G. KARÁCSONY GERGELY

# Inkább bízzunk a robotokban? A mesterséges intelligencia döntéseiért való emberi felelősség kritikája

A jogfejlődés következő lépését minden bizonnyal a technológiai fejlődés fogja vezérelni. Az elmúlt néhány év eseményei azt mutatják, hogy a jogalkotó szervek és szabályozó hatóságok egyaránt a mesterséges intelligencia, és az ennek nyomán megvalósuló gépi tanulás és emberi beavatkozás nélküli döntéshozatal kihívásait próbálják felderíteni, megérteni, és lehetőségekhez mérten kezelni.[1] Az egyik legnagyobb kérdést az jelenti, hogy a mesterséges intelligencia által hozott döntések következményeiért ki és milyen módon viseli a felelősséget. A szakirodalomban és a jogalkotási előkészítő anyagokban felvetődött annak lehetősége, hogy a mesterséges intelligencia egyfajta korlátozott jogi személyiséget nyerjen, és így saját neve alatt váljon jogok és kötelezettségek alanyává.[2] Ezzel szemben a napjainkban leggyakrabban alkalmazott, nemzetközi szinten elterjedt megoldás az emberi jogalany közvetlen vagy mögöttes felelősségét feltételezi a gépi döntés mögött. Tanulmányomban a mesterséges intelligencia alapú döntési mechanizmusok műszaki jellemzőire és az emberi viselkedés és észlelés korlátaira alapuló kritikai elemzését végzem el ennek a konstrukciónak.

[1] Az Európai Unió számos dokumentumban foglalkozik a kérdéssel. Ezek közül a legfontosabbak: az Európai Parlament 2017. február 16-i állásfoglalása a Bizottságnak szóló ajánlásokkal a robotikára vonatkozó polgári jogi szabályokról (P8_TA(2017)0051); a mesterséges intelligencia, a dolgok internete és a robotika biztonsági és felelősségi vonatkozásairól szóló jelentés (COM(2020) 64); a megbízható mesterséges intelligenciára vonatkozó etikai iránymutatás tervezete (2018). Magyarország kormánya 2020 szeptemberében fogadta el a tíz évre szóló Mesterséges Intelligencia Stratégiát, amely kiemelt helyen kezeli a szabályozási keretek megalkotását (ld. Kormany.hu: Elkészült a Mesterséges intelligencia stratégia, 2020). Az USA elnökének kezdeményezésére számos jogi és politikaalkotási iránymutatás készült a technológia fejlődésével kapcsolatban (például: Preparing for the future of artificial intelligence. Executive Office of the President National Science and Technology Council Committee on Technology, Washington, DC, USA, 2016).

[2] E megközelítés kritikáját lásd: Keserű, 2020, 47-51.

## I. 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. Ha a kötelező szabályokat, illetve a hatósági vagy bírósági döntések által kimunkált keretrendszert nézzük, szinte minden érintett esetben elvárásként jelenik meg az emberi beavatkozás lehetősége, és károkozás esetén elsődlegesen a felügyeletet gyakorló személyt vagy szervezetet mondja ki felelősnek a hatóság. Ez alól csak a nyilvánvaló üzemzavar vagy hibás működés esete jelent kivételt, amikor pedig a termék gyártója a felelős. Így van ez a térbeli kiterjedéssel nem bíró (szoftveres) mesterséges intelligenciák és a tárgyi világban is megjelenő gépi döntéshozók (robotok, önvezető autók) esetében is.

A robotok világát vizsgálva láthatjuk, hogy az önvezető autók esetében a működésüket egyáltalán megengedő államok jogalkotója minden esetben előírja a járművezető folyamatos kapcsolatát a gépjármű kezelőszerveivel (így a piacon elérhető önvezető gépkocsik érzékelik azt, 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 szavatolni 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 kapcsán a vizsgálatot folytató hatóság vagy bíróság szinte kivétel nélkül arra vezette vissza a balesetért való felelősséget, 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.[3] 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; itt a balesetben szerencsére csak kisebb sérüléseket szenvedtek az utasok.[4] 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 belső fedélzeti kamera képével igazolhatóan) az Uber önvezető módban haladó gépjárművének vezetője nem figyelt a vezetésre, és éppen oldalra nézett, amikor a baleset történt.[5] Mindegyik baleset azt mutatja, hogy a járművezetők nem tettek eleget a jogszabályok által előírt

[3] Beszámoló például itt: Theregister.com: Tesla death smash probe, 2017.
[4] Wired.com: Why Tesla's Autopilot Can't See a Stopped Firetruck, 2018.
[5] Bloomberg.com: Uber Self-Driving Car…, 2018.

(és az autók használati utasításában nyomatékosított) folyamatos készenléti és monitorozási kötelezettségüknek a baleset idején.

A szoftveres mesterséges intelligenciák megítélése hasonló. 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.[6] 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 meggyőződés, szakszervezeti tagság, genetikai vagy egészségi állapot, szexuális irányultság vagy nemi identitás alapján, illetve amelyek ilyen hatást kiváltó intézkedésekhez vezetnek.[7] 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 a kérdését, hogy mi a kötelezettsége az emberi beavatkozónak abban az esetben, ha az érintett vitatja a döntést, de az nem jogsértő vagy diszkriminatív.[8] 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 ún. „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.[9]

[6] GDPR 22. cikk (3) bekezdése.
[7] GDPR (71) preambulumbekezdés.
[8] Goodman – Flaxman, 2017, 6.
[9] Article 29 Working Party Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. WP 251 Adopted on 3 October 2017. 9-10.

A nem személyes adatokkal kapcsolatos kockázatok kezelése hasonlóan két irányba mutat az európai államokban, ahogyan azt az Európai Bizottság összefoglalója is leírja.[10] Egyfelől a fejlesztés vagy tervezés hiányosságaiból fakadó károk megtérítését a termékfelelősség és a termékbiztonság szabályai szerint a gyártótól követelheti a károsult. Az üzemeltető felelőssége szintén fennáll abban az esetben, ha konkrét biztonsági vagy használati előírások álltak rendelkezésére a mesterséges intelligencia alapú alkalmazás biztonságos üzemeltetése érdekében, azonban ezeket nem tartotta be. Az uniós megközelítés alapja az, hogy a károsult számára mindenképpen álljon rendelkezésre megfelelő lehetőség a kárigényének érvényesítésére.

Ez a megközelítés egyértelműen kihatással volt a mesterséges intelligencia alapú termékek felhasználásának, illetve ilyen szolgáltatások nyújtásának gyakorlatára is. A gyártók vagy szolgáltatók a felelősségük minimalizálása érdekében előírásokat adnak az üzemeltetőnek arra nézve, hogy az emberi beavatkozás biztosítsa a kontrollt az automatizmus által hozott döntések felett. Ha 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 menükből kapcsolhatóan jelezhetjük, bekapcsolva 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 oldalakon 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 a 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 azt, hogy a mesterséges intelligenciák „szabadon garázdálkodjanak", esetleg elszabadulva hibás vagy jogellenes döntések sorozatát hozzák 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.

[10] A mesterséges intelligencia, a dolgok internete és a robotika biztonsági és felelősségi vonatkozásairól szóló jelentés (COM(2020) 64), 3. pont.

## II. AZ EMBER FELÜGYELETI (MÖGÖTTES) FELELŐSSÉGÉNEK KRITIKÁJA

A nyilvánvaló jószándék és az ember középpontba való visszahelyezésének elismerése mellett is érdemes az emberi beavatkozás, mint a mesterséges intelligencia szabályozásának „Szent Grálja" jogalkotói szemléletmódját kritikával kezelni. Az, hogy az üzemeltetést végző személytől várjuk el végső soron a biztonság szavatolását és a „folyamatos kontroll" előírásával őt tesszük (teszik a gyártók, illetve hatóságok) felelőssé a bekövetkezett károkért, mind a mesterséges intelligenciák műszaki jellegzetességei, mind az emberi természet sajátosságai folytán problémásnak tekinthető.

Az első korlát, ami a szemünkbe ötlik, 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 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 az 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. Végső soron ez a megoldás a bekövetkezett károkért való felelősség telepítésében nyújt segítséget, azonban a kár bekövetkezésének érdemi megelőzése nem várható el tőle.

Nem lehet eltekinteni attól a ténytől sem, hogy a felülvizsgálatot végző személy egyéni percepciói és gondolkodása markánsan befolyásolhatják 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 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 algoritmus eredményként közöl velük. Ez a tudományosan is alátámasztott jelenség az automatizációs elfogultság (*automation bias*) nevet viseli. Egy viselkedéstani kutatásban[11] azt vizsgálták, hogy egy repülőgép-szimulátoron hogyan hoznak döntést a pilóták abban az esetben, ha számítógépes döntéstámogató rendszer működik a gépen, illetve 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, amire a gép külön nem figyelmeztette őket, valamint a józanésznek és saját kiképzésükben tanultaknak is ellentmondó döntést hozni akkor, ha a gép erre utasította őket. Ugyanezt a jelenséget figyelhetjük meg ak-

[11] Skitka et al., 1999, 51., 991-1006.

kor, amikor a szövegszerkesztő automatikus javítási funkciójára hagyatkozva egyre több nyelvhelyességi hibát hagyunk bent a szövegben, mert azokat nem jelezte a gép, vagy amikor a GPS utasításait követve a vezető belehajt egy tóba,[12] holott nagyon 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 az, 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.[13] 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.[14] 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 erőforrásfelhasználás útján lenne képes. A gépi tanulás, különösen annak megerősítéses tanulás *(reinforcement learning)* 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, illetve 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 azt, hogy 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. Ezen kívül pedig arra is szükség van,

[12] Nymag.com: Yet Another Person Listens..., 2018; Independent.co.uk: Woman follows sat nav..., 2016.
[13] Burton et al., 2020, 7-8.
[14] Thierer – O'Sullivan – Russell, 2017, 35-37.

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 fent írtak a diszkrét, egyedi döntéseket meghozó mesterséges intelligenciákra vonatkoznak. A folyamatos, valós idejű emberi szupervíziót mesterséges intelligenciára alapozott folyamatok esetében még ehhez képes is teljességgel illuzórikusnak tekinthetjük. Ennek két okát látjuk. 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.[15] 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 intelligenciák 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, ami 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.[16] Az alkalmazás viszonylag egyszerű hibáját a valós idejű megfigyelők az ésszerűen értelmezhető emberi reakcióidőn és döntéshozatali időn belül észlelték, és reagáltak arra, azonban 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 60 mérföldes óránkénti sebességnél a reakcióidőt és a teljes lefékezéshez szükséges távolságot együtt számolva közel 80 méterre[17] lett volna szükség az akadály észlelésétől számítva

[15] Például jelen kézirat lezárásának időpontjában a Google másodpercenként 86 346 keresési kérést dolgozott fel. Lásd: Internetlivestats.com, 2020.
[16] Bbc.com: High-frequendy trading and the $330m mistake, 2012.
[17] Rac.co.uk: Stopping distances made simple, 2017.

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, ami 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 ésszerűen gondolkodva 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 egyidőben 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: nem csak a forgalmi helyzetet kell folyamatosan figyelnie és megfelelő időben megfelelően reagálnia (vagyis autót vezetnie), hanem ezen felül monitoroznia 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.

## III. ÖSSZEGZÉS

Láthatjuk tehát, hogy az érdemi emberi beavatkozásnak számos morális, szociológiai, lélektani és nem utolsó sorban technológiai korlátja van. A mesterséges intelligenciát elsődlegesen azért hoztuk létre, és azért használjuk rendszeresen, hogy olyan feladatokat oldjon meg, ami 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, amit az emberi elme nem tudna felismerni. Olyan mesterséges intelligencia alapú alkalmazásokat használunk nap mint nap, amik a másodperc törtrésze alatt tudnak felismerni egy helyzetet, adekvát döntést hozni, és azt végre is hajtani (lásd a korábban említett nagysebes-

ségű tőzsdei algoritmusokat, amelyek a kereskedés apró rezdüléseit figyelik és kihasználják a csak néhány pillanatra nyitva álló előnyös lehetőségeket is). Szintén mesterséges intelligenciát találunk a nagymennyiségű, 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 nem csak életszerűtlen, hanem 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 csak.

A felhasználó vagy az üzemben tartó felelősségének kimondásával a mesterséges intelligencia szabályozása az objektív (kvázi veszélyes üzemi) felelősség felé közelít, ami azonban több okból sem kívánatos. Jogalkotói, jogpolitikai szempontból helytelen volna az objektív 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 életünk minden területét. Célszerűtlen és jogpolitikailag is megkérdőjelezhető döntés volna egy ennyire mindennapos jelenséget felelősségtani szempontból 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, a tevékenység eredményéből tud következtetni, vagyis a működési vagy következtetési hibákról is csak utólag értesül (a Knight Captial sem bukott volna akkorát, ha előre tudja, hogy az algoritmus fordított logikával fog kereskedni). Így a felhasználót vagy a felügyeletet gyakorló személyt olyasmiért tesszük felelőssé, 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, akkor nincs igazán félnivalójuk. A felhasználói vagy felügyelői felelősség megállapításával a cég kibújhat a termékfelelősség terhe alól, ezzel a jogalkotó nem presszionálja abba az irányba a mesterséges intelligencia-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,[18] 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,[19] 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 fejlesztő cégek önszabályozásában látja a megoldást. Megítélésem szerint azonban a fejlesztő cégek mindaddig nem tekintik ezt prioritásnak, amíg az algoritmus által hozott döntésért – az egyértelmű üzemzavartól eltekintve – a felhasználó vagy a felügyeletet gyakorló személy, végső soron az üzemeltető viseli a felelősséget. 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.[20]

[18] Cbcnews.com: Google Struggles..., 2010.
[19] Cath et al., 2017, 4-6.
[20] Wsj.com: Google and Facebook Likely to Benefit..., 2018.

## FELHASZNÁLT IRODALOM


- Burton, Simon – Habli, Ibrahim – Lawton, Tom – McDermid, John – Morgan, Phillip – Porter, Zoe (2020): Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. In: *Artificial Intelligence*. 279 (2020) 103201.
- Cath, Corinne – Wachter, Sandra – Mittelstadt, Brent – Taddeo, Mariarosaria – Floridi, Luciano (2017): Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. In: *Science and Engineering Ethics*. 24/2017.
- G. Karácsony Gergely (2019): A mesterséges intelligenciák szabályozásának közjogi kérdései. In: Glavanits Judit (szerk.): *A gazdasági jogalkotás aktuális kérdései.* Dialóg Campus Kiadó, Budapest, pp. 53-67.
- Goodman, Bryce – Flaxman, Seth (2017): European Union regulations on algorithmic decision-making and a "right to explanation". In: *AI Magazine*. Vol 38, No 3.
- Keserű Barna Arnold (2020): *A 21. századi technológiai változások hatása a jogalkotásra.* Dialóg Campus Kiadó, Budapest.
- Skitka, Linda et al. (1999): Does automation bias decision-making? In: *International Journal of Human-Computer Studies*. Vol. 51/1999.
- Thierer, Adam – Castillo O'Sullivan, Andrea – Russell, Raymond (2017*): Artificial Intelligence and Public Policy.* Mercatus Research, Mercatus Center at George Mason University, Arlington, VA.


## ONLINE FORRÁSOK


- Bbc.com: High-frequendy trading and the $330m mistake, 2012. (Elérhető: https://www.bbc.com/news/magazine-19214294. Letöltés ideje: 2020. 09. 19.).
- Bloomberg.com: Uber Self-Driving Car in Crash Wasn't Programmed to Brake, 2018. (Elérhető: https://www.bloomberg.com/news/articles/2018-05-24/uber-self-driving-system-saw-pedestrian-killed-but-didn-t-stop. Letöltés ideje: 2020. 09. 16.).
- Cbcnews.com: Google Struggles with Its „Do First, Ask Forgiveness Later" Strategy, 2010. (Elérhető: https://www.cbsnews.com/news/google-struggles-with-its-do-first-ask-forgiveness-later-strategy/. Letöltés ideje: 2018. 09. 27.).
- Independent.co.uk: Woman follows sat nav and drives straight into a lake, 2018. (Elérhető: https://www.independent.co.uk/news/world/americas/woman-following-sat-nav-canada-drives-straight-into-lake-huron-ontario-a7029131.html. Letöltés ideje: 2018. 09. 27.).
- Internetlivestats.com, 2020. (Elérhető: http://www.internetlivestats.com/one-second/#google-band. Letöltés ideje: 2020. 09. 28.).
- Kormany.hu: Elkészült a Mesterséges intelligencia stratégia, 2020. (Elérhető: https://www.kormany.hu/hu/innovacios-es-technologiai-miniszterium/hirek/elkeszult-a-mesterseges-intelligencia-strategia. Letöltés ideje: 2020.10.03.).
- Nymag.com: Yet Another Person Listens to GPS App and Drives Car Into Lake, 2018. (Elérhető: http://nymag.com/selectall/2018/01/waze-app-directs-driver-to-drive-car-into-lake-champlain.html. Letöltés ideje: 2018. 09. 27.).
- Rac.co.uk: Stopping distances made simple, 2017. (Elérhető: https://www.rac.co.uk/drive/advice/learning-to-drive/stopping-distances/. Letöltés ideje: 2020. 09. 28.).
- Theregister.com: Tesla death smash probe: Neither driver nor autopilot saw the truck, 2017. (Elérhető: https://www.theregister.co.uk/2017/06/20/tesla_death_crash_accident_report_ntsb/. Letöltés ideje: 2020. 09. 16.).

• Wired.com: Why Tesla's Autopilot Can't See a Stopped Firetruck, 2018. (Elérhető: https://www.wired.com/story/tesla-autopilot-why-crash-radar/. Letöltés ideje: 2020.09.16.).
• Wsj.com: Google and Facebook Likely to Benefit From Europe's Privacy Crackdown, 2018. (Elérhető: https://www.wsj.com/articles/how-europes-new-privacy-rules-favor-google-and-facebook-1524536324. Letöltés ideje: 2020. 09. 29.).