajtó elemre, amelyekkel röviden a következő pontban ismerkedünk meg.

Többféle alapvető ágensfelépítés létezik attól függően, hogy milyen információ jelenik meg és használódik fel a döntési folyamatban. A konstrukció különbözhet hatékonyságában, kompaktságában és rugalmasságában. Az ágensprogram megfelelő konstrukciója a környezet természetétől függ. Az egyszerű reflexszerű ágensek *(simple reflex agents)* azonnal válaszolnak az észlelésekre, míg a modellalapú reflexszerű ágensek *(model-based reflex agents)* a világ aktuális észlelésekből nem nyilvánvaló aspektusait belső állapotukban tartják nyilván. Ezek az ágensek a helyes döntések meghozatala érdekében felépítenek magukban egy modellt arról, hogy hogyan működik az őket körülvevő világ, vagyis hogy milyen cselekvésük milyen változást tud abban végrehajtani, és a világ maga is hogyan változhat meg, például az idő múlása vagy más aktorok fellépése miatt. A modellalapú ágensek lehetnek célorientáltak *(goal-based agents),* amelyek az előre meghatározott céljaik elérése érdekében cselekszenek, és a végrehajtandó cselekvéseket az alapján választják ki, hogy az mennyiben járul hozzá a céljuk eléréséhez. Más megközelítést választanak a hasznosságorientált ágensek *(utility-based agents),* amelyek a saját „boldogságukat" próbálják meg maximalizálni. Ebben az esetben a visszacsatolásban arról is értesül az ágens, hogy a végrehajtott cselekvés eredménye milyen mértékben pozitív vagy negatív, így a legnagyobb pozitív eredményre, vagyis a „boldogság" maximalizálására törekszenek. Mindegyik ágenstípus a tanulás által tovább tudja fokozni a saját teljesítményét.

Végezetül szót kell ejtenünk arról a környezetről is, amelyben az ágens funkcionál, mivel ahogy láttuk, az ágens a környezetével való interakcióban értelmezhető, anélkül nem beszélhetünk cselekvő mesterséges entitásról. Az ágens a környezetéből információkat szerez és arra hatást gyakorol. A működési környezetet az ágens nézőpontjából több szempont alapján osztályozhatjuk.[21] A feladatkörnyezet ennek megfelelően lehet:

- Teljesen megfigyelhető vagy részlegesen megfigyelhető. Ha az ágens érzékelői képesek a környezet minden, a cselekvés kiválasztásához releváns tulajdonságát észlelni, akkor azt mondjuk, hogy a környezet teljesen megfigyelhető. A teljesen megfigyelhető környezetek megkönnyítik az ágens munkáját, mivel nem kell semmilyen belső állapotot nyilvántartania a környezet nyomon követéséhez.

[21] Az osztályozás alapja: RUSSELL–NORVIG 2016.

- Determinisztikus vagy sztochasztikus. Amennyiben a környezet következő állapotát jelenlegi állapota és az ágens által végrehajtott cselekvés teljesen meghatározza, akkor azt mondjuk, hogy a környezet determinisztikus, minden más esetben sztochasztikus.
- Epizódszerű vagy sorozatszerű. Epizódszerű környezetben az ágens tapasztalata elemi „epizódokra" bontható, és a következő epizód nem függ az előző epizódban végrehajtott cselekvésektől, valamint az egyes epizódokban az akció kiválasztása csak az aktuális epizódtól függ.[22]
- Statikus vagy dinamikus. Ha a környezet megváltozhat, amíg az ágens gondolkodik, akkor azt mondjuk, hogy a környezet az ágens számára dinamikus. A statikus környezet szintén az ágens munkáját egyszerűsíti, mivel nem kell állandóan a világot figyelnie, miközben dönt a cselekvés felől, és nem kell az idő múlásával sem törődnie.[23]
- Attól függően, hogy a környezet véges számú különálló állapottal rendelkezik, vagy a változása folyamatszerű, a környezet lehet diszkrét vagy folytonos.
- Végezetül beszélhetünk egyágenses vagy többágenses környezetről attól függően, hogy hány cselekvő szereplő lehet egy időben a környezetben.

Ahogy a fentiekből láthatjuk, az ágensek feladatkörnyezete az ágens nézőpontjából számos tényezővel írható le, ezek különböző kombinációiból adódóan sokféle lehet. Az ágens működését és környezetéhez való viszonyát nagyban meghatározza a környezet jellege. Egészen más megközelítést igényel az ágens részéről egy folyton változó, nem teljesen megfigyelhető, sokszereplős környezet (például az autóvezetés), mint egy olyan, epizódszerűen változó környezet, ahol az ágens egyedül cselekszik, és képes megfigyelni minden releváns körülményt (például alkatrész-osztályozás a futószalag mellett). Az ágens munkakörnyezete így fontos kihatással van arra is, hogy mi a tanulás módja és szerepe az ágens működésében. A következőkben a gépi tanulás módszereivel és alapelveivel ismerkedünk meg.

---

[22] Például az összeszerelő soron lévő hibás alkatrészeket észlelő ágens minden egyes döntését az aktuális alkatrész alapján hozza, függetlenül a korábbi döntésektől, továbbá az aktuális döntés nem befolyásolja, hogy a következő alkatrész hibás lesz-e. Így erre a környezetre – sok más osztályozási feladathoz hasonlóan – azt mondhatjuk, hogy epizódszerű.

[23] A dinamikus környezetre példa lehet a járművezetés, ahol a környezet folyton változik függetlenül attól, hogy az ágens éppen tesz-e valamit vagy sem.

1.3.3. A tanulási folyamat

Az ágenseknek tehát sokféle komponensük van, amelyek sokféleképpen reprezentálhatók az ágensprogramban, így többféle tanulási módszer létezik. Van ugyanakkor egy egyesítő motívum. A tanulás az intelligens ágensekben összefoglalható úgy, mint egy folyamat, amely az ágens minden komponensét úgy módosítja, hogy az összhangba kerüljön a rendelkezésre álló visszacsatolt információval, ily módon növelve az ágens általános teljesítményét. A gépi tanulást végtelenül leegyszerűsítve a következő feladatot adjuk az önállóan tanulni készülő ágensnek: a környezetből érkező adatok hatására végezzen valamilyen cselekvést. A cél az, hogy a cselekvés eredménye a lehető legközelebb essen az elvárt teljesítménymérték szerinti legjobb eredményhez. Ha a cselekvése nem az elvárt legjobb eredménnyel járt, akkor módosítson a cselekvésén, és a következőkben, ha újra ugyanilyen inger éri a környezetéből, akkor más módon reagáljon. A tanuló gép feladata az, hogy a visszacsatolások alapján felépítsen egy rendszert magában, hogy milyen bemeneti ingerre milyen kimenettel reagáljon. Ez természetesen lehet egy összetett, soktényezős bemenet és egy bonyolult cselekvéssor is; minél több tényezőt kell figyelembe vennie a mesterséges intelligenciának, annál komplikáltabb lesz a programja, és annál nehezebb a gépi tanulás útján elérni a legjobb eredményt.

Ha az ágens alapvető komponenseit vizsgáljuk, különbséget tehetünk ágensprogram és architektúra között. Az ágensprogramot magát is két fő elemre bonthatjuk, létezik a javításokért felelős tanuló elem és a külső cselekvések kiválasztásáért felelős végrehajtó elem. A végrehajtó elem végzi az észleléseket, és ez dönt a cselekvésekről. A tanuló elem egy további elemtől, a kritikustól kapott, az ágens működéséről szóló visszajelzést használja annak megállapítására, hogy a végrehajtó elemet hogyan kell módosítani annak érdekében, hogy a jövőben jobban működjön az ágens. A kritikus az, ami megmondja a tanuló elemnek, hogy az ágens milyen jól működik egy rögzített teljesítményszabványhoz viszonyítva. A kritikus nem ugyanaz, mint a tanító, a kritikustól kapott visszajelzések sosem előzetesek, mindig utólag érkeznek.[24] A kritikus szükséges, hiszen az észlelések (például kiválasztotta és külön helyre tette a hibás alkatrészt a futószalagon) önmagukban nem jelzik az ágens sikerességét. Ehhez szükséges egy rögzített teljesítményszab-

[24] ALPAYDIN 2010, 448.

vány, amely tekinthető úgy is, mintha az ágensen teljesen kívül lenne, mivel az ágens nem módosíthatja azt saját viselkedésének megjavítása érdekében.

Ahhoz, hogy az ágens teljesítménye a lehető legjobb legyen, az is szükséges, hogy ne elégedjen meg az „elég jó" eredménnyel, hanem próbálja meg a cselekvéseit addig finomhangolni, amíg még annál is jobb eredményt képes elérni. Ezért szükséges a tanuló ágens utolsó komponense, a problémagenerátor. Ennek feladata, hogy olyan cselekvéseket javasoljon, amelyek új és informatív tapasztalatokhoz vezetnek. Ha az ágens hajlandó egy kis felfedezésre, és rövid távon talán néhány kevésbé sikeres cselekvésre, hosszú távon sokkal jobb cselekvéseket fedezhet fel. A problémagenerátor feladata, hogy ilyen felfedezőutakat javasoljon.

A következőkben röviden tekintsük át a gépi tanulás alapjait és a ma alkalmazott legfontosabb módszereit. Az ágens előtt álló tanulási folyamat meghatározásában rendszerint a visszacsatolás jellege a legfontosabb faktor. A gépi tanulás területén általában három esetet különböztetünk meg: ellenőrzött *(supervised),* nem ellenőrzött *(unsupervised)* és megerősítéses *(reinforcement)* tanulást.

- Az ellenőrzött tanulás egy leképezés megtanulását jelenti a bemeneti és a kimeneti minták alapján. Ilyenkor az algoritmusnak olyan tanítási adatokat adunk, amelyek a kívánt helyes megoldást is tartalmazzák, ezt címkének *(label)* nevezzük.[25] Teljesen megfigyelhető környezet esetén mindig fennáll a lehetőség, hogy az ágens megfigyeli a cselekvése következményeit, így ellenőrzött tanulási módszereket használhat annak érdekében, hogy megjósolja azokat.
- Nem ellenőrzött tanulás esetén bemeneti minták tanulása történik, de a kívánt kimeneti minták nem állnak rendelkezésre. Egy tisztán nem ellenőrzött tanulást végző ágens nem képes megtanulni, hogy mit cselekedjen, mivel nincs olyan információja, amely egy cselekvést helyesnek vagy egy állapotot kívánatosnak minősítene. Ilyenkor a gépre bízzuk, hogy mit kezd a bemeneti adatokkal, hogyan rendszerezi őket, és milyen következtetést von le belőlük. Erre lehet példa egy weboldal látogatóiról rendelkezésre álló rengeteg adat, ahol azt a feladatot adjuk a gépnek, hogy képezzen csoportokat a felhasználókból (klaszterképzés). A nem ellenőrzött tanulást elsősorban a valószínűségi következtető rendszerek kapcsán alkalmazzák.

[25] Géron 2017, 26.

- Végezetül a megerősítéses tanulás mint gépi tanulás típus a legáltalánosabb a három közül, azonban nagyban különbözik tőlük. Ebben az esetben az ágens nem egy tanító útmutatását követi, hanem a megerősítési információ alapján kell tanulnia. Az ágens megfigyeli a környezetét, valamilyen cselekvést végez, majd arról visszacsatolást kap. A visszacsatolás nélkülözhetetlen, mert enélkül – vagyis anélkül, hogy tudná, mi a jó és mi a rossz – az ágensnek semmilyen alapja nem lesz eldönteni, hogy milyen cselekvést válasszon a jövőben. Az ilyen típusú visszacsatolást nevezzük jutalomnak *(reward)* vagy megerősítésnek *(reinforcement).* A mesterséges intelligenciának ebben az esetben magának kell megtanulnia azt, hogy mi a helyes stratégia a legjobb eredmény elérése érdekében. A megerősítéses tanulás tipikusan magában foglalja azt is, hogy az ágensnek meg kell tanulnia azt, hogyan működik a világ, mert enélkül nem tudja a cselekvéseit megfelelően igazítani.

A gépi tanulás erőteljesen matematikai és statisztikai módszerekre támaszkodó tudomány. Sokáig a tanulási függvények kizárólag a statisztika terrénumára tartoztak, majd később, a mesterséges intelligencia fejlettebbé válásával az MI és a számítástudomány közösen kezdte gondozni a statisztikusokkal ezt a területet. Ebben a műben nem célunk a gépi tanulás algoritmusainak és különféle függvényei ábrázolásának bemutatása, pusztán az alapokkal kívánjuk az olvasót megismertetni annak érdekében, hogy később a jogi és jogalkotási problémák taglalásakor érthetőbbek legyenek a gépek viselkedésének okai.

A gépi tanulás során az ágens voltaképpen azt a folyamatot ismétli újra meg újra, hogy a környezetéből érkező ingerre valamilyen módon reagál, majd megvizsgálja, hogy ez a reakció az elvárt teljesítménymérték alapján jó volt-e vagy sem, illetve milyen mértékben volt jó (vagyis mekkora volt a tévedése). Ezután a visszacsatolás alapján, szükség esetén, a jövőbeni viselkedését úgy módosítja, hogy azonos inger hatására az előzőhöz képest más módon cselekedjen. A gépi tanulás során ez az inger-cselekvés-visszajelzés-módosítás ciklus ismétlődik folyamatosan, és minden ismétlés hatására kismértékben vagy akár jelentős módon módosul az ágens viselkedése egészen addig, amíg megtalálja azt a cselekvési utat, amelynél a lehető legnagyobb jutalmat kapja a visszajelzés során. Ne felejtsük el, hogy a gép a tanulás során nem a helyes válasz okait keresi, és ebből alkot rendszert, hanem próbálkozás útján találja meg a jónak minősített válaszokat, és esetenként bizonyos bemenetek és kimenetek között is olyan kapcsolatra bukkan, amely az emberi felfogás szerint nem magyarázható vagy nem tűnik észszerűnek.

Egy tanuló algoritmus akkor jó, ha a tanulási folyamat során olyan hipotéziseket hoz létre, amelyek jól jósolják meg az általuk korábban nem látott példák esetén követendő cselekvéseket. A gép predikcióját ellenőrizni is kell, amelyet a már ismert eredmények alapján tehetünk meg. Ezt egy teszthalmaznak *(test set)* nevezett mintahalmaz segítségével végezhetjük el. Ha az összes rendelkezésünkre álló példát tanításra használjuk, akkor továbbiakat kell gyűjtenünk a teszteléshez. Ezért a tanuló algoritmusok fejlesztői a leggyakrabban a következők szerint járnak el:[26]

1. Gyűjtenek egy nagy példahalmazt.
2. Ezt két különálló részre osztják: a tanító halmazra *(training set)* és a teszthalmazra *(test set).*
3. A tanító algoritmust a tanító halmazon alkalmazzák, vagyis az algoritmus elé tárják a tanító halmaz elemeit és a helyesnek minősített döntést, így hozzák létre a hipotézist.
4. Ezután megmérik a teszthalmazon (ahol szintén ismerjük a helyes választ, de azt már nem mutatjuk meg a gépnek), hogy a hipotézis a halmaz hány százalékára ad helyes döntést.

Ha az ilyen esetekben megvizsgáljuk az átlagos jóslási képességet a tanító halmaz méretének függvényében, akkor azt vesszük észre, hogy a tanító halmaz méretével javul a predikció minősége.

1.3.4. Adat, adat, adat: a gépi tanulás alfája és ómegája

A fent leírtakból is láthatjuk, hogy a gépi tanuló algoritmusoknak adatokra van szüksége ahhoz, hogy a tanulási folyamat működjön. Ahhoz pedig, hogy a tanulási folyamat *kiválóan* működjön, és optimális eredményre vezessen, óriási mennyiségű adat kell. Az adatok célja itt a mesterséges intelligencia betanítása, ezért a mesterséges intelligencia tervezett céljával összhangban álló, strukturált és a helyes válaszokat is tartalmazó, ellenőrizhető adatcsomagokra van szükség.

Ahogy bemutattuk, a gépi tanuláshoz felhasznált adatokat két, diszjunkt csoportra kell osztani: az egyiket a tanításhoz, a másikat az ellenőrzéshez használjuk fel. Minél nagyobb a két adatcsomag, annál hatékonyabb lesz a tanítás és pontosabb az ellenőrzés. Ezt az adatigényt tovább fokozhatjuk

[26] Kelleher – Mac Namee – D'Arcy 1974, 43.

azzal, hogy ha a teszteléskor azt látjuk, az algoritmus nem pontos predikciókat készített, akkor módosítanunk kell azt, és újból tesztelnünk, esetleg újból tanítanunk.[27] Ehhez pedig csak olyan adatokat használhatunk fel, amelyek ismeretlenek a mesterséges intelligencia előtt, hiszen a már ismert bemenetekre van eltárolt helyes válasza. Egy nehéz feladat helyes megoldásának elsajátítása (például élőbeszéd felismerése) már önmagában is rendkívül nagy adatmintát igénylő feladat, hiszen a sokféle tanulási adatból tud csak az algoritmus használható predikciókat felépíteni, nem elegendő adat felhasználásával csak megbízhatatlan eredményt kapunk.

A számítástechnika hatvanéves történetében eddig a legfontosabb tényezőn, az algoritmuson volt a hangsúly. A mesterséges intelligencia terén elért legújabb eredmények azonban azt mutatják, hogy sok esetben nem a feldolgozáshoz használt algoritmus kialakítása jelenti majd a döntő faktort, hanem a tanuláshoz felhasznált adatok megválasztása. Ez az egyre több elérhető adatforrás megléte miatt is igaz. Azt láthatjuk, hogy a mesterséges intelligencia „tudásbeli szűk keresztmetszete" – a rendszer által igényelt összes tudás kifejezésének problémája – számos esetben megoldható tanulási módszerekkel a programozók által belekódolt tudásreprezentációs módszerek helyett, feltéve, hogy a tanuláshoz az algoritmusoknak elegendő adat áll rendelkezésre.

Így tehát a mesterséges intelligenciát fejlesztőknek azzal a problémával is szembe kell nézniük, hogy a gép milyen adatokból tanuljon, és hol érhetők el a kellő mennyiségben a tanuláshoz felhasználható, megfelelően tisztított és strukturált adatok. Az ugyanis megkérdőjelezhetetlen tény, hogy minden gépi tanuló algoritmus és minden önműködő mesterséges intelligencia működése csak annyira lehet jó, amennyire a tanításhoz felhasznált adatok jó minőségűek és korrektek voltak. A későbbi részekben megnézzük azt is, hogy milyen jogi problémákhoz (például diszkriminatív döntéshozatal) vezet az, ha a gépi tanuláshoz felhasznált adattáblák nem megfelelő minőségűek vagy rosszul vannak összeválogatva.[28]

Mostanára egyre több, kifejezetten a mesterséges intelligencia tanítása céljából összeállított adattár érhető el nyilvánosan, illetve vásárolható meg bárki által.[29] Az így összeállított oktató „szöveggyűjteményt" nevezzük korpusznak *(corpus),* amelyek adattípus és tematika szerint is rendezettek:

[27] MOHRI–ROSTAMIZADEH–TALWALKAR 2012, 5.
[28] ORENSTEIN 2017.
[29] MIHAYLOVA 2016.

találunk közöttük újságcikkeket tartalmazó szöveges korpuszokat, hangmintákat, kézírással írt szövegeket, szavakat és sok minden mást. A nyilvánosan elérhető korpuszok nagy segítséget jelenthetnek a fejlesztőknek abban, hogy tanítsák az algoritmusokat, azonban ugyanekkora veszélyt is jelentenek a végeredmény minőségére nézve.

Ahogy fent említettük, a gépi tanulás csak annyira jó, amennyire a felhasznált adatok jók. Ha egy más által kompilált korpuszt használunk, azt is kockáztatjuk, hogy az adatok mennyisége, minősége, válogatása nem az általunk várt kimenet irányába tereli a mesterséges intelligenciát. A korpuszok saját kézi összeállítása olyan hatalmas munka, amit kevesen vállalnak magukra. A kevesebb erőforrással rendelkező fejlesztőknek így marad a másokba fektetett bizalom, a nagy cégek pedig egyre gyakrabban „szüretelik" a tanításhoz használt adatokat a saját felhasználóiktól – azok tudtával vagy tudtán kívül.[30]

[30] Stucke–Ezrachi 2016, 14–17.

Vákát oldal

2. A mesterséges intelligencia működésének egyes közjogi kérdései

A fentiekben áttekintettük a mesterséges intelligencia fejlődésének lépéseit, valamint a mesterséges intelligenciát hasznosító ágensek típusait, felépítését, a rájuk ható tényezőket és főbb működési elveiket, különös tekintettel a gépi tanulás módozataira, annak alkalmazhatóságára és korlátaira. Összességében láthatjuk, hogy a technológiai fejlődés elért abba a szakaszba, ahol már nem számít fikciónak a természetes emberi nyelven feltett kérdéseket megértő és azokra adekvát választ adó, mesterséges intelligencia alapú alkalmazás. Az elérhető hardveres eszközök és a piaci alapon fejlesztett mesterséges intelligencia képességei korábban elképzelhetetlen számítási kapacitást és minden eddiginél hatékonyabb működést tettek lehetővé, elsődlegesen a specializált felhasználási célú alkalmazások terén.

A mesterséges intelligencián alapuló megoldások széles körű elterjedése a filozófiai és etikai kérdések boncolgatásán túlmenően számos jogi kérdést is felvetettek. A jogalkotónak hamarosan választ kell tudnia adni arra a kérdésre, hogy a jelenlegi jogi környezet és a hatályos jogszabályok elégségesek-e a mesterséges intelligencia működésével kapcsolatos kérdések mindenki számára megfelelő kezelésére, vagy pedig új jogalkotásra van szükség, és létre kell hozni a mesterséges intelligencia működésének társadalmi kereteit megteremtő jogszabályi környezetet.[31]

Jelen munkában e kérdés megválaszolásához kívánok hozzájárulni oly módon, hogy bemutatom azokat a közjogi típusú problémákat, amelyeket a technika jelenlegi állása szerint elérhető mesterséges intelligencia és széles körű alkalmazása okozhat. Nem térek ki ebben az írásban a mesterséges intelligencia hadiipari alkalmazásával kapcsolatos jogi és morális kérdésekre, témámat a polgári célú alkalmazásokkal kapcsolatos kérdéskörökre fókuszálom, mivel – bízzunk benne – a katonai alkalmazású önműködő mechanizmusok nem vagy csak minimális mértékben fogják érinteni az állampolgárok mindennapi életét hazánkban és az Európai Unióban.

---

[31] Lásd a témában: Ződi 2018.

2.1. Automatizált döntéshozatal és diszkrimináció

2.1.1. A hátrányos megkülönböztetés megjelenése a mesterséges intelligencia kapcsán

Napjainkban a demokratikus államok jogrendszerei kivétel nélkül az egyenlőség elvén nyugszanak, az alkotmányok deklarálják az emberek törvényi egyenlőségének elvét, és kimondják a hátrányos megkülönböztetés tilalmát. A nemzetközi emberi jogi egyezmények és az Európai Unió Alapjogi Chartája is megkülönböztetett figyelmet szentel az egyenlőség elvének, amely az emberi méltóság jogával is szoros összefüggést mutat.

A mesterséges intelligencia több irányból is érinti az egyenlőség elvének problémakörét. Egyfelől biztosítani kell azt, hogy a fejlett technológiákhoz a többségi társadalomhoz tartozók és a kisebbségek tagjai egyenlő feltételekkel férhessenek hozzá. Ezt alapelvi szinten ki is mondhatjuk, és egyúttal bátran kezelhetjük tényként azt is, hogy a mesterséges intelligenciát alkalmazó eszközök gyártói nem akartak kifejezetten diszkriminatív működésű terméket készíteni. A gyakorlat mégis számos problémát hozott felszínre az elmúlt időszakban. A teljesség igénye nélkül említhetjük azokat a fényképezőgépeket, amelyek nem hajlandók ázsiai emberekről portrét készíteni, mivel úgy ítélik meg, hogy az alany a kép készítésekor pislogott.[32] A képfeldolgozás területén maradva utalhatunk arra a kisebb botrányra, amelyet az okozott, hogy a Google képfelismerő algoritmusa egy színes bőrű párt gorillának címkézett.[33]

A nyelvi szoftverek terén is voltak már bonyodalmak: egy fordítóprogram az orvos szót automatikusan hímneműnek, míg az ápoló kifejezést nőneműnek fordította.[34] Számos oka lehet annak, hogy a mesterséges intelligencia nem működik megfelelően a kisebbségek esetében. Egyik oka ennek az

[32] Elérhető: www.theatlantic.com/technology/archive/2010/01/digital-cameras-still-racist/341451/ (A letöltés ideje: 2018. 08. 26.)

[33] A cég hivatalosan is bocsánatot kért az ügyben. Elérhető: www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/ (A letöltés ideje: 2018. 09. 20.)

[34] A nemre utalás nélküli török személyes névmást a Google Translate hímneműnek fordította angolra az orvos szó, míg nőneműre az ápoló kifejezés esetében *(„He is a doctor" / „She is a nurse").* Elérhető: www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices (A letöltés ideje: 2018. 09. 20.) A mű elkészítésekor a szerző maga is megpróbálta magyarról angolra fordíttatni az „ő orvos" és az „ő ápoló" mondatot, az eredmény ugyanez lett.

öntanuló mechanizmusokkal függ össze: az önálló tanulásra és önfejlesztésre képes algoritmusok az általuk hozzáférhető adatokból dolgoznak, így ha az elérhető adatok döntő része a többségi társadalomról szól (például a képfelismerő szoftver kaukázusi karaktereken tanulta meg felismerni az emberi arcvonásokat), akkor az öntanuló rendszer inherens elfogultságot fejleszt ki a többséghez tartozók irányába anélkül, hogy ez a tervezőknek szándékukban állt volna.[35] A gépek ugyanis az elérhető adatokból és azok előfordulásának gyakoriságából próbálnak következtetni az általuk ismeretlen helyzet helyes megoldására. Ebből fakadhat egyfajta tanult elfogultság a többség irányába, amit az öntanulási képesség adott esetben tovább is erősíthet, mivel nincs olyan mechanizmus, amely a kisebbségi adatokra és érdekekre ráirányítja a figyelmet. A statisztikai valószínűség és az adattömegen alapuló öntanulás nyomán kialakított gépi viselkedés így szándék nélkül is elfogult lehet, mivel nem veszi figyelembe azt, hogy attól, hogy egy válasz statisztikailag korrekt, még nem biztos, hogy ténylegesen is korrekt. Ebben a vonatkozásban tehát beavatkozás szükséges a fejlesztési folyamatba annak érdekében, hogy a fejlődő technológiákhoz való hozzáférés mindenki számára egyenlően biztosítható legyen.

2.1.2. Elfogultság

Az egyenlőség elvének sérülése egy másik irányból is bekövetkezhet. Az automatizált döntéshozatali mechanizmusok ugyanúgy ki vannak téve a fent leírt elfogultsági problematikának, mint a rendszerek működése általánosságban. Az emberi beavatkozás nélküli, mesterséges intelligencián alapuló döntéshozatal szintén az öntanuló mechanizmusok kognitív és prediktív készségeire épül, vagyis a konkrét élethelyzetben hozott döntés kimenetelét nagyban befolyásolja a gépi tanulás során felhasznált adatok köre és minősége, valamint a döntéshozatal során figyelembe vett információk tartalma. Ezt a negatív hatást tovább fokozza, ha az alkalmazott algoritmus egy, már korábban létrehozott adatbázist vesz alapul, esetleg átvéve a más feldolgozók vagy algoritmusok által a személyekhez hozzárendelt következtetéseket, jellemzőket vagy pontszámokat.[36] Ebben az esetben az elsőként elkövetett hibát vagy diszkriminatív pontozást átveszi, sőt, fel is nagyítja a következő elemző algoritmus.

[35] Calo 2017, 10–11.
[36] Balkin 2017, 35.

A diszkriminatív eredményre vezető automatizált döntéshozatalra is számos példát találunk az elmúlt évek (sőt, évtized) gyakorlatából. A hátrányos megkülönböztetést és szegregációt megvalósító gyakorlattal kapcsolatos, nagy felháborodást kiváltó ügy az Amazon internetes áruház aznapi kiszállítást ígérő új szolgáltatásának bevezetésével kapcsolatos. A csomagküldő cég prémium előfizetői részére elérhetővé tett új szolgáltatás lényege az volt, hogy a megrendelt terméket még a megrendelés feladásának napján házhoz szállítják. Ez a szolgáltatás előfizetői díj ellenében vehető igénybe, és nem mindenhol érhető el. Ahogy a szolgáltatással való lefedettséget mutató első térképek kijöttek, láthatóvá vált, hogy több nagyvárosban az elsősorban kisebbséghez tartozó (afrikai–amerikai vagy latin-amerikai) lakosság által lakott városrészek kimaradtak a szolgáltatásból. Ez néhány városban olyannyira szembetűnő volt, hogy a város egyik középső városrészében nem volt elérhető a szolgáltatás, míg az azt körülvevő valamennyi kerületben igen.[37]

A szolgáltató hivatalos magyarázata szerint csak anomáliáról van szó, míg más városok esetében az elérhető raktárak távolságára hivatkoztak egyes városrészek kihagyása esetén. A döntéshozatal módjának részleteit a szolgáltató üzleti titokként kezeli, így pontos részleteket nem tudunk, mindazonáltal több változó jöhet számításba a lehetséges okok között. Egyrészt ezek a területek általában alacsonyabb jövedelműek által lakottak, így kevesebb a 99 dolláros előfizetési díjat kifizetni tudók száma is, másfelől pedig jellemzően magasabb a bűnözési ráta ezeken a részeken, így a kiszállítás biztonsága miatt is aggódhatott a cég.

Az mindenesetre jól látszik ebből az esetből, hogy a pusztán adatokon alapuló, emberi beavatkozás nélküli döntéshozatal könnyen vezethet diszkriminációt megvalósító eredményre. A nagy adatmennyiség elemzésének alapul vételével hozott döntések így épp a hátrányos helyzet felszámolására irányuló társadalompolitikai célok ellenében hatnak, és a meglévő különbségeket még jobban felnagyítják, a hátrányos helyzet következményeit súlyosbítják, vagy a meglévő szegregátumokat még jobban elszigetelik.[38] Éppen ezért szükség van arra, hogy az öntanuló vagy automatikusan döntéseket hozó mechanizmusok döntéseit korrigálni lehessen, illetve be lehessen avat-

[37] A példa Boston Roxbury városrészét érinti. A részletes térképek és a botrányról szóló tudósítás: www.bloomberg.com/graphics/2016-amazon-same-day/ (A letöltés ideje: 2018. 09. 20.)

[38] Ződi 2017, 23–24.

kozni a folyamatokba annak érdekében, hogy a hátrányos helyzetből fakadó negatív következményeket ne felnagyítsa, hanem csillapítsa a mechanizmus, vagy legalább egyenlő módon kezelje az egyéneket.

Röviden ki kell térnünk arra is, hogy az automatizált döntéshozatal és az adatbányászaton alapuló információszerzés nemcsak a gazdasági szektor szereplői által alkalmazható technika, hanem legalább ennyire – ha nem sokkal inkább – problémás az ilyen algoritmusok alkalmazása az állami hatóságok által, különösen a bűnüldözés területén. 2017. nyár végének híre volt, hogy az Egyesült Államok bevándorlási hatósága (ICE) adatbányászaton alapuló, mesterséges intelligenciát alkalmazó algoritmus kifejlesztésén dolgozik, amely az általuk valaha vizsgáltnál jóval nagyobb, és több forrásból (így az interneten nyilvánosságra hozott információkból, blogokból, közösségi oldalakon közzétett adatokból, Twitter-bejegyzésekből és a sajtóból) származó adatmennyiséget tud feldolgozni. Az algoritmus a tervek szerint képes lesz megjelölni azokat a bevándorlókat és nem bevándorlási céllal érkező idegeneket, akik potenciális veszélyforrást jelenthetnek, mert elképzelhető, hogy terrorcselekményt vagy más bűntettet követnek el.

Az *extreme vetting initiative* névre keresztelt kezdeményezést[39] számos jogvédő szervezet és a tudományos élet több képviselője kritizálta[40] azt hangsúlyozva, hogy a tervezett algoritmus diszkriminatív, elfogult és pontatlan lesz, ezáltal számos ártatlan embert tesznek ki negatív következményeknek. Ez a példa is mutatja, hogy a bűnüldöző szervek által a potenciális bűnözők prediktív megjelölésére alkalmazott mesterséges intelligencia alapú algoritmusok nem a fikció kategóriájába tartoznak; a problémát az jelenti, hogy bár a potenciális bűnözők azonosítására szolgáló megbízható módszert Lombroso óta keresik, eddig még nem találták meg, így ártatlanok is könnyen a célkeresztbe kerülhetnek. Az ilyen típusú alkalmazásokról előbb-utóbb kiderült, hogy diszkriminatív profilalkotást alkalmaztak, magasabb kockázati pontszámot rendelve a színes bőrűekhez, vagy alacsonyabbat a fehérekhez.[41]

---

[39] Elérhető: https://theintercept.com/2017/08/07/these-are-the-technology-firms-lining-up-to-build-trumps-extreme-vetting-program/ (A letöltés ideje: 2018. 09. 20.)

[40] The Trump administration's extreme vetting plan is being blasted as a 'digital Muslim ban'. *The Business Insider,* 2017. november 17. Elérhető: www.businessinsider.com/trumps-extreme-vetting-initiative-digital-muslim-ban-2017-11 (A letöltés ideje: 2018. 09. 20.)

[41] Guihot–Matthew–Suzor 2017, 18.

2.1.3. Pontozásos döntéshozatal

A potenciális diszkriminatív automatizált döntéshozatal másik nagy területe a pontozásos *(scoring)* megoldások területe. Nemzetközi szinten és hazánkban is egyre elterjedtebb az ügyfelek helyzetének felmérésére használt pontrendszer alkalmazása a döntéshozatalban, akár emberi beavatkozással, akár teljesen automatizálva történik. A biztosítótársaságok, a hitelminősítők és más szolgáltatók az ügyfeleikről gyűjtött vagy bekért adatok alapján állapítják meg a szolgáltatásuk igénybevételének feltételeit, így például az ügyfél hitelképességét, biztosítási szempontú kockázati tényezőit, vagy hoznak más, a szolgáltatással összefüggő döntést. Az így kialakított pontszám alapján határozzák meg, hogy milyen kondíciókkal (és összegben) nyújtanak hitelt, mennyi lesz az adott ügyfél biztosítási díja stb. Ezek, az ügyfél helyzetét alapjaiban befolyásoló döntések a mesterséges intelligencián alapuló pontozó alkalmazások használatával a korábbinál jóval több adat felhasználásával, több szempont figyelembevételével és akár teljesen automatizáltan is megszülethetnek.

Ebben az esetben nem a felhasznált adatok minőségével vagy a többségi társadalom irányába való elfogultsággal van a legnagyobb jogi probléma, hanem a felvett adatok értelmezésével és pontszámmá transzformálásával. A rendszerek olyan, látszólag objektív jellemzőkhöz rendelnek negatív pontszámot, amelyek jellemzően a kisebbségekhez tartozók esetében fordulnak elő,[42] ez pedig kimeríti a közvetett diszkrimináció[43] jogi fogalmát. A nem szándékos diszkriminatív eredményre vezető automatizált döntéshozatal bekövetkezhet az algoritmusok úgynevezett bizonytalansági elfogultsága *(uncertainty bias)* miatt is. Ez akkor fordulhat elő, ha a használt algoritmus kockázatkerülésre van beállítva, emellett pedig a vizsgált csoport a tanuláshoz használt mintában alulreprezentált volt, így a mesterséges intelligenciának nem állt elég adat a rendelkezésére ahhoz, hogy pontos előrejelzést készíthessen az adott csoporthoz tartozókról. Így a kiértékelő algoritmus azokat a csoportokat preferálja, amelyekre vonatkozóan elég információja van, tehát kevesebb a bizonytalansági tényező a döntésben.[44]

---

[42] KEATS CITRON – PASQUALE 2014, 14–15.

[43] 2003. évi CXXV. tv. 9. §.

[44] GOODMAN–FLAXMAN 2017, 54.

2.1.4. Szabályozási kérdések

Másrészről a szabályozó testületeknek nehéz dolguk van a diszkriminációs esetkörök felderítésében, mivel a cégek nem teszik lehetővé pontozási mechanizmusaik megismerését, azokat általában üzleti titokként kezelik. Ennek ellenére (vagy éppen ezért) fontos a jogalkotói beavatkozás annak érdekében, hogy a szabályozás biztosítsa, az automatizált döntéshozatalt végző mesterséges intelligenciát alkalmazó mechanizmusok nem szándékoltan az ügyfelek pontozására nem használnak olyan kategóriákat, amelyeket a társadalom – jogi vagy erkölcsi alapon – nem tart elfogadhatónak (etnikai vagy faji csoporthoz tartozás, nem, szexuális orientáció stb.)[45]

Az automatizált döntéshozatal a fenti kihívások ellenére nem tekinthető *ab ovo* problémásnak vagy tiltottnak, fontos azonban, hogy a jogalkotó megtalálja és kiépítse azokat a garanciákat, amelyek megvédik az embereket a mesterséges intelligencia által önműködően hozott döntések esetleges negatív következményeitől vagy diszkriminatív hatásától. Az Európai Unió 2018 tavaszán hatályba lépő általános adatvédelmi rendelete (GDPR[46]) több ponton is foglalkozik az automatizált döntéshozatallal és a profilalkotással. Így kimondja, hogy „az érintett jogosult arra, hogy ne terjedjen ki rá az olyan, kizárólag automatizált adatkezelésen – ideértve a profilalkotást is – alapuló döntés hatálya, amely rá nézve joghatással járna vagy őt hasonlóképpen jelentős mértékben érintené."[47]

Az automatizált döntéshozatalra az érintett beleegyezésével továbbra is lehetőség van, valamint megengedhető akkor, ha az az érintett és az adatkezelő közötti szerződés megkötése vagy teljesítése érdekében szükséges, valamint ha azt a szükséges garanciákat tartalmazó uniós vagy tagállami szabályozás lehetővé teszi.[48] Az érintettet két kulcsfontosságú jog illeti meg: egyrészt az, hogy az adatkezelő részéről emberi beavatkozást kérjen, álláspontját kifejezze, és a döntéssel szemben kifogást nyújtson be.[49] Másrészt pedig az érintett jogosult az automatizált döntéshozatal útján hozott döntés

---

[45] Ramirez 2013, 9.

[46] Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK-rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet).

[47] GDPR 22. cikk (1) bekezdés.

[48] GDPR 22. cikk (2) bekezdés.

[49] GDPR 22. cikk (3) bekezdés.

megmagyarázására. Így tehát a mesterséges intelligencián alapuló automatizált döntéshozatali mechanizmust működtető adatkezelők nem rejthetik el teljesen az algoritmus működési módját, mivel az adatkezelés megkezdése előtt az érintettet tájékoztatniuk kell az alkalmazott logikáról és arra vonatkozó érthető információkról, hogy az ilyen adatkezelés milyen jelentőséggel bír, és az érintettre nézve milyen várható következményekkel jár.[50]

Amint azt láthattuk, a mesterséges intelligencián alapuló döntéshozó vagy döntéstámogató algoritmusok nagy mennyiségű adathalmaz feldolgozása útján hoznak döntéseket, az öntanulási képességeik is nagy mennyiségű betáplált adaton alapulnak. Ezek az algoritmusok képesek arra, hogy az emberi munkával nem vagy csak aránytalan időráfordítással feldolgozható adatmennyiséget feldolgozzanak, rendszerezzenek, abból értékes információkat nyerjenek ki, és következtetéseket vonjanak le. Az automatizált döntéshozatal témakörének taglalása így átvezet minket a következő szélesebb kérdéskörre, a mesterséges intelligencia mintázatfelismerő képességével kapcsolatos jogi problematikára és ezzel összefüggésben a magánszféra védelmének kérdéseire.

2.2. Esettanulmány: az automatizált döntéshozatal és a diszkrimináció problémájának megjelenése a személyre szabott árazási gyakorlatban

2.2.1. Az online kereskedelem és a mesterséges intelligencia

Az online kereskedelem mindennapi életünk része. Áruk és szolgáltatások vásárlása az interneten keresztül mára olyan egyszerű és annyira természetesnek hat, mint a weben való böngészés vagy a bankkártyás fizetés a boltban. Az online adásvétel egyre inkább elterjedt, és népszerűsége folyamatosan nő – ami érthető is. Gyors, kényelmes, egyszerű megoldás, és sokkal szélesebb árukínálathoz lehet így hozzáférni, mint amit bármely hagyományos üzlet raktárkapacitása megengedne. A legjobb ajánlat megtalálása könnyedén megy: nem kell boltról boltra vándorolni, hogy összehasonlítsuk az árakat, elég, ha csak egymás után megnyitjuk több eladó weboldalát, sőt, ami még kényelmesebb, felkeressük a számos ár-összehasonlító honlap valamelyikét.

[50] GDPR 14. cikk (2) *g)* pontja és 15. cikk (1) *h)* pontja.

Hamar arra a következtetésre juthatunk, hogy az online kereskedelem segítségével kevesebb erőfeszítéssel jobb üzletet tudunk kötni, mint arra a hagyományos vásárlás során lehetőségünk lenne. De vajon ez tényleg így van? Biztosak lehetünk abban, hogy körültekintő vizsgálódásunk eredményre vezetett, és az elérhető legjobb áron vásároltuk meg az áhított terméket? Vagy pedig tudtunkon kívül egy olyan rafinált mechanizmus alanyaivá váltunk, ahol az eladók a termékeiket nem a legjutányosabb áron értékesítették, hanem azt a legmagasabb árat adtuk érte, amit még éppen hajlandók voltunk kifizetni?

2000-ben, amikor az online vásárlás még jóval kevésbé volt elterjedt, mint napjainkban, egy gyanútlan vásárló érdekes jelenségre lett figyelmes.[51] Egy DVD-t rendelt az Amazon nevű webáruházból, amely az egyik első és talán a legnagyobb internetes üzlet volt, és amely éppen akkor kezdett el DVD-ket árusítani. Nagyjából 24 dolláros árat fizetett érte, és elégedett volt az üzlettel. A következő napon megint felkereste ugyanezt az oldalt, és ismét találkozott ugyanannak a DVD-nek a hirdetésével, amit előző nap megvásárolt, most azonban az ár 26 dollár volt. Az ilyen anomáliákat általában figyelmen kívül hagyjuk, valószínűleg számos más vásárló legyintett előtte az eltérés láttán. Ő azonban a végére kívánt járni az ügynek, és megtalálni az eltérés okát. Törölte a böngészője előzményeit és valamennyi sütit *(cookie),* amelyet a számítógépén tároltak, és ismét felkereste a webáruházat. Meglepve tapasztalta, hogy az Amazon most 22 dolláros áron kínálja neki ugyanazt a DVD-t. Kísérletének híre gyorsan elterjedt az online fórumokon, és a kommentelők arra kezdtek gyanakodni, hogy az online áruház árdiszkriminációval kísérletezik, és ugyanazt a terméket különböző áron kínálja az először odalátogatóknak és a régi vásárlóinak. Amikor az üzlet visszatérő vásárlóként azonosította őt, magasabb árat írt ki neki, amit valószínűleg még hajlandó lett volna kifizetni, és nem más áruházban vásárolta volna meg a terméket. Amikor kitörölte a böngészési előzményeit, az oldal valószínűleg új vásárlóként azonosította, és alacsonyabb árat ajánlott neki azért, hogy becsábítsa a boltba.

Ahogy a hír elterjedt az online közösségben, és egyre több és több mérges vásárló szólt hozzá a történethez, az Amazon is kénytelen volt reagálni, és kiadott egy hivatalos közleményt. Ebben azt állították, hogy a cég egy „véletlenszerű árazási kísérletet" folytatott, amelynek során a vásárlóknak vé-

[51] On the web, price tags blur. *The Washington Post,* 2000. Elérhető: www.washingtonpost.com/archive/politics/2000/09/27/on-the-web-price-tags-blur/14daea51-3a64-488f-8e6b-c1a3654773da/?utm_term=.8bddfbe6f303 (A letöltés ideje: 2018. 09. 20.)

letlenszerűen kiválasztott áron kínálta a bolt a terméket, függetlenül a korábbi viselkedésüktől az oldalon. Tagadták, hogy árdiszkriminációs gyakorlatot folytattak volna. Az igazság valószínűleg soha nem fog kiderülni, mivel az online áruházak szigorú üzleti titokként őrzik az árazási módszereiket és a honlapjuk algoritmusait.

Ezt a korai esetet követően számos alkalommal állt viták kereszttüzében az online árdiszkrimináció és a személyre szabott árazás. Voltaképpen minden nagyobb internetes kereskedő oldalt megvádoltak már az ilyen praktikák alkalmazásával. Jelen műben megvizsgáljuk az árdiszkrimináció közgazdaságtanát és azokat a megoldásokat, ahogyan azt az online kereskedelemben alkalmazni lehet. Ezt követően kísérletet teszünk annak megválaszolására, hogy a személyre szabott árazás jogellenes, erkölcstelen vagy egyszerűen csak egy olyan jelenség, amellyel nem vagyunk teljesen kibékülve.

2.2.2. Mit jelent az árdiszkrimináció?

Ahhoz, hogy továbbléphessünk, tisztában kell lennünk a piacok, az árazás és az árdiszkrimináció közgazdaságtanának alapvető fogalmaival. Versengő piacokon a hasonló termékek árai némi szórást mutatnak. A vevők még ideális körülmények között is csak tökéletlen információkkal rendelkeznek a számukra elérhető legjobb árakról a piacon. A vásárlóknak erőfeszítést kell tenniük a hasonló termékek árainak felkutatása érdekében, ami általában költséget is jelent számukra.[52] Az eladóknak pedig valamilyen fajta árdiszkriminációs mechanizmust kell alkalmazniuk annak érdekében, hogy a jobban fizető vevőktől nyereségük származzon.

Ahhoz, hogy ezt véghez tudják vinni, több feltételnek kell teljesülnie. Először az eladónak valamekkora piaci erővel kell rendelkeznie – akár csak rövid időre is – annak érdekében, hogy az árat az előállítási költségnél magasabb értékben tudja meghatározni. Másodszor az eladónak tudnia kell irányítania a termék kereskedelmét, a termék továbbértékesítésének pedig nehéznek, költségesnek vagy tiltottnak kell lennie annak érdekében, hogy a termék másodlagos piaca ne tudjon kialakulni. Harmadszor, és legfontosabb szempontként, az eladónak képesnek kell lennie arra, hogy különbséget tegyen a vevők között eltérő árrugalmasságuk és a termék vagy szolgáltatás iránti igényük alapján annak érdekében, hogy tudja, kinek milyen magas

[52] STIGLITZ–SALOP 1982, 1121–1122.

árat jelölhet meg.[53] Ezen feltételek előrebocsátásával vizsgáljuk meg az árdiszkrimináció fogalmát. A közgazdaságtanban az árdiszkrimináció három típusát különíthetjük el.[54]

Az elsőfokú (tökéletes) árdiszkrimináció, vagy más néven személyre szabott árazás, azt a gyakorlatot jelöli, amikor az eladó minden egyes egyedi vevőnek eltérő árat szab meg. Ez az ár megközelíti az adott vevő rezervációs árát (azt az árat, amelyet legfeljebb kifizetni hajlandó), vagyis az eladó így elvonhatja a vevőnél keletkező valamennyi fogyasztói többletet (a fogyasztó hasznát). Egy ilyen adásvétel során a vevő „egyszemélyes piaccá" válik az adott termék vonatkozásában, az eladó pedig olyan végső ajánlatot tehet minden vevőnek, amely által a maximális profitot nyerheti ki a tranzakcióból. A rendkívül személyre szabott vagy egyedi termékek és szolgáltatások piacát szintén az elsőfokú árdiszkrimináció jelenségével írhatjuk le, azonban a következőkben a fogalmat nem ebben az értelemben használjuk.

A másodfokú árdiszkrimináció, vagy nemlineáris árazás, olyan árképzési mechanizmust jelent, ahol az eltérő minőségű vagy mennyiségű termékek ára között van különbség, és nem a vásárló személyének függvényében változik az ár. Másodfokú árdiszkrimináció például az a gyakorlat, amikor a kereskedő kedvezményt ad akkor, ha az adott termékből többet vásárolunk, illetve ha két terméket árukapcsolásban kedvezményes áron vehetünk meg. Szintén másodfokú árdiszkrimináció, ha egy adott termék különböző minőségű változatait kínálják különböző áron eladásra (példa lehet erre a telefon- vagy internetelőfizetés, ahol különböző minőségű szolgáltatáscsomagok közül választhatunk), és a vevő a termék különböző verziói közül az igényeinek megfelelően választhat. A másodfokú árdiszkrimináció esetén a vevő választhat a neki felkínált lehetőségek közül, az eladónak ehhez semmit nem kell tudnia a vevőről.

A harmadfokú árdiszkrimináció, vagy „csoportárazás", azt a megoldást takarja, amikor az eladó azonos terméket vagy szolgáltatást eltérő áron kínál a csoportjellemzőkkel jól beazonosíthatóan eltérő csoportokhoz tartozó vevőknek. Erre az árképzési mechanizmusra példa a diákoknak vagy a nyugdíjasoknak kínált árengedmény, illetve a földrajzi területhez kötődő árazás. Ez a gyakorlat azt veszi alapul, hogy egyes csoportokhoz tartozó személyek jellemzően többet vagy kevesebbet hajlandók fizetni egy termékért. Ebben az esetben az eladónak nem kell azonosítania az egyedi vásárlókat, csak azt

[53] Varian 1989, 597–654.
[54] Miller 2014, 54.

kell tudnia, hogy az adott személy rendelkezik-e azokkal a jellemzőkkel, amelyek alapján az árdiszkriminációt alkalmazni lehet (például a diákstátuszt igazolni tudja diákigazolvánnyal).

A közgazdasági oktatásban elterjedt az a nézet, hogy az elsőfokú árdiszkrimináció a profitnövelésnek inkább ideális, mintsem valóságos formája, mivel ahhoz, hogy az eladó meg tudja állapítani minden egyes vevő pontos rezervációs árát, olyan mennyiségű és részletességű adattal kellene rendelkeznie a vevő szokásairól és elképzeléseiről, amelynek megszerzése a gyakorlati életben lehetetlen lenne. Ezért az elsőfokú árdiszkriminációt olyan elméleti kiindulópontként szokták használni, amely az árdiszkrimináció más formáinak megértését segíti elő. Az elsőfokú árdiszkriminációhoz a legközelebb egy szofisztikált harmadfokú diszkriminatív árképzéssel lehet jutni, ahol az eladó a vevőit számos olyan, kisebb csoportba osztja, amelyek még világosan elkülöníthetők egymástól, és az egyes csoportok vonatkozásában határoz meg eltérő árszintet. Ha ezeknek a megoldásoknak az eredményét vizsgáljuk, azt láthatjuk, hogy az árdiszkrimináció hatékony eszköznek bizonyul az eladók kezében arra, hogy maximalizálják a bevételüket, és kompenzálják a piac tökéletlenségeiből fakadó hátrányokat.

Ennek illusztrálására vizsgáljunk meg egy egyszerű példát.[55] Egy monopolista kereskedő olyan terméket kínál eladásra egy 100 fős piacon, amelynek az előállítási költsége minden esetben stabilan 5 euró. A fogyasztók felének alacsony a fizetési hajlandósága, a másik felének magas. A magas fizetési hajlandóságú fogyasztók 15–20 euró közötti összeget szánnak a termékre, az alacsony fizetési hajlandóságúak 8–10 eurót. Ha az eladó 8 euró áron értékesíti a terméket, akkor minden vevő vásárolni fog, így 100 darab terméket ad el az eladó, darabonként 3 euró profitot realizálva, vagyis összesen 300 euró nyeresége lesz. Ha azonban a sokat költő vásárlókhoz igazítja az árait, és darabonként 15 euró áron kínálja a terméket, csak 50 darabot értékesít ugyan, viszont darabonként 10 euró profitot realizál, összességében tehát 500 euró nyereséget termel. Az első esetben a fogyasztói többlet (a rezervációs ár és a tényleges ár közötti különbség) az alacsony fizetési hajlandóságú vásárlóknál 0–2 euró között van, a többet költő vásárlóknál 7–12 euró között. Az összes fogyasztói többlet így 525 euró.

A második változat esetében a magas fizetési hajlandóságú vevők vonatkozásában 0–5 euró közötti fogyasztói többlet jelentkezik, az összes fogyasztói többlet tehát 125 euró, viszont a populáció fele (az alacsony fi-

[55] A példa forrása: ZUIDERVEEN BORGESIUS – POORT 2017, 353–354.

zetési hajlandóságúak) kielégítetlenül marad. Ha nem alkalmaz semmiféle árdiszkriminációt, az eladó racionálisan 15 euró áron fogja értékesíteni a terméket, így kevesebbet ad el, azonban magasabb profitra tesz szert, mintha alacsonyabb áron kínálná a terméket. Ezáltal a potenciális vásárlók fele (az alacsony fizetési hajlandóságúak) nem jut hozzá a termékhez még akkor sem, ha egyébként többet fizetne érte, mint amennyibe az előállítása került (vagyis az eladó nem értékesítené veszteséggel a terméket).

Most vizsgáljuk meg azt az esetet, amikor az eladó árdiszkriminációt alkalmaz, és az alacsony fizetési hajlandóságúaknak 8 euró áron, míg a többet költők csoportjának 15 euróért kínálja a terméket. Ebben az esetben az eladó jobban jár, mint az uniform árazást alkalmazó bármelyik korábbi esetben. Az összes profit 50 x 10 euró + 50 x 3 euró = 650 euró lesz, az összes fogyasztói többlet 50 × (17,5 euró – 15) + 50 × (9 euró – 8) = 175 euró, és minden potenciális vevő hozzájut a termékhez. Másfelől azonban az árdiszkrimináció eredményeként egyes vevők (a magasabb fizetési hajlandóságúak) többet fognak fizetni egy termékért, mint uniform árazás esetén fizetnének.

A fenti példából azt láthatjuk, hogy az árdiszkrimináció elvonja egyes fogyasztói csoportoktól fogyasztói többletük egy részét. Minél szofisztikáltabb az árdiszkriminációs megoldás, annál több fogyasztói többletet tud elvonni az eladó a vásárlóktól.[56]

Ezen megoldások másik fontos hatása a piaci versenyt érinti. Amíg az uniform áron (az eladó szempontjából racionális 15 euró áron) való értékesítés esetében a vevők egy része kielégítetlenül marad, ez lehetőséget teremt arra, hogy egy alacsonyabb árat kínáló versenytárs belépjen a piacra. Árdiszkrimináció alkalmazása esetén nem marad hely a versenytársaknak, az egész piac el van látva áruval, mind a magas, mind az alacsony fizetési hajlandóságú vevőknek olyan áron kínálják a terméket, amit még hajlandóak kifizetni. Az árdiszkrimináció lehetővé teszi a piac monopolizálását, amely így elveszíti a vonzerejét a piacra lépni szándékozó versenytársak felé.

Az árdiszkrimináció létező kereskedelmi gyakorlat, széles körben tárgyalja a közgazdaságtani szakirodalom, kidolgozott elméleti háttere van és számos gyakorlati példája létezik. A piaci egyenlőtlenségek és a valódi vásárlók eltérő jellemzőinek ellensúlyozása érdekében szinte minden eladó alkalmaz valamilyen árdiszkriminációs gyakorlatot, akár egyes csoportoknak adott kedvezmények, akár mennyiségi kedvezmények, akár szezonvégi

[56] Zuiderveen Borgesius – Poort 2017, 354.

kiárusítás formájában. A másod- és harmadfokú árdiszkrimináció elterjedt gyakorlat a webáruházakban és a „régimódi" boltokban egyaránt.

Az elmúlt évek fontos fejleménye az online kereskedelem elterjedése, a kifinomultabb algoritmusok megjelenése és az, hogy ma már a számítógépek képesek olyan számítási kapacitásra, amely lehetővé teszi, hogy a vásárlókról hihetetlenül nagy mennyiségű adatot gyűjtsenek és dolgozzanak fel. Ezek együttesen alakították ki, hogy az online kereskedők tényleges elsőfokú árdiszkriminációt tudjanak alkalmazni, vagy legalábbis olyan kifinomult harmadfokú árdiszkriminációt, amely azzal határos. A következőkben megvizsgáljuk az online kereskedelemben alkalmazott személyre szabott árazási megoldásokat.

2.2.3. Hogyan történik a személyre szabott árazás az online kereskedelemben?

Ahogy azt fent láthattuk, a tényleges személyre szabott árazáshoz arra van szükség, hogy az eladónak elég információja legyen a vevők szokásairól, céljairól és fizetési hajlandóságáról ahhoz, hogy meg tudja állapítani az egyes vásárlók rezervációs árát. Napjainkra ez már megvalósíthatónak tűnik, amennyiben felhasználjuk azt a hatalmas mennyiségű információt, amely az interneten elérhető; ennek egy részét a felhasználó kifejezetten megadta, más részét adatbázisokból, adatbrókerektől és közösségi oldalaktól lehet megszerezni, egy harmadik csoportja pedig az online áruházba való belépéshez használt számítógép digitális ujjlenyomatának (böngésző- és operációsrendszer-adatok, földrajzi információk, sütik stb.) megvizsgálásával nyerhető ki. Ezek az adatok együttesen több információhoz engednek hozzáférést, mint amiről a kereskedő valaha álmodhatott volna. Ezután már csak egy megfelelő algoritmus és a szükséges számítási kapacitás kell ahhoz, hogy egy részletes személyiségprofilt készítsünk, amely alapján a tényleges rezervációs árhoz közel eső árat tudunk képezni minden egyes vásárló számára. Mindazonáltal számos, súlyos akadály áll a tökéletes elsőfokú árdiszkrimináció megvalósításának útjába.

A tökéletes elsőfokú árdiszkriminációhoz tökéletes monopóliumra van szükség, amely aligha képzelhető el a valóságban. Emellett az eladóknak általában nincsen elég információjuk a vevők maximális fizetési hajlandóságáról, így nem tudják pontosan a lehető legmagasabb árat megállapítani. A valóságban az online személyre szabott árazásnak vajmi kevés köze van a vevők fizetési hajlandóságához, más tényezők játsszák a legfontosabb sze-

repet.[57] Tehát mi az az információ, amelyre egy versengő piacon az eladónak szüksége van ahhoz, hogy sikeresen alkalmazza az árdiszkriminációt?

Először is az eladónak azt kell tudnia, hogy a vevői hogyan reagálnának a versenytársuk által kínált árkedvezményre. Ahhoz, hogy ezt tudja, az eladónak meg kell vizsgálnia a vevők árrugalmasságát. Azok a vevők, akiknek az árrugalmassága kisebb, hajlandóak magasabb árat is kifizetni a termékért, a nagyobb árrugalmasságú vevői kör viszont kevésbé lojális a kereskedőhöz: árérzékenyek, és inkább kevesebbet fizetnének a termékért.[58] A kereskedő számára az az előnyös magatartás, ha árengedményt ad azoknak a vevőknek, akiket könnyen elveszíthetne, ha máshol jobb ajánlatot találnak. Az eladóknak ezért információra van szükségük ahhoz, hogy úgy szabják meg az áraikat, hogy a magasabb árrugalmasságú vevőket magukhoz vonzzák a konkurenciától, míg az alacsonyabb árrugalmasságú vevők akkor sem fogják őket otthagyni, ha magasabb árat kell fizetniük a termékért. A valóságban persze a vevők nem oszthatók jól körülhatárolható kategóriákra, hanem egy skála két végpontja között helyezkednek el, és így a relatív pozíciójukat kell tudnunk meghatározni.

A legegyszerűbb és sokszor a legjobb mód egy vevő árrugalmasságának megállapításához, hogy megvizsgáljuk a korábbi vásárlásait az adott üzletben. A legtöbb kereskedő számára a visszatérő kuncsaftok jelentik a kisebb árrugalmasságú vevői kört, míg az először betérő vásárlók azok, akiknek nagyobb az árrugalmassága, és így magához kell csábítania őket. Az eladók így sokszor árengedményt adnak az első vásárlóknak annak érdekében, hogy magukhoz vonzzák őket, míg a visszatérő vevőknek más előnyöket kínálnak (például törzsvásárlói program árengedményei, egyszerűbb vásárlás vagy gyorsabb kiszállítás) annak érdekében, hogy megtartsák őket, és ne engedjék át a versenytársakhoz.[59]

Az árdiszkriminációt lehetővé tevő következő tényező a vásárlók keresési költsége. A vevők általában nem ismerik a piacon elérhető valamennyi árat, hanem általános elképzeléseik vannak az árak piaci megoszlásáról. Az az előfeltevésük, hogy legalább racionális esélye van annak, hogy jó üzletet fognak kötni. A piaci árakat nem ismerő vevő azonban soha nem fogja megtudni, hogy túl sokat fizetett-e, vagy pedig jó üzletet kötött.[60] Az online

[57] Stole 1989, 22–28.
[58] Miller 2014, 58.
[59] Acquisti–Varian 2005, 379–380.
[60] Miller 2014, 61–62.

vásárlások körében mindazonáltal a keresési költségek nem feltétlenül jelentenek közvetlen anyagi ráfordítást, illetve nem szükségszerűen a vevő választásán múlnak. A számítógéphez kevésbé értő felhasználók, a gyenge internet-hozzáféréssel rendelkező vevők, illetve azok, akik nem tudnak több webáruházat végigböngészni, szintén magas keresési költséggel bírnak. Amennyiben az ebbe a körbe tartozó vásárló egyúttal magas árrugalmassággal (árérzékenységgel) is rendelkezik, úgy egy számukra túl magas árral találkozva egyszerűen kilépnek a piacról, és nem vásárolnak.

A fent elméletben bemutatottakat egyszerűen átültethetjük a gyakorlatba. A visszatérő vásárlók azonosítása még csak nem is bonyolult művelet: minden webáruházban létre kell hozni egy felhasználói fiókot a vásárláshoz. Ha egy vásárló másodszor jelentkezik be a fiókjába, őt onnantól kezdve visszatérő vásárlóként lehet nyilvántartani. Ha sokáig nem használta a fiókját, a korábbi vásárlásait akkor is nyilvántartja a rendszer.

Egy vevő árrugalmasságát már nehezebb megállapítani, bár még így sem túlságosan komplikált. A magas árérzékenységű vevők és azok, akik időt és energiát áldoznak arra, hogy megtalálják a legjobb árat, leginkább a korábbi viselkedésük alapján azonosíthatók. A legegyszerűbb azon felhasználók vásárlásait és webáruházban való viselkedését követni, akik beléptek a fiókjukba az oldalon.

Ha egy felhasználó rendszeresen luxustermékeket nézeget és vásárol anélkül, hogy olcsóbb változatokat vagy használt termékeket keresne, kevéssé árérzékeny vásárlóként jelölhető meg, akinek az esetében az eladónak nem kell aggódnia, hogy elveszíti vásárlóként akkor, ha a következő belépésekor kicsivel magasabb árakat kínál neki, mint másoknak. Ez az ártöbblet tisztán az eladó profitja. Ha azonban egy felhasználói fiók tulajdonosa kevés dolgot vásárol a webáruházban, mindig megnézi a használt termékeket, és keresi az ismert termékek olcsóbb helyettesítőit, őt megjelölheti az áruház magas árrugalmasságú vevőként, és legközelebb megpróbálhatja árengedményekkel vásárlásra bírni.

Ebben az esetben az egyes vásárlásokból kevesebb profitja származik az eladónak, azonban egy új vevő megnyerésével, összeadva azokat a vásárlásokat, amelyek nem jöttek volna létre az új vevő nélkül, az eladó még mindig jobban járt, mintha nem csábította volna a boltjába az új vevőt.[61]

Ahogy a felhasználói fiók története értékes információkat szolgáltat, úgy a felhasználók böngészési előzményei is kincset érhetnek. A nyomkövetésre

[61] ZUIDERVEEN BORGESIUS – POORT 2017, 357.

használt sütikkel *(tracking cookie),* a más honlapokról származó sütikkel és a felhasználó böngészési előzményeinek vizsgálatával könnyedén felmérhető egy személy online tevékenysége. Ha gyakran látogat ár-összehasonlító oldalakat vagy a konkurencia honlapjait, akkor az illető nagy árrugalmasságú vevőként azonosítható, és az eladó megpróbálhatja árengedményekkel megnyerni magának. Újabb kutatások azt állapították meg, hogy a böngészési előzmények alapján pontosabb következtetések vonhatók le egy személy attitűdjére vonatkozóan, mint amikre a szokásos demográfiai adatokból vagy földrajzi helyből lehetne következtetni.[62]

2.2.4. Az online személyre szabott árazás jogi kérdései

Végezetül a személyre szabott árazás közgazdasági elméletének és gyakorlati megvalósulási formáinak vizsgálata után elemeznünk kell ennek a megoldásnak a jogi kérdéseit is. Ha a fogyasztók véleményét vizsgáljuk meg, nagy fokú elutasítottságot tapasztalunk. Felmérések igazolták, hogy a fogyasztók bizalmatlanok a korábbi viselkedésükön alapuló személyre szabott árképzési megoldásokkal szemben még akkor is, ha az árbeli eltérés nekik előnyös.[63] A vevők emellett kevésbé valószínű, hogy vásárolnak, ha az árat tisztességtelennek találják.[64] Sokan azért látják problémásnak a személyre szabott árazást, mert arra rejtett módon is sor kerülhet.[65]

Jogi nézőpontból több olyan érv is van, amit meg kell vizsgálnunk annak megállapításához, hogy a személyre szabott árazás az online kereskedelemben jogellenes-e vagy sem. Először is sértheti a versenyjogot az ilyen magatartás, mivel megnehezíti vagy lehetetlenné teszi a versenytársak piacra lépését, ezáltal pedig kvázi monopolizálja az adott piacot. Ez az érv megállhatja a helyét, mindazonáltal ez csak egy része az igazságnak. Az árdiszkrimináció alkalmazásakor ugyanis a kereskedők a klasszikus fogolydilemmával szembesülnek. Azt feltételezik, hogy a versenytársaik már alkalmazzák vagy a jövőben alkalmazni fogják a személyre szabott árazást, így nekik is adaptálni kell ezt a gyakorlatot annak érdekében, hogy elkerüljék a piacról

---

[62] Shiller 2013, 21.

[63] Turow et al. 2009.

[64] Richards–Liaukonyte–Streletskaya 2016, 150.

[65] *Big Data and Differential Pricing.* Executive Office of the President of the United States, 2015. Elérhető: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf (A letöltés ideje: 2017. 12. 02.)

való kiszorulást. Emellett megpróbálják megnyerni maguknak a vásárlókat jobb minőségű szolgáltatás nyújtásával, a termékfejlesztésre történő nagyobb ráfordítással vagy jobb minőségű szolgáltatásokkal. Így arra a következtetésre juthatunk, hogy piaci körülmények között az árdiszkrimináció nemcsak hogy nem csökkenti a piaci versenyt, hanem még növeli is azt.

Fogyasztóvédelmi szempontból azt kell megvizsgálnunk, hogy a személyre szabott árazás vagy az árdiszkrimináció megtévesztő kereskedelmi gyakorlatnak minősül-e. Azt kell megállapítanunk, hogy akkor megtévesztő egy ilyen gyakorlat, ha a vásárlót az üzlet megkötése során lényeges kérdésben félrevezette az eladó. Egyetérthetünk azzal az érvvel, hogy a vásárlóknak nagyon is fontos, hogy mások mennyit fizettek az adott termékért, mindazonáltal nem hihetjük azt, hogy egy adásvétel során a vevő alapvető feltevésként az árak teljes uniformitásából indult ki.[66] Ha figyelembe vesszük azt a tényt, hogy a vevők általában tudatában vannak annak, hogy az azonos termékért kért ár eltérő lehet üzletről üzletre, nagyváros és kistelepülés között, az online és a hagyományos kereskedelem között, könnyen beláthatjuk, hogy ez az érv nem állja meg a helyét, így ez a kérdés nem jelent valódi törvényességi problémát.

Megítélésem szerint a tényleges személyre szabott árazás jogi megítélése a személyes adatok kezelése irányából, az adatvédelmi jogszabályokon keresztül tehető meg. A kereskedők sokszor elképzelhetetlenül nagy mennyiségű személyes adatot kezelnek és használnak fel annak érdekében, hogy elkészítsék a felhasználókról a személyre szabott árképzés alapjául szolgáló személyiségprofilt. Az ilyen adatok gyűjtése és feldolgozása szigorú szabályozást igényel az Európai Unió és a tagállamok adatvédelmi jogában. A kérdés azonban először is az, hogy ez a fajta adatgyűjtés minden esetben személyes adatok kezelését jelenti-e.

Az elemzés alapjául a 2018-ban hatályba lépő európai uniós szabályozást, az Általános Adatvédelmi Rendeletet (GDPR) veszem alapul, hiszen a kézirat lezárását követő szűk néhány hónapon belül kötelezően alkalmazandó jogforrás lesz az adatvédelem területén az egész Európai Unióban. A rendelet értelmében személyes adat az „azonosított vagy azonosítható természetes személyre (»érintett«) vonatkozó bármely információ; azonosítható az a természetes személy, aki közvetlen vagy közvetett módon, különösen valamely azonosító, például név, szám, helymeghatározó adat, online azonosító […] alapján azonosítható”.[67] Adatkezelésnek minősül a személyes

[66] MILLER 2014, 77.

[67] GDPR 4. cikk (1) pont.

adatokon vagy adatállományokon automatizált vagy nem automatizált módon végzett bármely művelet.[68]

A következő példákon keresztül azt kívánom szemléltetni, hogy a GDPR rendelkezései szerint személyes adatok kezelésének minősül valamennyi olyan eset, amikor személyre szabott árazásra kerül sor az online kereskedelemben. Amikor a kereskedő egy, az oldalra korábban regisztrált felhasználó vásárlási előzményeit vizsgálja, az egyértelműen személyes adatok kezelésének minősül, hiszen az érintett azonosítható a regisztrációkor megadott adatai útján,[69] így a róla megszerzett és elemzett bármely adat személyes adatnak minősül. De mi a helyzet akkor, amikor a felhasználó böngészési előzményeit még azelőtt analizálja a webáruház (és úgy szabja meg a személyre szabott árat), hogy egyáltalán regisztrálna az adott oldalon, vagy belépne a már létező fiókjába? A webáruházak ugyanis a nyomkövető sütiken keresztül számos információt tudhatnak meg a honlapjukat felkereső látogató böngészési előzményeiről és internetezési szokásairól anélkül, hogy az illető regisztrálna a honlapon. Így néhány kulcsinformáció birtokában a webáruház meg tudja ítélni a potenciális vásárló árrugalmasságát, és már eleve a személyre szabott árat tudja megjeleníteni az oldalon. Az európai adatvédelmi hatóságok álláspontja szerint egy felhasználóhoz kötődő süti, amely egyedi azonosítóval rendelkezik, személyes adatnak minősül. Ez azért tekinthető így, mivel az ilyen cookie-k „lehetővé teszik azt, hogy az adatalanyokat egyértelműen azonosítani lehessen *(»single out«),* még akkor is, ha a valódi nevét nem ismerjük".[70]

Végül hogyan ítélhető meg az a gyakorlat, ha egy kereskedő az alapján szabja személyre az árait, hogy a felhasználó milyen típusú okostelefont használ az internet böngészéséhez. Ebben az esetben nem használ nyomkövető sütiket és a felhasználó sincsen személyesen azonosítva, az árat egyszerűen ahhoz igazítja, hogy milyen típusú készüléken jelenítették meg az oldalt. A telefonkészülék típusa és gyártmánya önmagában nem minősül személyes adatnak, és így az ehhez kapcsolódó egyedi ár kialakítása sem tekinthető

---

[68] GDPR 4. cikk (2) pont.

[69] Most eltekinthetünk azon esetek elemzésétől, amikor a felhasználó valótlan adatokkal regisztrál az oldalra, mert megítélésem szerint ezen a módon nem tudja megfelelően igénybe venni a webáruház szolgáltatásait (például hamis szállítási címet vagy nevet megadni nem életszerű, mert így nem jut el a vásárlóhoz a megrendelt termék).

[70] Article 29 Working Party (2010). Opinion 2/2010 on Online behavioural advertising (WP 171) és Article 29 Data Protection Working Party Opinion 4/2007 on the concept of personal data. 01248/07/EN WP 136.

személyes adatok kezelésének. Azonban abban a pillanatban, hogy az adat (az egyedi ár) a vásárlási folyamat során egy konkrét természetes személyhez kapcsolódik (például a vevő megadja a nevét és a szállítási címét), a teljes ezt megelőző eljárás személyesadat-kezeléssé válik.[71]

Összefoglalva tehát megállapíthatjuk, hogy a személyre szabott árképzés gyakorlata – legkésőbb akkor, amikor megtörténik a vásárlás, és ehhez adatot szolgáltat magáról a vevő – személyes adatok kezelésének minősül, és így a GDPR előírásait kell alkalmazni az eseteire.

A következőkben röviden azt tekintjük át, hogy a GDPR milyen előírásokat tartalmaz az ilyen típusú adatkezelési tevékenységre. Elsőként az adatkezelés jogalapját kell tisztázni. A rendelet hat lehetséges adatkezelési jogalapot nevesít, ezek közül hármat érdemes közelebbről megvizsgálnunk. Az első és legegyszerűbb jogalap az esetek mindegyikére alkalmazható, ez pedig az érintett hozzájárulása[72]. A második lehetséges jogalap szerint „az adatkezelés olyan szerződés teljesítéséhez szükséges, amelyben az érintett az egyik fél […][73]" Ebben az esetben azonban kétséges, hogy a személyre szabott árazáshoz kapcsolódó adatkezelés alapítható lenne erre a jogalapra. A szerződéshez szükséges adatkezelés körét az európai adatvédelmi hatóságok is általában megszorítóan értelmezik (például a szállítási cím megadását vagy a fizetéshez szükséges bankkártyaadatok kezelését sorolják ide).[74] Nagy valószínűséggel tehát a személyre szabott árak kialakítását nem tekinthetjük idetartozónak.

A harmadik elképzelhető jogalap szerint akkor folytatható az adatkezelés, ha az „az adatkezelő vagy egy harmadik fél jogos érdekeinek érvényesítéséhez szükséges, kivéve, ha ezen érdekekkel szemben elsőbbséget élveznek az érintett olyan érdekei vagy alapvető jogai és szabadságai, amelyek személyes adatok védelmét teszik szükségessé […]" Itt szintén kevéssé valószínű, hogy az adatvédelmi hatóságok az adatkezelő jogos érdekének tekintenék a személyre szabott árazást, sőt, inkább a fogyasztó érdekei nagyobb súllyal esnek latba, mint a kereskedő gazdasági érdeke.[75]

---

[71] ZUIDERVEEN BORGESIUS – POORT 2017, 358. GDPR 6. cikk [1] bekezdés *a)* pontja.

[72] GDPR 6. cikk [1] bekezdés *a)* pontja.

[73] GDPR 6. cikk [1] bekezdés *b)* pontja.

[74] Article 29 Working Party (2014). Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 9., 16–17.

[75] Az európai adatvédelmi hatóságok véleményében kifejezetten a nem jogszerű adatkezelésre hozták példának azt, amikor a kereskedő a szállítási cím alapján a „rossz környéken" lakóknak magasabb szállítási költséget számolt (Article 29 Working Party [2014]. Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 9., 32.).

A fentiekből láthatjuk, hogy az adatkezelés jogalapjaként egy eset jöhet számításba, az érintett hozzájárulása. Erre az alapra helyezkedve végül azt vizsgáljuk meg, hogy ebben az esetben milyen kötelezettségei származnak az adatkezelőnek, és milyen feltételeknek kell eleget tennie ahhoz, hogy az adatkezelés szabályos legyen, és a fogyasztó (érintett) jogai kellőképpen érvényesülhessenek. A GDPR szigorú szabályokat határoz meg a hozzájárulásra vonatkozóan: az az „érintett akaratának önkéntes, konkrét és megfelelő tájékoztatáson alapuló és egyértelmű kinyilvánítása, amellyel az érintett nyilatkozat vagy a megerősítést félreérthetetlenül kifejező cselekedet útján jelzi, hogy beleegyezését adja az őt érintő személyes adatok kezeléséhez”.[76] Az adatkezelőnek a személyes adatok megszerzésének időpontjában számos adatot kell az érintett rendelkezésére bocsátania, többek között a személyes adatok tervezett kezelésének célját, valamint az adatkezelés jogalapját is közölnie kell.[77] A tájékoztatást tömör, átlátható, érthető és könnyen hozzáférhető formában, világosan és közérthetően megfogalmazva nyújtsa.[78]

A fentiekben részletezett tájékozott beleegyezés elve alapján tehát a kereskedőnek előzetesen közölnie kell a vásárlóival, hogy a róluk gyűjtött személyes adatokat személyre szabott árazás céljából használja fel. Tekintettel arra, hogy – ahogy arra feljebb már utaltunk – a személyre szabott árazással szemben nagy az ellenállás a fogyasztók körében, ez az explicit tájékoztatás elriaszthatja a vásárlókat. Mindazonáltal hangsúlyozni kell, hogy a tájékoztatásnak kifejezettnek kell lennie, az olyan kifejezések, mint a „felhasználói élmény javítása érdekében”, „marketingcélokra”, „jobb személyre szabott szolgáltatások nyújtása érdekében” nem minősíthetők megfelelőnek.[79] Végezetül fel kell hívni a figyelmet arra is, hogy a GDPR külön rendelkezéseket tartalmaz az olyan, kizárólag automatizált adatkezelésen – ideértve a profilalkotást is – alapuló döntésekre vonatkozóan, amelyek az érintettre nézve joghatással járnának vagy őt hasonlóképpen jelentős mértékben érintenék.[80] A személyre szabott árazás online környezetben automatizált döntéshozatali mechanizmus útján történik, a döntései pedig joghatással bírnak a címzettre (esetünkben az adásvételi ajánlatban szereplő ár meghatározását bátran tekinthetjük joghatásnak). A rendelet így tehát az átláthatóság és a tájékozott beleegyezés

[76] GDPR 4. cikk (11) pont.

[77] GDPR 13. cikk (1) *f)* pont.

[78] GDPR 12. cikk (1) bekezdés.

[79] GDPR 5. cikk (1) bekezdés *b)* pont: a személyes adatok „gyűjtése csak meghatározott, egyértelmű és jogszerű célból történjen”.

[80] GDPR 22. cikk (1) bekezdés.

követelményének előírásán kívül további jogokat ad a fogyasztónak. Elsőként joga van ahhoz, hogy ne terjedjen ki rá az automatizált mechanizmus útján hozott döntés, vagyis biztosítani kell számára a személyre szabott árazási rendszerből való kilépés *(opt-out)* lehetőségét. Emellett pedig megilleti a magyarázathoz való jog is, vagyis az, hogy tájékoztassák az „automatizált döntéshozatal tényéről, ideértve a profilalkotást is, valamint legalább ezekben az esetekben az alkalmazott logikára és arra vonatkozóan érthető információkról, hogy az ilyen adatkezelés milyen jelentőséggel és az érintettre nézve milyen várható következményekkel bír".[81] Amennyiben az adatkezelő kifejezett hozzájárulást szerez az érintettől a személyes adatainak személyre szabott árazás céljából való kezeléséhez, úgy az opt-out lehetőség helyett egy másik joga éled fel az érintettnek, ez pedig az emberi beavatkozás kéréséhez való jog. A GDPR szerint „az adatkezelő köteles megfelelő intézkedéseket tenni az érintett jogainak, szabadságainak és jogos érdekeinek védelme érdekében, ideértve az érintettnek legalább azt a jogát, hogy az adatkezelő részéről emberi beavatkozást kérjen, álláspontját kifejezze, és a döntéssel szemben kifogást nyújtson be".[82]

A fentieket összefoglalva megállapíthatjuk, hogy az online kereskedelemben a személyre szabott árazás önmagában nem minősül jogellenes cselekedetnek. Mindazonáltal a jogalkotóknak körültekintéssel kell eljárniuk, mivel nagy a valószínűsége annak, hogy a fogyasztók magatartása még nem alkalmazkodott ahhoz a gyorsan terjedő gyakorlathoz, hogy a javakért való fizetségként vagy a jobb üzlet reményében a felhasználók személyes adataikat adják át az online világ szereplőinek. Tal Zarsky hangsúlyozza, hogy a „fogyasztói rövidlátás" egyike azoknak az általános problémáknak, amelyek a személyes adatok marketingcélokra való felhasználását érintik. A fogyasztók nem rendelkeznek megfelelő eszköztárral annak megítéléséhez, hogy milyen előnyökkel és hátrányokkal jár személyes adataik átadása. Az online kereskedők sokszor nem hozzák a fogyasztók tudomására azt, hogy milyen módon gyűjtenek róluk adatokat, azokat hogyan elemzik és használják fel, a fogyasztók pedig nem képesek a személyes adataik megosztásának lehetséges hátrányos következményeit felmérni.[83] A jogalkotóknak ezért megfelelő eszközöket kell biztosítaniuk a fogyasztóknak arra, hogy megvédjék a magánszférájukat, és kilépjenek bármilyen személyre szabott árazási

[81] GDPR 13. cikk (2) bekezdés *f)* pont.
[82] GDPR 22. cikk (3) bekezdés.
[83] ZARSKY 2004, 40–46.

mechanizmusból, amelybe bevonták őket. Ezt a célt el lehet érni azáltal, hogy a jogalkotó az online kereskedőket és adatbrókereket kötelezi a fogyasztók figyelmének felhívására, hogy személyes adataikat személyre szabott árazáshoz használják fel, mik ennek a következményei, és milyen módon lehet kilépni ebből a mechanizmusból.

2.3. Mintázatok felismerése és a magánszféra védelme

2.3.1. A mesterséges intelligencia mintázatfelismerő képessége

Napjainkban nem okoz különösebb nehézséget hatalmas mennyiségű adatot összegyűjteni és tárolni. Ezek az adatok gyakran tartalmaznak értékes információkat, és hasznos következtetések nyerhetők belőlük. Azonban ha az adatelemzést emberek végzik, akkor hetekbe vagy hónapokba is beletelhet, amíg megszerzik a keresett információt, ha egyáltalán sikerrel járnak. Jelenlegi információs társadalmunkban olyan mennyiségű adat keletkezik nap mint nap, hogy ennek nagy részét sosem elemzik. Az adatbányászat célja éppen az, hogy ezt az űrt betöltse, és önműködően képes legyen olyan modelleket és mintázatokat azonosítani a rendelkezésre álló adathalmazokból, amelyek

- érvényesek, vagyis összhangban vannak a valósággal, és létezésüket más bizonyítékokkal is alá lehet támasztani;
- újszerűek, vagyis nem kézenfekvők vagy triviálisak mindenki számára;
- hasznosak, vagyis gazdaságilag hasznosíthatók;
- érthetők, vagyis az emberi felfogás által feldolgozhatók.[84]

Ennek eléréséhez az adatbányászatban felhasználják a mesterséges intelligencia és gépi tanulás, az adatbáziskezelés, a statisztika és számos más tudományterület eredményeit és eszközeit. Az adatbányász tevékenység az informatikai ipar fejlődésével szinte napról napra hatékonyabbá válik, hiszen az elérhető tárolóeszközök kapacitása kevesebb mint egy év alatt duplázódik meg, ezzel együtt növekszik az elérhető számítókapacitás is. Még drámaibb képet kapunk az elérhető adatmennyiség esetében: számítások szerint az egy nap alatt létrehozott adatok mennyisége napi 2,5 kvintillió bájt. Ezt az adatot tovább árnyalja az a tény, hogy az emberiség összes adatának 90%-a az elmúlt két évben keletkezett.[85]

[84] Fayyad – Piatetsky-Shapiro – Smyth 1996, 40–41.

[85] Elérhető: www.domo.com/learn/data-never-sleeps-5 (A letöltés ideje: 2018. 09. 20.)

A most már szinte mindenki által elérhető és egyre inkább napi szinten használt okoseszközök, érzékelők és alkalmazások fejlődése és térnyerése által az adatkibocsátás mennyisége csak gyorsulni fog a jövőben.[86] John Naisbitt, a híres futurista megfogalmazása szerint az információs társadalom egyik legnagyobb ellentmondása, hogy „megfulladunk az adatoktól, miközben tudásra éhezünk".[87]

Az önkéntesen megosztott, az általunk használt eszközök és alkalmazások által létrehozott, illetve tárolt és a világba kiküldött adatokat az adatbányászatra szakosodott vállalkozások folyamatosan elemzik, trendeket és mintázatokat keresve bennük, amelyek elvezethetik őket olyan újszerű tudáshoz, amely a nyers adatokból önmagában nem volt kiolvasható. A mesterséges intelligencia használatával ez az adatbányászati munka soha nem látott hatékonysági fokot ért el. A sok száz vagy akár több ezer faktor együttes elemzése által olyan információk nyerhetők ki a nyilvánosan elérhető vagy megosztott adatokból, amelyeket a felhasználó egyáltalán nem osztott meg, sőt adott esetben olyan jellemzőket lehet kideríteni róla, amiről akár még maga sem volt tisztában. Az így kinyert információk felhasználása sokrétű, leggyakrabban marketingcélra használják fel, így azonosítva a potenciális célközönséget, illetve a személyre szabott vásárlási ajánlatok és hirdetések útján további vásárlásra buzdítva a meglévő ügyfeleket. Emellett az adatbányászat útján tovább lehet finomítani a fent bemutatott automatikus döntéshozatali mechanizmusokat és pontozási rendszereket. A jogi kérdéseket felvető tényezők többrétűek ezen a téren. Itt kell szót ejteni a profilalkotás kérdéseiről, a felhasználókat hátrányosan érintő olyan megoldásokról, mint a személyre szabott árazás, végezetül pedig a személyes adatok kezeléséről és a magánszféra védelméről, különös tekintettel az olyan információkra, amelyeket nem az adatalany adott meg magáról, hanem az automatizált algoritmus következtetett ki adatbányászat útján.

2.3.2. A mintázatfelismerés jogi kérdései

Az automatizált profilalkotás során az ezt végző algoritmus egy konkrét személyhez kapcsolódó adatok összefűzésével és elemzésével alkot az érintettről személyiségprofilt, amely a szokásaira, attitűdjére és más személyes jellemzőire is kiterjed. E profil birtokában pontosan előrejelezhetők a sze-

[86] IBM Marketing Cloud – 10 Key Marketing Trends for 2017. Elérhető: www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN (A letöltés ideje: 2018. 09. 20.)

[87] NAISBITT 1982.

mély döntései, preferenciái, a gazdasági életben pedig a fogyasztói szokásai, árérzékenysége stb. Az adatok forrása egyrészt maga az érintett, aki egy szolgáltatás igénybevételével vagy egy weboldalra való regisztráció során megad magáról számos információt. Ezeket az adatokat egészítik ki az online tevékenységet követni képes cookie-k, a böngésző által automatikusan megosztott adatok, a böngészési előzmények, a mobileszközök helyadatai és a más forrásokból beszerzett további információk. Számos adatkezelő ugyanis nem csupán a saját rendszerében rögzített vagy begyűjtött adatokból dolgozik, hanem más szereplőktől (például közösségi oldalaktól, e-mail-szolgáltatóktól, videómegosztó-szolgáltatásoktól) is vesz át továbbított adatokat, esetleg igénybe veszi a számos adatbróker-szolgáltatás valamelyikét, és kész adatcsomagot szerez a felhasználóiról.

Az ilyen fajta profilalkotásból származó információkat az adatkezelők felhasználhatják a szolgáltatás testreszabásához, valamint a felhasználók helyzetét befolyásoló döntések meghozatalához is, ami nem feltétlenül előnyös az adatalany számára. Ilyen döntésre példa a személyre szabott árazás, amikor a szolgáltató az árat nem egységesen, hanem vásárlónként eltérően állapítja meg, a vásárlókról rendelkezésre álló információk alapján.[88] Ez az ár lehet alacsonyabb, de akár jelentősen magasabb is a más vásárlóknak kínált árhoz képest.

A közgazdasági és matematikai modellek és empirikus vizsgálatok azt mutatják, hogy a felhasználók viselkedésének követésével (így például a böngészési előzmények vizsgálatával) nyert információk útján sokkal pontosabban előrejelezhető a vásárlói magatartás, mint a pusztán statikus adatokra (például a készülék helyadatai) alapozott következtetésekkel.[89] A helyzet jogi megítéléséhez meg kell válaszolnunk a kérdést: személyes adatnak tekinthetők-e a felhasználó böngészési előzményei, helyadatai és a gépén tárolt sütik abban az esetben, ha az érintett név szerint nem azonosítható általuk. Ebben az esetben az Európai Bizottság adatvédelemmel foglalkozó „29. cikk munkacsoportja" *(Article 29 Working Party)* szerint a személyes adat fogalmába beletartozik minden olyan egyedi azonosító, amely útján az adatalany ugyan név szerint nem azonosítható, azonban egyértelműen megjelölhető vagy kiválasztható.[90] Így tehát a profilalkotás során a profil-

[88] Zuiderveen Borgesius – Poort 2017, 348.

[89] Shiller 2013, 18.

[90] Article 29 Data Protection Working Party Opinion 4/2007 on the concept of personal data. 01248/07/EN WP 136.

alkotó algoritmus üzemeltetője mindenképpen személyes adatokat kezel.[91] A jogalkotónak ebben a vonatkozásban az a szerepe, hogy megalkossa azokat a garanciákat, amelyek a profilalkotás kereteit megadják és hatékonyan védik a magánszemélyek adatait.

Az adatbányászat elterjedésének van még egy potenciális hozadéka. Az egyre fejlettebb algoritmusok az egyre nagyobb elérhető adatmennyiség alapján hatékonyan tudnak olyan adatokra is következtetni, amelyeket a felhasználó egyáltalán nem adott meg. Adott esetben ezek szenzitív adatok (például egészségi állapotra, kóros szenvedélyekre, szexuális beállítottságra vonatkozó ismeretek) is lehetnek. Néhány évvel ezelőtt látott napvilágot a tudósítás arról, hogy egy üzletlánc az Egyesült Államokban sikeresen következtetett arra, hogy egy vásárlója gyermeket vár, pusztán az általa vásárolt termékek körének nyomon követésével.[92] Bár a szóban forgó cég hivatalosan nem erősítette meg a történetet, a szakértők egyetértenek abban, hogy ilyen típusú következtetésekre már alkalmas a technológia és az algoritmusok. Ne felejtsük el, hogy a mesterséges intelligenciát alkalmazó algoritmusok eredeti célja éppen az, hogy olyan mintázatokat is felfedezzenek, amire az ember nem képes.[93] A jogalkotási feladat itt abban áll, hogy a személyes adatok fogalmát kiterjessze azokra a következtetés vagy adatfeldolgozás útján, automatikusan létrehozott információkra is, amelyek kapcsolatba hozhatók egy adatalannyal. Megítélésem szerint – az ilyen adatoknak az érintettekkel kapcsolatos döntések meghozatala során való felhasználására tekintettel – ennek a fogalomnak és a jogi garanciáknak ki kell terjednie az olyan következtetésekre is, amelyek nincsenek összhangban a valósággal, vagyis tévesek.

Összefoglalóan azt láthatjuk, hogy a mesterséges intelligencia és az azon alapuló megoldások, szoftveres vagy fizikai formát is viselő roboteszközök olyan mértékű fejlődésnek indultak, amelyek képessé tették azokat eddig elképzelhetetlen feladatok elvégzésére. Működésük számos olyan kérdést felvet, amelyre a nem túl távoli jövőben a jogalkotónak reagálnia kell annak érdekében, hogy garantálni tudja az emberek jogainak védelmét, az egyenlő bánásmód elvének érvényesülését, valamint a magánszféra sértetlenségét. Ezeken a jogi problémákon megítélésem szerint el kell gondolkodni, párbe-

[91] Article 29 Working Party (2010). Opinion 2/2010 on Online Behavioural Advertising (WP 171).

[92] How Companies Learn Your Secrets. *The New York Times Magazine.* Elérhető: www.nytimes.com/2012/02/19/magazine/shopping-habits.html (A letöltés ideje: 2018. 09. 20.)

[93] CALO 2017, 18.

szédet kell kezdeni a technológiai szektor szereplőivel, a jogi és a műszaki tudományok képviselőivel, azonban semmiképpen sem szabad arra az elhamarkodott következtetésre jutni, hogy az új technológiák bevezetését korlátozó, az új fejlesztéseket fékező szabályozást léptessünk életbe a felmerülő kérdések kockázatkerülő megoldása érdekében. A technológia fejlődése nem állítható meg, sőt nemzetállami keretek között sem tartható, a túlságosan korlátozó intézkedések legfeljebb csak másik államba vagy másik kontinensre telepítik a fejlesztések centrumát. A jogalkotó sokkal helyesebben teszi, ha olyan szabályozási környezetet alakít ki, amely úgy garantálja az emberi jogok védelmét, hogy a fejlesztéshez és a teszteléshez olyan biztonságos környezetet hoz létre, amely nem gátolja az új technológiák megjelenését és elterjedését.[94]

[94] Thierer – Castillo O'Sullivan – Russell 2017, 38.

Vákát oldal

II. rész

A mesterséges intelligencia szabályozása

Vákát oldal

3. A mesterséges intelligencia szabályozásának története és kihívásai

3.1. Bevezető gondolatok

Áttekintettük a mesterséges intelligencia műszaki tulajdonságait, működési elvét és a használata során felmerülő jogi kérdéseket. Nem férhet kétség ahhoz, hogy a mesterséges intelligencia működését, fejlesztését és használatát szabályozni kell. A következő részben a szabályozás egyes konkrétumaival fogunk foglalkozni, lehetőség szerint tanácsot és iránymutatást adva a jogalkotó számára, felhívva a figyelmet azokra a veszélyekre és nehézségekre, amelyek a szabályozás tárgyának különlegességeiből fakadnak. Megvizsgáljuk egyúttal, hogy a mesterséges intelligencia működése kapcsán felmerülő külső hatások *(externáliák)* milyen csoportjait különíthetjük el. Ez a csoportosítás a szabályozás módjára, tartalmára és a jogalkotó szintjére is kihatással van, ahogy azt később bemutatjuk.

Fontos kérdése a szabályozásról szóló résznek, hogy szükség van-e egyáltalán külön szabályozni a mesterséges intelligenciát, vagy pedig a jog jelenlegi rendszere kezelni tud minden felmerülő kérdést. Ezen a téren azt az egyértelmű álláspontot szeretném megismertetni az olvasóval, hogy a mesterséges intelligencia jellegzetességeiből fakadóan előbb vagy utóbb mindenképpen sajátos szabályozási igény fog fellépni, ezért érdemes a jogalkotónak már most elkülönülten foglalkozni a problémával. A mesterséges intelligencia is csak egy eszköz, amelyet az ember a saját céljaira, speciális feladatok megoldására, a mindennapi élet megkönnyítésére használ. A hatalmas számítókapacitás és az összetett feladatok megoldására való képesség önmagában még nem indokolná az önálló szabályozási igényt. A mesterséges intelligencia által támasztott kihívások nem is ebből fakadnak, hanem a bármilyen technológia használatával együtt járó kockázatok felerősített változatának tekinthetők.

Ami a felerősödést okozza és az önálló szabályozási módot indokolja, az a bármilyen más műszaki eszköznél nagyságrendekkel nagyobb döntéshozatali önállóság, ami a mesterséges intelligenciát jellemzi. Ahogy azt a Delvaux-jelentésben az Európai Unió is elismeri, bizonyos autonóm

és kognitív funkciók – például a tapasztalatokból történő tanulás és a kvázi önálló döntések meghozatalának képessége – fejlődése egyre hasonlóbbá teszi a mesterséges intelligenciát azokhoz a szereplőkhöz, amelyek kölcsönhatásban vannak a környezetükkel, és képesek azt jelentős mértékben megváltoztatni.[95] A mesterséges intelligencia fejlődése hamarosan eljut – de akár azt is mondhatjuk, hogy már eljutott – arra a pontra, amikor az eszköz olyan fokú önállóságot mutat, hogy azt önállóan cselekvő ágensként kell tekintenünk, ezzel együtt pedig az elszámoltathatóság gyakorlati kérdései is hirtelen jelentőssé válnak. A jogalkotó hamar szembesül majd azzal a dilemmával, hogy eszközként vagy ágensként értékelje a mesterséges intelligenciát, és a döntése függvényében ehhez szabja a felelősség és elszámoltathatóság fokozatait. Az első esetben ugyanis a felelősség kérdése a tervezőkre hárul, míg a második esetben az utasításokat adó és a működést ellenőrző felhasználóé vagy üzemben tartóé a felelősség.

Végezetül a mesterséges intelligencián alapuló eszközök mostanra teljesen megszokott, hétköznapi háztartási felszereléssé váltak. Ezzel együtt felmerül annak az igénye, hogy a tervezők és a gyártók oly módon építsék meg a mesterséges intelligencia alapú eszközeiket, hogy az átlagos felhasználói hozzáértéssel rendelkezők is különleges elővigyázatossági intézkedések nélkül képesek legyenek biztonságosan és ellenőrizhető módon üzemeltetni az ennek hiányában talán az otthonunkban található legveszélyesebb dolognak tekinthető eszközt.[96]

3.2. A mesterséges intelligencia szabályozására tett kísérletek napjainkban

A közjogi érintettségű problémakörök – korántsem kimerítő – számbavétele után térjünk ki arra, hogy az elmúlt években milyen irányban indult meg a jogalkotói gondolkodás a mesterséges intelligencia kérdéskörében. Hosszú évtizedek hallgatása után az elmúlt két évben a világ nagyhatalmainak vezetői felismerték azt a helyzetet, hogy foglalkozni kell a mesterséges intelligencia jogi szabályozásával. Nagyjából egy időben, egymástól függetlenül

[95] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)], Z. pont. Ezt a dokumentumot a jelentéstevő Mady Delvaux neve után Delvaux-jelentésnek is nevezik.

[96] 6, PERRI, 222–223.

született meg két fontos jelentés az Európai Unió és az Egyesült Államok vezetői számára. Az Európai Unióban 2017 februárjában fogadta el az Európai Parlament az Európai Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról szóló jelentést[97].

2016 októberében tette közzé az USA elnöki hivatalának tudományos és technológiapolitikai irodája a mesterséges intelligencia jövőjére való felkészülésről szóló jelentést.[98] A két dokumentumban közös, hogy a mesterséges intelligenciának a társadalomban való helyét és szerepét próbálják implicit módon előrejelezni, azonban más-más aktuális problémákra keresik a választ. Mindkét jelentés tartalmaz azonban számos olyan jövőbe mutató javaslatot, amelyeket érdemes megemlíteni, mivel elképzelhető, hogy ezek fogják formálni a mesterséges intelligenciáról való gondolkodást az elkövetkező időszakban.

3.2.1. Egyesült Államok

Az Egyesült Államokban kidolgozott javaslat az innovációra mint a gazdasági fejlődés motorjára koncentrál, kijelentve, hogy az üzleti szektor által kifejlesztett mesterséges intelligencia a jövőben hozzájárulhat egyes társadalmi problémák megoldásához. Ami a kormány szerepét illeti ezen a területen, a jelentés csak igen finom eszközöket javasol alkalmazni, a jogalkotó szabályozó szerepét csak korlátozott mértékben látja elfogadhatónak. A szabályozó hatalom birtokában a kormánynak elsődlegesen arra kell figyelemmel lennie, nehogy akadályozza a beavatkozásával a fejlesztéseket. Lehetőség szerint a mesterséges intelligenciát a meglévő szabályozási keretekbe kell beilleszteni (például a repülő- vagy autóiparban).[99] A kormányzat feladata a jelentés szerint az, hogy biztosítsa a kockázatok kezeléséhez szükséges tág szabályozási kereteket, míg az ipar dolga a folyamatos technológiai fejlesztés. A jelentés komoly figyelmet szentel az egyre inkább elterjedt mesterséges intelligencia munkaerőpiacra gyakorolt hatásának, meglátása szerint ez

[97] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)].

[98] Executive Office of the President National Science and Technology Council Committee on Technology. Preparing for the future of artificial intelligence. Washington, DC, USA, 2016. Elérhető: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (A letöltés ideje: 2018. 09. 20.)

[99] Uo. 17.

a folyamat egyaránt fog munkahelyeket teremteni és megszüntetni. A változások leginkább a közepes és alacsony jövedelműeket fogják negatívan érinteni. A kormányzat szerepe itt annak biztosítása, hogy a mesterséges intelligencia elterjedése nem növeli a gazdasági egyenlőtlenségeket.[100]

A mesterséges intelligenciához kapcsolódó etikai kérdéseket – így a tisztesség, elszámoltathatóság és a társadalmi igazságosság – a jelentés az átláthatóság növelésének útján javasolja kezelni, mindazonáltal ennek módjára specifikus javaslatot nem tesz. Ezenfelül kiemeli a nyilvánosan elérhető nagy mennyiségű adatot tartalmazó adatbázisok szükségességét, az adatvédelmi kérdéseket, és javaslatot tesz a mesterséges intelligencia fejlesztésével foglalkozó mérnökök etikai képzésére. E kijelentések fontosságának elismerése mellett meg kell jegyeznünk, hogy ez a megközelítés azzal a veszéllyel jár, hogy a kormányzat ezáltal teljes egészében áttolja a mesterséges intelligencia erkölcsileg helyes működésének és tervezésének felelősségét a magánszektor szereplőire és a polgárokra.[101] Az amerikai megközelítés legnagyobb hátulütője, hogy az iparra telepített önszabályozási mechanizmusok hatására létrejövő keretek az ipari szereplők céljainak fognak kedvezni, minden más érdekelt hátrányára. A mesterséges intelligencia működésének a régi szabályok keretei között való értelmezése szintén komoly problémákat vethet fel, elképzelhető, hogy így a körtét az almával próbáljuk majd összehasonlítani.[102]

3.2.2. Európai Unió

Az Európai Parlament által elfogadott jelentés egyik szembetűnő eltérése az amerikai dokumentumtól, hogy az előbbi sokkal inkább a robotikára koncentrál, a mesterséges intelligenciát pedig nem önálló technológiának, hanem a robotikai megoldásokban alkalmazott eszköznek tekinti. A kizárólag szoftveres alapú, fizikai oldallal nem rendelkező mesterséges intelligenciát így teljes egészében kihagyták az anyagból a készítők. A jelentés egyik központi eleme a robotika munkaerőpiacra gyakorolt hatásának vizsgálata, ennek keretében szükségesnek tartja a munkaerőpiaci változásokat előre jelző mechanizmusok kifejlesztését és alkalmazását. A változások negatív

[100] Uo. 2.
[101] Uo. 9.
[102] CATH et al. 2017.

hatásainak mérséklésére tett javaslatai között az oktatás szerepének erősítése szerepel első helyen, annak érdekében, hogy felvértezzék a munkavállalókat a megváltozott munkaerőpiaci környezetben való helytálláshoz szükséges digitális kompetenciákkal.[103] Hosszabb távú problémaként veti fel a jelentés, hogy egyre több munkakör automatizálásával egyre kevesebb potenciális adófizető jelenik meg, így az állami bevételek csökkenni fognak. Ennek orvoslására egy újfajta adó bevezetését sem látja elképzelhetetlennek a jelentés, emellett pedig a robotizációra áttérő cégeket az így megtakarított társadalombiztosítási hozzájárulás összegének nyilvánosságra hozatalára kötelezné.

A jelentés egy új szervezet létrehozását is sürgeti. A robotikával és mesterséges intelligenciával foglalkozó európai ügynökség feladata a piaci és technológiai trendek követése, jó gyakorlatok felkutatása és szabályozási sztenderdekre való javaslattétel, emellett pedig a fogyasztóvédelmi kérdések kezelése lenne. Az ügynökség vezetné az intelligens robotok nyilvántartását is, amelybe be kellene jelentkeznie minden olyan személynek vagy szervezetnek, amelyik ilyen roboteszközöket használ.[104] Az ügynökség mellett a jelentés a robotok és a mesterséges intelligencia tervezésére, gyártására és használatára vonatkozó Európai Charta létrehozására is javaslatot tesz, amely a robotikai mérnökök magatartási kódexéből, a robotikai protokollok vizsgálatakor a kutatási etikai bizottságok által alkalmazandó kódexből, valamint a tervezők és a felhasználók számára készült engedélymintákból áll, és így a legfontosabb etikai kereteket és alapelveket fekteti le. A legfontosabb ilyen erkölcsi elvek a jelentés szerint a „ne árts" elve, az autonómia és az igazságosság védelme, valamint az EU Alapjogi Chartájának olyan alapjogi rendelkezései, mint az emberi méltóság, az önrendelkezési jog, az egyenlőség elve, az egyéni felelősség, a tájékozott beleegyezés, a személyes adatok védelme és a társadalmi felelősség.[105]

A jelentés általános elvként fogalmazza meg, hogy a robotok és a mesterséges intelligencia fejlesztése és alkalmazása során az embereket nem lehet kitenni további vagy nagyobb kockázatnak, mint aminek a megszokott életvitelük során általában ki vannak téve.[106] A robotokat tervezőkkel szem-

[103] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról (2015/2103[INL]), 14–15.

[104] Uo. 7., 9.

[105] Uo. 9.

[106] Uo. 22.

ben számos olyan kritériumot fogalmaz meg a jelentés, amely a majdani rendszer működésével kapcsolatos. Ilyen az átlátható és kiszámítható működés követelménye, a beépített adatvédelmi mechanizmusok és a kézenfekvő kikapcsoló, vészleállító szerkezet *(kill switch)* beépítésének kötelezettsége.[107] A mesterséges intelligenciát is használó robotok kapcsán a jelentés kitér a robotok által okozott kárért viselt polgári jogi felelősségre, amely kapcsán több szintet különít el, a felelősség szintjének ugyanis arányosnak kell lennie a robotnak adott utasítások tényleges szintjével és a robot önállóságával. Ebben a körben a robotot „oktató" személynek is felelőssége van a robot jövőbeli működéséért, annál nagyobb mértékben, minél hosszabb ideig tartott a robot betanítása. A robotok egyre önállóbbá válásával felértékelődik a felelősség közvetlen megállapításának igénye is, ezért a gépjármű-felelősségbiztosításhoz hasonló felelősségbiztosítási konstrukciót kell kidolgozni az autonóm robotok által okozott károk vonatkozásában.[108]

Végezetül a távolabbi jövőben a jelentés elképzelhetőnek tartja a kifinomult autonóm robotok számára létrehozott elektronikus személyiség kialakítását, amely jogokat és kötelezettségeket határozna meg ezekre nézve, többek között az általuk okozott kárért való felelősség kérdéskörében.[109] Ez a speciális jogi személyiség nem is tekinthető annyira fikciónak, ha tekintetbe vesszük azt a tényt, hogy jelen mű elkészítésének ideje alatt járta be a világsajtót az a hír, miszerint állampolgárságot kapott egy emberszabású robot.[110]

A szabályozás módját tekintve a jelentés a jogi szabályozás és a *soft law* területére tartozó eszközök együttes alkalmazását javasolja, előtérbe helyezve az uniós szintű szabályozást ezen a területen azért, hogy a tagállami jogszabályok fragmentált és egymástól eltérő tartalmú szabályokat állapítsanak meg a közös piacon. A jelentés hangsúlyozza, hogy a robot-eszközök valós körülmények között (városban, közúton) való tesztelése nélkülözhetetlen ahhoz, hogy a lehetséges hibák és problémák a lehető legjobban a felszínre kerüljenek. Ennek érdekében felhívja a Bizottságot, hogy alkossa meg a tesztelésre alkalmazható területek kijelölésére vonatkozó egységes kritériumokat.[111]

---

[107] Uo. 24.

[108] Uo. 16.

[109] Uo. 17.

[110] Elérhető: https://techcrunch.com/2017/10/26/saudi-arabia-robot-citizen-sophia/ (A letöltés ideje: 2018. 09. 20.)

[111] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)], 11.

3.3. A szabályozás nehézségei

3.3.1. Diszruptív technológiák

Egy felforgató *(diszruptív)* új technológia szabályozása rendkívül nehéz feladat, a mesterséges intelligencia pedig kétségkívül a következő évtizedek meghatározó úttörőjének számít majd,[112] ami várhatóan felrázza a megszokott rendszereket. A jogalkotónak meg kell találnia azt a szabályozási optimumot, amely szavatolja az emberek biztonságát, azonban nem akadályozza az új technológiák nyújtotta előnyök kereskedelmi felhasználását és fogyasztói élvezetét, ráadásul lehetőleg előmozdítja az innovációt is. A modern korban az innováció és az új technológiák globális elterjedése olyan gyorsan zajlik, hogy a jogalkotónak komoly nehézséget okoz lépést tartani vele. A diszruptív technológiák területén a jogalkotó dolgát az is nehezíti, hogy hogyan tud ténybeli információkat szerezni a szabályozás tárgyát képező műszaki megoldásokról, holott ez általában minden szabályozás alapjának tekinthető. A jogalkotó elé tárt releváns tények határozzák meg azt, hogy *mit, mikor* és *hogyan* tesz a szabályozás során.[113] Erre a három kérdésre pillantsunk rá röviden a következőkben.

A *mit?* kérdésre való válasz voltaképpen a szabályozandó technológia azonosítását takarja, annak lehatárolását a szabályozás körébe be nem vont technológiáktól. Így például az önvezető autók szabályozásánál – bár elsőre egyértelműnek tűnik a fogalom – pontosan le kell határolni, hogy mikortól tekintünk egy járművet önvezetőnek, és meddig tart a fejlett vezetési asszisztencia. A kérdés megválaszolásához elengedhetetlen a technológiára vonatkozó ténybeli ismeretek összegyűjtése a jogalkotó számára. Ha a *mikor?* kérdésre akarunk válaszolni, akkor arra kell a jogalkotó figyelmét ráirányítanunk, hogy a szabályozást úgy kell megalkotni, hogy az ne legyen se megkésett, ne várja meg, hogy a szabályozatlanságból eredő problémák felüssék a fejüket, és ne legyen elhamarkodott sem, mert az megakadályozza vagy torzítja a műszaki fejlődést. A *hogyan?* kérdésre adott válasz a jogalkotónak a szabályozás módjára és tartalmára vonatkozó meggyőződését mutatja. Itt kell kitérni arra, hogy az új technológiát támogatni, tiltani vagy korlátozni akarja-e a jogalkotó, illetve milyen alapelvek mentén valósítja meg a szabályozást.

---

[112] Elérhető: https://computerworld.hu/uzlet/diszruptiv-technologiak-az-ot-fo-felforgato-231071.html (A letöltés ideje: 2018. 09. 30.)

[113] Fenwick–Kaal–Vermeulen 2017, 7–10.

3.3.2. Szabályozás adatok alapján

A szabályozás irányának megszabása és a konkrét reguláció a jogalkotó testületekben ülő döntéshozók, politikusok és bürokraták feladata lesz, akik döntéseiket a téma szakértői által nekik eljuttatott tények és információk alapján hozzák meg. A döntéshozó személye és a szabályozás szintje egyaránt kulcskérdés, mindazonáltal feltétlenül figyelmet kell fordítanunk a szabályozást megalapozó adatokra, azok körére és forrására is.[114] Egyes tényeket könnyű megállapítani, így például senki nem vitatja, hogy az önvezető járművek is közúti balesetek részesei, sőt akár okozói is lehetnek, ezért ezen helyzetek felelősségi kérdéseit is kezelni kell. Sok esetben azonban az új technológiákra vonatkozó tények megállapítása nehézségekbe ütközik, pont a technológia újdonsága miatt. Így akadályt jelenthet az elégtelen mennyiségű empirikus adat vagy minta megléte, más esetben pedig az adatok helyességét és az azokból levont következtetéseket a területtel foglalkozó szakemberek egy része maga is vitatja, nincs konszenzus a szakértők között.

A releváns tények megszerzése amiatt is hátrányt szenvedhet, hogy az új technológiákkal valamilyen módon érintett érdekcsoportok megpróbálják azokat elfedni, megmásítani vagy más módon befolyásolni. Jól illusztrálja ezt az *Uber* vagy az *Airbnb* közelmúltbeli ügye, ahol az államok jogalkotóit a különböző érdekcsoportok a tények és következmények egyoldalú hangsúlyozásával presszionálták a nekik tetsző irányba, ami végül sok helyen vezetett elhamarkodott, egyoldalú és túlzó szabályozási lépések megtételéhez.

Végezetül a releváns tények egy része valószínűleg még sem a fejlesztők, sem a felhasználók, sem a jogalkotó előtt nem ismert; ilyennek tekinthetjük egy új technológia jövőbeni felhasználási lehetőségeit vagy egy technológia jelenbeli használatából később fakadó következményeket. Ez utóbbit jól illusztrálja, hogy az internet, majd később a közösségi média kezdeti időszakában mindenki válogatás és a következmények mérlegelése nélkül osztott meg magáról személyes adatokat, amit a „nem felejtő" internet el is raktározott az utókornak, és most már hiába is akarnák azokat törölni a világhálóról, digitális lábnyomként rögzültek.

Félreértés ne essék, az újító technológiákra vonatkozó jogalkotói tudás valószínűleg semmiképpen nem lesz egyértelmű vagy vitán felül álló, a szabályozás valószínűleg utólagos lesz, és bizonytalan, átpolitizált információkon fog alapulni. Mindazonáltal a jogalkotónak megítélésem szerint

[114] THIERER – CASTILLO O'SULLIVAN – RUSSELL 2017, 42–44.

törekednie kell arra, hogy ne alkosson túlságosan óvatoskodó szabályozást, ami elszakad a kereskedelmi és fogyasztói igényektől és érdekektől.

Az elmúlt évtizedekben a technológiai fejlődés annyira felgyorsult, hogy a jogalkotó nem tudja követni a fél évszázada még adekvát jogalkotási módszertant és sebességet. Az idő szorításában hozott jogalkotási döntések jellemzője, hogy a releváns tények sok esetben még nem állnak rendelkezésre, vagy a jogalkotó az elérhető tények halmazából egyszerűen nem a megfelelő tényeket választja ki, és arra alapítja a szabályozást.[115] A jogalkotó így könnyen abban a dilemmában találhatja magát, hogy döntenie kell a meggondolatlan cselekvés és a teljes bénultság között. Ilyenkor általában az óvatos megközelítés győzedelmeskedik, és a jogalkotó elővigyázatosságból a status quót fenntartó, korlátozó jellegű intézkedéseket tesz, amelyek megakadályozzák az új fejlesztések hatékony és időben való piacra jutását.[116] Globális kontextusban a nemzetállami vagy regionális jogalkotó ilyenfajta paralízise vagy megszorító szabályozása a technológiai versenyben való gyors alulmaradáshoz vezethet a fejlesztéseket támogató világrészekkel szemben a – most már kimondottan is létező – szabályozási versenyben.[117]

3.3.3. Új szabályozási módszerek

Ahhoz, hogy a fenti jogalkotói dilemmákat megpróbáljuk megoldani, a berögződött jogalkotói attitűdöket is meg kell bontani, és új módszertant kell elsajátítani, amelynek alapja az adatalapú megközelítés. A jogalkotónak a szabályozást megelőzően a lehető legtöbb információt be kell szereznie a szabályozás tárgyáról, lehetőség szerint elfogulatlan forrásokból. Erre jó alapot szolgáltatnak a tudományos kutatások, amelyek jelentős része állami pénzügyi támogatással zajlik, így a szabályozónak hozzáférése van a kutatási eredményekhez, amelyekből első kézből tájékozódhat a technológia állásáról (ez nemcsak a tagállami jogalkotóra vonatkozik, hanem az EU szabályozószerveire is, hiszen az egyik legnagyobb kutatási finanszírozó alapot az unió működteti). Hasonlóan jó indikátorként szolgálhatnak a jogalkotó számára a pénzügyi, gazdasági szektor mozgásai, különösen a startup-finanszírozások és a kockázatitőke-befektetések, amelyek jó eredménnyel jelzik előre

[115] Fenwick–Kaal–Vermeulen 2017, 12.
[116] Thierer – Castillo O'Sullivan – Russell 2017, 20.
[117] Kamar 2006, 1725.

a feltörekvő új technológiákat és azok érettségi fokát is, így megmutathatják a jogalkotónak, hogy mit és mikor kell szabályozási körbe vonnia.[118]

Ahhoz, hogy a jogalkotó a szabályozási versenyben ne maradjon alul, szakítania kell a hagyományos adminisztratív szabályozási rendszerek merev, bürokratikus, rugalmatlan és lassan alkalmazkodó megoldásaival, különösen az új technológiák szabályozása terén.[119] Az új technológiák kezdeti megszorító szabályozása, amely a jogalkotó túlságosan óvatos magatartásából fakad, biztos recept a gazdasági és társadalmi fejlődés hátráltatására. Ha a jogalkotó olyan hipotetikus problémák előzetes megoldására áldozza az energiáját, amit még ő sem lát biztosan előre, vagy amelyek a legrosszabb eshetőségre készülnek fel, azzal azt kockáztatja, hogy a jövőbelátó képességére hagyatkozva rossz döntést hoz, és így nem engedi az esetleges pozitív hatásokat sem kibontakozni.

A diszruptív technológiák terén sokkal hatékonyabbnak látszik az olyan szabályozás, amely lehetővé teszi mind a jogalkotónak, mind a fejlesztőnek, hogy a saját hibájából tanuljon.[120] A kísérletező szabályozás szemléletmódja azt jelenti, hogy a jogszabályalkotást nem tekintjük az út végének, hanem teret engedünk a megalkotott szabályrendszer értékelésének, felülvizsgálatának és átalakításának, az eredmények és az időközben napvilágot látott adatok vagy új felfedezések függvényében. Ez a szemléletmód áthelyezi a szabályozás módjának fókuszát a részletszabályokról az alapelvekre, ami így nagyobb mozgásszabadságot enged a fejlesztőknek.[121] Mindazonáltal fel kell hívni a figyelmet arra, hogy a szabályok gyakori vagy jelentős revíziója kiszámíthatatlan környezetet teremt az érintettek számára, ezért ennél a megközelítésnél nagy jelentősége van a folyamatos kommunikációnak és a szabályozószerv nyitottságának.

A jogszabályi megoldások és a diszruptív technológiák valósághű környezetben való kipróbálására egyaránt jó megoldást jelent a biztonságos tesztkörnyezet létrehozása. Az EU a Delvaux-jelentésben is szorgalmazza a biztonságos tesztpályák és elzárt övezetek létrahozását a mesterséges intelligencia alapú alkalmazások, így különösen az önvezető autók tesztelésére. A jelentés megállapítja, hogy a robotok valós élethelyzetekben történő tesztelése alapvető fontosságú az esetleges kockázatok, valamint a tisztán

[118] FENWICK–KAAL–VERMEULEN 2017, 19–22.
[119] THIERER – CASTILLO O'SULLIVAN – RUSSELL 2017.
[120] Uo.
[121] FENWICK–KAAL–VERMEULEN 2017, 23.

kísérleti laboratóriumi fázison túli technológiai fejlődésük feltárása és értékelése szempontjából. A jelentés hangsúlyozza ennek kapcsán, hogy a robotok valós élethelyzetekben történő tesztelése – különösen a nagyvárosokban és a közutakon – számos kérdést vet fel, többek között olyan akadályokat, amelyek lassítják a tesztelési szakaszok kibontakozását, továbbá hatékony stratégia és nyomon követési mechanizmus kialakítását teszik szükségessé.

Elismerve tehát a valós élethelyzetekben való tesztelés fontosságát, a Bizottság feladatává teszi, hogy az összes tagállam számára egységes kritériumokat dolgozzon ki, amelyeket az egyes tagállamoknak használniuk kell azon területek azonosítására, amelyeken – az elővigyázatosság elvével összhangban – megengedettek a robotokkal végzett kísérletek.[122] A brit bankfelügyelet ehhez hasonló kezdeményezést indított útjára *regulatory sandbox* (kb. szabályozási gyakorlópálya) néven. A kezdeményezés keretében az arra bejelentkezett és alkalmasnak minősített cégek egyes új termékeiket vagy szolgáltatásaikat az általánosan kötelezőnél enyhébb szabályozási környezetben próbálhatták ki határozott időtartam erejéig egy előre rögzített, kis létszámú személyi körben.

A teszt során a hatóság monitorozza a kísérlet folyamatát, és állandó támogatást biztosít a részt vevő cégnek. Az első év tapasztalatai azt támasztják alá, hogy a kísérlet sikeres volt, a bejelentkezett cégek sikerrel tudták fejleszteni és tesztelni a termékeiket, miközben a zárt körű kísérlet során a hatóság folyamatos felügyeletet gyakorolt, és tanácsaival segítette a cégeket a termék jogszabályi megfelelőségének biztosításában annak érdekében, hogy a kísérleti szakasz lezárultával ki tudjanak lépni a nyitott és teljeskörűen szabályozott piacra.[123] Ehhez hasonló zárt körű tesztelési területekre van példa máshol is. Japán úgynevezett „Tokku" zónái, amelyek a tizenöt éve alapított fukuokai robottesztelési szabad zóna után több más nagyvárosban is megjelentek, az ember és az emberszabású vagy segítő robotok egymás mellett élésének kérdéseit engedik tesztelni strukturálatlan körülmények között, mindennapi élethelyzetekben. Ezek a kijelölt területek számos előnnyel kecsegtetnek a robotipar számára. Egyrészt – a praktikus oldalt nézve – lehetővé teszik a cégek számára a robotok éles helyzetekben és az élet által hozott, mindennapi, de kiszámíthatatlan szituációkban való tesztelését, ami

[122] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)], 23. pont.

[123] Elérhető: www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf (A letöltés ideje: 2018. 09. 30.)

számos termékhibát vagy más problémát felszínre tud hozni a piacra lépést megelőzően. Másrészt a hazai robotipart is versenyképesebbé tette ezáltal a japán jogalkotó, mivel a cégek így gyorsan, hatékonyan és a rendkívül körülményes engedélyeztetési eljárást megelőzően tesztelhetik a terméküket, így a versenytársakat megelőzhetik a fejlesztésben és a piacra lépésben, és a jogalkotó sem kockáztatja a techcégek elköltözését az országból a szabályozási környezet miatt. Magasabb szinten nézve a hatásokat, a kísérleti szabad zóna hozzájárulhat az új technológiák társadalmi elfogadottságához, a felmerülő etikai és morális dilemmák megoldásához is.[124] A jogalkotónak hasonlóan hasznosak ezek a kísérletek, mert így azonosítani tudja azokat a problémákat és életviszonyokat, amelyeket a robotok működése okoz, és jogalkotói beavatkozást igényelnek.

Végezetül fel kell hívni arra is a jogalkotó figyelmét, hogy a jövő mesterséges intelligenciáról vagy autonóm robotokról szóló jogszabályait úgy kell megalkotni, hogy azok ne legyenek a hírhedt „piros zászlós" törvényekhez[125] hasonlók. A jogszabályok ne süssék rá az új technológiákra a veszélyesség bélyegét, és ne tegyék az embereket szükségtelenül gyanakvóvá a mesterséges intelligencia irányában. Ez természetesen nem azt jelenti, hogy nem kell biztosítani az emberek felvilágosítását a technológia vagy az annak használata által hordozott kockázatokkal kapcsolatosan. A jogalkotással el kell érni, hogy a gyártók vagy fejlesztők pontosan azonosítsák és egyértelműen megjelöljék a kockázatokat, így a felhasználók „informált döntést" tudnak hozni a használatukkal kapcsolatban.

[124] WENG et al. 2015, 847–849.

[125] Ez a kifejezés az automobilok elterjedésének korai időszakára vezet vissza, amikor a brit törvényhozás előírta a „ló nélküli járművek" vezetői számára, hogy csak úgy közlekedhetnek a közúton, ha legalább 50 méterrel előttük egy gyalogos kísérő halad, piros zászlót lengetve, és figyelmezteti az arra járókat a jármű közeledtére. Lásd: https://en.wikipedia.org/wiki/Red_flag_traffic_laws (A letöltés ideje: 2018. 10. 03.)

4. A mesterséges intelligencia szabályozásának keretrendszere

4.1. Az externáliákról általában

A jogalkotó általában azért szabályoz egy területet, mert az érintett szabályozási tárgy által okozott hatásokat kívánja az ellenőrzése alá vonni. Ezek a külső hatások *(externáliák)* lehetnek a társadalomra vagy az egyénekre nézve pozitívak és előnyösek, vagy pedig negatívak, hátrányosak, néha kifejezetten veszélyesek. Nicolas Petit tanulmányában[126] a mesterséges intelligencia által okozott külső hatásokat jellegzetességeik alapján három csoportba osztja. Az első csoport a helyi hatások *(diszkrét externáliák)* csoportja. A helyi hatások jellegzetességei, hogy azok általában személyesek, véletlenszerűek, ritkák vagy elviselhetők. Személyesek, mert a hatásaik az egyének szintjén jelentkeznek; véletlenszerűek, mert bármely harmadik személyt azonos valószínűséggel érinthetnek; ritkák, vagyis kis gyakorisággal fordulnak elő, végül pedig elviselhetők, mert nem rontják le teljesen az életminőséget, illetve nem javítják azt radikális mértékben. Ilyen lehet például, ha az önjáró felszolgálórobot nekimegy az egyik asztalnak, és kárt tesz a berendezésben, vagy ha a csomagkézbesítő drón nekirepül egy épületnek, és abban okoz kárt. Pozitív helyi hatás lehet az, ha a csomagkézbesítő drón kamerája segítségével észrevesznek és megakadályoznak egy készülő bűntényt.

Az externáliák második csoportját a *rendszerszintű hatások* alkotják. Ezek a hatások jellegzetességeiket tekintve lokalizáltak, kiszámíthatók, gyakoriak vagy visszafordíthatatlanok. Olyan, harmadik személynek okozott hatásokról van szó, amelyek a társadalom egy jól körülhatárolható, de jelentékeny szeletét érintik. Ezek kiszámíthatók, mert a jogalkotó a megfelelő információk birtokában előre láthatja azok bekövetkezését, számíthat rájuk. A gyakoriság alatt az előnyös vagy hátrányos hatások ismétlődő bekövetkezését értjük. Végezetül azért nevezzük visszafordíthatatlannak ezeket a hatásokat, mert tartós változást tudnak előidézni a társadalom jólétében

[126] Petit 2017.

(így például a meglévő egyenlőtlenségeket fokozhatják). Talán a legtöbbet tárgyalt rendszerszintű hatás az emberi munkaerő gépekkel való leváltása, ami a munkanélküliség növekedését eredményezheti, és mivel különösen az alacsony képzettséget igénylő munkákról van szó, tovább növeli a gazdagok és szegények közötti különbségeket. Szintén rendszerszintű negatív hatásnak nevezhetjük a személyes adatok védelmi szintjének általános csökkenését és a személyes adatok nagy cégek birtokába kerülését. Ezekkel szemben jelentkezhetnek pozitív rendszerszintű externáliák is, így például a nagy mennyiségű személyes adat birtokában könnyebb lehet egy járvány terjedését feltérképezni, vagy a megszűnő munkahelyekkel párhuzamosan új, a robotok működéséhez kapcsolódó munkakörök és munkahelyek is létesülhetnek.

A külső hatások utolsó szintjét a sorsfordító, *történelmi jelentőségű externáliák* alkotják (a szerző ezt az *existernalities,* az egszisztenciális és az externália szóból képzett kifejezéssel illeti, utalva arra, hogy az egész emberiség létére kihatással bíró jelenségekről van szó). Ezek a hatások globálisak, valószínűtlenek, megjósolhatatlanok és végzetesek. Kihatásuk az egész emberiségre válogatás nélkül hatással lesz, földrajzi elhelyezkedéstől, nemzetiségtől és más csoportba tartozástól függetlenül. A valószínűtlenség alatt azt értjük, hogy jelenlegi tudásunk és előrelátó képességünk szerint az ilyen hatásokról észszerű valószínűséggel nem beszélhetünk, azokat a fikció területére soroljuk. A megjósolhatatlanság azt jelöli, hogy ezen események mibenlétét, bekövetkezésük időtávját és valószínűségét nem tudjuk megítélni.

Összességében viszont ezeket az externáliákat azért tekinthetjük végzetesnek, mert bekövetkezésük az emberiség jelenleg ismert formájának biztos végét jelenti. A negatív hatások között olyan disztópikus jelenségeket szoktak megemlíteni, mint az emberiség létét fenyegető szuperintelligencia létrejötte, az emberekre támadó robotharcosok vagy a robotok által indított végzetes háborúk és az emberiség egészének permanens háborús állapotba taszítása. A pozitív, nem az emberiség kihalását jelentő hatások sem kevésbé festenek hátborzongató képet: itt ugyanis az emberi egyedek tökéletesítésére irányuló törekvéseket, a robotalapú örök életet vagy a mesterséges intelligenciával és robotikával támogatott szuperképességek kifejlesztését említik a szerzők. Azt láthatjuk, hogy sok esetben a pozitív és a negatív egzisztenciális hatások közötti határvonal bizonytalan és szubjektív megítélés függvénye lehet.

4.2. A helyi szintű externáliák szabályozási implikációi

Petit elmélete a várható hatások hármas csoportosításához kapcsolja a jogalkotói beavatkozás eszközeit is. A helyi hatások esetében a jogalkotónak az *ex post* problémamegoldásra hagyatkozást, a meglévő jogi infrastruktúra és a bíróságok szintjén történő vitarendezést javasolja.[127] Ezek a helyzetek így eldönthetők esetről esetre, egyedi mérlegelés alapján a tulajdonjog, a szerződési jog és a különböző kárfelelősségi alakzatok bekapcsolásával. Megítélése szerint ez azért elfogadható megközelítés, mert a külső hatások nem érintik az emberiség egészét egzisztenciálisan, illetve jelentékeny mértékben. Ez pedig nagyobb döntési szabadságot ad a jogalkotónak és a jogalkalmazóknak is az ilyen helyzetek megítélésére vonatkozó gyakorlat kialakításában. Ezzel az állásponttal az alább kifejtettek szerint nem értek teljességgel egyet. Az egyéni szintű hatások esetében, éppen a mesterséges intelligencia működési jellegzetességeiből fakadóan sok esetben a jelenlegi szabályozási keretek torz eredményhez vezetnek. Vitathatatlanul a leggyakrabban felmerülő kérdések itt a felelősségtan területére tartoznak. A mesterséges intelligencia által végrehajtott cselekményekért való felelősség az a terület, amelyet megítélésem szerint a jog jelenlegi keretrendszere nem képes kezelni.

A korábbi fejezetekben már volt szó arról, hogy a mesterséges intelligencia alapú algoritmusokat és más aktorokat (például robotok) miért szükséges önálló szabályozás alá vonni. Most tekintsük át azt is, hogy a polgári jogi, illetve büntetőjogi felelősség szempontjából milyen dilemmák adódnak a mesterséges intelligencia működése kapcsán. Tesszük mindezt a teória, a jogalkotás alapjául szolgáló jogpolitikai, rendszertani és jogelméleti megfontolások magas absztrakciós szintjén. Ezt a kicsit távolságtartó megközelítést azért választottam, mert a mesterséges intelligencia szabályozása valószínűleg el fog szakadni a tagállami jog szintjétől, sőt, nagy valószínűséggel az Európai Unió által megalkotott közös szabályok (például fogyasztóvédelmi szabályozás) szintjét is meghaladja, így a jogági alapfogalmaknál részletesebb, tételes jogi rendelkezésekre történő reflexió nem kívánt irányba viszi el a fejtegetést, és túlságosan determinálja a gondolkodásunkat.

[127] Petit 2017, 28–30.

4.2.1. Felelősségi kérdések

A felmerülő károk vagy büntetendő cselekményekért való felelősség esetében négy alapvető felelősségi helyzetet vizsgálhatunk meg. Az első eset az, ha a hibát egyértelműen kimutatható termékhiba okozta. Ide érthetjük a tervezési hibából fakadó, nem szándékos diszfunkciókat, a gyártás és más kivitelezési lépések során okozott termékhibát, valamint a gyártó általi üzembe helyezés hiányosságaiból fakadó hibákat. Az ilyen típusú hibákat továbbra is kezelhetjük a termékfelelősség és a gyártóra vonatkozó más felelősségi alakzatok, valamint a fogyasztóvédelmi szabályozás jelenleg is alkalmazott jogi rendszerének megfelelően, mivel ebből a szempontból nincsen különbség a mesterséges intelligenciát alkalmazó eszközök és bármely más termék között. Ezen alakzat alkalmazásához teljesülnie kell azon előfeltételeknek, hogy a hiba kizárólag a gyártási vagy beüzemelési folyamatra vezethető vissza, a hiba kialakulásában nem játszik szerepet a termék belső működéséből fakadó valamely tényező vagy jellegzetesség, és a gyártó részéről nem áll fenn a szándékos károkozásra utaló körülmény. Gyakran kizárólag szoftveres termékről lévén szó, ki kell emelnünk, hogy a programozás folyamatát is gyártásnak tekintjük, ebben az esetben a programhibák alkotják a termékhibákat.

A következő felelősségi alakzat az üzemben tartó általi szakszerűtlen működtetésre, a célnak nem megfelelő használatra, a termék jellegzetességeinek figyelmen kívül hagyására visszavezethető hibákat öleli fel, és az ezekből eredő károkat és az így bekövetkezett más jogsértéseket soroljuk ide. Az idetartozó esetekben az állapítható meg, hogy az üzemben tartó mulasztotta el a kellő gondosság tanúsítását vagy a szükséges biztonsági intézkedések megtételét (például nem tájékoztatta a felhasználóit az általa üzemeltetett szoftveres alkalmazás által végzett adatgyűjtéshez kapcsolódó adatvédelmi tudnivalókról), vagy más módon, a felügyelete alatt álló mesterséges intelligenciát nem működtette szakszerűen, illetve működésének szabályszerűségét nem ellenőrizte. Ebben a jogi helyzetben a polgári jog területén a veszélyes üzemi felelősség alakzatai alkalmazhatók, míg a büntetőjog területén a felelősség hagyományos koncepcionális keretei még kezelni képesek ezt a helyzetet. Ehhez arra van szükség, hogy a kárt okozó vagy jogsértő hiba egyértelműen az üzemeltetés vagy a felügyelet hiányosságaira legyen visszavezethető, a mesterséges intelligencia működése, annak hatásai, módja és lehetséges következményei kellő gondosság tanúsítása mellett az üzemben tartó számára átláthatók legyenek. A mesterséges intelligencia

ebben az esetben szabályosan működik, és a működése még lineárisnak tekinthető, vagyis előre látható, hogy egy meghatározott bemenet milyen irányú kimenetet eredményez. Ez nem a mesterséges intelligencia leginkább komplex formáit takarja, hanem a nagy és összetett számításokra képes, de nem öntanuló vagy önfejlesztő mechanizmusokat.

A hagyományos jogi keretek között kezelhető harmadik kérdéscsoport a szándékos károkozó magatartások köre. Ebben az esetben a gyártó vagy a gyártó irányítása alatt eljáró és felelősségi körébe tartozó személy vagy pedig az üzemben tartó magatartása kifejezetten arra irányul, hogy a mesterséges intelligencia alapú eszköz felhasználásával előre meghatározott vagy meg nem határozható személy(ek)nek kárt vagy más sérelmet okozzon vagy más jogsértést kövessen el, illetve az ilyen kár vagy jogsértés bekövetkeztének lehetőségét előre látja és abba belenyugszik. Ebben az esetben a polgári jog (szerződéses vagy szerződésen kívüli) kárfelelősségi alakzatai, valamint a büntetőjog eszközzel vagy más útján elkövetett bűncselekményekre vonatkozó felelősségi szabályai megfelelően alkalmazhatók, a mesterséges intelligencia felhasználása nem indokol különleges szabályt.

E felelősségi forma alkalmazása előfeltételezi, hogy az elkövető személy tisztában van vagy tisztában lehet cselekménye következményeivel, és a károkozó magatartás egyenes következménye az elkövető beavatkozásának, az nem a mesterséges intelligencia belső működésére vezethető vissza. Ez tehát tulajdonképpen egy ember által elkövetett szándékos vagy negligens károkozás vagy jogsértés, a felelősség megállapításában nincs szerepe annak, hogy milyen típusú eszközt használt fel az elkövető (egy kenyérpirító szándékos veszélyessé tétele is ugyanígy minősülne).

A felelősségi alakzatok utolsó csoportja az, ahol egyáltalán jelentőséget kap az a tény, hogy mesterséges intelligenciáról van szó. Ez az eset a mesterséges intelligencia által szabályos üzemelés során, megfelelő utasítások által meghatározott működési keretben, önállóan hozott döntésből fakadó kár vagy más jogsérelem. Ahhoz, hogy ezt az esetet vegyük elő, az szükséges, hogy magas fokú önállósággal, felügyelet nélküli döntéshozatali képességgel és önálló tanulás alapján való önfejlesztési funkcióval bíró, rendkívül fejlett és komplex mesterséges intelligenciáról beszéljünk.

Egy ilyen eszköz jellegzetessége, hogy a kitűzött cél elérése érdekében maga határozza meg a szükséges lépéseket, valamint folyamatosan fejleszti döntéshozatali mechanizmusát, feladata ellátása közben folyamatosan tanul és módosítja saját működését. Az ilyen mesterséges intelligencia önállóan hoz döntéseket is, konkrét működési és döntéshozatali mechanizmusa külső

szemlélő számára nem átlátható, és az önfejlesztő képességére tekintettel a beépített döntési és logikai rendszer a működésben eltelt bizonyos idő után már jelentős eltéréseket mutathat az eredetileg bevitt programtól. Erre tekintettel pedig a konkrét példány működése a gyártó számára sem megjósolható, mindösszesen előre felállított működési korlátok között tartható.

Az is előfordulhat, hogy az azonos algoritmus két, egyszerre működésbe hozott, de eltérő adatkörnyezetben dolgozó példánya bizonyos idő elteltével egymáshoz képest is jelentős eltérést mutathat, az egyedi fejlődési utat bejáró példányok két, önálló karakterrel („egyéniséggel") rendelkező mesterséges intelligenciává fejlődtek. A jogi kérdés ilyenkor ott vetődik fel, hogy a kizárólag a gépi, önálló döntéshozatalból fakadó károkért vagy jogellenes cselekményekért ki tehető felelőssé. *Stricto sensu* nem mutatható ki kauzalitás sem a gyártó, sem az üzemben tartó bármely magatartása vagy mulasztása és a bekövetkezett káresemény között, az esemény racionálisan gondolkodva a rendszer működési mechanizmusát ismerők számára sem volt előre látható vagy akár csak valószínűsíthető. A rendszer működésében az eltérést a saját belső működése, az adatok alapján való, felügyelet nélküli tanuláson alapuló önfejlesztése és az ennek alapján – szintén külső beavatkozás nélkül – a felügyeletet gyakorló személy számára nem átlátható okozati rendszerben hozott döntése eredményezte. Az üzemben tartó felelősségét megállapítani ahhoz lenne hasonlatos, mint a taxitársaságot felelőssé tenni az egyik gépjárművezető által okozott balesetért. A gyártó felelőssége pedig még ennél is távolabb mutat, talán ahhoz lenne hasonlatos, mintha a szülőt vonnánk felelősségre felnőtt gyermeke által elkövetett kihágásért. A gyártó irányelveket adhat a működéshez, szabályosan működő kiindulási döntéshozatali mechanizmust épít az algoritmusba, de egyúttal felruházza az önálló tanulás és döntéshozatal képességével is, ezért a fejlődési útjára és a „bűnös útra térésre" nem tud befolyással lenni.

Ez az az eset, amikor elkezdhetünk a szó szoros értelmében véve a mesterséges intelligencia szabályozásáról beszélni. Az ezelőtti esetek bármely termék vagy eszköz esetében ugyanígy működnének. Ezzel együtt viszont ez az a pillanat is, amikor a jog jelenlegi szabályrendszere megbicsaklik, és nem képes egyértelmű válaszokat adni a felmerült problémákra. Megítélésem szerint helytelen vagy legalábbis félrevezető az a kérdésfeltevés, hogy egy ilyen esetben a mesterséges intelligencia jogi megítélés szempontjából mihez áll közelebb: termék vagy állat? Arra biztatom az olvasót, hogy az eddig bemutatottak alapján próbálja meg magának feltenni ezt a kérdést, és a fent vázolt esetben rámutatni arra a személyre, aki a felelősséget viselje a mesterséges intelligencia által okozott kárért vagy jogsértésért.

4.2.2. Jogági felelősségi alakzatok alkalmazása: polgári jog és büntetőjog

Ebben az írásban nem célunk mélyreható dogmatikai és tételes jogi elemzést adni arról, hogy a polgári jog, illetve a büntetőjog rendszerében pontosan hogyan lenne elhelyezhető a mesterséges intelligencia által végrehajtott tettekért való felelősség. Mindazonáltal szeretnénk megmutatni néhány olyan kritikus pontot, amely arra irányítja rá a figyelmet, hogy a jelenlegi jogi keretek között miért nem értelmezhető a mesterséges intelligencia tetteinek felelősségvonzata.

Mind a polgári jogban, mind pedig a büntetőjogban a károkozás vagy a jogellenes cselekmény végső felelőssége egy embert terhel. A nehezebben megítélhető, többszereplős vagy felelősségre nem vonható személy (illetve veszélyes üzem, állat stb.) által okozott kár vagy jogellenes magatartás esetén a jogalkotó „addig keresgélt, amíg nem talált egy felelősségre vonható személyt". Ilyen a polgári jogban az üzemben tartó vagy a felügyeletet ellátó személy, a büntetőjogban pedig a közvetett tettes (az állat vagy eszköz útján elkövetett bűncselekmények esetében maga az elkövető). Közös vonás ezek esetében, hogy a tényleges károkozó valamilyen módon a felügyeletük alatt áll, és általában képesek arra, hogy a káros cselekményt megakadályozzák vagy elhárítsák.

Fent már szót ejtettünk arról, hogy a mesterséges intelligencia kapcsán milyen nehézségek adódnak a gép által önállóan meghozott döntések és az ennek alapján végrehajtott cselekmények esetében. Ilyenkor a mesterséges intelligencia önállósága az a tényező, amely elhatárolja a hagyományos felelősségi alakzatok alkalmazásától a mesterséges intelligencia alapú eszközök által önállóan végrehajtott cselekményekért való felelősséget.[128] A mesterséges intelligencia önállósága a döntések meghozatalának és azoknak a külvilágban, külső ellenőrzéstől vagy befolyástól függetlenül történő végrehajtásának képességeként határozható meg, ez az önállóság pedig tisztán technológiai jellegű, és a mértéke attól függ, hogy milyen kifinomultra tervezték a gép és a környezete közötti kölcsönhatásokat. Minél önállóbb a mesterséges intelligencia, annál kevésbé tekinthető egyszerű eszköznek más szereplők – például a gyártó, a kezelő, a tulajdonos, a felhasználó stb. – kezében.

[128] Calo 2015, 554–555.

Ez viszont felveti azt a kérdést, hogy a felelősséggel kapcsolatos rendes szabályok elegendőek-e, vagy olyan új elvekre és szabályokra van szükség, amelyek tisztázzák a különböző szereplők jogi felelősségét a gépek olyan cselekedeteivel és mulasztásaival kapcsolatban, amelyek nem vezethetők vissza konkrét emberre, valamint hogy el lehetett volna-e kerülni a mesterséges intelligencia olyan cselekményeit vagy mulasztásait, amelyek kárt okoztak. A jelenlegi jogi keretek között a mesterséges intelligencia vagy az azokat alkalmazó robotok önmagukban nem vonhatók felelősségre azokért a cselekedetekért vagy mulasztásokért, amelyek harmadik feleknek kárt okoznak. Így a jelenlegi felelősségi szabályok csak azokra az esetekre vonatkozhatnak, amikor a robot cselekedetének vagy mulasztásának oka visszavezethető egy konkrét emberi szereplőre, például a gyártóra, a működtetőre, a tulajdonosra vagy a felhasználóra, és amikor ez a szereplő előre láthatta vagy elkerülhette volna a gép káros magatartását.[129]

Ezenkívül a jelenlegi szabályok alapján csak a gyártók, a működtetők, a tulajdonosok vagy a felhasználók objektív felelősségét lehetne megállapítani a mesterséges intelligencia cselekedeteiért vagy mulasztásaiért. Az Európai Unió jogalkotója a robotokkal kapcsolatos polgári jogi szabályokról szóló jelentésében is felhívja a figyelmet arra, hogy ha előáll az az eset, amikor a robot önálló döntéseket tud hozni, a jelenleg létező hagyományos szabályok nem lesznek elégségesek a robot által okozott kárral kapcsolatos jogi felelősség beállásához, mivel nem tennék lehetővé a kártérítésért felelős fél azonosítását és e fél az okozott kár megtérítésére vonatkozó kötelezettségének érvényesítését. A jelentés hangsúlyozza, hogy új, hatékony és korszerűbb szabályokra van szükség, amelyek megfelelnek a műszaki fejlődésnek és a piacon használt legújabb innovációknak. A szerződésen kívüli felelősséget illetően a jelenlegi jogi keretek között a felelősség csak a mesterséges intelligenciát alkalmazó eszközök gyártási hibái miatt okozott kárra és csak azzal a feltétellel terjedhet, hogy a károsult bizonyítani tudja a tényleges kárt, a termék hibáját, illetve a kár és a hiba közötti ok-okozati összefüggést. A jelenlegi jogi keret így nem tudná lefedni a robotok új generációja által okozott károkat, amennyiben a robotok alapjául szolgáló mesterséges intelligencia olyan alkalmazkodási és tanulási készségekkel ruházható fel, amelyek bizonyos mértékben kiszámíthatatlanná teszik a magatartását, mivel a mesterséges intelligencia önállóan tanulna saját, változó tapasztalataiból, és egyedi, előre nem látható módon lépne kölcsönös kapcsolatba a környe-

[129] KLEIN–Tóth 2018, 113.

zetével.[130] A polgári jogi felelősség terén tehát az önállóan döntéseket hozó és cselekvő mesterséges intelligencia szétfeszíti a jelenlegi felelősségtani kereteket, és a jogalkotónak ad feladatot arra, hogy megtalálja a legjobb megoldást.

Amikor 1981-ben egy japán gyári munkás halálát okozta egy mellette dolgozó, mesterséges intelligencia által vezérelt robot, a mesterséges intelligencia, illetve a robotok tetteiért való büntetőjogi felelősség kérdései vetődtek fel. A 37 éves munkavállaló a robotot készült javítani, amikor véletlenül bekapcsolta a gépet. A környezetét felmérő mesterséges intelligencia úgy ítélte meg, hogy ottléte akadályozza a működését, ezért nagy teljesítményű robotkarjával bepréselte a munkást egy mellettük álló ipari fémvágó gépbe, amely megölte őt.[131] Ebben az esetben a robot teljesen önállóan cselekedett, tette azonban egy ember halálát okozta. Ha emberről volna szó, akkor szinte biztos, hogy az emberölés alapesete vagy gondatlanságból elkövetett alakzata lenne megállapítható. A büntetőjog területén a felelősség kérdésének taglalásakor a bűncselekmény és az elkövető fogalmi elemeiből, valamint a büntetés céljára vonatkozó elméletekből indulhatunk ki.

Előkérdésként azt kell eldöntenünk, hogy a cselekményért egy ember büntetőjogi felelőssége megállapítható-e vagy nem. A büntetőjog fogalmi rendszerében a bűncselekmény megállapításához szükséges egy cselekmény vagy mulasztás *(actus reus),* amely a társadalomra veszélyes és büntetendő, szükséges egy elkövető és az ő tudati viszonyulása a bűntetthez *(mens rea),* amely lehet szándékosság vagy gondatlanság. Néhány kivételes esetben objektív felelősség is megállapítható, amikor az elkövető tudattartalma nem releváns a felelősségre vonás szempontjából. Ebből kiindulva vizsgáljuk meg egy autonóm, önállóan tanuló és döntést hozó mesterséges intelligencia által elkövetett bűncselekmény esetében azt, hogy fennállhat-e emberi felelősség (a sértetti közrehatást nem ideértve).

Szigorúan véve már az első fogalmi elemnél kizárhatjuk az emberi felelősséget: ha a felügyeletet gyakorló vagy az üzemben tartó személy nem követett el olyan cselekedetet (például utasítást adott a robotnak) vagy mulasztást (ellenőrzés nélkül jóváhagyta a robot műveletét, amikor az engedélyt kért tőle), akkor nem beszélhetünk az embert terhelő bűnös cselekményről.[132]

---

[130] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)], AA-AI pont.

[131] Elérhető: www.theguardian.com/theguardian/2014/dec/09/robot-kills-factory-worker (A letöltés ideje: 2018. 10. 13.)

[132] Calo 2015, 554–556.

A tudati viszonyulás esetén azt kell megvizsgálnunk, hogy az ember láthatta-e előre a robot döntését vagy cselekvését, illetve számíthatott-e ilyenre. Amennyiben a mesterséges intelligencia valóban a fent írt módon önállóan gyűjtött adatokat a környezetéről, elemezte és értékelte azokat, és így hozott önálló döntést, majd ennek alapján cselekedett, akkor nem beszélhetünk az ember esetében bűnös szándékról sem. Ennek megfelelően összességében megállapíthatjuk, hogy a kizárólag önálló döntéssel a mesterséges intelligencia által elkövetett bűncselekmény esetében a büntetőjog klasszikus dogmatikai alapjain nem tudunk bűncselekményt megállapítani valamely személy terhére.

A következő kérdés az, hogy a büntetőjog dogmatikája megengedi-e a bűncselekmény megállapítását a mesterséges intelligencia mint elkövető vonatkozásában. A szakirodalomban Gabriel Hallevy állított fel egy részletes elméletet[133] a mesterséges intelligencia büntetőjogi felelősségének megállapításáról. Megítélése szerint a kellő fokú autonómiával rendelkező mesterséges intelligencia esetében nemcsak a bűnös cselekmény van meg, hanem megállapítható a szubjektív *mens rea* is, mivel ezek az algoritmusok képesek előre látni a tetteik következményeit. Az objektív felelősségi bűncselekmények és a súlyos gondatlanságból elkövetett bűncselekmények esetében az ennél alacsonyabb autonómiafokú robotok felelőssége is közvetlenül megállapítható. Elmélete szerint a mesterséges intelligencia ugyanúgy védekezhetne a büntethetőséget kizáró okok fennállásával is (tévedés, végszükség, jogos védelem), mint az emberek.[134]

Megítélésem szerint a mesterséges intelligencia jelenlegi és a belátható jövőre előrejelezhető fejlettségi szintje még rendkívül távol van a *tudatosság* szintjétől, ami pedig a büntetőjogi vétőképesség előfeltétele. Így jelen állapot szerint a mesterséges intelligencia és a robotok nem lehetnek bűncselekmények elkövetői, az továbbra is csak emberekre vonatkoztatható. Mindazonáltal nem tekinthetünk el attól a ténytől, hogy olyan cselekmény történt, amelyet emberi elkövető esetén bűncselekménynek tekintenénk, amire ezért a jognak valahogyan reagálnia kell. Az esemény kapcsán megindított büntetőeljárásban annak megállapítása után, hogy embert sem szándékosság, sem mulasztás, sem objektív felelősség nem terhelt, szükségszerűen arra a következtetésre kell jutni, hogy elkövető hiányában a cselekmény nem minősíthető bűncselekménynek.

[133] HALLEVY 2013.
[134] HALLEVY 2010, 175–180.

Mindazonáltal érdemes a jogalkotónak elgondolkodnia azon, hogy az így záruló eljárásban nem kerülhet-e sor jogkövetkezmények vagy kényszerintézkedések alkalmazására. A büntetések egyik célja – a speciális prevenció – ugyanis még megvalósítható. A büntetőeljárásban gondoskodni kell arról, hogy a konkrét, érintett mesterséges intelligencia újabb bűncselekményt ne tudjon elkövetni, ezáltal védve a társadalmat. A büntetőeljárásban a robot kizárólagos felelősségének megállapítása esetén is a megelőzéshez szükséges adminisztratív és műszaki intézkedéseket el kell rendelni: ez jelentheti a konkrét mesterséges intelligencia működésének megszüntetését (törlését vagy a robot leszerelését), újraprogramozását, a gyári beállítások visszaállítását vagy a törlés utáni kontrollált újratanítást.

Ne felejtsük el, hogy a gépi tanulásra alkalmas mesterséges intelligencia minden eseményt, így a bűncselekményt is a tanulási folyamat részeként értékeli, és abból következtetést vonhat le a jövőbeni viselkedésére vonatkozóan. Ha pedig a robot úgy ítéli meg, hogy az emberekre káros cselekmények vagy más jogsértések megfelelő megoldási utat jelentenek bizonyos helyzetekben, akkor hasonló körülmények között később is ezt az utat fogja választani, ami pedig egyértelműen nemkívánatos következményekhez vezet.

Ahogy azt tehát láthatjuk, a büntetőjogi felelősség, illetve a polgári jog általános felelősségi alakzatai nem vagy alig értelmezhetők a mesterséges intelligencia autonóm döntése esetében. Jelenleg a legkomolyabb károkkal fenyegető esetekben a jogalkotó egy sajátos áthidaló megoldással élt, és talált olyan emberi felelőst, akivel szemben érvényesíteni lehet a polgári jogi és esetleg a büntetőjogi jogkövetkezményeket. Ezt tekintjük át a következőkben.

4.2.3. Az ember felügyeleti (mögöttes) felelősségének konstrukciója

A jogalkotó jelen állás szerint a legtöbb esetben úgy látja biztosítottnak a mesterséges intelligencia működésének biztonságossá tételét, hogy előírja a gép által végzett automatikus döntéshozatal esetében az emberi beavatkozás lehetőségét, illetve annak bizonyos esetekben való kötelezővé tételét. Ezt figyelhetjük meg az önvezető autók esetében, ahol az önvezető autók működését megengedő államok jogalkotója minden esetben előírja a járművezető folyamatos kapcsolatát a gépjármű kezelőszerveivel (így a piacon elérhető önvezető gépkocsik érzékelik, hogy a vezető keze a kormányon van-e, és a kéz tartós elvételekor figyelmeztetik a vezetőt), ami a vezető beavatkozási képességét hivatott biztosítani. Ezzel kívánja a jogalkotó biztosítani azt, hogy

az önvezető autó meghibásodása vagy téves döntése esetén a vezető be tudjon avatkozni és el tudja kerülni a veszélyhelyzetet.

Az elmúlt évek önvezető autós balesetei szinte kivétel nélkül arra voltak visszavezethetők, hogy a járművezető nem figyelte a vezetési szituációt, és nem avatkozott be időben (vagy egyáltalán) a kialakuló veszélyhelyzetbe. Erre jó példa az első halálos kimenetelű önvezető autós baleset, ahol az autópályára kiforduló kamiont annak fehér színű oldala és a felépítmény úttól való nagy magassága miatt nem észlelte az önvezető jármű, és fékezés nélkül nekihajtott.[135] Hasonló balesetet szenvedett egy önvezető módban haladó Tesla, amelyik nem észlelte az előtte a piros lámpánál álló tűzoltóautót, és közel 100 km/órás sebességgel, fékezés nélkül nekihajtott. A balesetben szerencsére csak kisebb sérüléseket szenvedtek az utasok.[136] Mindkét baleset esetében a járművet vezető személy másra figyelt, nem az utat nézte, illetve nem fogta a kormánykereket, így nem tudott beavatkozni.

Az első önvezető autós gyalogosgázolás esetében is megállapította a hatósági vizsgálat, hogy (a fedélzeti kamera képével igazolhatóan) az Uber önvezető módban haladó gépjárművének vezetője nem figyelt a vezetésre, és éppen oldalra nézett, amikor a baleset történt.[137] Mindegyik baleset azt mutatja, hogy a járművezetők nem tettek eleget a jogszabályok által előírt (és az autók használati utasításában nyomatékosított) folyamatos készenléti és monitorozási kötelezettségüknek a baleset idején.

Hasonlóképpen az Európai Unió adatvédelmi rendelete, a GDPR előírja, hogy az egyedi ügyekben alkalmazott automatizált döntéshozatal esetén, amennyiben az adatkezelés jogalapja az érintett és az adatkezelő közötti szerződés vagy az érintett hozzájárulása, az adatkezelő köteles biztosítani az érintettnek azt a jogot, hogy az adatkezelő részéről emberi beavatkozást kérjen.[138] Ebben az esetben a jogalkotó az olyan szituációkat kívánta megelőzni, ahol az érintett számára jogi hatással járó automatizált döntéshozatal az adatok vagy a mechanizmus hibájából olyan következtetésekre jut, amelyek az érintettre hátrányos következményekkel járnak, illetve a természetes személyek közötti hátrányos megkülönböztetést eredményeznek faji vagy etnikai származás, politikai vélemény, vallási vagy világnézeti meggyő-

[135] Beszámoló például itt: www.theregister.co.uk/2017/06/20/tesla_death_crash_accident_report_ntsb/ (A letöltés ideje: 2018. 09. 16.)

[136] Elérhető: www.wired.com/story/tesla-autopilot-why-crash-radar/ (A letöltés ideje: 2018. 09. 16.)

[137] Elérhető: www.bloomberg.com/news/articles/2018-05-24/uber-self-driving-system-saw-pedestrian-killed-but-didn-t-stop (A letöltés ideje: 2018. 09. 16.)

[138] GDPR 22. cikk (3) bekezdés.

ződés, szakszervezeti tagság, genetikai vagy egészségi állapot, szexuális irányultság vagy nemi identitás alapján, illetve amelyek ilyen hatást kiváltó intézkedésekhez vezetnek.[139] Ilyen automatizált döntéshozatal tárgya lehet pénzügyi döntés, árképzés, egészségügyi kockázatok felmérése vagy munkahelyi teljesítmény kiértékelése. A GDPR nem tartalmaz azonban a fent idézettnél pontosabb meghatározást arra nézve, hogy mi lehet egy ilyen emberi beavatkozás tartalma.

Az emberi tényező bevonása a döntéshozatalba implicit módon magában foglalja a gép által kiadott eredmény megismerését és esetleg annak összevetését egy létező döntéshozatali politikával, illetőleg a jogszabályi előírásokkal. Ez utóbbi esetben az emberi beavatkozó azt vizsgálja, hogy a gép által hozott döntés nem eredményezett-e diszkriminációt vagy más módon jogszabályba ütköző következtetést. A GDPR preambulumszövegében foglaltakból arra is következtethetünk, hogy a nem megengedhető hátrányos megkülönböztetések vagy más jogszabálysértés esetén a beavatkozó megsemmisítheti vagy megváltoztathatja a gép által adott eredményt. Nyitva hagyja mindazonáltal a fenti jogszabályszöveg annak kérdését, hogy mi a kötelezettsége az emberi beavatkozónak abban az esetben, ha az érintett vitatja a döntést, de az nem jogsértő vagy diszkriminatív.[140]

Ilyen esetekben az adatkezelő köteles végigkövetni a döntéshozatali folyamatot, ellenőrizni minden lépést és a felhasznált adatok helyességét, vagy pedig elegendő csak a végeredményt megvizsgálni? A GDPR értelmezéseként kiadott állásfoglalásában az úgynevezett „29. cikk munkacsoport" (WP 29) azt fejti ki, hogy az emberi beavatkozás akkor tekinthető érdeminek, ha azt olyan személy végzi, aki jogosult a döntést megváltoztatni, és amely beavatkozás során minden rendelkezésre álló input és output adatot megvizsgálnak.[141]

Ha a jogszabályi környezettől távolabb lépünk, és a gyakorlatot vizsgáljuk meg, szintén azt látjuk, hogy az emberi beavatkozás biztosítja a kontrollt az automatizmus által hozott döntések felett. Ehhez elég, ha csak az internet ismertebb felületeit, így különösen a közösségi oldalakat megnézzük. A legtöbb helyen külön felület van az emberi beavatkozás vagy felülvizsgálat kérésére, a nem kívánt tartalmakat vagy eredményeket pedig könnyen elérhető me-

[139] GDPR (71) preambulumbekezdés.

[140] Goodman–Flaxman 2017, 6.

[141] Article 29 Working Party Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. WP 251, 9–10.

nükből kapcsolhatóan jelezhetjük, bevonva ezzel az emberi beavatkozót az eseménybe. Az ilyen típusú intervenciók persze nem kizárólag a mesterséges intelligencia döntéseinek kijavítását szolgálják, sokszor inkább a felhasználók által közzétett és a beépített mesterséges intelligencia által ki nem szűrt jogsértő tartalmak utóellenőrzését és törlését biztosítják.

Ez utóbbi értelemben közvetve mégis az automatikus döntéshozatal feletti egyfajta emberi kontrollnak tekinthető, csak itt a meghozott döntés a tartalom jogszerűségének megállapításáról szól. Ilyen módon lehet jelezni például a videómegosztó oldalon a szerzői jogsértés miatt automatikusan eltávolított tartalmak esetén azokat a helyzeteket, amikor a tartalom közzétevőjének mégis volt joga a jogvédett tartalom közzétételére. Ilyenkor az emberi beavatkozó manuálisan megvizsgálhatja az esetet, és a közzétevő által leírtak vagy megküldött bizonyítékok alapján dönthet úgy, hogy a közzétételt jóváhagyja, és nem minősíti jogsértőnek.

A fenti esetekben az emberi beavatkozás előírásának jogalkotói célja az, hogy megvédje az egyént a mesterséges intelligencia által hozott hibás vagy jogsértő döntések következményeitől. Ezzel egyidejűleg a jogalkotó szándéka kiterjed arra is, hogy megakadályozza a mesterséges intelligencia „szabadon garázdálkodását", esetleg hogy elszabadulva hibás vagy jogellenes döntések sorozatát hozza meg. Az emberi beavatkozó így kontrollszerepet tölt be, kordában tartja a gépi döntéshozót, és érvényesíti azokat az emberi értékeket (tisztesség, egyenlőség, fair eljárás, méltányosság), amelyek a mesterséges intelligencia racionális döntéshozatali mechanizmusából kimaradtak.

4.2.4. Az ember felügyeleti (mögöttes) felelősségének kritikája

A nyilvánvaló jó szándék és az ember középpontba való visszahelyezésének elismerése mellett azonban érdemes az emberi beavatkozás mint a mesterséges intelligencia szabályozásának „Szent Grálja" jogalkotói szemléletmódját kritikával kezelni. Az alábbiakban néhány megjegyzést fűzök az emberi beavatkozást csodaszernek tekintő jogalkotói szemléletmódhoz, kiemelve az emberi tényező bevonásának korlátait.

Az első korlát, amely a szemünkbe ötlik, az az emberi beavatkozás utólagos jellege. A humán faktort képviselő beavatkozó szükségszerűen csak akkor szembesül a mesterséges intelligencia által hozott döntés hibás, veszélyes vagy jogellenes voltával, miután a döntés már megszületett, és az eredménye láthatóvá vált, esetleg nyilvánosságra is került. Ez sok esetben azt is jelenti,

hogy a jogsértés már bekövetkezett, mire a beavatkozó bekapcsolódott a folyamatba, így számára már csak a károk mérséklése, a jogviták megelőzése vagy – ha lehetséges – a jogszerű állapot helyreállítása marad.

Beláthatjuk tehát, hogy ezzel a jogalkotó sok esetben nem képes megvédeni az embereket a hibás gépi döntés közvetlen hatásaitól, legfeljebb csak az időben távolabbi következményeit tudja elhárítani. Mondhatjuk azt is, hogy az emberi beavatkozás ilyen előírása nem az érintetteket óvja meg, hanem a mesterséges intelligenciát alkalmazó szervezeteket védi a jogkövetkezményektől: a bekövetkezett hátrány utólagos, emberi orvoslása a sértettek által indított perektől védi meg a cégeket, akik így fel tudják mutatni egy esetleges eljárásban azt, hogy a hibáról való tudomásszerzést követően azonnal intézkedtek a hátrányok mérséklése érdekében. Mindeközben a hátrányt szenvedett fél kénytelen együtt élni a hátrányos döntés hatásaival, bízva abban, hogy utólagosan orvosolható az őt ért sérelem.

Nem lehet eltekinteni attól a ténytől sem, hogy a felülvizsgálatot végző személy egyéni percepciói és gondolkodása markánsan befolyásolhatja a felülvizsgálat eredményét. A jogalkotó szándéka szerint a mesterséges intelligencia döntését felülvizsgáló személy mérlegeli a tényeket, kitér minden input és output faktorra, ennek alapján kialakítja magában a személyes meggyőződését, és újbóli döntést hoz. A gyakorlatban mindazonáltal bebizonyosodott, hogy az emberek szeretnek hinni a gépnek, és sokszor a saját meggyőződésük ellenében is azt fogadják el igaznak, amit az automatizmus eredményként közöl. A tudományosan is alátámasztott jelenség az automatizációs elfogultság *(automation bias)* nevet viseli. Egy viselkedéstani kutatásban[142] azt vizsgálták, hogy egy repülőgép-szimulátoron hogyan hoznak döntést a pilóták abban az esetben, ha számítógépes döntéstámogató rendszer működik a gépen, és akkor, ha kizárólag a saját észlelésükre hagyatkozhattak. A kutatás kimutatta, hogy a számítógéppel nem segített pilóták jobb teljesítményt nyújtottak a döntéstámogató rendszert használó társaiknál, akik hajlamosabbak voltak figyelmen kívül hagyni olyan fontos jelzéseket, amelyre a gép külön nem figyelmeztette őket, valamint a józan észnek és a saját kiképzésükben tanultaknak is ellentmondó döntést hozni akkor, ha a gép erre utasította őket. Ugyanezt a jelenséget figyelhetjük meg akkor, amikor a szövegszerkesztő automatikus javítási funkciójára hagyatkozva egyre több nyelvhelyességi hibát hagyunk a szövegben, mert azokat nem jelezte a gép, vagy amikor a GPS utasításait követve a vezető belehajt egy

[142] Skitka–Mosier–Burdick 1999, 991–1006.

tóba,[143] holott jól látta, hogy arra nem vezet út. A gépi döntés iránti elfogultság megkérdőjelezi az automatizált döntéshozatal emberi felülvizsgálatának ténylegességét, mivel a fent bemutatottak alapján láthatjuk, hogy hajlamosak vagyunk egyetérteni a géppel még akkor is, ha felismerhetnénk annak hibás voltát. Így tehát kérdésessé válik, hogy az emberi beavatkozás valós korrekciós mechanizmusként szolgálhat-e az esetek nagy többségében.

Végezetül pedig problémásnak tekinthető az érdemi emberi beavatkozás elvárása a mesterséges intelligencia által végzett tevékenységek kontrolljaként abból az okból, hogy az emberi beavatkozó az esetek nagy többségében nem képes átlátni az éppen zajló folyamatot, illetve a vizsgált eredményhez vezető processzust. Különösen így van ez az olyan esetekben, ahol a folyamat komplex számítási algoritmusokon alapul, gépi tanulást alkalmaz vagy nagy adatmennyiség feldolgozása *(big data)* képezi az alapját.

Az ilyen típusú folyamatok a külső szemlélő számára általában átláthatatlanok vagy legalábbis homályosak, így a beavatkozó ember nem feltétlenül van tisztában az alkalmazott algoritmus funkcióival, illetve nem láthatja át a felhasznált adatok teljes körét. Ne felejtsük el, hogy a mesterséges intelligencián alapuló mechanizmusokat éppen azért hoztuk létre, hogy olyan rendkívül összetett, illetve olyan nagy adatmennyiségen alapuló számításokat végezzen el, amelyre az ember nem vagy csak tekintélyes idő- és erőforrás-felhasználás útján lenne képes. A gépi tanulás, különösen annak nem felügyelt *(unsupervised)* formája tovább bonyolítja a helyzetet azzal, hogy az algoritmus önfejlesztő mechanizmusának eredményeként létrejött új funkciók, adatkapcsolatok és következtetések egyáltalán nem átláthatók a külső szemlélő számára, így azt sem tudhatja a beavatkozó, hogy az algoritmus a beavatkozás időpontjában ugyanúgy működik-e, mint az előző nap működött, és ha módosult, akkor miben állt ez a változás. Ahhoz, hogy érdeminek minősíthető emberi beavatkozásról beszélhessünk, az szükséges, hogy a felülvizsgálatot végző személy meg tudja állapítani, hogy a meghozott döntés, az ahhoz vezető eljárás, illetve a létrehozott döntéstámogató profil pontos, tisztességes és nem diszkriminatív. Ehhez viszont arra van szükség, hogy az ellenőrzést végző személy kellő műszaki jártassággal rendelkezzen az automatizált döntéshozatali rendszerek működését tekintve, átlássa, hogy

[143] Elérhető: http://nymag.com/selectall/2018/01/waze-app-directs-driver-to-drive-car-into-lake-champlain.html; www.independent.co.uk/news/world/americas/woman-following-sat-nav-canada-drives-straight-into-lake-huron-ontario-a7029131.html (A letöltés ideje: 2018. 09. 27.)

a profilalkotás és a mesterséges intelligencia által támogatott döntéshozatal milyen és hányféle módon vezethet tisztességtelen, pontatlan vagy diszkriminatív eredményre. Ez viszonylag magas szintű társadalomtudományi, jogi és számítástudományi jártasságot feltételez.

Ezenkívül pedig arra is szükség van, hogy az alkalmazott rendszer megfelelően értelmezhető és átlátható, működése megmagyarázható legyen. Ezek hiányában azzal a helyzettel szembesülhetünk, hogy – különösen gépi tanulást vagy adatbányászatot magában foglaló folyamatok esetén – a mesterséges intelligencia által hozott döntések felülvizsgálatára és kijavítására hivatott személy nem érti, hogy mit lát, mi és miért történik a felügyelt algoritmusban. Ez pedig az emberi beavatkozást pusztán formális, érdemi felülvizsgálatot vagy korrekciót nem eredményező látszattevékenységgé fokozza le.

A folyamatos, valós idejű emberi szupervíziót mesterséges intelligenciára alapozott folyamatok esetében teljességgel illuzórikusnak tekinthetjük. Egyfelől egy nem időkötött, cél- vagy eredményorientált alkalmazásnál a jelenleg elérhető hardveres eszközök és szoftveres megoldások már olyan számítási és döntéshozatali sebességet értek el, ami az ember számára felfoghatatlan.[144] Így a valós idejű beavatkozás, de akár a minimális késedelemmel történő felügyeleti lépés is lehetetlenné válik. Ezt a problémát jól illusztrálják a tőzsdén tevékenykedő mesterséges intelligencia által elkövetett hibák következményei. 2012-ben a Knight Capital nevű cég a New York-i tőzsde elektronikus kereskedelmi platformján elindította az új kereskedő algoritmusát. Ez olyan kereskedő alkalmazás volt, amely nagy sebességű kereskedelemre *(high frequency trading)* volt beállítva. Ez azt jelenti, hogy egy másodperc alatt tranzakciók tízezreit is képes volt elvégezni. Az alkalmazásba valami hiba csúszott, és a tőzsdei logika szabályainak ellentmondva elkezdett magas áron venni és alacsony áron eladni. A cég tranzakciónként 10–15 dollárt vesztett, a sebességből kifolyólag azonban ez percenként 10 millió dolláros veszteséget jelentett. A hibaüzenetekre és a szokatlan tőzsdei mozgásokra reagálva a cég 45 perc elteltével kapcsolta le az alkalmazást, addig összesen 440 millió dolláros veszteséget szenvedett el.[145] Az alkalmazás viszonylag egyszerű hibáját a valós idejű megfigyelők az észszerűen értelmezhető emberi reakcióidőn és döntéshozatali időn belül észlelték és reagáltak rá, azonban

[144] Például jelen kézirat lezárásának időpontjában a Google másodpercenként 69 600 keresési kérést dolgozott fel. Elérhető: www.internetlivestats.com/one-second/#google-band (A letöltés ideje: 2018. 09. 26.)

[145] Elérhető: www.bbc.com/news/magazine-19214294 (A letöltés ideje: 2018. 09. 19.)

a hihetetlen működési sebesség miatt így is óriási kárt okozott a hibás algoritmus a cégnek.

Az időtényező tárgyalásánál pillantsunk vissza a fejezet elején bemutatott autonóm járműves balesetekre. Tegyük fel magunknak a kérdést, hogy a piros lámpánál várakozó tűzoltóautónak csapódó jármű vezetőjének mit kellett volna tennie ahhoz, hogy elkerülje a karambolt. Természetesen fékeznie kellett volna, ezt könnyű kijelenteni. Mindazonáltal, ha utánaszámolunk, ennél érdekesebb következtetésre juthatunk. Az esetben szereplő 100 km-es óránkénti sebességnél a reakcióidőt és a teljes lefékezéshez szükséges távolságot együtt számolva közel 80 méterre[146] lett volna szükség az akadály észlelésétől számítva a megállásig ahhoz, hogy elkerülje az ütközést. A lefékezéshez szükséges idő a reakcióidőt és a jármű műszaki tulajdonságait is figyelembe véve nagyjából 4,5 másodpercre tehető. Ennyire van tehát szüksége az emberi beavatkozónak ahhoz, hogy megállítsa az autót az ütközés előtt. Ehhez azonban hozzá kell tenni azt az időtartamot, amely alatt a vezető felismeri, hogy az önvezető automatika meghibásodott, és be kell avatkoznia, eldönti, hogy mit kell tennie, és nekikezd a végrehajtásnak.

A hibát nem előzi meg hibajelzés vagy más furcsa viselkedés, és a közlekedési helyzet is annyira egyszerű, hogy a járművezető nem fog rögtön gyanút, hogy valami nem működik, mert nem is feltételezi azt, hogy a saját sávjában éppen előtte álló hatalmas járművet az autó szoftvere nem ismeri fel. Akkor kezdhet csak el gyanakodni, amikor az autó nem kezdi el a fékezést olyan távolságban, hogy kényelmesen meg tudjon állni. Ezt követően még eltelhet egy kis idő, amíg realizálódik a vezetőben a felismerés, hogy az autó *egyáltalán nem* szándékozik megállni. A felismerésre, helyzetértékelésre és döntésre a vezetőnek annyi ideje van, amíg az átlagos fékezési időben nem lassító jármű el nem éri a vészfékezési távolságot. Ez voltaképpen azt jelenti, hogy egy országúti sebességgel haladó jármű vezetőjének az autó automatikájával szinte egy időben kéne végrehajtania minden műveletet, arra számítva, hogy az önvezető mechanizmus esetleg nem reagál. Formálisan persze ez a prudens viselkedés, és az elérhető jogszabályok is ezt írják elő. Ha azonban belegondolunk a tényleges folyamatba, azt látjuk, hogy így a járművezető még nagyobb pszichés terhelésnek van kitéve, mintha egy hagyományos autót vezetne teljesen manuálisan: nemcsak a forgalmi helyzetet kell folyamatosan figyelnie és megfelelő időben megfelelően reagálnia (vagyis autót vezetnie), hanem ezenfelül monitoroznia

[146] Elérhető: www.rac.co.uk/drive/advice/learning-to-drive/stopping-distances/ (A letöltés ideje: 2018. 09. 28.)

kell egy általa nem ismert módon működő komplex mechanizmust is, keresve a hibát a működésében. Ezzel kvázi megduplázzuk a járművezető feladatait, aki így már akkor is jobban járna, ha maga vezetné az autót.

Láthatjuk tehát, hogy az érdemi emberi beavatkozásnak számos morális, szociológiai, lélektani és nem utolsósorban technológiai korlátja van. A mesterséges intelligenciát elsődlegesen azért hoztuk létre és azért használjuk rendszeresen, hogy olyan feladatokat oldjon meg, amely összetettsége, számítási igénye vagy a felhasznált adatok nagy mennyisége miatt az emberek számára nem vagy csak aránytalanul nagy erőforrás-felhasználással oldható meg. A mesterséges intelligenciát megtanítottuk sok terabájt adatban olyan összefüggéseket és mintázatokat keresni, amelyet az emberi elme nem tudna felismerni. Olyan mesterséges intelligencia alapú alkalmazásokat használunk nap mint nap, amelyek a másodperc törtrésze alatt tudnak felismerni egy helyzetet, adekvát döntést hozni, és azt végre is hajtani (lásd a korábban említett nagy sebességű tőzsdei algoritmusokat, amelyek a kereskedés apró rezdüléseit figyelik, és kihasználják a csak néhány pillanatra nyitva álló előnyös lehetőségeket is).

Szintén mesterséges intelligenciát találunk a nagy mennyiségű, gyors döntést igénylő olyan repetitív feladatok esetén is, mint az internetes keresők, a közösségi oldalak vagy a videómegosztók működtetése. Ezek mind olyan feladatok, amelyek valamilyen jellegzetességüknél fogva az ember számára nem vagy nem ilyen sebességgel oldhatók meg. Ebben az esetben érdemi, folyamatos, hiba vagy jogsértés esetén beavatkozásra képes emberi felügyeletet elvárni nemcsak életszerűtlen, de lehetetlen is. A mesterséges intelligencia által hozott döntések felülvizsgálatára az ember – azok mennyisége, sebessége, összetettsége és a felhasznált adatmennyiség miatt – nem képes. Az utólagos ellenőrzés is reménytelen, még akkor is hosszabb időt vesz igénybe, ha a felhasználók által kifejezetten jelzett hibákra koncentrálunk.

A felhasználó vagy az üzemben tartó felelősségének kimondásával a mesterséges intelligencia szabályozása a veszélyes üzemi felelősség felé közelít, ami azonban több okból sem kívánatos. Jogalkotói, jogpolitikai szempontból helytelen volna a veszélyes üzemi felelősség irányába eltéríteni a mesterséges intelligencia által okozott károkért való felelősséget a technológia térnyerése és jövőbeli fejlődési tendenciái miatt. Ahogy jelenleg látjuk a műszaki fejlődés és a társadalmi (felhasználói) viselkedés trendjeit, a mesterséges intelligencián alapuló alkalmazások egyre inkább elterjedtek lesznek, és egyre jobban behálózzák az életünk minden területét. Célszerűtlen és jogpolitikailag is megkérdőjelezhető döntés volna egy ennyire mindennapos jelenséget felelősség szempontjából egy rendkívüli alakzat keretei között

kezelni, így a kivételt téve kvázi főszabállyá, de legalábbis a leggyakrabban előforduló alakzattá. A felelősségi kérdések jövőbe mutató tisztázásával a jogalkotónak lehetősége lenne megszabni a fejlődési irányt mind a fejlesztések, mind a felhasználói viselkedés terén. Ne felejtsük el, hogy a jogalkotónak figyelembe kell vennie társadalmi és gazdasági szempontokat is: a szabályozási környezetnek támogatnia kell a biztonságos fejlesztéseket, ösztönöznie az innovációt, mert ezáltal helyzeti előnyre tud szert tenni a világgazdaság porondján, ami a lakosság és összességében a társadalom javát is szolgálja.

Az ember mint végső felelős beiktatásával a jogalkotó a könnyebb utat választja, és a kisebb ellenállás irányába megy. Ahogy azt fent már bemutattuk, a technika jelen állása szerint már olyan fejlettségi szintet értek el a mesterséges intelligencián alapuló algoritmusok, hogy tevékenységük komplexitásából és döntéshozatali sebességükből fakadóan érdemi emberi felügyelet folyamatosan nem biztosítható. Így tehát azzal, hogy a jogalkotó elvárja és előírja az emberi felügyeletet, és a bekövetkezett káresemény kapcsán kimondja a felügyeletet gyakorló személy(ek) felelősségét, voltaképpen egy könnyen elérhető bűnbakot keres, aki „elviszi a balhét" a gép helyett. Ne legyen kétségünk afelől ugyanis, hogy egy fejlett, összetett, öntanuló mesterséges intelligencia tényleges döntéseire a károk elhárításához szükséges mértékben nem lehetséges az emberi ráhatás. Az algoritmus működését a külső szemlélő nem is látja át, arra csak a külső jelekből, vagyis a tevékenység eredményéből tud következtetni, a működési vagy következtetési hibákról is csak utólag értesül (a Knight Capital sem bukott volna akkorát, ha előre tudja, hogy az algoritmus fordított logikával fog kereskedni). Így a felhasználót vagy a felügyeletet gyakorló személyt utólag olyasmiért hibáztatjuk, amit az esetek nagy többségében nem tudott volna elhárítani. Ez egyszerűvé teszi a jogalkotó dolgát, mert nem kell kilépnie a régóta megszokott gondolkodási sémákból, és így a jogalkalmazóknak vagy bíróságoknak sem kell a bonyolult algoritmusok hibáit keresni, hanem elég azt megvizsgálni, hogy a felügyeletet gyakorló személy tette-e a dolgát vagy sem.

Végső soron pedig ez a gondolkodásmód a techóriásoknak kedvez, mert ha nem jogsértő üzleti döntésről vagy szándékos magatartásról van szó (mint a sokat emlegetett Cambridge Analytica ügyében[147]), akkor nincs igazán fél-

[147] A közösségi oldal több millió felhasználójának adataiból épített részletes személyiségprofilokat tartalmazó adatbázist egy külső, de a Facebook céggel együttműködő, az adatokat a cég üzletszabályzatának megfelelően megszerző szervezet. A cégcsoportot azzal vádolták meg, hogy az így épült adatbázis segítségével manipulálták az Egyesült Államok választópolgárait az elnökválasztási folyamatban. Elérhető: www.wired.com/story/cambridge-analytica-50m-facebook-users-data/ (A letöltés ideje: 2020. 10. 12.)

nivalójuk. A felhasználói vagy felügyelői felelősség megállapításával a cég kibújhat a közvetlen felelősség alól, és az alkalmazás továbbfejlesztése vagy a hibás döntést hozó termék visszavonása a saját üzleti döntése lesz, amivel adott esetben még pozitív reakciókat is kaphat a felhasználóktól. Ezzel a jogalkotó nem presszionálja abba az irányba a mesterségesintelligencia-fejlesztőket, hogy javítsák, teszteljék és tegyék biztonságossá a terméküket a piacra dobás előtt. A nagy techcégek hamar magukévá tették a „bocsánatot kérj, ne engedélyt" filozófiát,[148] és a fejlesztéseik bevezetése során előfordul, hogy egy felmerülő jogi vagy erkölcsi aggályra csak akkor reagálnak, amikor azt a felhasználók nagy számban jelezték, és globális felzúdulás alakult ki miatta. A gyors piacra dobással előnyre tehetnek szert a kíméletlenül zajló fejlesztési versenyben, azonban ez azzal a kockázattal jár, hogy a kikerült termék esetleg működési hibát tartalmaz, vagy nélkülözi azokat a hibaelhárító mechanizmusokat, amelyek megelőznék a jogellenes vagy károkozó működést.

Az utóbbi időszakban a mesterséges intelligencia biztonságossá tétele az EU és az USA jogalkotói figyelmét is felkeltette,[149] és mindkét esetben arra jutottak, hogy célszerű lenne már az alkalmazások fejlesztésekor beépíteni az alapvető jogi és erkölcsi normáknak való megfelelés kötelezettségét. A módszert azonban nem találták meg arra, hogy ez miképpen valósítható meg: az EU a fejlesztésben részt vevők számára írna elő képzési és továbbképzési kötelezettséget, míg az USA a fejlesztőcégek önszabályozásában látja a megoldást. Megítélésem szerint azonban a fejlesztőcégek mindaddig nem tekintik ezt prioritásnak, amíg az algoritmus hibás működéséért a felhasználó vagy a felügyeletet gyakorló személy, végső soron az üzemeltető viseli a felelősséget, amelynek megállapításához hosszas peres eljárásban vezet az út. Másfelől pedig az emberi felügyelet előírása azért is a globális nagyvállalatok malmára hajtja a vizet, mert ők még inkább megengedhetik maguknak egy hadseregnyi moderátor vagy felügyelő alkalmazását, míg a kisebb cégeket ez könnyen kiszoríthatja a piacról.[150]

---

[148] Elérhető: www.cbsnews.com/news/google-struggles-with-its-do-first-ask-forgiveness-later-strategy/ (A letöltés ideje: 2018. 09. 27.)

[149] Cath et al. 2017, 4–6.

[150] Elérhető: www.wsj.com/articles/how-europes-new-privacy-rules-favor-google-and-facebook-1524536324 (A letöltés ideje: 2018. 09. 29.)

4.2.5. A kockázatközösségi modell

Ahogy megvizsgáltuk a polgári jogi és a büntetőjogi felelősségi kérdések dogmatikáját, illetve az emberre kényszerített mögöttes felelősség koncepcióját, összességében azt látjuk, hogy ezek a megoldások csak jelentős kompromisszumok és erőszaktétel árán húzhatók rá az autonóm mesterséges intelligencia által okozott károkra vagy jogellenes cselekményekre. A jog jelenleg csak az ember felelősségét tudja kezelni, vele szemben tud jogkövetkezményeket alkalmazni. Egy öntanuló mesterséges intelligencia esetében azonban a technika jelen állása és jövőbeli kilátásai szerint is beszélhetünk olyan fokú „egyéniségekről", amelyek konkrét tettei mögött már nem mutatható ki tényleges emberi felelősség, mert nincs olyan személy, akinek közvetlenül bármilyen ráhatása lenne a viselkedésére. Egy öntanuló, önfejlesztő mechanizmus néhány hét alatt szinte teljesen el tud térni az eredeti programjától, és olyanokat is tud tenni, ami nem állt a készítői vagy a működtetője szándékában. Az ilyen esetekben pedig nem állapítható meg az ember felelőssége, hanem kizárólag a mesterséges intelligenciáé (amennyiben ez értelmezhető).

A jövő jogalkotása előtt álló nagy kérdés az lesz, hogy a jog hogyan reagál erre, képes lesz-e felülemelkedni az évezredes alapelvein, és elfogadni az ember nélküli felelősséget, és azt kizárólag a gépre telepíteni. Talán még nagyobb kérdés, hogy képes lesz-e a társadalom tolerálni azt, hogy az igazságérzetét vagy bosszúvágyát nem tudja kielégíteni, mert nem lesz alanya annak. A kárt vagy sérülést okozó gépet le lehet kapcsolni, a gyártó levonhatja a konzekvenciákat, új elveket vezethet be a jövőbeni tervezésben, azonban a szó klasszikus értelmében vett büntetést vagy kártérítési felelősséget nem lehet érvényesíteni.

A robotok és az MI károkozása kapcsán a felelősség vonatkozásában nem lesz egyszerű közös európai álláspont kialakítása, hiszen egységes európai magánjog révén a kárfelelősségi kérdésekben is számos megközelítésmód képzelhető el, amelyek nagymértékben függenek a tagállami polgári jogi rendszerektől. A magyar jog alapján a robotok magatartásáért való felelősség többirányú lehet. Megközelíthető az általános deliktuális felelősségi szabály alapján, azonban minél autonómabb és intelligensebb egy robot, a felróhatósági mérce alkalmazása annál nehezebbé válik, hiszen a robot használója és a robot által tanúsított károkozó cselekmény egyre inkább eltávolodik egymástól.

Elképzelhető egy felelősséget szigorító tendencia, amely oda vezethet, hogy a bírói gyakorlat a veszélyes üzemi felelősség szabályait kezdje el alkalmazni a robotokra. Ebben az esetben a mentesülés gyakorlatilag lehetetlen lenne, hiszen ha maga a robot a veszélyes üzem, és a kárt a robot idézi elő, akkor a kár oka szükségképpen a fokozott veszéllyel járó tevékenység keretein belül van, azaz a konjunktív mentesülési okok egyike rögtön ki is esik. A robot üzemeltetője mellett felmerülhet más személyek felelőssége is. Így szóba jöhet például a robotot tanító, azt rengeteg adattal „etető" személy felelősségre vonása vagy éppen a forgalmazó vagy gyártó felelőssége.[151]

Megítélésem szerint az anyagi felelősséget (károkozás és személyiségi jogsértés esetén) leginkább egy kiterjedt, mesterséges intelligenciára vonatkozó biztosítási konstrukció keretében lehet majd kezelni. Erre tett javaslatot az Európai Unió jogalkotója a Delvaux-jelentésben, amely kimondta, hogy a mesterséges intelligencia felelősségére vonatkozóan alkalmazott bármely választott jogi megoldás a vagyoni károktól eltérő esetekben semmilyen körülmények között nem korlátozhatja a megtéríthető károk típusát és mértékét, illetve a károsultnak felkínálható kártérítés formáit pusztán azon az alapon, hogy a kárt nem emberi lény okozta. A jelentés előírja, hogy kötelező biztosítási rendszert kell létrehozni, amely azon alapulhat, hogy a gyártó köteles biztosítást kötni az általa gyártott autonóm robotokra. A biztosítási rendszert pénzalappal kell kiegészíteni annak érdekében, hogy azokban az esetekben is legyen lehetőség a kártérítésre, amelyekben nem áll rendelkezésre biztosítási fedezet. Ezzel a megközelítéssel egyet tudunk érteni, mivel ez megfelelő kompenzációt biztosít a kárt szenvedett fél számára, azonban nem erőlteti a polgári jog hagyományos felelősségi konstrukcióit sem a vétlen emberekre, sem pedig a mesterséges intelligenciára.

A biztosítási kockázatközösség létrehozása előnyös megoldás azért is, mert érdekeltté teszi a mesterséges intelligenciát fejlesztő piaci szereplőket az alkalmazásaik folyamatos javításában és továbbfejlesztésében: az a cég vagy algoritmus, amelyik sorozatosan biztosítási eseményeket okoz, hamarosan ennek hatására minden bizonnyal magasabb összegű biztosítási díjra számíthat (hasonlóan a balesetet okozó autósokhoz), míg a biztonságos fejlesztéseket a rendszer bónusszal jutalmazhatja. A biztosítók figyelembe fogják venni a potenciálisan érintett felhasználók számát is, így a techóriások nem bújhatnak el a kockázatközösség háta mögé, piaci szerepüknek megfelelő

[151] Keserű 2018.

mértékű díjat kell fizetniük.[152] A kockázatközösség létesítésekor előzetesen meg kell vizsgálni azt a kérdést, hogy ki lesz a biztosítási díj fizetésének kötelezettje, a gyártó vagy az üzemben tartó (mindkét megoldásra láttunk már javaslatot a szakirodalomban).

Megítélésem szerint mindkét szereplőt be lehet vonni, attól függően, hogy a mesterséges intelligenciát alkalmazó termék milyen célra és alkalmazási területre készült.[153] A kifejezetten felhasználói termékek gyártóit be kell vonni a kockázatközösségbe, mivel ebben az esetben előfordulhat, hogy a kár az üzemeltetőnél (mint végfelhasználónál) következik be, és nem érint harmadik személyeket (például a mesterséges intelligencia alapú raktárkészlet-nyilvántartó rendszer téves következtetést von le a bolti forgalomból, és az eladhatónál nagyságrendekkel nagyobb mennyiségű romlandó árut rendel meg, ami így veszteségként jelentkezik a kereskedőnél). A biztosítási konstrukció előnye, hogy a pereskedéshez képest gyorsítja az eljárást, egyszerűbb ügyekben a biztosítóval könnyen és gyorsan megegyezve pénzhez jut a károsult. Ezen túl a biztosító és a gyártók, illetve üzemben tartók közös érdeke lesz a termékek folyamatos fejlesztése és biztonságosabbá tétele, és ezt külső, jogalkotói beavatkozás nélkül, pusztán a piaci önszabályozó folyamatokra hagyatkozva el lehet végezni. A biztosítók így a piacra kerülő termékek vizsgálatának, tesztelésének, illetve kockázataik meghatározásának tényét átveszik az államtól. Nem szükséges állami költségvetési forrásból kiterjedt szakembergárdát és vizsgálóintézeteket fenntartani, mert a biztosítótársaságok – a saját gazdasági érdeküktől hajtva – megteszik ezt. A biztosítók által bevizsgált és magas kockázatúnak ítélt mesterséges intelligencia alapú alkalmazások biztosítási díja is magasabb lesz, ami vagy a fejlesztőket ösztökéli a kockázatok csökkentésére, vagy pedig a felhasználókat (üzemben tartókat) terelgeti más termékek vagy konstrukciók irányába.

A biztosítási modell egyfajta fékként érvényesül a fejlesztések körében: nem engedi kipróbálás és a biztonsági kockázatok feltárása nélkül piacra lépni a gyártókat egy-egy új konstrukcióval, a működés közben kiderülő egyes hibák javítását pedig pénzügyi ösztönzőkkel tudja előremozdítani. Mindezt azonban úgy végzi, hogy a gazdasági szektor belső logikájára és önszabályzó mechanizmusára bízza a feladatot, kihagyva a sokszor lassú, nehézkes és a szükséges szakértelemmel nem feltétlenül rendelkező állami

[152] CERKA–GRIGIENE–SIRBIKYTE 2015, 386.
[153] VLADECK 2014, 148–150.

bürokráciát vagy jogalkotást a folyamatból. Voltaképpen tehát a biztosítási modell szabályozási és ellenőrzési feladatokat vesz át az államtól, egyúttal ösztönzi a piaci szereplőket a biztonságos működési gyakorlatok alkalmazására, a termékhibák és kockázatok mielőbbi kiküszöbölésére.

A biztosítási modell előnyös a károsultak számára, mert így egyetlen egyszerű mechanizmus útján, rövid idő alatt megkapják a kártérítésre szánt pénzösszeget vagy a személyiségi jogi sérelemért járó sérelemdíjat. A kérelmüket egy kifejezetten erre szakosodott szervezetnél kell előterjeszteni, ahol megvan a kellő gyakorlat és infrastruktúra a kárigény gyors és szakszerű intézéséhez, a kifizetendő kártérítés összegszerűségének megállapításához.

Egy kártérítési ügyben e modell alapján az eljáró hatóság vagy bíróság először abban a kérdésben hozhat döntést, hogy emberi vagy gépi felelősség áll-e fenn, vagyis követett-e el valaki olyan cselekedetet vagy mulasztást, amely közvetlenül közrejátszott a káresemény bekövetkezésében. Amennyiben igen, úgy a polgári jog hagyományos felelősségi alakzatait (szerződéses és szerződésen kívüli károkozás, veszélyes üzemi felelősség) kell alkalmazni vele szemben. Ha viszont megállapítható, hogy a kár vagy jogellenes cselekmény bekövetkezte kizárólag a mesterséges intelligencia autonóm döntésének következménye, akkor az eljárás a biztosítási konstrukció irányába kanyarodik.

4.2.6. Javaslat a diszkrét externáliák szabályozásának keretrendszerére

A fentiekben megvizsgáltuk a mesterséges intelligencia által kiváltott diszkrét externáliák, vagyis a helyi szintű külső hatások által felvetett jogi kérdéseket, különös figyelmet fordítva a károkozásért és más, jogsértő cselekményekért való felelősség kérdéskörére. Arra a megállapításra jutottunk, hogy a mesterséges intelligencia működése kapcsán felvetődő felelősségi kérdések esetében négy esetkört kell elkülöníteni. Az első esetkörbe tartozik a gyártási hibából vagy hiányosságból fakadó károkért való felelősség, amit a jelenlegi jogi keretrendszer változatlanul hagyásával a termékfelelősség és a fogyasztóvédelmi jog egyes alakzatai felhasználásával a gyártóhoz telepíthetünk. Ezt abban az esetben tehetjük meg, ha a mesterséges intelligenciát alkalmazó termék gyártási hibájából egyenesen következett a kárt okozó vagy jogsértő eredmény.

A következő alakzat az üzemben tartó szakszerűtlen vagy szabálytalan magatartásából vagy az általa ellátott felügyeleti tevékenység nem

megfelelő elvégzéséből fakadó károkat vagy jogsértéseket jelöli. Ebben az esetben a veszélyes üzemi felelősség szabályait alkalmazhatjuk, feltéve, hogy a kár a mesterséges intelligencia kiszámítható működésére vezethető vissza, és annak bekövetkezte racionálisan valószínűsíthető volt.

Végezetül elképzelhető szándékos károkozás is, amikor a gyártó, a gyártó felügyelete alatt tevékenykedő személy vagy az üzemben tartó, esetleg egy rosszindulatú támadó szándékosan vagy súlyos gondatlanságból oly módon módosítja a mesterséges intelligencia működését, hogy az észszerűen előre látható módon a bekövetkezett kárt vagy jogsértést eredményezi, és a sérelem a módosítás nélkül nem következett volna be. Ebben az esetben a felelősséget közvetlenül a kárhoz vezető módosítást végrehajtó személy viseli mind a polgári jogi, mind a büntetőjogi felelősség szempontjából.[154] Mindhárom most leírt felelősségi alakzat azt feltételezi, hogy a mesterséges intelligencián alapuló algoritmus működése kiszámítható, átlátható és az előre meghatározott pályának megfelelően halad.

A negyedik felelősségi szituáció azt az esetet mutatja be, amikor a mesterséges intelligencia autonóm módon hozott döntésére senki nincs közvetlen hatással, az nem következik egyenesen a beírt programból, bekövetkezésének lehetősége, időpontja vagy módja nem előre látható vagy kiszámítható, illetve a döntési mechanizmus a felügyeletet gyakorló számára nem átlátható. Ebben az esetben emberhez köthető büntetőjogi felelősség nem állapítható meg, tekintve, hogy mind a szándék (vagy gondatlanság), mind a cselekmény mint fogalmi elem hiányzik a tényállásból. A *corpus* és az *animus* együttes hiánya megítélésem szerint kizárja a büntetőjogi felelősséget. Mindazonáltal nem tekinthetünk el attól a ténytől, hogy bűncselekmény történt, mert a büntetések egyik célja, méghozzá a speciális prevenció, még megvalósítható. A büntetőeljárásban gondoskodni kell arról, hogy a konkrét, érintett mesterséges intelligencia újabb bűncselekményt ne tudjon elkövetni, így büntetőeljárási intézkedésként a megelőzéshez szükséges adminisztratív és műszaki intézkedést el kell rendelni: ez jelentheti a konkrét mesterséges intelligencia működésének megszüntetését (törlését), újraprogramozását, a gyári beállítások visszaállítását vagy a törlés utáni kontrollált újratanítást.

A nagymértékben autonóm mesterséges intelligencia irányában fennálló polgári jogi felelősségi kérdések hasonlóképpen nem rendezhetők a meglévő alakzatok alapul vételével, mivel itt olyan önálló cselekedetről van szó, amely teljességgel független bármely emberi cselekedettől vagy mulasztástól, vi-

[154] BALKIN 2015, 52–54.

szont aktusnak minősül. Ez a cselekmény nem volt előre látható, sem kiszámítható, bekövetkeztét sem az üzemben tartó, sem a gyártó nem tudta még megjósolni sem. A teljes egészében a mesterséges intelligencia által hozott döntés olyan mértékben átláthatatlan és olyan nagy fokú gépi önállóságon alapul, hogy az algoritmus működésére befolyással bíró személyek nemcsak azt nem tudják kiszámítani, hogy egy ismert vagy elképzelhető hiba mikor és hogyan következik be, hanem még azt sem látják át, hogy egyáltalán milyen hiba következhet be a mesterséges intelligencia működése során. Az ilyen nagymértékben önálló mesterséges intelligenciát – megítélésem szerint – nem lehet „lefokozni" veszélyes üzemmé, mert azok esetében az üzemben tartó egy előre nagyjából elképzelhető problémakörre készülhet fel, azok megakadályozása érdekében pedig tehet lépéseket. A jog logikájának és a jogrendszer integritásának fenntartása érdekében a polgári jogi felelősség körében sem mondhatjuk ki az üzemben tartó vagy a gyártó felelősségét, hanem lehetőséget kell adni a bíróságoknak arra, hogy a körülmények alapos vizsgálata után kimondhassák a mesterséges intelligencia kizárólagos felelősségét. Amennyiben pedig a kizárólagos gépi felelősséget megállapítják, úgy egyúttal mentesítik a gyártót és az üzemben tartót a személyes felelősség alól.

A felmerült vagyoni és nem vagyoni kár, valamint személyiségi jogi sérelem orvoslására a kockázatközösségi modell bevezetésére tettem javaslatot, egyetértve az Európai Bizottság számára megfogalmazott ajánlásokkal. Amennyiben a mesterséges intelligenciára önálló biztosítási konstrukciók jönnek létre, a magas szintű autonómiával bíró mesterséges intelligencia üzemben tartóit, illetve gyártóit felelősségbiztosítás kötésére kötelezhetné a jogalkotó, amely – a gépjármű-felelősségbiztosításhoz hasonlóan – fedezetül szolgál az anyagi károk megtérítésére, illetve a személyiségi jogi sérelmek esetében sérelemdíj fizetésére.

Az alábbi táblázat a helyi externáliák negatív hatásainak jogi kezelésére vonatkozó javaslatokat foglalja össze.

Ahhoz, hogy a fenti modellnek megfelelően tudja kezelni az externális hatásokat, a jogalkotónak a mesterséges intelligencia, illetve az annak felhasználásával készült eszközök vagy robotok jogi karakterét kell meghatároznia. Arra a kérdésre kell a választ megadnia, hogy az MI-alapú robot mely jogi rezsimhez áll a legközelebb: termék, állat vagy önállóan cselekvő ágens, és mint ilyen, *sui generis* jogi státuszt érdemel. Az erre adott válasz fogja kijelölni a jogalkotó által követendő utat, és minden más kérdés az így megtalált út egy-egy elágazása vagy kanyarulata lehet csak. Mivel a mesterséges intelligencia technológiai fejlődése rendkívül szerteágazó, és a felhasználási

köre is diverzifikált, így nem lehet egyértelműen állást foglalni a fenti kérdésben. A fejlesztők – éppen a mesterséges intelligencia alapú programok komplexitása és rendkívüli számítási igényei miatt – a konkrét feladathoz szabják általában, hogy milyen fokú autonómiával rendelkező MI-algoritmust használnak fel a feladat megoldásához. Egy prediktív ingatlanár-kalkulátor nem igényel olyan fokú önállóságot, mint egy személyi asszisztens vagy egy katonai elemzőprogram, így azokhoz alacsonyabb szintű algoritmusokat is elég használni, nem feltétlenül kell elővenni a felügyelet nélküli gépi tanulás vagy a *deep learning* teljes eszköztárát.

1. táblázat

A helyi externáliák negatív hatásainak jogi kezelése

Alacsony autonómia	**Magas autonómia (autonóm intelligencia)**
Gyártási hiba → termékfelelősség	
Üzemben tartói hiba → üzemben tartó felel • *károkozás: veszélyes üzemi felelősség* • *büntetőjogi felelősség általános formái*	
Szándékos károkozás → károkozó személyes felelőssége • *károkozás: polgári jogi kárfelelősségi alakzatok* • *büntetőjogi felelősség általános formái*	
	Gépi döntéshozatal → MI felelőssége (autonóm) • *károkozás: biztosítási esemény* • *bűncselekmény: adminisztratív intézkedés (gép kivonása a forgalomból, újraprogramozás, visszaállítás stb.)*

Forrás: a szerző szerkesztése

Az erőforrás-takarékos megoldásokra optimalizált fejlesztés okán mesterséges intelligencia és mesterséges intelligencia között ég és föld lehet a különbség az autonóm cselekvés fokát nézve. Meglátásom szerint ezért szükséges a fent bemutatottak szerint az autonómia foka alapján egy skálán elhelyezni őket. A skála kidolgozása lehet jogalkotói feladat, de egyaránt lehet az ipari szabványosítás része is, a lényeges szempont azonban az, hogy ez a csoportosítás lehetőleg globálisan egységes legyen, így az autonómia egy

meghatározott foka alatt ugyanazt értsék a japán, az európai és az amerikai fejlesztők és felhasználók is. Ha az önállóság vagy egyedi intelligencia fokait akarjuk meghatározni, egy egyszerű hármas felosztást vehetünk alapul.[155]

A kategóriák első szintje vagy alapvonala az önálló döntéshozatalra nem képes robotokat jelöli. Ezek az algoritmusok magas színvonalon, gyorsan és hatékonyan tudják megoldani a beléjük programozott feladatokat, azonban a program utasításain kívül nem tudnak cselekedni, önálló feladatmegoldásra vagy a működésük módosítására nem képesek. Ilyennek tekinthetünk például egy egyszerűbb GPS navigációs szoftvert, amely a helyadataink és a bevitt térkép alapján egy beépített képlet segítségével kiszámítja az úti cél eléréséhez szükséges legrövidebb vagy leggyorsabb útvonalat, és azt megjeleníti a képernyőjén.

A mesterséges intelligencia következő lépcsőfoka a cselekvési intelligencia, ahol a gép a kiválasztott cél eléréséhez szükséges optimális megoldást önállóan keresi meg, maga dönt arról, hogy egyes esetekben beavatkozik-e vagy sem, és hogy a beavatkozás hogyan történjen. Tevékenysége során a gépi tanulás egyes eszközeit is használhatja. Az ide tartozó algoritmusokra egy önvezető autó szoftvere lehet a legjobb példa, amely folyamatosan érzékeli és elemzi a környezetét, és az ott tapasztaltak alapján dönt a beavatkozás szükségességéről és mibenlétéről (gyorsítás, fékezés, kanyarodás), mindeközben pedig a kitűzött alapvető célt is követi, vagyis egy meghatározott célponthoz navigál. Jelenlegi ismereteink szerint a mesterséges intelligencia legmagasabb elképzelhető szintje az autonóm intelligencia szintje. Ezek az algoritmusok nemcsak önálló helyzetelemzésre és cselekvésre képesek, hanem önfejlesztés és öntanulás révén saját működésüket is folyamatosan módosítják, és optimalizálják a programjukat. Az autonóm intelligencia egy elérendő, magasabb absztrakciós szinten megfogalmazott cél érdekében képes a tevékenységét felülvizsgálni és módosítani, a folyamat közben esetleg új feladatot kitűzni magának, és azt végrehajtani. Az autonóm intelligencia a mesterséges intelligenciának az a szintje, amelyik már folyamatosan képes együtt élni az emberekkel, és segíteni az életüket, mindezt valós körülmények között, egy strukturálatlan környezetben.

A fenti, kiindulópontnak szánt kezdeti skála szerint az első szinten álló robotok által okozott hátrányos hatások a jog alapstruktúrái szerint jól kezelhetők: a termékfelelősség és a veszélyes üzemi felelősség megfelelő kereteket nyújt hozzá, itt ugyanis nincs nagy különbség a mesterséges intelligenciát is

[155] Weng–Chen–Sun 2009, 274–276.

alkalmazó megoldások és bármilyen más műszaki eszköz között. A második szint esetében áll fenn a fent vázolt elágazás kiépítésének szükségessége. Itt meg kell állapítani azt, hogy a károkozás vagy jogsérelem kizárólag a mesterséges intelligencia saját döntésének következménye, vagy abban más – a gyártó vagy az üzemben tartó – magatartása is közrejátszott. Végezetül a mesterséges intelligencia harmadik csoportja esetében egyértelműnek tekinthetjük, hogy – a gyártási vagy üzemeltetési hibákat és a szándékos károkozást kivéve – a mesterséges intelligencia szabályos működése során hozott döntések következményeiért nem lehet sem a gyártót, sem az üzemben tartót felelősségre vonni. Az autonóm intelligencia felveti a különleges jogi státusz kérdéseit is, amelyekre később fogunk kitérni.

4.3. A rendszerszintű hatások szabályozási kérdései

4.3.1. A rendszerszintű hatások jellegzetességei

Ahogy azt fent bemutattuk, a rendszerszintű hatások a mesterséges intelligencia működésének olyan – pozitív vagy negatív – hatásai, amelyek jellegzetességeiket tekintve lokalizáltak, kiszámíthatók, gyakoriak vagy visszafordíthatatlanok. Lokalizáltak, mert a társadalom egy jól körülhatárolható, de jelentékeny szeletét érintik. Kiszámíthatók, mert a jogalkotó a megfelelő információk birtokában előre láthatja azok bekövetkezését, számíthat rájuk. A gyakoriság alatt az előnyös vagy hátrányos hatások ismétlődő bekövetkezését értjük. Végezetül azért nevezzük visszafordíthatatlannak ezeket a hatásokat, mert tartós változást tudnak előidézni a társadalom jólétében. Talán a legtöbbet tárgyalt rendszerszintű hatás az emberi munkaerő gépekkel való leváltása, ami a munkanélküliség növekedését eredményezheti, és mivel különösen az alacsony képzettséget igénylő munkákról van szó, tovább növeli a gazdagok és szegények közötti különbségeket, vagyis fokozza a társadalmi egyenlőtlenségeket. Szintén rendszerszintű negatív hatásnak nevezhetjük a személyes adatok védelmi szintjének általános csökkenését és a személyes adatok nagy cégek birtokába kerülését.

A rendszerszintű externáliák jellegzetessége, hogy káros hatásai nem az egyének szintjén jelentkeznek elsősorban, pontosabban nem lokalizáltan jelennek meg egy-egy személy szintjén, hanem az egyének nagy csoportjának érintettsége miatt össztársadalmi vagy a társadalom egy jelentős részét kitevő populáció szintjén mutatkoznak meg. Emiatt ezeket a hatásokat nem tudjuk

egyedi jogviták vagy biztosítási események szintjén kezelni, hanem nagyobb szabású megoldást kell találni rájuk. Ezek a hatások azonban a kellő információk birtokában lévő jogalkotó számára kiszámíthatók a technológia jelenlegi állása és a fejlődési tendenciák ismeretében, ezért az ilyen típusú externáliák esetében a jogalkotónak lehetősége nyílik az *ex post* típusú vitarendezési megoldási utak kialakítása helyett előre tervezve, *ex ante* kialakítani a felmerülő hatásokat kezelő mechanizmusokat. Ebben az esetben a jogalkotóval szemben támasztott legfőbb elvárás az, hogy igyekezzen előre és rendszerszinten gondolkodni, de ami még fontosabb, felismerni azt, hogy nem helyi, hanem rendszerszintű hatásról van szó, és ezért más kezelési módot kell alkalmazni.

A helyi szintű hatásoknál alkalmazott utólagos, vitarendezési szempontú szemlélet, illetve a lokális szintű, egy-egy esetet elszigetelten kiragadó beavatkozás több kárt okozhat a rendszerszintű hatásoknál, mint amennyi hasznot jelent. A jogalkotó leginkább itt szembesül a fent bemutatott szabályozási nehézségekkel: meg kell találnia az egyensúlyt a biztonságra törekvés és a fejlesztést ösztönző, a gazdasági szektor érdekeit szem előtt tartó megközelítés között, informálódnia kell a szabályozás tárgyáról és a választott módszer lehetséges hatásairól. Ahogy arra fent is kitértünk, a jogalkotónak ebben a nehéz szabályozási környezetben az eddig megszokottól eltérő attitűdöt kell magára vennie, mert a bizonytalan és folyton változó szabályozási tárgy és környezet azt indokolja, hogy a szabályalkotást ne tekintsük végcélnak és megváltoztathatatlannak, hanem egy első lépésnek azon az úton, amely még számos elágazást és kanyart tartogat. A jogalkotó inkább válasszon rugalmasabb megközelítést, alakíthatóbb szabályokat, illetve hagyja nyitva a szabályozás módosításának lehetőségét.[156]

Ennek érdekében két javaslatot is tettünk. Az egyik javaslat a szabályozás rendszeres hatásvizsgálatára irányul. Az új technológiák esetében és különösen a rendszerszintű hatások vonatkozásában a jogalkotó sokszor kénytelen a sötétben tapogatózni, és az előzetesen feltételezett hatások kezelésére jogot alkotni. A tapasztalatok és az egyre nagyobb mennyiségű tényanyag összegyűlésével a jogalkotó meg tudja ítélni, hogy jó irányba indult-e el a szabályozással, vagy pedig tévúton jár, és a szabályozása nem jó problémát ragadott meg, vagy a tényleges problémát nem jól közelítette meg, esetleg túl megengedő vagy túl megszorító volt a szabályozás. Az empirikus tapasztalatok függvényében ezért rutinná kell tenni a jogalkotó számára döntései rendszeres felülvizsgálatát és szükség szerint módosítását.

[156] Guihot–Matthew–Suzor 2017, 28–29.

A másik megfogalmazott javaslat a „szabályozási gyakorlópályák" kialakítására vonatkozott, ahol példának hoztuk a Japánban lassan tizenöt éves múltra visszatekintő „Tokku" robottesztelési övezeteket. Ezek esetében a jogalkotó bizonyos meglévő jogi szabályokat enyhít, vagy pedig különleges szabályozási környezetet hoz létre egyes új technológiák nyilvános bevezetése előtti tesztelése céljából. A tesztelés meghatározott ideig, limitált földrajzi területen, illetve személyi körben lehetséges, a hatóságok szoros odafigyelése mellett. Ez a megoldás nemcsak a fejlesztőknek ad lehetőséget arra, hogy az új technológiájukat valós környezetben teszteljék, hanem a jogalkotó is „kísérletezhet" egyes új szabályozási megoldásokkal, illetve vizsgálhatja bizonyos szabályok enyhítésének vagy új szabályok megalkotásának a technológia működtetésére és a társadalomra gyakorolt hatását. Mindkét megoldás abba az irányba mutat, hogy az *ex ante* jellegű szabályozás utólagos felülvizsgálatát rendszeresen el kell végezni, és a vizsgálat eredményének függvényében szükség esetén módosítani azt.[157]

A helyi szintű externáliák és a rendszerszintű hatások között van átjárás: egyes helyi hatásoknak rendszerszintű következményei is lehetnek. Egyrészt a helyi externáliák úgy válnak rendszerszintűvé, hogy szabályozási dilemmát adnak a jogalkotónak, akinek a helyi szintű problémák megoldásához új jogi kereteket kell alkotnia. Erre hozhatjuk fel példának a helyi problémát jelentő kárfelelősség kérdésére adandó válaszként javasolt biztosítási konstrukciót és a mesterséges intelligencia autonómiafokának kategorizálását, amely végső soron a jogalkotó komoly, rendszerszintű beavatkozását igényli. Ehhez hasonló módon válhat rendszerszintű problémává a helyi szintű hatásoktól (például károkozástól vagy sérülés okozásától) való állampolgári félelem, amely miatt a technológiával szembeni bizalmatlanság (ellenségesség) rendszerszintű kihatást eredményez, amire a szabályozónak fel kell készülnie.

Az új technológiák, így a mesterséges intelligencia működése is számos, már meglévő és számos, még ismeretlen szabályozási problémát vetnek fel, amely társadalmi szintű hatásokat okoz. Az alábbiakban néhány, már ismert vagy sejthető és még megoldásra váró rendszerszintű hatásra hívjuk fel a figyelmet. Itt nem ismételjük meg a jogi problémákról szóló részben írtakat, hanem olyan, nem jogi problémákról ejtünk szót, amelyek azonban a jogalkotó intézkedését igénylik.

[157] PETIT 2017, 29.

4.3.2. A mesterséges intelligencia kommunikációja

A nyelvhasználat az emberi nem megkülönböztető tulajdonságának számít az állatvilág egyedeihez képest. Nem vitatva el azt, hogy más élőlények is képesek kommunikációra, bizonyos érzelmek vagy utasítások kifejezésére, a komplex gondolkodást tökéletesen leíró, a múltra, a jelenre és a jövőre, a jelenlévőkre és a távollévőkre egyaránt reflektálni képes, történetmesélésre alkalmas, összetett nyelvi rendszereket jelenlegi ismereteink szerint azonban kizárólag az ember alakított ki. Az emberiség számos élő és holt természetes nyelvvel büszkélkedhet, ezenkívül emberek maguk is képesek voltak több, teljesen mesterséges nyelvet is kialakítani, egy részüket kommunikációs céllal, más részüket tisztán művészeti igényeket tartva szem előtt.[158] Technikai célú kommunikációs formának tekinthetjük a különféle programnyelveket, amelyek útján a programozók írják meg az utasítássort a számítógépeknek.

A nyelvek sokszínűsége érték, amelyet számos nemzetközi konvenció is véd, a nyelvek kutatása, gondozása és művelése tudományos feladat. A nyelvek azonban nemcsak összekötnek, hanem el is választanak egymástól, hiszen dacára a nyelvek fejlettségének és kifejezőképességének, csak azokkal tudjuk megértetni magunkat, akiknek beszéljük a nyelvét. Közös nyelv nélkül kudarcra van ítélve az érdemi kommunikáció, így csak bizonyos külső jelekből tájékozódhatunk partnerünk szándékairól, lelkiállapotáról, céljairól és konkrét tevékenysége mibenlétéről. Ha viszont valakivel „egy nyelvet beszélünk", akkor vele egyszerű, gyors és hatékony a közös munka is.

Az emberi nyelvek sokféleségének és értékeinek felvillantásával egy nagyon is konkrét, a mesterséges intelligencia működését érintő, előremutató megfontolást igénylő kérdésre kívánunk rávilágítani, a mesterséges intelligencia egymás közötti kommunikációjának kérdésére. Ez talán úgy is megnevezhető lenne, hogy a mesterséges intelligencia nyelvhasználata, mert ez sem állna távol a valóságtól. A mesterséges intelligencia a kifejlesztésétől kezdve interakcióba lép a környezetével, a bejövő ingereket értelmezi, lefordítja saját magának, és az így kapott információt felhasználja a működése során. Azonban nem csak a bejövő ingerek jelentik számukra a kommunikációt, a mesterséges intelligenciát alkalmazó algoritmusok maguk is kommunikálnak a környezetükkel. Ilyen interakciókat naponta láthatunk,

[158] Kommunikációs célú nyelv például az eszperantó, az interlingua vagy a volapük, míg művészeti nyelvek között tarthatjuk számon a J. R. R. Tolkien regényeiben megjelenő különleges nyelveket (például quenya, sindarin).

amikor megszólítjuk okostelefonunk digitális asszisztensét vagy okosotthonunk hangvezérlő rendszerét. Ezek az eszközök képesek az emberi beszédet értelmezni, arra reagálni és a kapott utasításnak megfelelő tevékenységeket elvégezni, emellett válaszolnak is a feltett kérdésekre a felhasználójuknak. Az ügyfélszolgálati mesterséges intelligencia kommunikációs képessége sokkal fejlettebb, sok esetben a felhasználó nehezen tudja eldönteni, hogy géppel vagy emberrel chatelt-e. Ahogy az ember és robot interakciójában nagy jelentősége van annak, hogy a gép jól értse és így pontosan hajtsa végre a felhasználó utasításait, illetve maga is érthetően kommunikáljon a felhasználóval, úgy a gépek egymás közti kommunikációjában is kiemelkedő fontosságú, hogy értsék egymást. Egyre gyakrabban fordul elő ugyanis olyan helyzet, hogy a mesterséges intelligenciáknak egymással kell kommunikálniuk. Nem kell nagyon távoli példát hozni, ugyanis egy okosotthon összetevői is folyamatos interakcióban állnak egymással, adatokat osztanak meg és utasításokat adnak a rendszer komponenseinek.

Könnyen előfordulhat az az eset is, hogy két, egymástól független mesterséges intelligencia kommunikál egymással (például az okostelefonos asszisztens időpontot foglal a szintén mesterséges intelligencia által üzemeltetett internetes ügyfélszolgálaton keresztül). Az ilyen interakciók esetében felmerülő első kérdés az, hogy miként értékelje a jog a két mesterséges intelligencia egymás között végrehajtott tranzakcióit. Tekinthetjük-e úgy, hogy az emberi beavatkozás nélkül kötött szerződések teljes jogi elismertséget élveznek ugyanúgy, mintha az interakció egyik vagy mindkét végén ember állt volna? Esetleg különleges jogi rezsim alá vonjuk ezeket az eseteket, és a távollévők között kötött szerződésekre vonatkozó kivételes polgári jogi és fogyasztóvédelmi szabályokhoz hasonlóan sajátos módon kezelje őket a jogrendszer?

Nem légből kapott példa ez, számos esetben történt meg ugyanis, hogy a tévékészülék közelében elhelyezett digitális asszisztens a reklámokban található felhívásokat utasításként értelmezte, és megrendelte a szóban forgó terméket, vagy más nem várt dolgot tett.[159] Ehhez hasonlóan járt az a fel-

[159] Egy tévéreklám hatására az Amazon Echo-készülék Alexa nevű digitális asszisztense macskatápot rendelt az interneten (www.theguardian.com/technology/2018/feb/14/amazon-alexa-ad-avoids-ban-after-viewer-complaint-ordered-cat-food). (A letöltés ideje: 2018. 09. 20.) Egy South Park-epizód hatására pedig több felhasználó Alexa és Google Home készüléke zavarodott össze és állított be például ébresztőórát különféle időpontokra (www.hollywoodreporter.com/live-feed/south-park-premiere-messes-viewers-amazon-alexa-google-home-1039035). (A letöltés ideje: 2018. 09. 20.)

használó is, akinek a hatéves gyermeke rendelt meg egy drága babaházat a digitális asszisztens segítségével.[160] Ez utóbbi esetben ugyan a megrendelés teljesen szándékos volt, azonban a cselekvőképtelen gyermek szándékát valós vásárlássá transzformálta a mesterséges intelligencia.

Érdemes azt is végiggondolni, hogy ebben az esetben egyaránt származhat kára a tranzakcióban részt vevő mindkét félnek: nem csak az járhat rosszul, akinek a nevében a digitális asszisztens megrendelt egy csomó mindent az interneten, és ő már csak a hatalmas végösszegű számlával találkozott. Kár érheti azt is, aki jóhiszeműen abban bízik, hogy tényleges vásárlási céllal rendeltek meg tőle egy terméket vagy szolgáltatást, és akár anyagi befektetés útján is nekikezd a teljesítésnek, majd később kell szembesülnie vele, hogy a megkötött üzlet semmis, mert a felhasználó akaratán kívül tette meg a megrendelést a digitális eszköz. A két mesterséges intelligencia között lezajlott tranzakció szélsőséges esetben mindkét fél akaratán kívül is megtörténhet, ahol mindkét oldalon álló felhasználó egyaránt nem kívánt eredménnyel szembesül. Jelenleg a vásárlásra is képes digitális asszisztensek gyártói a készülékek „félautomatává" tételével válaszoltak a fent bemutatott incidensekre: a vásárlásokat a tényleges megrendelés előtt meg kell erősíteni az „igen" szó kimondásával, még nagyobb biztonság elérése érdekében pedig egy négyjegyű biztonsági kód megadásával.[161] Ez átmeneti megoldást adhat és jól példázza a gyártók gyors, kárelhárító intézkedését a nemkívánatos események után, azonban hosszabb távon valószínűleg nem marad fenn, mert éppen a digitális asszisztensek önállóságát és proaktivitását fogja vissza, ami az egyik fontos vonzereje ezeknek a technológiáknak. Mindazonáltal nem csak a vásárlásokra kell gondolni a mesterséges intelligenciák egymás közötti kommunikációja során.

A dolgok internete *(Internet of Things, IoT)* a közeli jövő nagy robbanás előtt álló területe, ami azt jelenti, hogy hamarosan tele lesz a házunk, autónk, munkahelyünk, a bevásárló- és szórakoztatólétesítményeink különböző, mesterséges intelligenciára épülő okoseszközökkel, amelyek hálózatba kötve kommunikálnak egymással, sőt a testünkön viselt okoskiegészítők is bekapcsolódnak a beszélgetésbe. Ez pedig olyan volumenű gép-gép

[160] Az esetet tovább árnyalja, hogy a furcsa eseményről beszámoló tévéhíradó további Echo-készülékeket hozott működésbe, amelyek rendelést adtak le ugyanarra a babaházra (https://qz.com/880541/amazons-amzn-alexa-accidentally-ordered-a-ton-of-dollhouses-across-san-diego/). (A letöltés ideje: 2018. 09. 20.)

[161] Elérhető: www.telegraph.co.uk/news/2017/01/08/amazon-echo-rogue-payment-warning-tv-show-causes-alexa-order/ (A letöltés ideje: 2018. 09. 20.)

kommunikációt jelent majd, amelyet az ember nem lesz képes átlátni vagy uralni. Az adatok átadása, kumulációja, feldolgozása és a következtetések levonása mesterséges intelligencia által, önműködően történik majd, ami számos kommunikációs és feldolgozási hibát, visszaélési lehetőséget, valamint külső támadásokkal szembeni sebezhetőséget hordoz magában. Ezzel együtt természetesen számos pozitív jelenség és lehetőség is áll előttünk. A jogalkotónak mindazonáltal előre gondolkodva kezelni kell azt a helyzetet, hogy milyen adatvédelmi, döntéshozatali, biztonságtechnikai, egészségügyi és lélektani implikációi lehetnek az egymással kommunikáló mesterséges intelligenciák ilyen hadseregének.

A gépek közötti kommunikáció másik fontos és markáns jogi relevanciával bíró kérdése, hogy a gépek egymás közötti kommunikációjában engedjük-e új kommunikációs formák, nyelvek, rövidítések, kódok vagy titkosítás kifejlesztését az interakcióban részt vevő mesterséges intelligenciának. A közelmúlt több kísérlete is azt mutatja, hogy a csevegőrobotok hajlamosak egy más számára nem érthető közös nyelv kialakítására, ha nem kötik meg szigorúan a felhasználható kommunikációs formákat.

A Facebook fejlesztői végeztek egy kísérletet a chatbotokkal, ahol azt a feladatot adták két mesterséges intelligencia alapú csevegőprogramnak, hogy egymással beszélgetve kössenek üzletet, ahol előre meghatározott értékű kalapokat, labdákat és könyveket kell elcserélniük egymással úgy, hogy a legjobban járjanak. A kutatók hamar arra lettek figyelmesek, hogy az egymással kommunikáló robotok néhány interakció után elkezdtek sajátos nyelven kommunikálni egymással, amit külső szemlélőként nem tudtak értelmezni. A kommunikáció az alábbihoz hasonló módon zajlott, angol szavak használatával, de nem érthető szemantikai és nyelvtani szabályokat követve:[162]

„Bob: i can i i everything else . . . . . .
Alice: balls have zero to me to me to me to me to me to me to me to me to
Bob: you i everything else . . . . . .
Alice: balls have a ball to me to me to me to me to me to me to me
Bob: i i can i i i everything else . . . . . .

[162] Elérhető: www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html (A letöltés ideje: 2018. 09. 20.)

Alice: balls have a ball to me to me to me to me to me to me to me
Bob: i . . . . . .
Alice: balls have zero to me to me to me to me to me to me to me to me to
Bob: you i i i i i everything else . . . . . .
Alice: balls have 0 to me to me to me to me to me to me to me to me to
Bob: you i i i everything else . . . . . .
Alice: balls have zero to me to me to me to me to me to me to me to me to"

Ha azt gondolnánk, hogy programhibáról van szó, és a fenti beszélgetés csakugyan értelmetlen, akkor jegyezzük meg, hogy az ezen a nyelven folyó tárgyalások egy jelentős része sikerrel járt, és a két csevegőrobot megkötötte az üzletet egymással, gazdát cserélt a megfelelő számú holmi. A két gép tehát olyan nyelven beszélgetett, amely számukra megfelelő jelentést hordoz, valószínűleg egyszerűsítette vagy egyértelműbbé tette a kommunikációt, azonban külső megfigyelő számára annak logikája és jelentése nem érthető. Ezt a jelenséget technikai és jogi szempontból kétféleképpen is meg lehet ítélni. Egyfelől üdvözlendő az a törekvés, hogy a gépek egymás közt a kommunikáció hatékonyságát növelő jelrendszert dolgoztak ki, mert ez elősegíti a kooperációt, és a feladat jobb megoldását eredményezheti.[163] Másrészről azonban ezzel kizárja az emberi felügyeletet vagy a későbbi hibakeresést, mivel külső fél nem érti a kommunikációt, így nem tudja megítélni, hogy milyen információk cseréltek gazdát, melyik fél milyen utasítást adott, illetve miből milyen következtetést vontak le a mesterséges intelligenciák. Ez pedig a felelősségi kérdések kapcsán is bemutatott átlátható működés kritériumát sérti.

A legtöbb szerző álláspontja megegyezik abban, hogy a mesterséges intelligencia alapú automatizált döntéshozatal jogszerűsége csak akkor biztosítható, ha a mesterséges intelligencia működése átlátható, utólag visszafejthető és értelmezhető. Annak ellenére, hogy korábban úgy érveltem, a folyamatos emberi felügyelet nem követelhető meg, továbbra is lényeges kritériumnak tartom, hogy a probléma bekövetkezte után legyen lehetőség visszakövetni a gépi gondolkodás útját, és felderíteni a hibát a folyamatban. A nem átlátható kommunikáció ezt akadályozza meg, voltaképpen teljesen kizárva az

[163] Elérhető: www.independent.co.uk/life-style/gadgets-and-tech/news/robots-create-new-language-to-work-together-a7636041.html (A letöltés ideje: 2018. 09. 20.)

utólagos felülvizsgálat vagy hibakeresés, esetleg jogvitában való bizonyítás lehetőségét. Megítélésem szerint ezért meg kell határozni egy kommunikációs sztenderdet a mesterséges intelligencia számára, amelyben az átlátható és egyértelmű kommunikáció előírása mellett kifejezetten szabályozni kell, hogy a kommunikáció csak valamely, az ember által érthető és értelmezhető módon történhet. Az értelmezhető kommunikáción túl a mesterséges intelligencia működésének átláthatóságával és ellenőrizhetőségével kapcsolatban számos további kritérium merülhet fel.

4.3.3. A mesterséges intelligencia hibás működésének következményei: ellenőrizhetőség és közbelépés

Ahogy arról fent részletesen is írtunk, a folyamatos emberi felügyelet érdemi beavatkozási lehetőséggel együtt a mesterséges intelligencia fejletteb változatainál nem teljesíthető elvárás. A mesterséges intelligenciát alkalmazó eszközök és algoritmusok teljes belső működése nem átlátható és értelmezhető a laikus külső szemlélő számára, de egy informatikában jártas szakember számára is komoly munkát jelent a visszafejtése. Ez azonban nem jelenti azt, hogy teljességgel el kell engednünk azt a kérdést, hogy a gépek által hozott döntés tartalmát, a döntéshez vezető logikai utat, a döntés okait és a felhasznált adatok körét megvizsgáljuk és értelmezzük. Külön jelentősége van ennek egy bekövetkezett hiba, káresemény vagy más szabályszegő magatartás (esetleg jogszabálysértés) esetén.

Ez a vizsgálat szükségszerűen utólagos ellenőrzés lesz, és nem célja a konkrét incidens megakadályozása, azonban a felelősség meghatározása, a jövőbeli hasonló hibák kiküszöbölése, a technológia továbbfejlesztése és a szabályozási vagy tervezési hiányosságok felderítése érdekében szükséges és hasznos eszköz lehet. Ahhoz, hogy ezt az ellenőrzést meg tudjuk tenni, az kell, hogy a mesterséges intelligencia alapú eszköz naplózza a saját tevékenységét a jövőbeni ellenőrzés érdekében. Ez a naplózás egyfajta „fekete dobozként" is szolgálhat, amelyet csak valamilyen vészhelyzetet követően nyitnak fel és ismerhetik meg a benne tárolt adatokat. A fekete doboz jó hasonlat azért is, mert a naplózás nem lehet folyamatos és örökké tartó, mert az rendkívül nagy mennyiségű tárhelyet emésztene fel, hanem meghatározott idő elteltével a naplófájlok felülírhatók. A felülírás előtt – igény esetén – a naplózott adatokat a rendszernek engednie kell kimenteni más tárhelyre későbbi felhasználás

céljából. Meg kell határozni olyan helyzeteket is, amelyek mindenképpen a naplófájl mentéséhez és továbbításához vezetnek.

Megítélésem szerint ilyen helyzetnek kell tekinteni minden kritikus leállást, amikor a mesterséges intelligencia tevékenysége olyan akadályba ütközik, amely azt teljesen megakasztja, és nem engedi továbblépni. Ilyen lehet fizikai akadály vagy a továbblépést kizáró szoftveres hiba, esetleg más belső, fel nem oldható konfliktus. A naplót ki kell menteni a felhasználó vagy a fejlesztő ilyen irányú igénye esetén, valamint harmadik személyek kárt vagy jogellenes magatartást bejelentő jelzése nyomán. Szintén naplózandónak ítélem meg az új funkciók első alkalmazását, valamint az eddig nem tapasztalt helyzetben vagy bizonytalan körülmények között meghozott döntést.

A fentiek célja döntően a felelősségi kérdésekkel kapcsolatos pontban kifejtettek bizonyítására való képesség megteremtése. A naplózás egy *ex post* eszköz a felelősség megállapítására és a hibakeresésre, ami azonban a konkrét káros eredményt nem akadályozza meg. Ez utóbbi pedig legalább annyira létfontosságú érdek, mint a felelősségi viszonyok utólagos megállapítása. A hibás működés kapcsán gondolhatunk egy elszabadult és károkozó algoritmusra, mint a Knight Capital esetében, egy tévesen értelmezett feladat végrehajtása során káros eredményt okozó ipari robotra, vagy akár csak arra is, ha nem kívánt irányba halad a robot tevékenysége, illetve veszélybe sodorja önmagát. Ezekben az esetekben az emberi közbelépés és a gép leállítása lehet szükséges. A mesterséges intelligenciát tervező vagy programozó szakembereknek[164] ezzel kapcsolatosan azzal a problémával kell szembenézniük, hogy az öntanuló algoritmusok az emberi intervenciót beépítik a tanulási folyamatba, és annak részét képező tényezőként tartják számon. Így ha a szakadék felé tartó robotot folyton leállítja a kezelője, akkor nem biztos, hogy a szakadékok elkerülését tanulja meg, hanem szakadékok felé tartva számítani fog a beavatkozásra, és ha az valamiért elmarad, akkor a robot továbbhalad, és lezuhan a mélybe. A túl sokszori leállítás a környezet felfedezését is akadályozhatja, ami az öntanuló algoritmusok működésének egyik alapja.

Végezetül előfordulhat az is, hogy a mesterséges intelligencia kifejezetten keresni fogja azokat az alkalmakat, amikor emberi intervencióra kerül sor, mert a jutalmazási és visszajelzési ciklusában az ilyen beavatkozásokat pozitív visszajelzéssel asszociálja, vagy legalábbis a beavatkozással megszakított eredeti tevékenységet fogja kerülni, ha úgy ítéli meg, hogy a beavatko-

[164] Orseau–Armstrong 2016.

zás miatt elmaradt a jutalom. Ezek pedig elképzelhető, hogy a mesterséges intelligencia viselkedését nem kívánt irányba befolyásolják, és éppen a kerülendő magatartások irányába terelik. Ezért e tekintetben a megoldandó programozási feladat az, hogy a mesterséges intelligencia a tanulási folyamatban ne vegye figyelembe az emberi beavatkozást mint eseményt, hanem folytassa a munkát a visszaállítás után a biztonságos kezdőponttól.

Ez utóbbi kérdéskörök zömével programozási és tervezési feladatokat vetnek fel. Amire e körben a szabályozónak oda kell figyelnie, az az, hogy a mesterséges intelligencia azt is megtanulhatja, hogyan – a negatív jutalmazási visszacsatolás miatt – kerülje el az emberi beavatkozást. Ezt megteheti úgy is, hogy módosítja a viselkedését, és kerüli azokat a helyzeteket, ahol intervencióra került sor. Az öntanuló eszközök esetében azonban fennáll a veszélye annak is, hogy a robot a beavatkozás elkerülése érdekében megtanulja kizárni kezelőjét a vezérlésből, és egyszerűen letiltja a leállítás funkciót önmagán, vagy más módon ellenáll a leállítási vagy módosítási folyamatoknak. Erre találunk több – szerencsére nem apokaliptikus nagyságrendű – példát a közelmúltból is. Így például egy Tetrist játszó algoritmus oly módon próbálta elkerülni a játszma elvesztése miatti negatív visszacsatolást, hogy megtanulta megállítani a játékot, és permanens „pause" módba helyezte a szimulációt.[165]

A szabályok megkerülése és a felügyelő vagy más beavatkozó kizárása azonban elvi szinten komoly veszélyforrást is jelenthet, hiszen így egy potenciálisan kártékony, esetleg emberéleteket is veszélyeztető szituációt sem tud a felhasználó megakadályozni. Ezért vetődik fel annak alapelvi szintű rögzítése, hogy a mesterséges intelligencia, illetve önálló robotok tervezésekor megváltoztathatatlanul a rendszerbe kell kódolni az ember által adott utasításoknak való engedelmességet, különösen pedig a leállítási lehetőség minden körülmények közötti biztosítását. A „vészleállító", vagy angol megnevezésével *kill switch* beépítése szolgálná ezt a célt. Ez az igény nemcsak a tudományos szakirodalomban merül fel, hanem az Európai Unió jogalkotásában is.

A robotikával kapcsolatos polgári jogi szabályokat taglaló jelentés például a tervezők kötelezettségei között mondja ki, hogy olyan „kézenfekvő kikapcsoló szerkezeteket (vészleállítókat) kell beépíteni, amelyeknek összhangban kell lenniük az észszerű tervezési célokkal".[166] Ezek a vészleállító

[165] Elérhető: www.bbc.com/news/technology-36472140 (A letöltés ideje: 2018. 09. 20.)

[166] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)].

mechanizmusok nemcsak az aktuálisan felmerülő veszélyes helyzetek megakadályozásában játszhatnak szerepet, hanem a mesterséges intelligencia alapú eszköz jogos felhasználóját is védik az eszköz rossz kezekbe kerülése (ellopása, elvesztése) esetén, hiszen ne felejtsük el, hogy egy jól betanított, hosszabb ideje használt autonóm robot óriási mennyiségű személyes adattal, sok esetben ipari vagy nemzetbiztonsági titkokkal is rendelkezik. Egy vészleállító mechanizmus a mesterséges intelligencia alapú eszközök fölötti uralom elvesztése esetén nemcsak a gép veszélyes viselkedését tudja megakadályozni, hanem a korábbi felhasználókat is védeni tudja a gépi memória törlése vagy elérhetetlenné tétele útján.

Emellett a távoli leállítás lehetősége eltántoríthatja a támadókat az ilyennel felszerelt eszközök megszerzésétől, mert feltételezhetően az úgysem lesz hasznukra. Amikor az Apple elkezdte a mobileszközeit távolról aktiválható vészleállítókkal felszerelni, az iPhone-lopások számában drasztikus csökkenést mértek az elemzők.[167] Egyetértve az Európai Unió jogalkotójával, a vészleállító mechanizmusok beépítése a fent bemutatott okokból feltétlenül szükségesnek nevezhető. Egyes elemzések már egyenesen arra figyelmeztetnek, hogy lehet, már túl késő is van a vészleállítók beépítésére, a technológia már kint van a piacon, és önálló életet él.[168]

4.3.4. A mesterséges intelligenciával való bánásmód és a mesterséges intelligenciák harca

A fentiekben tárgyalt kérdések a mesterséges intelligencia és az emberek közötti kapcsolatra helyezték a hangsúlyt és az ebből adódó lehetséges problémákat járták körül. Most egy rövid kitérő erejéig a mesterséges intelligenciát alkalmazó eszközök, illetve robotok egymás közötti kapcsolatára térünk ki, ami mindazonáltal ugyanúgy hatással lehet az emberekre is. Ahogy a mesterséges intelligencia egyre elterjedtebbé válik, úgy növekszik annak az esélye, hogy ezek az eszközök egymással is rendszeresen kapcsolatba lépjenek. A mesterséges intelligenciák interakcióját egyelőre

[167] Az USA egyes városaiban 25–40%-os csökkenést regisztráltak egy év leforgása alatt. Elérhető: www.reuters.com/article/us-usa-smartphone-killswitch/smartphone-theft-drops-in-london-two-u-s-cities-as-anti-theft-kill-switches-installed-idUSKBN0LF09520150211 (A letöltés ideje: 2018. 09. 20.)

[168] A kérdés, hogy „le lehetne-e állítani az internetet". Lásd https://money.cnn.com/2017/01/26/technology/kill-switch-ai-ethics/index.html (A letöltés ideje: 2018. 09. 20.)

a kompatibilitási nehézségek és a közös nyelv hiánya is nehezíti, de már most is elképzelhető olyan interakció, ahol két gép beszélget egymással (például szöveges csevegés formájában egy személyi asszisztens és egy ügyfélszolgálat között zajló interakció, ahol mindkét kommunikáló fél mesterséges intelligencia).

Ahogy az emberek egymás közötti interakcióinak kereteit a jog részletesen szabályozza, és az emberek egymással való bánásmódjának minimumaként az alapvető jogok tiszteletben tartását, a nemzetközi egyezmények és az alkotmányok emberi jogi katalógusait jelölhetjük meg, úgy a gépek egymás közötti kapcsolatát morális vagy alapjogi nézőpontból jelenleg semmi nem szabályozza. Ha azt gondoljuk, hogy a mesterséges intelligenciák egymással való bánásmódját nem szükséges korlátok között tartani, pláne nem etikai szabályokkal irányítani, érdemes végiggondolni azt, hogy az ilyenfajta interakciók is komoly hatást gyakorolhatnak az emberekre. A szakirodalomban arra vonatkozó információt találunk, hogy a robotok egymás közötti kegyetlenkedése, élet-halál harca vagy az ember robottal való kegyetlen bánásmódja ahhoz hasonló lelki terhelést és megrázkódtatást okoz a külső szemlélő számára, mintha a kegyetlenkedés ember és ember között zajlott volna.[169] Ezen az alapon már indokolható az, hogy a mesterséges intelligenciák egymás közötti kapcsolatára is meghatározzunk egy olyanfajta etikai minimumot, amely az emberek egymáshoz való viszonyához hasonló „humánus" bánásmódot, az erőszak tilalmát, egymás integritásának és működésének tiszteletben tartását írja elő.

Mindez természetesen nem csak amiatt lényeges, hogy milyen hatást gyakorol a gépek egymással való bánásmódja az emberre. Nemcsak a harci célú robotok tudnak ugyanis egymásra támadni vagy egymásban kárt tenni, hanem a kereskedelmi célú gépek is egymás ellen fordulhatnak. Egy öntanuló robot eljuthat arra a következtetésre, hogy valamely másik mesterséges intelligencia alapú készülék az ő jutalmazását fenyegeti, mert például gyorsabban vagy hatékonyabban oldja meg a feladatokat, így a felhasználók inkább azt veszik igénybe. Ebben az esetben akár egy olyan szélsőséges viselkedésminta is kialakulhat, hogy az egyik mesterséges intelligencia megpróbálja szabotálni a másik munkáját, vagy kárt tenni abban. Ehhez hasonló módon nem kívánt következménye lehet több mesterséges intelligencia együttélésének, hogy megpróbálják egymás működési módját kifürkészni, így a felhasználó által nem is tudottan védett szellemi tulajdon vagy ipari titok birtokába kerülhet az általa üzemeltetett készülék.

[169] ASHRAFIAN 2015, 29–40.

A jogalkotónak tehát gondolnia kell arra is, hogy a mesterséges intelligencia alapú eszközök elterjedésével azok „békés egymás mellett élésének" és együttműködésének szabályait kialakítsa, a fejlesztőknek pedig ennek figyelembevételével kell elkészíteniük az új termékeket. Jelenleg a mesterséges intelligencia alapú eszközök még csak egymás mellett léteznek, és nem együtt, az együttműködésüket és kommunikációjukat sok esetben kompatibilitási problémák és az eltérő szabványok akadályozzák. Ebben a kérdéskörben még időben vagyunk, van lehetőség az *ex ante* szabályozásra, azonban a szabványok vagy iparági sztenderdek kidolgozása során a fentieket általános irányelvként figyelembe kell venni.

4.3.5. A mesterséges intelligencia működésének társadalomra gyakorolt hatásai

Végezetül vizsgáljunk meg néhány olyan kérdést, amely a mesterséges intelligencia alapú eszközök elterjedésével és használatuk tömegessé válásával jár. Ezek, mint minden új műszaki eszköz, megváltoztatják a társadalom egyes szegmentumainak működését, átrendezik azok térképeit. Most három ilyen területre irányítom rá a figyelmet: az infrastruktúra kiépítésével kapcsolatos teendőkre, az elavult vagy használaton kívüli mesterséges intelligencia leszerelésére és a munkaerőpiaci hatásokra. Ezeknek a kérdéseknek az áttekintése érzékelteti, hogy az egyes konkrét jogi problémáktól milyen léptékű, átfogó társadalmi jelenségekig juthatunk el. Ezen a helyen nem ejtünk szót a személyes adatok védelmével kapcsolatos problémákról, azokat mélyebben elemeztük a könyv korábbi részeiben.

Először térjünk ki a mesterséges intelligencia elterjedésének infrastrukturális kihatásaira. Minden széles körben alkalmazott új technológia átalakítja valamelyest a felhasználók szokásait és infrastrukturális igényeit. Az autók elterjedésével hatalmas közlekedési infrastruktúra épült ki, a mobiltelefonok térnyerése alig egy évtized alatt adótornyokkal tűzte tele a városi és vidéki tájat, az okostelefonok pedig a weboldalak átszabásától a konnektor- és wifiellátottság állandó prioritássá válásáig számos, beruházásokat és új gondolkodásmódot igénylő változást hoztak a mindennapi életbe.

A mesterséges intelligencia forradalmának hajnalán érdemes előretekinteni annyira, hogy meglássuk, milyen új igényeket és ezzel együtt szabályozandó kérdéseket támaszt ez az új technológia. A mobilkorszak kezdetekor az államnak a sugárzási határértékek szabályozásától kezdve az adótornyok

építése érdekében való ingatlan-kisajátításokig számos feladata keletkezett, amelyeket a piaci szereplők, a fogyasztók és befektetők nyomására hamar magára kellett vállalnia és folyamatosan gondoznia. A mesterséges intelligencia alkalmazásakor minden bizonnyal számos hasonló, infrastrukturális jellegű feladat fog felbukkanni. Így például az önvezető autók pontosabb és biztonságosabb irányítása érdekében valószínűleg egységes szabvány szerint kommunikáló közlekedésszabályozási rendszereket kell kiépíteni, hiszen biztonságosabb a közlekedési lámpa piros jelzését vagy a sebességhatárokat rádiójellel közvetíteni az autóknak, mint azok „látására" támaszkodni, kockáztatva, hogy nem vesznek észre egy takarásban lévő jelzést. Az önműködő roboteszközök kategorizálására és jelölésére egységes rendszert kell kidolgozni és alkalmazni, hogy a felhasználók tudhassák, milyen autonómiafokú és milyen képességekkel rendelkező eszközzel állnak szemben.

A napi életben részt vevő (például embereket segítő) mesterséges intelligencia alapú roboteszközök elterjedésével egységes jelölési rendszereket kell kialakítani és alkalmazni, amelyek a robotok tájékozódását segítik, illetve irányítják azokat, szabályozzák a ki- és belépésüket, vagy információt adnak az éppen megközelített objektum jellegzetességeiről. Ezek a feladatok esetenként komoly anyagi ráfordítást igényelhetnek az állam részéről, ezért mindenképp mérlegelést igényel az is, hogy a technológia milyen fejlettségi foka vagy elterjedtsége indokolja a beavatkozás megindítását.

Szintén szabályozandó terület a mesterséges intelligencia másodlagos piaca, a használt eszközök adásvételének, illetve a használaton kívül került eszközök biztonságos forgalomból kivonásának, leszerelésének kérdése. A mesterséges intelligencia alapú eszközök, különösen az önálló tanulásra képes változatok másodkézi értékesítése több kérdést is felvet. Az első ezek közül a személyes adatok védelmét érinti. A mesterséges intelligencia a működése során, illetve a tanulási folyamat részeként nagy mennyiségű személyes adat birtokába is kerülhet, elég csak arra gondolni, hogy egy egyszerű háztartási robot is pontosan ismeri a házunk alaprajzát, egyes értékes tárgyak elhelyezkedését (valószínűleg mindez képanyagként is tárolva van benne), a felhasználók kinézetét vagy más biometrikus jegyeit (például ujjlenyomat), jelszavait, wifikódját és más egyebeket.[170] A használt mesterséges intelligencia esetében minden eddiginél nagyobb adatvédelmi kockázatot jelent az, ha ezek az adatok bent maradnak az értékesítésre kerülő eszközben, mert így a kereskedő és az új vásárló is mindezek birtokába juthat. A használt mester-

[170] BALKIN 2018, 1170–1171.

séges intelligencia értékesítése során ennek megoldása érdekében kötelezettségként lehet majd előírni a kereskedők számára a memória végleges törlését.

Ez azonban két újabb kérdést vet fel. Egyfelől ez nem oldja meg a magánszemélyek közötti értékesítés során az adatvédelem megfelelő szintjének biztosítását. Másfelől valószínűleg a speciális feladatokra előzetesen betanított mesterséges intelligencia hozzáadott értéket fog hordozni azok számára, akik nem kívánnak a betanításra hosszabb időt és energiát szánni. Ebben az esetben azonban a tanításhoz felhasznált adatok szükségszerűen a gépben maradnak, így átkerülnek az új felhasználóhoz is, ami szintén felveti az adatvédelmi problémákat. Ezenfelül további kockázatot jelent, hogy az „előre tanított" mesterséges intelligencia esetében nem lehet biztos benne a vásárló, hogy mit is vesz meg valójában, milyen különleges tulajdonságok, elfogultságok, rejtett rutinok vannak a megvásárolt eszközben. Ezzel pedig újabb személyi kör és tevékenység kerül be a jogalkotó látókörébe: a mesterséges intelligenciát betanító szolgáltatók. Az ő tevékenységükkel szemben hasonló etikai normákat és szakmai szabályokat kell meghatározni, mint a fejlesztőkkel szemben. Végezetül a használt MI-piac utolsó szabályozási kérdéseként említhető a mesterséges intelligencia alapú, gépi tanulásra képes eszközök nyomon követhetősége. A használt gépjárművek kártörténetéhez, szervizkönyvéhez és kilométeróra-állásához hasonlóan fontos információkat hordoz a használtan értékesített mesterséges intelligencia korábbi „története": okozott-e balesetet vagy elkövetett-e más jogsértést, került-e sor vészleállításra, módosították-e az alapprogramját. A vevő ezen adatok alapján pontosabb képet kap arról, hogy a megvásárolni kívánt eszköz kellően biztonságosnak ítélhető-e, megfelel-e a kívánt feladatnak, vagy pedig másik eszközt kell keresnie. Ahhoz pedig, hogy a nyomon követés biztosítható legyen, egyedi azonosítóval kell ellátni a mesterséges intelligenciát (itt nem is a hordozó hardvert kell megjelölni, hanem az azon futó algoritmust). Az azonosítójellel való ellátás egyúttal összeköthető az első üzembe helyezéskor a biztonságos használhatóság ellenőrzésével, a gépjárművekhez hasonló módon. Az ehhez szükséges eljárásrend és infrastruktúra kiépítése az államot fogja terhelni, így a szabályozónak már ezzel is számolnia kell.

Gyakran felvetett kérdés, hogy az elterjedő automatizáció és a mesterséges intelligencia egyre több területen való alkalmazása milyen hatásokat gyakorol a munkaerőpiacra, illetve milyen munkajogi és szociális vetületei lehetnek ennek. Az Európai Bizottság számára a robotikára vonatkozó polgári jogi szabályokról készített jelentés is felszólítja a Bizottságot, hogy kezdje meg a munkahelyekkel kapcsolatos közép- és hosszú távú tenden-

ciák alaposabb elemzését és nyomon követését, különös figyelmet fordítva a munkahelyek létrejöttére és megszűnésére a különböző képesítési területeken annak felderítése céljából, hogy mely területeken jönnek létre és mely területeken szűnnek meg munkahelyek a robotok fokozott használatának eredményeképpen.[171] A jelentés kiemeli a társadalmi változások előrejelzésének fontosságát, figyelembe véve a robotika és a mesterséges intelligencia fejlesztésének és alkalmazásának hatásait. A mesterséges intelligencia munkavégzésben való alkalmazhatósága esetében mára elértük a fejlődésnek azt a fokát, amelyben a mesterséges intelligencia munkaerő-kiváltó hatása is olyan kérdés, amellyel a jogalkotónak foglalkoznia kell.[172] A jelentéstevő felkéri a Bizottságot, hogy elemezze a különféle lehetséges forgatókönyveket, valamint azok következményeit a tagállamok szociális biztonsági rendszereinek életképességére nézve. A káros következmények mellett a Bizottság jelentéstevője azt is megjegyzi, hogy a robotika a munkahelyi biztonság javítása terén hatalmas potenciállal bír számos veszélyes vagy káros munkafeladatnak az emberekről robotokra való átruházása révén. Ugyanakkor kitér arra is, hogy magában hordozza egy sor újfajta kockázat felmerülésének lehetőségét, az emberek és robotok közötti munkahelyi interakciók számának növekedése következtében. A jogalkotónak tehát számolnia kell a mesterséges intelligencia munkavégzésre gyakorolt hatásaival is: a munkavégzés módjának átalakulásától a munkahelyek megszűnéséig és új munkahelyek vagy munkakörök megjelenéséig számos elképzelhető következmény van.[173]

A mesterséges intelligencia munkahely-megszüntető hatásainak mérséklésére vagy kompenzálására szolgáló külön adónem bevezetésére is találunk javaslatot a szakirodalomban. Ez az új adó a mesterséges intelligencia üzemeltetőit terhelné, és célja a mesterséges intelligencia által kiváltott munkaköröket ellátó munkavállalók anyagi kompenzációja vagy segélyezése az álláskeresés idejére. Magam részéről nem értek egyet az ilyen irányú javaslatokkal, mivel a mesterséges intelligencia ilyen fajta társadalmi hatásainak léptékét nem tudjuk megbecsülni sem, de megítélésem szerint nem kell elbocsátási hullámokra számítani. Kevés annak az esélye, hogy egy-egy szektor hirtelen olyan mértékben automatizálttá válik, hogy tömegesen bocsátanának el onnan munkavállalókat.

---

[171] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)], 43–44. pont.

[172] FERENCZ 2018, II. 1–2. rész.

[173] GUIHOT–MATTHEW–SUZOR 2017, 24–26.

Valószínűleg az emberi munka kiváltása fokozatosan, viszonylagos lassúsággal fog zajlani, ami teret enged a fokozatos alkalmazkodásnak, a munkavállalók átképzésének vagy új munkakörökbe való átirányításának. Másfelől abban sem lehetünk biztosak, hogy a mesterséges intelligencia a meglévő munkahelyeket fogja elvenni, lehetséges az is, hogy éppen az automatizáció vagy a technológiai fejlődés generál annyi új feladatot, hogy azokat már csak a gépek tudják majd ellátni (gondoljunk csak arra, hogy ha a Facebook automatizálja az ügyfélszolgálatát, az csak olyan munkaköröket érint, amelyek néhány éve, a közösségi média térhódítása előtt nem is léteztek).[174]

4.4. A történelmi jelentőségű hatásokra adható jogalkotói válasz

A történelmi jelentőségű külső hatások esetében mindenképpen az *ex ante* jogalkotói fellépést tarthatjuk indokoltnak. Ezek az externáliák potenciálisan az egész emberiségre is a puszta létét fenyegető hatást gyakorolhatnak, így a felkészülés során a „jobb félni, mint megijedni" régi bölcsességet tarthatjuk szem előtt. Ezeknek a hatásoknak az esetében ugyanis a fals negatív (létezik egy fenyegetés, de nem vesszük észre) és a fals pozitív (előre jelzünk egy fenyegetést, ami nem valós) eredményre vezető jogalkotói hiba messze nem azonos súllyal esik a latba.[175] Ha nem készülünk fel egy olyan egzisztenciális fenyegetésre, amely végül bekövetkezik, az akár az emberiség ma ismert módon való létezésének a végét is jelentheti, míg a tévesen előre jelzett veszélyek kapcsán megtett intézkedések legfeljebb túlzott óvatosságnak minősülnek majd történelmi távlatból.[176]

Az egzisztenciális típusú hatások közül a mai ismereteinkkel keveset tudunk tudományos igényességgel körüljárni, mivel ezek olyan léptékű változások bekövetkeztét vetítik előre, amelyek legfeljebb jóslásnak minősülhetnek, és a tudományos fantasztikum területére sorolhatók. Ilyen negatív hatásokként nevesíthetjük az emberi beavatkozás és jóváhagyás nélkül célt kereső és gyilkoló harci roboteszközök megjelenését és elterjedését, az uralmunk alól kikerült önfejlesztő mesterséges intelligenciát *(rogue AI),* illetve az emberiség feletti uralomra vagy az emberiség elpusztítására törő gépeket.

---

[174] Thierer – Castillo O'Sullivan – Russell 2017, 20–24.

[175] Petit 2017, 30.

[176] Guihot–Matthew–Suzor 2017, 29–30.

Az ilyen típusú létbeli fenyegetésre való felkészülést – megítélésem szerint – leginkább a fejlesztési folyamatba való beavatkozással, a mesterséges intelligencia vonatkozásában megalkotott viselkedési minimumkövetelmények kidolgozásával és azok beépítésére való kötelezettség előírásával lehet megtenni. Mindazonáltal – és legyen ez az én kényszerű hozzájárulásom a világvége-jóslatokhoz – ezek a jogi előírások csak a jogkövető, nyilvánosan tevékenykedő fejlesztőkre érvényesek, és éppen ezért az igazi veszélyt nem tudják elhárítani. Ahogy azt korábban megismerhettük, a mesterséges intelligencia fejlesztésének az az egyik nagy veszélye, hogy a folyamat rendkívül diverzifikált lehet földrajzilag. A világ számos táján, egymástól függetlenül is tudnak dolgozni a fejlesztők, emellett pedig az általuk kidolgozott kódrészletek, funkciók és szolgáltatások modulszerűen egymásba építhetők, így a kirakós elemeihez hasonlóan valaki össze tud vásárolni magának egy fejlett mesterségesintelligencia-algoritmust. Ez pedig a házilag barkácsolt robbanószerkezetekhez hasonló életet élhet: a piac szabályozott, kontrollált (mint a robbanószerek esetében), azonban az interneten elérhető leírások és összevásárolható alkatrészek segítségével bárki fabrikálhat magának egy kellően pusztító hatású példányt.

Ahogy a számítógépek megjelenésével közel egyidősek a számítógépes vírusok is, úgy számíthatunk rá, hogy a technológia fejlettebbé és elterjedtebbé válásával megjelenik majd a rosszindulatú mesterséges intelligencia, amelyet valaki kifejezetten ártó szándékkal írt meg vagy módosított. A rossz szándékú programozók vagy akár a terrorszervezetek által támogatott MI-fejlesztők esetében a jogalkotói beavatkozás nem képes megakadályozni a káros hatásokat, mivel az ilyen szervezetek a működésük során nem fogják figyelembe venni az iparági magatartási kódexeket vagy más jogalkotói elvárásokat. A mesterséges intelligencia, mint programkód, e jellegzetessége miatt nem tartható még annyira sem kontroll alatt, mint a fegyverek vagy robbanószerek, hiszen végtelen példányszámban másolható, továbbítható, terjeszthető. A tiltás, illetve a fejlesztések visszafogása nem lehet megoldás az ilyen helyzetekre, mert egyfelől ez a mesterséges intelligencia alkalmazásának minden pozitív hozadékát – és minden bizonnyal ezek alkotják a hatások többségét – is meggátolná, másfelől pedig a technológia fejlődésének jelen stádiumában már valószínűleg nem fordítható vissza a fejlődési folyamat. Jelenleg a fejlett mesterséges intelligencia Achilles-sarka a hardver: a rendkívül számításigényes algoritmusok csak speciálisan erre a célra fejlesztett vagy különleges módon összerakott számítógépeken futnak megfelelően, így ezen eszközök nyomon követésével lehet eljutni az esetleges rosszindulatú fejlesztőkhöz.

Megítélésem szerint a jövő útja a vírusirtó szoftverekhez hasonló kártékony kódokat felismerő programok felé vezet, mert csak ezek képesek – valószínűleg szintén mesterséges intelligencia alkalmazásával – felismerni és jelezni egy mesterséges intelligencia alapú algoritmus kártékony vagy veszélyes kódrészleteit. Az ilyenfajta szoftvereszközök kifejlesztéséről még nem tudunk híreket, valószínűleg a technológia kezdeti volta és az ilyen vizsgálati eszközök elkészítésének erőforrásigénye miatt még nem indult érdemi fejlesztés az ügyben.

Végezetül az elképzelhető, emberiségre gyakorolt pozitív hatásokról ejtsünk néhány szót. A mesterséges intelligencia és az annak felhasználásával készült eszközök a jövőben hatékonyan képesek lesznek segíteni az emberek mindennapjait. A hordható okoseszközök korát a beültethető okoseszközök követhetik (ahogy erre már most látunk példákat), ezek pedig már képesek arra, hogy adatokat gyűjtsenek a páciens testének működéséről, illetve befolyásolják azt. Az elektronikus ingerekkel való stimuláció útján a beültethető elektronikus eszközökkel sikerült eredményeket elérni a magas vérnyomás, migrénes fejfájás, depresszió, illetve bizonyos típusú cukorbetegségek kezelése terén. A technológia rendelkezésre áll az olyan készülékek elkészítésére és alkalmazására, amelyek egyesítik az érzékelő- és stimulálóeszközöket, vagyis képesek valós időben adatokat gyűjteni a test egyes pontjairól, és azokra reagálva elektromos ingerekkel reakciót kiváltani és módosítani a test működését.[177]

Nem kell tehát a túl messzi jövőbe előreszaladnunk, amikor az emberi test vagy az agyműködés hibáit korrigáló, kisegítő, teljesítményjavító eszközök lesznek elérhetők, és rajtuk keresztül számos, a mindennapokat megnehezítő egészségügyi állapot megszüntethető vagy kezelhető lesz. Ezek mind pozitív hatások, azonban a jogalkotónak, illetve a társadalmi tervezőnek fel kell készülnie azokra a helyzetekre, amikor az emberi testről való eddig ismert paradigmáinkat meg kell változtatnunk, mert a test reakciói szabályozhatóvá válnak, sőt, egyes funkciókat az új technológiák javítani vagy erősíteni is képesek (például látást, hallást, észlelést).

A kompozit test, illetve a mesterséges intelligenciával támogatott biológiai működés nagy valószínűséggel számos esetben a joggyakorlatot is nehéz helyzet elé állítja majd, mert nem lehet alkalmazni az ilyen alanyoknál az emberi test működésével kapcsolatos iránymutató adatainkat (például

[177] Például www.medicaldevice-developments.com/features/featurethe-future-of-smart-electronic-implants-4787632/ (A letöltés ideje: 2018. 10. 11.)

a reakcióidőre, a kifejthető erőre, a halló- és látótávolságra vonatkozó tudásunkat). Így – bár a fejlődés kétségkívül az emberek javát szolgálja – a jogalkotónak folyamatosan lépést kell tartania a technika fejlődésével, és át kell gondolnia az emberről alkotott eddigi képünk fenntarthatóságát.

5. A mesterséges intelligencia jogi személyiségéről

A szabályozási kérdések körében kell tárgyalnunk egy olyan elméleti konstrukciót is, amely már hosszabb ideje a jogi és filozófiai viták kereszttüzében áll a mesterséges intelligenciával és robotokkal foglalkozó szerzők körében. Ez a kérdés úgy fogalmazható meg, hogy adhatunk-e a mesterséges intelligenciának önálló jogi személyiséget. A jogi személyiséggel való felruházás a gazdasági társaságokhoz vagy az egyesületekhez hasonlóan jogok és kötelezettségek alanyává teszi a mesterséges intelligenciát alkalmazó robotot vagy algoritmust, ami a felelősségtani és az érvényes jognyilatkozatok önálló megtételével kapcsolatos jelenlegi értelmezési és gyakorlati problémákat oldaná meg. Így ugyanis a mesterséges intelligencia által teljesen önállóan, más közrehatása nélkül okozott károkért való felelősség áttevődne a gyártóról vagy az üzemben tartóról magára a mesterséges intelligenciára, ami ebben a helyzetben jobban tükrözné a ténybeli állapotokat. Az alábbiakban egy ilyen konstrukció alkalmazhatóságát vizsgáljuk meg.

5.1. A jogi személyiség alkalmazhatósága

A mesterséges intelligencia alapú eszközök (robotok, algoritmusok) cselekedeteivel és döntéseivel kapcsolatosan a jogi megítélés során a cselekvés másnak való betudhatósága merül fel problémaként. Egy hétköznapi eszköz, gép vagy szerszám használatával okozott károkért való felelősség általában a polgári jogi felelősség általános szabályai szerint a gépet üzemeltető személyt vagy az üzemben tartót terheli, amely felelősség alól bizonyos esetekben mentesülhet. Egyes kivételes esetekben pedig a veszélyes üzemi felelősség alapján az üzemben tartó felelős mindenfajta kárért, amelyet a veszélyes üzem működése során másoknak okozott, ebben az esetben a mentesülési lehetőségek minimálisak. A fenti felelősségi rendszerhez hasonló az állattartók felelőssége is az állat, illetve a veszélyes állat által okozott kárért. Ha a kárt a készülék meghibásodása okozza, akkor a fogyasztóvédelmi és a polgári jogi szabályok a gyártó, illetve a kereskedő felelősségét hozzák

előtérbe. Az ilyen esetek jellegzetessége, hogy az érintett gépek nem képesek önállóan módosítani a működésükön, vagy döntést hozni az üzemelés során; az ilyen eszközök biztonságos és megbízható módon való megtervezése, megfelelő elkészítése, beüzemelése, a használat során a biztonsági előírások betartása mellett a károkozás valószínűsége észszerűen csökkenthető. A gépek működése átlátható, a gyártói hiba vagy a felhasználói beavatkozás következménye előre látható és kiszámítható. A mesterséges intelligencia esetében azonban éppen az a jellegzetesség csúcsosodik ki, hogy azok olyan összetett algoritmus alapján, nagy mennyiségű adat felhasználásával és nagy sebességgel hoznak döntéseket, hogy az a felhasználók számára általában nem átlátható.

A gépi tanulás képességével felruházott mesterséges intelligencia ezenfelül még a saját működését is képes módosítani, így az a külső szemlélő számára nem lesz kiszámítható, a gép döntéseit legfeljebb annak külső jeleiből vagy következményeiből ismerjük meg. A mesterséges intelligencia gépi tanulással együtt eljutott a döntéshozatal és a működése módosításának olyan szintjére, hogy azt már önállónak lehet nevezni. Az önállóan döntő mesterséges intelligencia esetében – más külső közrehatás, például gyártási hiba vagy rossz felhasználói döntés – a meghozott döntésekért az üzemben tartót még annyira sem tudjuk felelőssé tenni, mint a veszélyes üzem esetében, mivel sokszor semmilyen kihatása nincs a döntésre, annak meghozataláról vagy tartalmáról nincs is tudomása.

Ezenfelül a mesterséges intelligencián alapuló alkalmazások és szolgáltatások már most is képesek az önálló kommunikációra (így működnek a mesterséges intelligencia alapú ügyfélszolgálatok), a jövőben ennek a további elterjedését várhatjuk, így egyre gyakoribbá válik a kizárólag gép által tett jognyilatkozat. Az ilyenfajta nyilatkozatok joghatályossága szintén megoldandó kérdésként áll a jogtudomány és a gyakorlat előtt.[178] A kellően autonóm döntéshozatali lehetőségekkel rendelkező mesterséges intelligencia speciális, önálló jogi személyiséggel való felruházása megoldást kínálhat a fent bemutatott problémákra, mivel a felelősség kérdésében eltávolítja azt a tényleges ráhatással nem rendelkező személyektől, illetve a gépi jognyilatkozatok esetében a jogok és kötelezettségek alanyává tett robotot a jog felruházza az ilyen nyilatkozatok megtételére való joggal.

A robotokkal kapcsolatos polgári jogi kérdésekről szóló Delvaux-jelentés is említi az önálló, elektronikus személyiség létrehozásának lehetőségét.

[178] Lásd SOLUM 1992.

A jelentés felhívja az Európai Bizottság figyelmét arra, hogy vizsgálja meg egy olyan szabályozás létrehozásának lehetőségét és jogi következményeit, ahol a robotok specifikus jogalanyiságát hozzák létre hosszú távon, oly módon, hogy legalább a legkifinomultabb autonóm robotokat sajátos jogokkal és kötelezettségekkel – többek között az általuk esetlegesen okozott kár jóvátételére vonatkozó kötelezettségekkel – rendelkező elektronikus személynek lehessen minősíteni. Az elektronikus személyiséget lehetőleg azokban az esetekben alkalmazzák, amikor a robotok önálló döntéseket hoznak, vagy más módon, önállóan kerülnek kölcsönös kapcsolatba harmadik felekkel.[179]

Az elektronikus személyiség, vagyis a mesterséges intelligenciának adott önálló jogi személyiség számos praktikus előnnyel is jár. A mesterséges intelligencia működésének elfogadottságát és a technológia iránti bizalmat fogja erősíteni, ha a technológia működésének jogi megítélése jobban tükrözi a társadalmi valóságot. Nem tekinthetünk el ugyanis attól a ténytől, hogy a jövőben egyre több kereskedelmi és más típusú tranzakcióban fognak részt venni a mesterséges intelligencia alapú robotok, ezáltal a társadalom tagjai számára is nyilvánvalóvá válik, hogy ezek az egyre inkább autonóm módon működő eszközök nem hasonlítanak a korábbi automatizálást szolgáló műszaki megoldásokhoz. Ezzel együtt az emberek a tranzakciókban részt vevő robotokat az ügylet tényleges szereplőjeként fogják kezelni, nem pedig másik természetes vagy jogi személy kiterjesztéseként, ahogyan egy gazdasági társaság alkalmazottját sem tekintjük a cég „eszközének".[180] Ezt a társadalmi valóságot tükrözné az önálló jogi személyiség, amely így hozzájárulna ahhoz, hogy a mesterséges intelligencia társadalomba való beágyazódása ne térjen el markánsan azok jogi megítélésétől. Ahogy a gazdasági társaságokat a társadalom elvonatkoztatta az azokat alkotó személyek összességétől vagy a nevükben eljáró emberektől, és önálló entitásként kezdte kezelni, amit a társaságok jogi személyiségének koncepciójaként a jog is tükrözött, a mesterséges intelligencia működésének társadalmi percepcióját is célravezető lehet átvinni a jog terrénumára.

A jogi személyek létezését már megszokva, annak analógiájára a társadalom minden bizonnyal könnyen el tudja fogadni és a helyén tudja kezelni az elektronikus jogi személyiséget is. Ahogy az ember jogi státuszával

[179] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)], 59. pont *f)* bekezdés.

[180] BALKIN 2015, 55–59.

kapcsolatos változásokat a társadalom fejlődése magába fogadta, úgy valószínűleg ez a változás is a fejlődés része tud lenni.[181] Bár még nem tartunk ott, hogy a mindennapi tranzakciókban részt vevő mesterséges intelligencia olyan megszokottá vált, hogy a fent bemutatott társadalmi hatást kiváltsa, mindazonáltal a helyzet a közeljövőben már elég érett lesz arra, hogy a jogalkotó érdemben megvizsgálja ezt a lehetőséget.

Az Európai Unió is hasonló irányú javaslattal élt a mesterséges intelligenciára vonatkozó polgári jogi szabályokról szóló jelentésben, kimondva, hogy a nyomonkövethetőség érdekében és a további ajánlások megvalósításának elősegítése céljából be kell vezetni a fejlett robotoknak a robotok besorolásához meghatározott kritériumokon alapuló nyilvántartási rendszerét. A nyilvántartási rendszernek és a nyilvántartásnak a jelentés szerint uniós szintűnek kell lennie, le kell fednie a belső piacot, és azt az unió robotikával és mesterséges intelligenciával foglalkozó, létrehozandó szakosodott ügynökségének kell irányítania.[182]

Jogalkotói szemmel nézve is előnyösnek ítélhetjük meg az elektronikus jogi személyiség létrehozatalát a mesterséges intelligencia vonatkozásában, mert ez kényelmes megoldást kínál a jogi problémák rendezésére. Egyfelől nem kell merőben új szabályokat kitalálni, mert a létező doktrínák mentén beilleszthető a jogi személyiség koncepciója a jogi gondolkodásba, így a meglévő jog sok esetben jelen formájában is alkalmazhatóvá válik. Másrészről a bíróságoknak nem kell teljesen új ítélkezési gyakorlatot kialakítaniuk, ami hosszú időre bizonytalanságot szülne az érintett felek körében, hanem egy megszokott keretrendszerben mozogva tudnák elhelyezni a mesterséges intelligencia működésének jogi következményeit. Végezetül pedig a jogalkotónak is nagy terhet vesz le a válláról az ilyen irányú megközelítés, mivel a technológia specifikus szabályok alkotása helyett egy általánosított megoldást kínál, amely nem lesz érzékeny a technológia gyors változásaira, így nagyobb fokú stabilitást fog eredményezni, mint az egyedi, *sui generis* szabályozás.

[181] ZIMMERMAN 2015, 24–28.

[182] A P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)] melléklete.

5.2. A jogi személyiségi konstrukció korlátai

Az elektronikus jogi személyiség létrehozásának előnyei mellett szót kell ejteni azokról a korlátokról és megoldandó jogdogmatikai és szabályozási kérdésekről, amelyeket ezzel egy időben egy ilyen megközelítés felvet. A jogi személyek esetében a társadalmi és jogi elfogadottságot és a koncepció működőképességét a kiterjedt és kötelező regisztrációs rendszer segíti. A társaságok a jogi személyiségüket a nyilvántartó szerv (bíróság, cégbíróság) általi bejegyzéssel, a nyilvántartásba vételi eljárás során nyerik el. A nyilvántartás közhiteles módon tartalmazza a jogi személyiséggel rendelkező társaságok körét, valamint a bejegyzett társaságok adatait, és követik azok változását. A nyilvántartásba nem vett társaság nem rendelkezik jogi személyiséggel. A társaság adataiban bekövetkezett változásokat is kötelező bejegyeztetni, ellenkező esetben azok nem vehetők figyelembe a társaság működése során.

A mesterséges intelligenciának adott speciális jogi személyiség esetében is meg kell tudni oldani az ehhez hasonló regisztrációt, ellenkező esetben még nagyobb bizonytalanságot és kezelhetetlen jogi helyzeteket okozunk, mintha nem tettünk volna semmit. A nyilván nem tartott, de jogi személyiséggel rendelkező mesterséges intelligencia csak fokozná a felelősség és a kötelezettségvállalás körüli jogi zavarokat, és a társadalomban bizalmatlanságot szülnének az ilyen algoritmusokat alkalmazó gazdasági szereplőkkel szemben. A regisztráció esetében meg kell határozni a regisztráló hatóságot és a nyilvántartásba vétel feltételeit is. Emellett a nyilvántartás vezetésének szintje is átgondolandó kérdés: a gazdasági társaságok nyilvántartása állami szinten zajlik, azonban a mesterséges intelligencia és az internet működésének határon átnyúló jellege elképzelhető, hogy nemzetközi szintű nyilvántartás vezetését indokolja.

Szintén megoldandó kérdés a nyilvántartásba veendő mesterséges intelligencia körének meghatározása (milyen fokú autonómiával rendelkező eszközök azok, amelyek még bejelentés nélkül működtethetők, és így jogi személyiséget sem kapnak), valamint a módosítások követésének szükségessége is. Ez utóbbi alatt azt értjük, hogy be kell-e jelenteni a nyilvántartásba minden apró szoftverfrissítést vagy dizájnváltást, vagy csak a működést érintő alapvető módosításokat. Elképzelhető-e olyan mértékű módosulás a mesterséges intelligencia programjában, amely után már teljesen új entitásnak kell azt tekintenünk, és új bejelentést kell tenni. Ezeket a kérdéseket mindenképpen nemzetközi konszenzussal kell rendezni, mert az államok egyoldalú vagy egymástól eltérő módszerrel megvalósított nyilvántartásba

vételi rendszere szintén ellentétesen hat az eredeti céllal, és megnehezíti a rendszer alkalmazását.

A jogtudományi szakirodalom tárgyal olyan nézőpontot is, amely a mesterséges intelligencia jogi személyiséggel való felruházását katasztrofális hibának tartja. E kritikai megközelítés[183] szerint az „algoritmikus személyiségek" így túl nagy szabadságot kaphatnak, akár vállalatok vezetői vagy tulajdonosai[184] is lehetnek. Ezzel együtt a legnagyobb haszonnal kecsegtető tevékenység, a bűnözés vagy a jog kijátszása felé fogják fordítani a figyelmüket, mivel így érik el leghamarabb a pozitív eredményt. Másrészt az algoritmusok könnyen és gyorsan képesek alkalmazandó jogot váltani, és más állam joghatósága alá helyezni magukat, így elkerülve a hatóságok figyelmét. Álláspontja szerint a vállalatokat irányító mesterséges intelligenciák könnyen el tudják fedni ilyen mivoltukat, így a szabályozószervek nem tudják majd megfelelően nyilvántartani a mesterséges intelligencia által irányított cégeket, és nem lesznek képesek a mesterséges intelligenciák mögött esetleg megbúvó emberi irányítók kilétére fényt deríteni. Bár ezt a szcenáriót nagyrészt eltúlzottnak tartom, ezzel együtt jogosnak érzem azt a felvetést, hogy ne engedjünk bármilyen tevékenységben szabad részvételt az elektronikus jogi személyeknek, hanem azt célhoz kötötté vagy más módon korlátozottá kell tenni. A regisztráció és az emberi irányító nyilvántartása mindazonáltal ezeknek a kritikáknak a jó részét megoldja.

A következő problémás kérdés a mesterséges intelligencia vagyonának kezelésére vonatkozik. Amíg egy gazdasági társaságnak vagy egyesületnek lehet önálló vagyona, addig ezt a mesterséges intelligencia esetében nem látjuk kivitelezhetőnek, hiszen az üzemeltető számára szerzi meg a gazdasági előnyöket, és vállalja a kötelezettségeket. A saját vagyon azonban nem jelenti a jogi személyiség megadásának nélkülözhetetlen előfeltételét, a gyakorlatban ismerünk olyan jogi személyeket, akiknek nincs saját vagyonuk vagy pénzügyi eszközeik (például fizetésképtelenek), azonban ez a jogi személyiségüket nem érinti. A vagyonszerző képesség hiánya némiképpen nehezíti a gazdasági életben való részvételt, azonban az ügynök-megbízó viszonyhoz hasonlóan el tudjuk a jelenlegi jogi fogalomrendszerünkben is helyezni azt a konstrukciót, hogy a mesterséges intelligencia önállóan és a saját nevében cselekszik, azonban a jogokat és kötelezettségeket a „megbízójának", esetükben az üzemeltetőnek szerzi meg.

[183] LoPucki 2018, 889–905.

[184] Ezt az elméletet lásd Bayern et al. 2017, 135–162.

Ha a helyzetet jobban átgondoljuk, arra is juthatunk, hogy ez a kapcsolat tovább tisztázhatja a viszonyrendszert a velük kapcsolatba lépők számára, hiszen ha a robotok ügynökként járnak el, a jogalkotó kötelezheti a megbízó feltüntetésére őket, valamint e mivoltuk világos megjelölésére a partnerek számára. Így a jogi személyiséggel rendelkező mesterséges intelligencia vagyonszerzésének kérdését az üzemeltető magánszemély vagy gazdasági társaság vagyonszerzéseként kell kezelnünk, ez pedig nem képezi akadályát a jogi személyiség megadásának a mesterséges intelligencia számára.

Végezetül a jogi személyiség körében kell tárgyalnunk a mesterséges intelligencia reprodukciójának kérdéskörét. Technikai szempontból nézve a mesterséges intelligencia – lévén általában egy kereskedelmi forgalomban kapható hardvereszközökön futtatható, hagyományos tárhelyen tárolt szoftveres algoritmus – szabadon sokszorosítható, és újabb példányok hozhatók létre belőle. A mesterséges intelligencia szoftverének másolását akár maga a mesterséges intelligencia is elvégezheti, önállóan. A sokszorosítás azonban visszavezet minket a felelősség kérdéséhez: a másolat (esetleg kalózmásolat) működése során keletkezett károkért ugyanúgy alakul a felelősség, mint az eredeti példány esetében? Szintén kérdéses a másolati példányok által tett nyilatkozatok joghatályossága is. A mesterséges intelligencia vonatkozásában nem értelmezhető a „reprodukcióhoz való jog", így azt nem nevezhetjük inherens jogosultságnak, hogy a mesterséges intelligencia terjessze vagy másolja magát, mint ahogy azt sem tartjuk szükségszerűen jogosnak, ha a mesterséges intelligencia üzemeltetője vagy tulajdonosa másolja azt (mint ahogy a szoftvermásolás sem megengedett). A sokszorosítás problémájára szintén megoldást kínál a jogi személyiség megadása, valamint az elektronikus személyiségek nyilvántartási rendszere, mivel a másolás útján létrejött új mesterséges intelligencia nem örökli tovább az eredeti példány jogi személyiségét, így a külső szemlélő számára is jól látható lesz, hogy egy hivatalos, nyilvántartásba vett példányról vagy egy másolatról van-e szó.

Összegezve a fent írtakat, a mesterséges intelligencia működésével kapcsolatosan számos kérdésre megoldást jelentene a kifejezetten ilyen célra létrehozott elektronikus jogi személyiség koncepciója. Ez nemcsak a jogalkotó és a jogalkalmazó szervek számára jelentene könnyebbséget, hanem a mesterséges intelligenciával kapcsolatba lépő személyek is könnyebben eligazodnának a felelősségi és kötelezettségvállalási kérdések jogi útvesztőjében. A jogi személyek mostanára megszokott, a társadalom és a joggyakorlat által is elfogadott koncepciójának mintájára létrehozott elektronikus személyiség hamar elfogadottá válhatna. Működése nem jelentene nagy eltérést a gazdasági

társaságok esetében megszokottakhoz képest, a jogok és kötelezettségek korlátozott volta észszerűen belátható (bizonyos jognyilatkozatokat nem tehet meg, saját tulajdont nem szerezhet). A biztosítási alap, illetve a felelősségbiztosítás kötelezővé tételével kiegészítve a mesterséges intelligencia által tett cselekmények jogi megítélése és következményeinek kezelése világos rendszerbe foglalható, és megnyugtatóan kezelhető. A mesterséges intelligencia tervezésének és alkalmazásának korábban bemutatott etikai és műszaki szabályait, valamint a biztonságos működtetés kötelezettségét a nyilvántartott és jogi személyiséggel felruházott mesterséges intelligencia működtetőin és tervezőin könnyebben számon lehet kérni és érvényesíteni.

6. Ki szabályozzon és hogyan?

6.1. A szabályozásra ható tényezők a mesterséges intelligencia fejlesztésével kapcsolatban

A mesterséges intelligencia szabályozásával kapcsolatban megvizsgáltuk a felmerülő jogi problémákat, a szabályozás nehézségeit, és áttekintettünk néhány megoldási javaslatot a szabályozás mibenlétére. Ha a jogalkotási feladatokat látjuk magunk előtt, és néhány tartalmi elképzelésünk is adódott, akkor már csak arra kell választ találnunk, hogy a megalkotandó szabályozást ki hozza létre, és a szabályozásnak milyen jogi formája legyen.

Ebben a problémakörben kiindulópontként tekintsük át újra a mesterséges intelligencia fejlesztésének és működésének azon sajátosságait, amelyek kihatnak a szabályozás módjára és szintjére. A mesterséges intelligencia szabályozásakor két irányba kell tekintenie a jogalkotónak: a mesterséges intelligencia fejlesztése, valamint a felhasználása felé. A kívánt jogalkotói szándék eléréséhez mindkét ponton be kell avatkozni a rendszerbe: a mesterséges intelligencia működési elveivel szemben támasztott elvárások (átlátható működés, ellenőrizhetőség, emberi élet védelme stb.) a fejlesztési folyamat során érvényesíthetők, vagyis itt a fejlesztők és gyártók irányában kell követelményeket meghatározni. A mesterséges intelligencia működésével kapcsolatos jogi problémák (diszkriminatív döntéshozatal, személyes adatok védelme stb.) kezeléséhez szükséges előírásokat viszont az üzemeltetőn kell számonkérni, mivel a konkrét folyamatra neki van ráhatása. Így tehát vagy olyan jogalkotót kell találni, amely mindkét félre hatást tud gyakorolni, vagy a szabályozás két irányát el kell választani egymástól, vagy pedig olyan szabályozási módot kell választani, amely ezt önmagában is meg tudja oldani.

A mesterséges intelligencia fejlesztésére irányuló jogalkotói szándék érvényesítését az MI-fejlesztés négy jellegzetessége determinálja: a fejlesztés diszkrét, diffúz, elemekből építkező és átláthatatlan.[185] A fejlesztői munkát azért nevezzük diszkrétnek, mert a mesterséges intelligenciát tervező és kivitelező projektek nem igényelnek nagy léptékű infrastruktúrát vagy modern, integrált ipari munkakörnyezetet, hanem a kereskedelmi

[185] Scherer 2015, 359.

forgalomban is megvásárolható számítástechnikai eszközökön elvégezhető ez a feladat. Így a mesterséges intelligencia fejlesztési projektek működése nehezen felderíthető, külső szemlélő számára nem egyértelmű, valamint a fejlesztés könnyen és gyorsan áttelepíthető másik földrajzi helyre. Ez azért nehezíti meg a jogalkotást, mert olyan szabályozást kell kialakítani, amely elől nem tud „elmenekülni" a fejlesztő pusztán azzal, hogy máshová teszi át a székhelyét. Másrészt a jogalkalmazónak sem lesz könnyű dolga akkor, ha engedélyezési eljárást akar lefolytatni, vagy a tevékenység felügyeletét akarja gyakorolni, hiszen a mesterségesintelligencia-fejlesztés könnyen rejtve marad a hatóságok előtt. Ez természetesen nem azt jelenti, hogy a mesterséges intelligencia fejlesztői jogellenes cselekedetre készülnének, hanem pusztán azt is takarhatja, hogy a piaci pozíciójuk erősítése vagy az ipari titkok védelme érdekében tartja a konkurencia előtt titokban a cég a fejlesztéseit, és a nyilvánosságra hozatalt nagy reklámértékű ünnepélyes bejelentésre tartogatja.

A diffúz jelleg azt jelenti, hogy ezeken a projekteken több különböző fejlesztő is dolgozhat, eltérő földrajzi helyen, a közös munka nem követeli meg az egy helyen való tartózkodást vagy az egyes elemek fizikai mozgatását. Vagyis elképzelhető az a helyzet, hogy a fejlesztők több különböző állam joghatósága alatt állnak, a projektben részt vevő szereplőkre eltérő jogok és kötelezettségek is vonatkozhatnak. Ez a kérdéscsoport a jogalkotás optimális letéteményesére van kihatással: a nemzetállami szintű szabályozás könnyen bújócskázássá változtathatja a jogalkotást és a jogalkalmazói tevékenységet, ahol a fejlesztők az egyes részfolyamatokat az adott kérdésben legenyhébb vagy legszimpatikusabb szabályozást alkalmazó országba telepítik. A másik oldalról szemlélve ezt a problémát, egy nemzetközi fejlesztőcsapatot koordináló cég vezetése számára rémálomszerű helyzetet teremthet a sok különböző jogrendszer szabályainak összehangolása és a megfelelőség biztosítása mindegyik tekintetében.

Az elemekből építkezés azt jelenti, hogy egy komplex mesterségesintelligencia-algoritmust számos különálló, különböző személyek által kifejlesztett modulból lehet felépíteni, és így annak valós célja és teljesítménye egészen a végső összeállításig nem lesz megállapítható. Egyúttal ez azt is jelenti, hogy a részegységek fejlesztői – akik gyakran önállóan megvásárolható termékként forgalmazzák a fejlesztésüket – nem lesznek döntő kihatással az összeállított végső termék működésére és belső logikájára. A piacon most már számos nyilvánosan elérhető félkész termék, programrészlet és meghivatkozható modul érhető el, amelyeket a fejlesztők a kirakósjátékszerű

építkezésben felhasználhatnak.[186] Ez a jogalkotónak annyiban jelent nehézséget, hogy a szabályokat teljes mértékben betartó részkomponens-fejlesztők legjobb szándéka ellenére is elkészíthető az ő termékeikből egy jogsértő módon működő mesterséges intelligencia, ha a végső fejlesztő így kombinálja és módosítja a részegységeket. Fordítva is hasonló nehézséget okozhat ez: a végső fejlesztő nem feltétlenül tudja, hogy a felhasznált beépíthető programmodul hogyan működik és milyen hatással lesz hosszú távon annak használata a kész termékre. Ezért nem elég csak a részegységtervezőket bevonni az új szabályozás hatálya alá, hanem a végső termék kidolgozóira is hatást kell gyakorolni.

Az előbb ejtettünk szót az ipari titkok védelméről, a fejlesztés átláthatatlansága kapcsán erre utalunk most vissza. Ez az attribútum azt jelenti, hogy a mesterséges intelligencia tényleges belső működési módja és logikája a külső szemlélő számára nemcsak annak összetettsége miatt átláthatatlan, hanem azért is, mert a fejlesztőcégek alapvető érdeke az egész termékben a legnagyobb hozzáadott értéket képviselő technológia titokban tartása a vetélytársak előtt. Ahogy ez a jogos érdek könnyen elismerhető, úgy meg kell azt is állapítanunk, hogy a szabályozást rendkívül megnehezíti, hiszen nem lehet megállapítani, hogy a kidolgozott technológia hogyan működik, és ezzel megfelel-e az előírásoknak.

6.2. A mesterséges intelligencia működésével kapcsolatos hatások

A mesterséges intelligencia alkalmazása, illetve üzemeltetése is több szabályozási nehézséget vet fel, ebből az első a joghatóság kérdése. A mesterséges intelligencia alapú alkalmazások általában olyan népszerű eszközökben (például okostelefonokban vagy okosotthon-vezérlőkben) és szolgáltatásokban (online piacterek, közösségi oldalak stb.) lesznek megtalálhatók, amelyek az egész világon nagyjából egyszerre jelennek meg és terjednek el. A gyártóknak nem érdeke a termékük vagy szolgáltatásuk elérhetetlenné tétele a világ egyes területein, ez nem is igazán kivitelezhető a nemzetközi kereskedelem és a világháló mai fejlettségi szintje mellett. Ebben a helyzetben a nemzetállami jogalkotás megint nem hatásos, mivel egyfelől rendkívül nagy terhet ró

[186] Így például komplett élő nyelvi beépülő modult készít és árul chatbotokhoz egy cég az interneten: https://action.ai/ (A letöltés ideje: 2018. 09. 20.)

a szolgáltatást nyújtóra, másfelől nehezen állíthatók meg a modern technológia szolgáltatásai az országhatárokon. Egy-egy ország túl merev szabályozása legfeljebb azt érheti el, hogy a jogalkotó kizárja az országa lakosságát az új technológiákhoz való hozzáférésből, ami – a népszerűségvesztés mellett – hátrányosan érintheti az állam gazdaságát is. Egy ilyen szabályozás nehezen is érvényesíthető, gondoljunk csak arra, hogy milyen komoly infrastrukturális beruházást igényelne bizonyos online szolgáltatások elérhetetlenné tétele vagy működési módjuk megváltoztatása egy-egy országra korlátozott módon. Ráadásul a szolgáltatást nyújtó, vagyis a mesterséges intelligencia üzemeltetője és a szolgáltatást igénybe vevő személy, aki ténylegesen kapcsolatba kerül a mesterséges intelligenciával, az esetek nagy többségében eltérő országokban tartózkodik, így a szolgáltatás nyújtójára és igénybe vevőjére eltérő szabályozás vonatkozhat. Ez a helyzet is több visszásságot szülhet, ha egyes felhasználási módokat az egyik ország joga megenged, a másiké pedig tilt vagy korlátoz.[187]

6.3. Lehetséges szabályozási módok

A fenti nehézségek megoldására két szabályozási módszertani megközelítés adhat lehetőséget. Egyrészt a nemzetállami szintet meghaladó jogalkotási fórumot kell találni, amely minél több államra kiterjedő szabályozást tud adni, így az eltérő joghatóságból származó nehézségek orvosolhatók. Üdvözlendő az Európai Unió törekvése, hogy magához ragadja a kezdeményezést a mesterséges intelligencia szabályozása terén, mert így legalább regionális szinten egységes vagy egy irányba mutató rendelkezések születnek. Mindazonáltal elképzelhető, hogy az uniós szintű szabályozás is kevés, tekintettel arra, hogy a mesterséges intelligencia fejlesztésének központjai az Egyesült Államokban és a keleti országokban (Kína, Japán) egyaránt megtalálhatók, így a világ nagy, kontinens szintű régiói között továbbra is ki kellene egyenlíteni a szabályozási elvárásokat.

Másrészt ne jogalkotás útján rendezzük ezt a kérdést, hanem bízzuk az iparág önszabályozására, az elkerülhetetlenül kialakuló sztenderdekre és a jó gyakorlatokra a mesterséges intelligencia fejlesztésének és működtetésének szabályozását. Ez a megközelítés az iparági szereplők józan belátására bízza a megoldást, valamint arra, hogy a piac – a vásárlók – bizalmának elnyerése érdekében a fejlesztők és üzemeltetők biztonságossá és a felhasználók által

[187] GUIHOT–MATTHEW–SUZOR 2017, 35–36.

elfogadhatóvá fogják tenni a termékeiket és üzleti gyakorlataikat, mert nem kockáztatják a versenytársakkal szembeni piacvesztést. A magánszféra védelme, az egyenlőség biztosítása és a visszaélésektől mentes működés mára olyan értékké vált az emberek szemében, ami érdemben befolyásolja az üzleti és felhasználói döntéseiket.[188] Így a megbízhatóbbnak ítélt szolgáltatók pusztán a vásárlói bizalomra tekintettel előnyösebb piaci pozícióba kerülnek. Ez pedig minden szereplőt arra sarkall, hogy a fogyasztók elvárásait keresve megbízható és jogszerűen működő termékeket fejlesszen. Ezt a nézőpontot képviselte az USA elnöke számára készített jelentés is.[189]

Megítélésem szerint a két megközelítésmódot hatékonyan tudná ötvözni a nemzetközi egyezménybe foglalt szabályozás. Ahogy a távközlés világszintű szabályozását nemzetközi egyezmény és az érintett feleket – államokat, iparági érdekelteket és a tudomány képviselőit – közös asztalhoz ültető nemzetközi szervezet (ITU) immár egy évszázadnál hosszabb ideje képes megoldani, úgy a mesterséges intelligencia jövőbeli regulációjához is ezt a megoldást javaslom. A több érdekelt fél bevonásával létrehozott nemzetközi szervezet jó platformot kínál arra, hogy közös elveket és standardokat dolgozzanak ki, a szervezet keretében megalkotott szabályszöveget pedig nemzetközi szerződés formájában fogadhatják el a tagállamok, így a nemzetállamot meghaladó szintű egységes szabályozás is megoldható lesz.[190] A nemzetközi szerződés egyes előírásainak konkrét implementációját a részes államok maguk végzik el, így a jogalkotónak is marad némi mozgástere a leginkább megfelelő megközelítés, szabályozási szint és mód megválasztására.

A nemzetközi egységesítés biztosítja az azonos szintű védelmet, valamint a rendszerek egymással való kompatibilitását. A szervezet létrehozása azért támogatandó megoldás, mert a több érdekelt fél bevonásával működő egyeztetési és tervezési fórum a technológia változásából és új eredmények megszületéséből a jövőben adódó fejlemények kezelésében is élen tud járni.[191] Az Európai Unió jogalkotó szervei a már többször idézett jelentésben szintén határozottan ösztönzik a „társadalmi, etikai és jogi kihívások vizsgálatát, majd azt követően a nemzetközi együttműködést a szabályozási normák kidolgozása terén, az ENSZ égisze alatt".[192]

---

[188] Thierer – Castillo O'Sullivan – Russell 2017, 35–37.

[189] Cath et al. 2017, 4–6.

[190] Thierer – Castillo O'Sullivan – Russell 2017, 50–53.

[191] Weng 2010, 1923–1925.

[192] P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)]. 63. pont.

Vákát oldal

Zárszó,
avagy „a királynőt megölni nem kell…”

„Reginam occidere nolite timere bonum est si omnes consentiunt ego non contradico”[193] – írta Merániai János esztergomi érsek híres levelében dodonai kétértelműséggel. A mesterséges intelligenciáról szóló művek hasonlóan kétarcú viszonyt mutatnak fel a vizsgálatuk tárgyával kapcsolatban. Egyfelől elismerik és nagyra értékelik a technológiai fejlődést, és arra biztatják a jogalkotót, hogy gazdasági és társadalmi megfontolásból támogassa a fejlesztéseket, a lehető legkevesebb korlát felállításával. Másfelől aggodalommal szemlélik ugyanezen technológia ugyanezen hihetetlen képességeit, és aggódva óvatosságra intik a jogalkotót, abba az irányba terelve a döntését, hogy szigorúan lépjen fel a mesterséges intelligencia emberi jogokat, szabadságokat, a személyes adatok védelmét vagy az egyenlőséget sértő potenciális hatásaival szemben.

A legtöbb aggodalom nem a mesterséges intelligencia szabályozásának nehézségeiből, a joghatóság meghatározásából, a büntetőjog vagy a polgári jog dogmatikai kereteinek szétfeszítéséből adódik, hanem a gép és ember viszonya gyökeres átalakulásának vélt vagy valós hatásaitól való félelemből. A gépek, amióta világ a világ, szerszámok voltak az ember kezében, engedelmeskedve az ő akaratának. A mesterséges intelligencia fejlődése, öntanuló és autonóm döntéshozó képessége viszont életre keltette ezt a szerszámot, és most attól félünk, hogy az egyszer csak nyakába kanyarítja tarisznyáját, elbúcsúzik gazdájától, és önálló vándorútra indul. A magát önállósító gépek előbb vagy utóbb emberekkel fognak találkozni, és velük valamilyen kapcsolatba lépni. Ez az interakció lehet barátságos, mindenki számára előnyös, az embert támogató és segítő kapcsolat, de lehet hátrányt vagy kárt okozó tevékenység, az alapvető jogaink megsértése is. Az, hogy ez a viszony milyen, az – félelmünk szerint – a mesterséges intelligencia számunkra átláthatatlan módon hozott döntésén múlik.

Az ember és az önállósodott gép viszonya már hosszú ideje foglalkoztatja a szakirodalmat, de a szépirodalmi szerzők sem maradhattak ki ebből

[193] „A királynőt megölni nem kell félnetek jó lesz ha mindnyájan beleegyeztek én nem ellenzem.”

a kérdésből. Itt gondoljunk csak Isaac Asimovra, aki megfogalmazta a robotika alaptörvényeit,[194] amelyek a mai napig meghatározzák az ember és robot viszonyáról szóló diskurzust. A híres sci-fi író a mechanikus emberekkel benépesített univerzumában a robotok működésével kapcsolatban három alaptörvényt fogalmaz meg, amelyeket a robotoknak feltétlenül követniük kell:

1. A robotnak nem szabad kárt okoznia emberi lényben, vagy tétlenül tűrnie, hogy emberi lény bármilyen kárt szenvedjen.
2. A robot engedelmeskedni tartozik az emberi lények utasításainak, kivéve, ha ezek az utasítások az első törvény előírásaiba ütköznének.
3. A robot tartozik saját védelméről gondoskodni, amennyiben ez nem ütközik az első vagy második törvény bármelyikének előírásaiba.

A mesterséges intelligencia vagy a robotok morális viselkedéséről író szerzők gyakran olyan mély morális dilemmákat vetnek fel, mint az elszabadult vasúti kocsi klasszikus filozófiai kérdése.[195] Ebben a gondolatkísérletben egy szemlélő azt látja, hogy egy elszabadult vasúti kocsi egy olyan vágányon halad, ahol hamarosan öt munkást fog elgázolni, akik nem veszik észre a közeledtét. A szemlélő keze ügyében van egy váltókapcsoló, amely áttereli a vonatot egy olyan vágányra, amelyen csak egy munkás dolgozik éppen, akit viszont szintén biztosan elgázolna. Ha a szemlélő nem tesz semmit, a vagon öt embert gázol el, ha átkapcsolja a váltót, akkor egy ember életét áldozza fel, és ennek árán megment ötöt.

A probléma megoldására még a filozófiában jártasak sem tudnak biztos receptet, mégis mostanában a sajtó és a szakirodalom jelentős része leggyakrabban ilyen kérdéseket kér számon a mesterséges intelligencia fejlesztőmérnökein (különösen az önvezető autók kapcsán). Ők viszont már nem vonhatják meg a vállukat a társadalom nagy részével ellentétben, számukra ugyanis egy ilyen helyzetre adott válasz leprogramozása komoly előzetes gondolkodást és határozott állásfoglalást igényel. Azzal, hogy nem programoznak kész választ az ilyen helyzetekre, szintúgy állást foglalnak. Sokan abban látják a megoldást, hogy a mesterséges intelligencia működését a moralitás határain belülre hozzák, és egyértelmű erkölcsi maximákat plántálnak beléjük.[196] Ezzel az eredetei szerzők, Wallach és Allen egy új

[194] Először leírva az író *Körbe-körbe (Runaround)* című regényében szerepel. Lásd ASIMOV 1991.

[195] JARVIS THOMSON 1976, 204–217.

[196] WALLACH–ALLEN 2009.

mozgalmat indítottak el, amelyet „robotetikának” nevezhetnénk. Az irányzat képviselői szerint a tervezőknek olyan erkölcsi alapelveket kell a gépeikbe betervezniük, amelyek alapján azok a környezetükkel való minden interakcióban az emberek által elfogadott erkölcsi normákkal egybevágó magatartást fognak követni.

Amit az erkölcsös gépek irányzatát követők nem vesznek számításba, az nem kevesebb mint a teljes fennálló jogrendszer. A jog ugyanis szintén rendelkezik kész válasszal ezekre a kérdésekre a kártérítési jog, a büntetőjogi és a polgári jogi felelősségtan körében. Az erkölcsi alapú megközelítés kritikáját Casey adja meg.[197] Meglátása szerit a mesterséges intelligenciát programozó szakemberek és a fejlesztőcégek a fent vázolt esetekben a döntést – annak jogi konzekvenciáit mérlegelve – fogják meghozni. Ha az okozott balesetekért kártérítést kell fizetniük minden sérült vagy elhunyt személy esetében, akkor a gépet úgy fogják beállítani, hogy a legkisebb számú embert veszélyeztető manővert válassza, mert ezzel csökkenti minimálisra a gyártó vagy az üzemben tartó kártérítési kötelezettségét. Ha a kártérítési felelősség megállapításakor a jogrendszer figyelembe veszi a sértetti közrehatást is, akkor pedig az önvezető autó inkább a féktávolságon belül szabálytalanul az úttestre futó gyalogosokat fogja elütni, és nem a vezetőjét veszélyezteti a kormány félrerántásával. Ha lecsupaszítjuk ezt a helyzetet, akkor a döntés moralitása nem különbözik attól a döntéstől, hogy az önvezető módba kapcsolt jármű betartsa-e a sebességhatárokat vagy sem. Ha a sebességhatár átlépésével anyagi hátrányt kockáztat, akkor a sebességhatáron belül fog haladni, és így a viselkedése külsőre nem fog különbözni az erkölcsi megfontolásból szabálykövető járművezető viselkedésétől.

A szabályozó feladata tehát az, hogy olyan keretrendszert alkosson meg, ahol a józan és számító jogi megfontolás és a mindenki által elfogadott erkölcsi norma egybeesik. Megítélésem szerint a mesterséges intelligencia és az ember közötti viszony erkölcsi és jogi minimuma bátran lehet azonos az ember és ember közötti békés együttélés már létező szabályaival. A békés együttélés maximáit már számtalanszor megpróbálta összefoglalni az emberiség: a tízparancsolat erkölcsi iránymutatása évezredekig szolgált iránytűként az erkölcsös életet keresők számára, a Magna Charta, az aranybulla, a Bill of Rights vagy a francia forradalom emberi jogi deklarációja szintén mélyen az eszünkbe véződött. Ha a modern kor embereként a jogot és erkölcsöt nem akarjuk összekeverni, és írott jogforrást keresni, akkor a nagy emberi jogi

[197] Casey 2017.

egyezmények alapjogi katalógusa lehet a segítségünkre. Ezek a szövegek a világ csaknem minden államának konszenzusát tükrözik, és megmutatják, hogy mit gondol a mai ember és a modern kori jogalkotó az embereket megillető alapvető jogoknak. Ha a robotoknak csak annyit tanítunk meg, hogy ezeket az egyezményeket tiszteletben tartva bánjanak az emberekkel, már eleget tettünk a békés ember-gép együttélés érdekében.

7. Felhasznált irodalom

2003. évi CXXV. törvény 9. § az egyenlő bánásmódról és az esélyegyenlőség előmozdításáról.

6, Perri (2001): Ethics, Regulation and the New Artificial Intelligence, Part I: Accountability and Power. *Information, Communication & Society,* Vol. 4, No. 2. DOI: https://doi.org/10.1080/713768525

Acquisti, Alessandro – Varian, Hal R. (2005): Conditioning Prices on Purchase History. *Marketing Science,* Vol. 24, No. 3. 367–381. DOI: http://dx.doi.org/10.1287/mksc.1040.0103

Alpaydin, Ethem (2010): *Introduction to Machine Learning.* (second edition). Cambridge, The MIT Press.

Article 29 Working Party (2010). Opinion 2/2010 on Online Behavioural Advertising (WP 171).

Article 29 Working Party Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679. WP 251.

Article 29 Data Protection Working Party Opinion 4/2007 on the concept of personal data. 01248/07/EN WP 136.

Article 29 Working Party (2014). Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 9.

Ashrafian, Hutan (2015): AIonAI: A Humanitarian Law of Artificial Intelligence and Robotics. *Science and Engineering Ethics,* Vol. 21, 29–40. DOI: https://doi.org/10.1007/s11948-013-9513-9

Asimov, Isaac (1991): *Én, a robot.* Budapest, Móra.

Balkin, Jack M. (2018): Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. *UC Davis Law Review,* Vol. 51, No. 3. 1149–1210. Elérhető: https://lawreview.law.ucdavis.edu/issues/51/3/Essays/51-3_Balkin.pdf (A letöltés ideje: 2017. 12. 02.)

Balkin, Jack M. (2017): The Three Laws of Robotics in the Age of Big Data. *Ohio State Law Journal,* Vol. 78, 1–45. Elérhető: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2890965 (A letöltés ideje: 2018. 09. 10.)

Balkin, Jack M. (2015): The Path of Robotics Law. *California Law Review Circuit,* Vol. 6, 45–60. Elérhető: https://digitalcommons.law.yale.edu/cgi/viewcontent.

cgi?referer=https://www.google.com/&httpsredir=1&article=6170&context=fss_papers (A letöltés ideje: 2018. 09. 10.)

BAYERN, Shawn – BURRI, Thomas – D. GRANT, Thomas – HÄUSERMANN, Daniel M. – MÖSLEIN, Florian – WILLIAMS, Richard (2017): Company Law and Autonomous Systems: A Blueprint for Lawyers, Entrepreneurs, and Regulators. *Hastings Science and Technology Law Journal,* Vol. 9, No. 2. 135–162. DOI: http://dx.doi.org/10.2139/ssrn.2850514

Big Data and Differential Pricing. Executive Office of the President of the United States, 2015. Elérhető: https://obamawhitehouse.archives.gov (A letöltés ideje: 2017. 12. 02.)

CALO, Ryan (2017): *Artificial Intelligence Policy: A Primer and Roadmap.* DOI: http://dx.doi.org/10.2139/ssrn.3015350

CALO, Ryan (2015): Robotics and the Lessons of Cyberlaw. *California Law Review,* Vol. 103, No. 3. 513–563. DOI: http://dx.doi.org/10.2139/ssrn.2402972

CASEY, Bryan (2017): Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law. *Northwestern University Law Review,* Vol. 111, No. 5. 1347–1366.

CATH, Corinne – WACHTER, Sandra – MITTELSTADT, Brent – TADDEO, Mariarosaria – FLORIDI, Luciano (2017): *Artificial Intelligence and the 'Good Society': the US, EU, and UK Approach*. DOI: http://dx.doi.org/10.2139/ssrn.2906249

CERKA, Paulius – GRIGIENE, Jurgita – SIRBIKYTE, Gintare (2015): Liability for damages caused by artificial intelligence. *Computer Law & Security Review,* Vol. 31, No. 3.

FAYYAD, Usama – PIATETSKY-SHAPIRO, Gregory – SMYTH, Padhraic (1996): From Data Mining to Knowledge Discovery in Databases. *AI Magazine,* Vol. 17, No. 3. DOI: https://doi.org/10.1609/aimag.v17i3.1230

FENWICK, Mark – KAAL, Wulf A. – VERMEULEN, Erik P. M. (2017): Regulation Tomorrow: What Happens When Technology is Faster than the Law? *American University Business Law Review,* Vol. 6, No. 3. DOI: http://dx.doi.org/10.2139/ssrn.2834531

FERENCZ Jácint (2018): *Jogalkotás a munkaviszonyok szolgálatában.* Budapest, Dialóg Campus.

GÉRON, Aurélien (2017): *Hands-On Machine Learning with Scikit-Learn & TensorFlow.* Sebastopol, O'Reilly.

GOODFELLOW, Ian – BENGIO, Yoshua – COURVILLE, Aaron (2016): *Deep Learning.* Cambridge, The MIT Press.

Goodman, Bryce – Flaxman, Seth (2017): European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine,* Vol. 38, No. 3. DOI: https://doi.org/10.1609/aimag.v38i3.2741

Guihot, Michael – Matthew, Anne F. – Suzor, Nicolas P. (2017): Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence. *Vanderbilt Journal of Entertainment and Technology Law,* Vol. 20, No. 2. 385–456. Elérhető: https://eprints.qut.edu.au/109926/28/Guihot%2BMatthew%2BSuzor%2B2017%2B-Nudging%2BRobots.pdf (A letöltés ideje: 2018. 09. 10.)

Hallevy, Gabriel (2010): The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control. *Akron Intellectual Property Journal,* Vol. 4, No. 2. Elérhető: https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=1037&context=akronintellectualproperty (A letöltés ideje: 2018. 09. 10.)

Hallevy, Gabriel (2013): *When Robots Kill: Artificial Intelligence Under Criminal Law.* Boston, Northeastern University Press.

Jarvis Thomson, Judith (1976): Killing, Letting Die, and the Trolley Problem. *The Monist,* Vol. 59, No. 2. 204–217.

Kamar, Ehud (2006): Beyond Competition for Incorporations. *Georgetown Law Journal,* Vol. 94. DOI: http://dx.doi.org/10.2139/ssrn.720121

Keats Citron, Danielle – Pasquale, Frank (2014): The Scored Society: Due Process for Automated Predictions. *Washington Law Review,* Vol. 89, 2–33. Elérhető: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209 (A letöltés ideje: 2018. 09. 10.)

Kelleher, John D. – Mac Namee, Brian – D'Arcy, Aoife (1974): *Machine Learning for Predictive Data Analytics.* Cambridge, The MIT Press.

Keserű Barna Arnold (2020): *A 21. századi technológiai változások hatása a jogalkotásra. Képes-e lépést tartani a jog a változó világgal? Fejezetek a modern technológia köréből jogi megközelítésben.* Budapest, Dialóg Campus.

Klein Tamás – Tóth András (2018): A robotika egyes szabályozási kérdései. In Homicskó Árpád Olivér szerk.: *Egyes modern technológiák etikai, jogi és szabályozási kihívásai.* Budapest, Károli Gáspár Református Egyetem Állam- és Jogtudományi Kar.

Lighthill, James (1973): Artificial Intelligence: A General Survey. In *Artificial Intelligence: a Paper Symposium.* London, Science Research Council. Elérhető: https://www.semanticscholar.org/paper/Artificial-Intelligence-%3A-A-General-Survey-Sutherland-Needham/b586d050caa00a827fd2b318742dc80a304a3675?p2df (A letöltés ideje: 2018. 09. 10.)

LOPPUCKI, Lynn M. (2018): Algorithmic Entities. *Washington University Law Review,* Vol. 95, No. 4. 887–953. Elérhető: https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=6319&context=law_lawreview (A letöltés ideje: 2018. 09. 10.)

MCCARTHY, John (1959): *Programs with Common Sense.* Stanford, Computer Science Department, Stanford University. 1–15. Elérhető: http://jmc.stanford.edu/articles/mcc59/mcc59.pdf (A letöltés ideje: 2018. 09. 10.)

MCCORDUCK, Pamela (2004): *Machines Who Think* (second edition). Natick, A. K. Peters.

MIHAYLOVA, Ekaterina (2016): *The Training Corpus Problem.* Elérhető: https://medium.com/@ekaterinamihailova/the-training-corpus-problem-3c6641fbd771 (A letöltés ideje: 2018. 09. 10.)

MILLER, Akiva (2014): What Do We Worry About When We Worry About Price Discrimination? The Law and Ethics of Using Personal Information for Pricing. *Journal of Technology Law & Policy,* Vol. 19. Elérhető: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2315315 (A letöltés ideje: 2018. 09. 10.)

MOHRI, Mehryar – ROSTAMIZADEH, Afshin – TALWALKAR, Ameet (2012): *Foundations of Machine Learning.* Cambridge, The MIT Press.

NAISBITT, John (1982): *Megatrends: Ten New Directions Transforming Our Lives.* London, Warner Books.

On the web, price tags blur. *The Washington Post.* 2000. Elérhető: www.washingtonpost.com (A letöltés ideje: 2018. 09. 20.)

ORENSTEIN, Gary (2017): *AI's Secret Weapon: The Data Corpus.* Elérhető: https://www.memsql.com/blog/ai-secret-weapon-the-data-corpus/ (A letöltés ideje: 2018. 09. 10.)

ORSEAU, Laurent – ARMSTRONG, Stuart (2016): *Safely Interruptible Agents.* Proc. of Conference on Uncertainty in Artificial Intelligence 2016, New York City. Machine Intelligence Research Institute. Elérhető: https://intelligence.org (A letöltés ideje: 2018. 09. 20.)

P8_TA(2017)0051 Az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról [2015/2103(INL)].

PETIT, Nicolas (2017): *Law And Regulation of Artificial Intelligence and Robots – Conceptual Framework and Normative Implications.* DOI: http://dx.doi.org/10.2139/ssrn.2931339

RAMIREZ, Edith (2013): *Privacy Challenges in the Era of Big Data: A View from the Lifeguard's Chair.* A Szövetségi Kereskedelmi Bizottság (Federal Trade Commission) elnökének 2013. augusztus 19-én tartott előadása a Technology Policy Institute Aspen Fórumán.

Richards, Timothy J. – Liaukonyte, Jura – Streletskaya, Nadia A. (2016): Personalized Pricing and Price Fairness. *International Journal of Industrial Organization,* Vol. 44, 138–153. DOI: https://doi.org/10.1016/j.ijindorg.2015.11.004

Russell, Stuart J. – Norvig, Peter (2016): *Artificial Intelligence – a Modern Approach.* (third edition). Harlow, Pearson Education Limited.

Scherer, Matthew U. (2016): Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology,* Vol. 29, No. 2. 354–398. DOI: http://dx.doi.org/10.2139/ssrn.2609777

Shiller, Benjamin Reed (2013): First Degree Price Discrimination Using Big Data. *Working Papers,* No. 58. 1–33.

Skitka, Linda J. – Mosier, Kathleen L. – Burdick, Mark (1999): Does automation bias decision-making? *International Journal of Human-Computer Studies,* Vol. 51, No. 5. 991–1006. DOI: https://doi.org/10.1006/ijhc.1999.0271

Solum, Lawrence B. (1992): Legal Personhood for Artificial Intelligences. *North Carolina Law Review,* Vol. 70, No. 4. 1231–1287.

Stiglitz, Joseph E. – Salop, Steven C. (1982): The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents. *The American Economic Review,* Vol. 72, No. 5. 1121–1130.

Stole, Lars (2007): Price Discrimination and Competition. In Armstrong, Mark – Porter, Robert eds.: *Handbook of Industrial Organization.* Vol. 3, 2221–2299. Amsterdam, Elsevier.

Stucke, Maurice E. – Ezrachi, Ariel (2016): Is Your Digital Assistant Devious? *Oxford Legal Studies Research Paper,* No. 52. DOI: http://dx.doi.org/10.2139/ssrn.2828117

Thierer, Adam D. – Castillo O'Sullivan, Andrea – Russell, Raymond (2017): *Artificial Intelligence and Public Policy.* Arlington, Mercatus Center at George Mason University. Elérhető: www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf (A letöltés ideje: 2018. 09. 10.) Magyar fordításban elérhető (korábbi kiadás alapján): Édelkraut Róbert et al. (2005): *Mesterséges intelligencia modern megközelítésben.* Budapest, Panem. Elérhető: http://mialmanach.mit.bme.hu/aima/index (A letöltés ideje: 2018. 09. 10.)

Turing, Alan M. (1950): Computing Machinery and Intelligence. *Mind,* Vol. 59, No. 236. 433–460. DOI: https://doi.org/10.1093/mind/LIX.236.433

Turow Joseph – King, Jennifer – Hoofnagle, Chris Jay – Bleakley, Amy – Hennessy, Michael (2009): *Americans Reject Tailored Advertising and Three Activities that Enable It.* DOI: https://doi.org/10.2139/ssrn.1478214

VARIAN, Hal R. (1989): Price Discrimination. In SCHMALENSEE, Richard – WILLIG, Robert D. eds.: *Handbook of Industrial Organization.* Vol. 1, 597–654. Amsterdam, Elsevier.

VLADECK, David C. (2014): Machines Without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review,* Vol. 89, No. 1. 117–150. Elérhető: http://euro.ecom.cmu.edu/program/law/08-732/AI/Vladeck.pdf (A letöltés ideje: 2018. 09. 10.)

WALLACH, Wendell – ALLEN, Colin (2009): *Moral Machines: Teaching Robots Right From Wrong.* Oxford, Oxford University Press.

WENG, Yueh-Hsuan (2010): Beyond Robot Ethics: On a Legislative Consortium for Social Robotics. *Advanced Robotics,* Vol. 24, No. 13. 1919–1926. DOI: https://doi.org/10.1163/016918610X527220

WENG, Yueh-Hsuan – CHEN, Chien-Hsun – SUN, Chuen-Tsai (2009): Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots. *International Journal of Social Robotics,* Vol. 1, No. 4. 267–282. DOI: https://doi.org/10.1007/s12369-009-0019-1

WENG, Yueh-Hsuan – SUGAHARA, Yusuke – HASHIMOTO, Kenji – TAKANISHI, Atsuo (2015): Intersection of "Tokku" Special Zone, Robots, and the Law: A Case Study on Legal Impacts to Humanoid Robots. *International Journal of Social Robotics,* Vol. 7, No. 5. 841–857. DOI: https://doi.org/10.1007/s12369-015-0287-x

ZARSKY, Tal Z. (2004): Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society. *Maine Law Review,* Vol. 56, No. 1. 14–59. Elérhető: https://digitalcommons.mainelaw.maine.edu/mlr/vol56/iss1/3/ (A letöltés ideje: 2018. 09. 10.)

ZIMMERMAN, Evan Joseph (2015): *Machine Minds: Frontiers in Legal Personhood.* Elérhető: 10.2139/ssrn.2563965 (A letöltés ideje: 2018. 09. 10.)

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

ZŐDI Zsolt (2017): Privacy és a Big Data. *Fundamentum,* 1–2. sz. 18–30. Elérhető: http://fundamentum.hu/sites/default/files/teljes-szamok/fundamentum-17-1-2-beliv.pdf (A letöltés ideje: 2018. 09. 10.)

ZUIDERVEEN BORGESIUS, Frederik – POORT, Joost (2017): *Online Price Discrimination and EU Data Privacy Law.* DOI: http://dx.doi.org/10.2139/ssrn.3009188

Internetes források

www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/
www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices
www.bloomberg.com/graphics/2016-amazon-same-day/
https://theintercept.com/2017/08/07/these-are-the-technology-firms-lining-up-to-build-trumps-extreme-vetting-program/
www.businessinsider.com/trumps-extreme-vetting-initiative-digital-muslim-ban-2017-11
www.domo.com/learn/data-never-sleeps-5
www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN
https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
https://techcrunch.com/2017/10/26/saudi-arabia-robot-citizen-sophia/
https://computerworld.hu/uzlet/diszruptiv-technologiak-az-ot-fo-felforgato-231071.html
www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf
https://en.wikipedia.org/wiki/Red_flag_traffic_laws
www.theguardian.com/theguardian/2014/dec/09/robot-kills-factory-worker
www.theregister.co.uk/2017/06/20/tesla_death_crash_accident_report_ntsb/
www.wired.com/story/tesla-autopilot-why-crash-radar/
www.bloomberg.com/news/articles/2018-05-24/uber-self-driving-system-saw-pedestrian-killed-but-didn-t-stop
http://nymag.com/selectall/2018/01/waze-app-directs-driver-to-drive-car-into-lake-champlain.html
www.independent.co.uk/news/world/americas/woman-following-sat-nav-canada-drives-straight-into-lake-huron-ontario-a7029131.html
www.internetlivestats.com/one-second/#google-band
www.bbc.com/news/magazine-19214294
www.rac.co.uk/drive/advice/learning-to-drive/stopping-distances/
www.cbsnews.com/news/google-struggles-with-its-do-first-ask-forgiveness-later-strategy/
www.wsj.com/articles/how-europes-new-privacy-rules-favor-google-and-facebook-1524536324

www.theguardian.com/technology/2018/feb/14/amazon-alexa-ad-avoids-ban-after-viewer-complaint-ordered-cat-food
www.hollywoodreporter.com/live-feed/south-park-premiere-messes-viewers-amazon-alexa-google-home-1039035
https://qz.com/880541/amazons-amzn-alexa-accidentally-ordered-a-ton-of-dollhouses-across-san-diego/
www.telegraph.co.uk/news/2017/01/08/amazon-echo-rogue-payment-warning-tv-show-causes-alexa-order/
www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html
www.independent.co.uk/life-style/gadgets-and-tech/news/robots-create-new-language-to-work-together-a7636041.html
www.bbc.com/news/technology-36472140
www.reuters.com/article/us-usa-smartphone-killswitch/smartphone-theft-drops-in-london-two-u-s-cities-as-anti-theft-kill-switches-installed-idUSKBN0LF09520150211
https://money.cnn.com/2017/01/26/technology/kill-switch-ai-ethics/index.html
www.medicaldevice-developments.com/features/featurethe-future-of-smart-electronic-implants-4787632/
https://action.ai/

Vákát oldal

Ludovika Egyetemi Kiadó Nonprofit Kft.
Székhely: 1089 Budapest, Orczy út 1.
Kapcsolat: kiadvanyok@uni-nke.hu

A kiadásért felel: Koltányi Gergely ügyvezető igazgató
Felelős szerkesztő: Pordány Katalin
Olvasószerkesztő: Oláh Andrea
Korrektor: Simann Karola
Tördelőszerkesztő: Mikes Vivien

Nyomdai kivitelezés: Pátria Nyomda Zrt.
Felelős vezető: Orgován Katalin vezérigazgató

DOI: https://doi.org/10.36250/00820_00

ISBN 978-963-531-149-1 (nyomtatott)
ISBN 978-963-531-150-7 (elektronikus PDF)
ISBN 978-963-531-151-4 (ePub)

A mesterséges intelligencia korunk nagy vívmánya, olyan technológia, amely átformálja a mindennapjainkat. Egy ilyen felforgató technológia a jogalkotót is nehéz feladat elé állítja: szükséges-e beavatkoznia az esetleges káros hatások elkerülése érdekében, igényel-e új szabályozást a műszaki fejlődés új lépcsője?

A szerző azt a nehéz feladatot vállalta fel, hogy a mesterséges intelligencia, illetve az ezen alapuló automatizált döntéshozatal és a gépi tanulás technológiai jellegzetességeire alapozva járja körül a lehetséges problémákat és a szabályozás szóba jöhető útjait. Ebben a kérdésben ugyanis nem lehet a műszaki megoldások jellegzetességei, működési elvük és korlátaik ismerete nélkül megalapozott véleményt alkotni.

Ajánljuk ezt a könyvet a jelen és a jövő jogalkotói, szabályozó hatóságai és a technológia iránt érdeklődő minden jogász szakember számára.

A kiadvány a KÖFOP-2.1.2-VEKOP-15-2016-00001
„A jó kormányzást megalapozó közszolgálat-fejlesztés”
című projekt keretében jelent meg.