

## **Az automatizálástól az autonóm rendszerekig: A jogi fenomenológia az elszámoltathatóság problémáival**

**Ugo Pagallo**

Torinói Egyetem, Olaszország, jogi  
kar ugo.pagallo@unito.it

### **Absztrakt**

Az elmúlt évtizedekben jelentős mennyiségű munkát szenteltek a robotok és a mesterséges intelligencia rendszerek autonómiájának és intelligenciájának fogalmának: az alkalmazástól függően számos szabványt javasoltak az "automatizáltsági szintekre". Bár a jelenlegi mesterséges intelligencia-rendszerek akár egy hűtőszekrény vagy egy kenyérpíró intelligenciájával is rendelkezhetnek, néhány ilyen autonóm rendszer máris megkérdőjelezte a társadalom és a jog alapvető pilléreit, például azt, hogy a halálos erőszakot valaha is engedélyezni kell-e a "teljes automatizálás". Ennek a tanulmánynak az a célja, hogy megmutassa, hogy a mesterséges intelligencia normatív kihívásai az elszámoltathatóság különböző típusait vonják maguk után, amelyek kéz a kézben járnak a technológiai függőség, a kognitív feladatok delegálása és a bizalom választásaival. Minél erősebb a társadalmi kohézió, annál nagyobbak azok a kockázatok, amelyek a mesterséges intelligencia rendszerekre és mesterséges ágensre bízott feladatok és döntések nem teljesen kiszámítható következményeinek normatív értékelése révén társadalmilag elfogadhatók.

Ennek a keretnek a tükrében az elemzés következtetése arra irányul, hogy megvilágítsa, milyen hozzáállással kell kezelnünk ezt a versenyt, valamint azokat az ígéreteket, amelyek időnként szükségesek a jog területén.

### **1 Bevezetés**

A tanulmány tömör fenomenológiát nyújt arról, hogy az automatizálás és a mesterséges intelligencia ("AI") rendszerek fejlődése hogyan hatott a jog pilléreire. A cél hármas: a 3. szakaszban az autonómia fogalmára összpontosít, mint a mai multidiszciplináris vitában a félreértések forrására. A 4. szakasz rávilágít a technológia normatív kihívásaira, valamint arra, hogy az automatizálás és a mesterséges intelligencia kiválthat-e kikapukat a jog területén. Az 5. szakasz felhívja a figyelmet a jogi elszámoltathatóság különböző típusaira, amelyek kéz a kézben járnak a technológiai függőség, a kognitív feladatok delegálása és a bizalom választásaival. A jog és a technológia közötti kapcsolatot úgy kell felfogni, mint a versengő szabályozási rendszerek közötti kölcsönhatást, amelyek nemcsak egymást erősíthetik vagy egymással, hanem további szabályozási rendszerekkel, például a piaci erőkkel és a társadalmi normákkal szemben is harcolhatnak.

## 2 Egy jogi fenomenológia

A szoftverügynökök és mesterséges intelligencia-rendszerek formájában megjelenő kognitív automatákról folyó jelenlegi vita a német tudósok félig-meddig az 1800-as évek végén tett, az automatizálásról és a jogról szóló megjegyzéseire vezethető vissza. A technológia és a jogrendszerek együttes fejlődése eddig három nagy lépést ismert: ezek a mai vita őseire, a mesterséges intelligencia hajnalára és az utóbbi 2000-es évek eleji fordulópontjára vonatkoznak. A következő fejezetekben e lépések mindegyikét külön-külön vizsgáljuk.

### 2.1 Ősök

A tudósok elkezdték vizsgálni az automatizálás jogi hatásait, pl. automaták, az 1800-as évek utolsó évtizede óta. Gondoljunk Günther *Das Automatenrecht* (1892), Schels *Der strafrechtliche Schutz des Automaten* (1897), Schiller *Rechtsverhältnisse des Automaten* és Ertel *Der Automatenmissbrauch und seine Charakterisierung als Delikt* című műveire, mindkettő 1898-ból, egészen Neumond 1899-es *Der Automat* című művéig. Polgári jogi - nem pedig büntetőjogi - szempontból kezdetben arról volt szó, hogy a szerződő felek akarata milyen szerepet játszhat az automatizálásban. Büntetőjogi szempontból a vita arról szólt, hogy az automatizálás folyamata a büntetőjog területén új generációs kiskapukat hozhat-e létre, és ha igen, milyen mértékben, ami a jogalkotót beavatkozásra kényszeríti. A polgári joggal ellentétben, ahol az analógia gyakran döntő szerepet játszik az egyéni felelősség megállapításában, az egyén csak kifejezett büntetőjogi norma alapján vonható büntetőjogi felelősségre a magatartásáért. Ez az elv a jogállamiság egyik alaptételéhez kapcsolódik, amelyet a kontinentális Európában a legalitás elvének formulájával foglalnak össze: "nincs bűncselekmény, sem büntetés büntetőjog nélkül". Több mint egy évszázaddal később ez a fajta vita még mindig tart. Gondoljunk a 2016-os 679. sz. uniós adatvédelmi rendeletre, vagyis a "GDPR-ra", amely szerint az egyéneknek joguk van a magyarázathoz, amely az adatkezelők értesítési kötelezettségeiből ered, annak érdekében, hogy az érintettek számára minden olyan információt megadjanak, amely a tisztességes és transzparens adatkezelés biztosításához szükséges. Egyrészt a GDPR 13. cikke (2) bekezdésének f) pontja és 14. cikke (2) bekezdésének g) pontja értelmében ez a tájékoztatás "a 22. cikk (1) és (4) bekezdésében említett automatizált döntéshozatal létezésére, beleértve a pro- filinget is, és legalább a

ezekben az esetekben az érintett logikára, valamint az ilyen adatkezelés jelentőségére és várható következményeire vonatkozó érdemi információk az érintett számára." Másrészt a 22. cikk (1) bekezdése szerint "az érintettnek joga van ahhoz, hogy ne terjedjen ki rá olyan, kizárólag automatizált adatkezelésen - ideértve a profilalkotást is - alapuló döntés hatálya, amely rá nézve joghatással jár, vagy őt hasonlóképpen jelentős mértékben érinti". Eltekintve e cikk egyes formuláinak trükkös jelentésétől - amelyek vagy általános záradékokra utalnak, mint például arra, hogy az "jelentősen" érintheti az érintettet, vagy homályos fogalmakra, hogy mi az, ami "joghatásokat idéz elő" -, továbbra sem világos, hogyan kell értelmezni az emberi részvétel szintjeit a döntéshozatali folyamatban, például hogyan kell értelmezni a "kizárólag automatizált adatkezelésen alapuló döntés" formulát. Vajon egy olyan automatizált feldolgozás eredménye, amelyet nem értékel aktívan egyetlen ember sem, de formálisan neki tulajdonítanak, kívül esik-e a 22. cikk (1) bekezdésének hatályán?

Van azonban egy lényeges különbség az automatizálásról szóló, 1890-es évekbeli jogi vita és az automatizált feldolgozásról szóló jelenlegi viták között. A technológiai ugrás az ilyen automatizált feldolgozás "logikájával" kapcsolatos. Ez utóbbi egyre inkább az algo- ritmusok egy bizonyos osztályára vonatkozik, amelyek vagy kiegészítik, vagy helyettesítik az emberi elemzést és döntéshozatalt, ahogyan az a gépi tanulás tudományágában történik, azaz olyan algoritmusokra, amelyek képesek önállóan meghatározni vagy módosítani a döntéshozatali szabályokat. Fenomenológiánk második lépése tehát a mesterséges intelligencia területéhez kapcsolódik, pontosabban az automatizálásról a mesterséges autonómiára való döntő elmozduláshoz.

## 2.2 A mesterséges intelligencia hajnala

Kétféleképpen értékelhetjük a mesterséges intelligencia hatását a jelenlegi jogrendszerre. Az első út a mesterséges intelligenciával és a joggal kapcsolatos jelenlegi munkák nagyapjára - nevezetesen a nagy német polihisztor Gottfried W. Leibnizre - és az ő álmára vonatkozik, hogy a jogi érvelést és végrehajtást automatikussá tegye a kombinatorikus analízis, a valószínűségszámítás és a bináris aritmetika alkalmazásával [Pagallo, 2005 és 2016a]. Ebből az álláspontból következik, hogy a jog a logikai alapú mesterséges intelligencia kutatás gazdag tesztkörnyezeteként és fontos alkalmazási területeként fogható fel, ami a következőket illeti: (i) a logika alkalmazása a jogi szabályok reprezentációjára, ahol a jogi következtetések ebből a reprezentációból következnek, dedukció útján; (ii) a logika alkalmazása a jogi szabályokra, amelyeket nem csak alkalmaznak, hanem amelyek érvelés és diskurzus tárgyát képezik; és (iii) mind az értelmező érvelés az ügy tényeinek fényében, mind a bizonyító erejű érvelés e tények megállapítására (mint például a törvényi szabályok alkalmazása előre nem látható körülmények között). A "jogi eljárás és a jogi eljárásokban a többszereplős interakció formális modelljei" alapján az általános elképzelés az, hogy a logika hatóköre a dedukciótól az információáramlásra, az érvelésre és az interakcióra terjed ki [Prakken és Sartor, 2015].

Másrészt, fel kell hívni a figyelmet az eredeti

a mesterséges intelligencia végső terve, nevezetesen olyan gépek tervezése és beállítása, amelyek (szintén, de nem csak) olyan kognitív funkciókat utánoznak, amelyeket az emberek saját intelligenciájukhoz társítanak, mint például a tanulás.

és érvelés, tervezés és problémamegoldás. Itt arra összpontosítunk, hogy a mesterséges intelligencia rendszerek és alkalmazások, például az intelligens szoftverügynökök, a mesterséges intelligencia robotok vagy más autonóm rendszerek hogyan befolyásolhatják a jog követelményeit és funkcióit, vagyis azt, hogy a jognak milyennek kell lennie (követelmények), és mit kell tennie (funkciók). Igaz, a mesterséges intelligencia 1950-es évekbeli megszületése és az ilyen jellegű kutatások alapító atyáinak és vezetőinek nagyszabású elvárásai után többet kellett várnunk Simon "25 événél" vagy Minski "jelenlegi generációjánál" ahhoz, hogy az egyszerű automatizálásról a robusztus autonóm rendszerek felé mozduljunk el [Simon, 1965; Minski, 1967]. Bár még mindig nem kaptunk sem olyan gépeket, amelyek képesek lennének bármilyen munkát elvégezni, amit az emberek el tudnak végezni, sem megoldást a megfelelő mesterséges intelligencia létrehozásának problémájára, egyre inkább olyan rendszerekkel foglalkozunk, amelyek a környezetükben élő élőlényekkel való saját interakciójukból szereznek ismereteket vagy készségeket, így az AI-alkalmazás állapot-átmeneti rendszerében összetettebb kognitív struktúrák jelennek meg. Más szóval, egyszerű gépek helyett, amelyek mechanikusan képesek működtetni vagy irányítani egy folyamatot, fokozatosan valódi mesterséges ágensekkel lépünk kapcsolatba.

### 2.3 A fordulópont

Az elmúlt évtizedben a jelenlegi AI-technológiák megvalósíthatósága, fontossága és méretezhetősége együtt járt a "négy önerősítő trend" gyors fejlődésével, amelyek a kifinomultabb statisztikai és valószínűségi módszerek fejlesztését, a nagy mennyiségű adat és az olcsó, hatalmas számítási teljesítmény egyre nagyobb mértékű rendelkezésre állását, egészen a helyek és terek IT-barát környezetté alakításáig, pl. az intelligens városok és a domotika [Cath et al., 2016]. Hogy mit jelent ez a hatalmas átalakulás jogi szempontból, azt itt két lényeges pontban foglalhatjuk össze.

Az első megjegyzés visszavezet bennünket Leibnizhez, pontosabban G.W.F. Hegel munkásságának kritikájához. Amint azt az *Előadások a filozófia történetéről* (III. rész, C, 1) hangsúlyozza, Hegel csodálta Leibniz "individualitás" és "megkülönböztethetlenség" elveit, mégis azt állította, hogy "ez egy olyan mesterséges rendszer, amely a megértés egy kategóriáján, az absztrakt individualitás abszolútumán alapul" [Hegel, szerk. 1892-6]. Egyszerűbben fogalmazva, amit Leibniz elmulasztott, az a saját elképzeléseinek és terveinek a való világra gyakorolt hatása volt, vagyis az, hogy a jogi érvelés automatikussá tétele hogyan változtatta volna meg a jog követelményeit és funkcióit egyaránt. Visszatérve a mesterséges intelligencia jelenlegi fejlődésére, ugyanez a megjegyzés igaz. Minél inkább eléri a mesterséges intelligencia saját célját, hogy intelligens módon működő rendszereket hozzon létre, annál kevésbé marad érintetlenül a világunk, annál inkább oda kell figyelniük ennek a mélyreható átalakulásnak a tendenciáira és hatásaira. Erre az álláspontra az alábbi 4. szakaszban térünk vissza.

A második lényeges pont a mai viták egyik népszerű témájával, valamint a tudósok és a politikai döntéshozók félreértéseinek egyik fő forrásával kapcsolatos, nevezetesen

az automatizálásról a mesterséges intelligencia-rendszerek autonómiája felé történő, fent említett elmozdulással. Az autonómia fogalmát valóban három különböző módon ragadhatjuk meg. Az első jelentést meg kell különböztetni az alkalmazkodóképesség és az interaktivitás további fogalmaitól.

ty [Allen et al., 2000]. Ebből a szempontból egy mesterséges intelligencia rendszer akkor tekinthető autonóm rendszernek, ha külső ingerek nélkül módosítja belső állapotait vagy tulajdonságait, és ezáltal az ember közvetlen beavatkozása nélkül gyakorol ellenőrzést a cselekvései felett [Floridi és Sanders, 2004]. Egy ilyen tulajdonság természetesen kiegészítheti a rendszer interaktivitását és alkalmazkodóképességét is. Az első esetben a rendszer érzékeli a környezetét, és saját tulajdonságainak vagy belső állapotainak értékeinek megváltoztatásával reagál az ingerekre. A második esetben egy mesterséges intelligencia rendszer akkor adaptálható, ha képes javítani azokat a szabályokat, amelyek révén saját tulajdonságai vagy belső állapotai változnak. Ezek a kritériumok együttesen kiemelik, hogy miért kell egy mesterséges intelligencia-rendszert "ágensként" emlegetni, nem pedig az emberi interakció egyszerű eszközeként.

Az autonómia második jelentése egyetlen szóval foglalja össze a cselekvőképesség fogalmának korábbi tulajdonságait. Vegyük például az Egyesült Királyság Védelmi Minisztériumának 2011. márciusi közös doktrínajegyzetét a "pilóta nélküli repülőgép-rendszerekről". Az autonómia fogalma ott egy olyan rendszerhez kapcsolódik, amely "képes megérteni a magasabb szintű szándékot és irányítást". Ennek megfelelően az autonómiának és az automatizálásnak különböző szintjei lehetnek. Például az Egyesült Államok Közlekedési Minisztériuma által 2016 szeptemberében elfogadott *szövetségi Autómatized Vehicles Policy*-ben ez utóbbi hat szintet különböztet meg aszerint, hogy ki mit és mikor csinál, azaz a 0. szinttől, ahol az emberi vezető mindent elvégez, az 5. szintig, ahol "az automatizált rendszer képes minden olyan vezetési feladatot elvégezni, minden olyan körülmények között, amit egy emberi vezető is elvégezhetne".

Az autonómia végső jelentése a vitatottabb, mivel a mesterséges intelligencia autonómiáját az emberi autonómiához hasonlítja, beleértve időnként a "szabad akaratot" és az "erkölcsi cselekvőképességet", ahogyan azt az emberi döntéshozatal leírására használják. Bár ez az álláspont meglehetősen science fictionnek tűnik, a jelenlegi vitákban újra és újra találkozunk vele, akár filozófiai okokból, akár politikai indíttatásból, pl. a "Kampány a gyilkos robotok megállításáért", amelyet 2013 áprilisában indított egy-egy neves civil szervezet, például a *Human Rights Watch* és a Harvard International Human Rights Clinic. Félretéve ebben az összefüggésben ez utóbbi vita politikai értelmezését [Burri, 2016], a következő szakaszban külön vizsgáljuk meg a mesterséges intelligencia autonómiájának antropomorfista jelentését.

### 3 Önálló vagy, vagy nem?

Az elmúlt években jelentős mennyiségű munkát szenteltek a teljesen autonóm AI-rendszerek és robotok normatív kihívásainak tanulmányozására, vagyis az autonómia harmadik jelentése szerint, ahogyan azt a korábbi fejezetben bemutattuk. A tudósok olyan foratókönyveket vizsgáltak, mint például az önfenntartási érdekeltséget kifejlesztő mesterséges intelligencia-rendszerek, vagy az embereknek ártó, saját hasznukra szolgáló rendszerek, vagy a legkifinomultabb vezérlő kódrendszerek feltörésére képes kvantumszámítógépekhez való hozzáférés. A tudósok emellett megvitatták, hogy a mesterséges intelligencia-

rendszereknek rendelkezniük kell-e saját jogokkal, például a fájdalomtól és szenvedéstől való mentességgel, a szabad szólásszabadsággal stb. kapcsolatos jogokkal, és ha igen, milyen mértékben. Ezt neveztem el néhány évvel ezelőtt a Robotfelszabadítási Front téziseinek [Pagallo, 2013].

A mesterséges intelligenciával kapcsolatos bűncselekmények és jogok új generációjának legelkötelezettebb szószólói közül elég, ha Gabriel Hallevy munkásságát említjük. Szerinte az AI-technológia "alkalmas arra, hogy teljesítse a büntetőjogban a tudatossági követelményeket" [Hallevy, 2015], valamint "a szándékos bűncselekmények és a gondatlansági bűncselekmények mentális elem- re vonatkozó követelményeit". Ez nemcsak azt jelenti, hogy az AI-rendszerek vagy közvetlen elkövetőként felelhetnek bűncselekményekért, vagy gondatlanságból elkövetett bűncselekményekért, vagy szigorú felelősség alapján, és így tovább. Sőt, az önkontroll elvesztése, az elmezavar, a részegség vagy a ténybeli és jogi tévedések általános védekezése is védelmet nyújthat az ilyen mesterséges ágenseknek. Amint a mentális elem követelménye egy robot esetében teljesül, nem lenne ok arra, hogy a büntetés és az ítélethozatal általános céljai, azaz a megtorlás és az elrettentés, a rehabilitáció és a cselekvőképtelenség megszüntetése, egészen a halálbüntetésig, ne legyenek alkalmazhatók a mesterséges intelligenciával rendelkező gépekre.

Bár elfogadhatjuk Lawrence Solum érvelését, miszerint "fogalmi alapon nem zárhatjuk ki előre annak lehetőségét, hogy a mesterséges intelligenciáknak alkotmányos személyiségi jogokat kell adni" [Solum, 1992] - vagy, ha már itt tartunk, hogy a büntetőjogászok hagyományos eszközeit megfelelően ki lehetne terjeszteni a mesterséges intelligencia rendszerek szabályozására -, két probléma van ezzel a fajta hozzáállással. Egyrészt, ha elismerjük, hogy léteznek olyan mesterséges intelligenciával működő gépek, amelyek minden releváns szempontból képesek az emberekéhez hasonló autonóm döntésekre, a következő lépés annak elismerése lenne, hogy a "személy" jogi jelentése, a szándékosság, a gondatlanság, a szigorú felelősség stb. bűncselekményeivel együtt radikálisan megváltozik. Még Solum is elismeri, hogy "az életformában bekövetkezett ilyen változás miatt a személyről alkotott fogalmunk oly módon változhat meg, hogy az ember és a személy között szakadék keletkezik". Hasonlóképpen, [Hildebrandt et al., 2010] figyelmeztetnek arra, hogy "az az empirikus megállapítás, hogy az új típusú entitások valamiféle öntudatot fejlesztenek ki és szándékos cselekvésre válnak képessé, ésszerűnek tűnik, amennyiben szem előtt tartjuk, hogy az ilyen entitások megjelenése valószínűleg megköveteli, hogy újragondoljuk a tudat, az öntudat és az erkölcsi cselekvés nem-~~at~~". Végül is senki sem tudja, hová vezethet ez a forgatókönyv. Elfogadná-e például egy erős mesterséges intelligenciával rendelkező robotjogász Hallevy érvelését, miszerint "a gonoszság nem része a büntetőjogi felelősség összetevőinek" [Hallevy, 2015]? Mi van akkor, ha az AI-ügynök a jelenlegi kizárólagos jogi pozitívizmus híve helyett inkább a természetjogi hagyomány követője?

Másrészt, kockáztatva, hogy a reakcióért való bírálatot... tionárius antropocentrizmus mellett egy másfajta prioritást is el kell ismernünk. Ahelyett, hogy az emberhez hasonló tulajdonságokkal - például szabad akarattal vagy erkölcsi érzékkel - felruházott mesterséges ágensekről szóló sci-fi forgatókönyveket vitatnánk meg, inkább arra kellene felhívni a figyelmet, hogy egyes mesterséges intelligencia-

rendszerek viselkedése már most is a jog kiskapui közé tartozik, és olyan nehéz esetek új generációját idézi elő, amelyek a jogalkotók beavatkozását teszik szükségessé mind nemzeti, mind nemzetközi szinten. Más szóval nem kell megvárnunk az automatizálás valamelyik legmagasabb szintjét, például az amerikai *szövetségi automatizált járművekre vonatkozó politika* 5. szintjét ahhoz, hogy megállapítsuk, hogy egyes jelenlegi mesterséges intelligencia-rendszerek már most is kihívást jelentenek a társadalom és a jog alapvető pillérei számára. A jogászoknak végül is pragmatikusnak kell lenniük.

## 4 Felelősségteljes mesterséges intelligencia

Amint azt a 2.1. szakaszban említettük, különbséget kell tennünk a büntetőjog és a polgári jog között annak érdekében, hogy értékelni tudjuk a mesterséges intelligencia technológia normatív kihívásait és azok hatását a jelenlegi jogrendszerekre. Míg a polgári jog területén az analógia döntő szerepet játszhat az elszámloltathatóság és a személyes felelősség kérdéseinek meghatározásában, addig a büntetőjogi felelősség csak kifejezett norma alapján állapítható meg. A következő szakaszok éppen ezt a különbséget szemléltetik néhány példán keresztül a büntetőjog (4.1. szakasz) és a polgári jog (4.2. szakasz) területén. Ezt követően az 5. szakasz azt vizsgálja, hogy a jogalkotóknak és a politikai döntéshozóknak hogyan kellene kezelniük a mesterséges intelligencia normatív kihívásait.

### 4.1 Büntetőjogi felelősségre vonás

Ahogy a jogi automatizálás és az AI & law nagyapja, Wilhelm Leibniz mondta, "minden elmének van egy horizontja a jelenlegi szellemi kapacitását tekintve, de nem a jövőbeli szellemi kapacitását tekintve" (idézi [Coudert 1995]). Leibniz bölcsességének fényében itt megkockáztatok egy kivetítést. A forgatókönyvet egy igaz történet ihlette: 2014 májusában Vital, az Aging Analytics UK által kifejlesztett robotot a Deep Knowledge nevű japán kockázati tőkebefektető cég igazgatósági tagjává nevezte ki, hogy előre jelezze a sikeres befektetéseket. Mint egy sajtóközleményben igyekeztek tájékoztatni, Vitalra azért esett a választás, mert képes "az emberek számára nem azonnal látható" piaci trendeket felismerni, ami az időskori betegségek terápiáival kapcsolatos döntéseket illeti. A mesterséges intelligencia előrejelzéseiből kiindulva, az ilyen tendenciák, miszerint az emberek a létfontosságú kognitív feladatokat autonóm mesterséges ágensekre bízzák, a belátható jövőben többszörösen is előfordulnak majd. De mi a helyzet egy robot helytelen értékelésével, amely a tőkeemelés elmaradásához és ezáltal a vállalat család csődjéhez vezet?

Ez utóbbi esetben az alternatíva a "gondatlansági bűncselekmények" és a mesterséges intelligencia vállalati felelősségének hipotézise között látszik. Ami a gondatlansági bűncselekményeket illeti, a felelősség a kellő gondosság hiányától függ, tehát attól, hogy az ésszerű személy nem véd meg másokat az előrelátható károktól. Ez a fajta felelősség a büntetőjog hagyományos "természetes-valószínű-következmény" felelősségi modelljén alapul, amely a felelősség két különböző típusát foglalja magában. Egyfelől képzeljük el a programozókat, vagy a gyártókat, vagy a felhasználókat, akik a mesterséges intelligencia rendszerükön keresztül szándékoznak elkövetni egy bűncselekményt, de az utóbbi eltér a tervtől, és más bűncselekményt követ el. Másrészt gondoljunk olyan emberekre, akiknek nincs szándékukban bűncselekményt elkövetni, de gondatlanul jártak el egy mesterséges intelligencia alkalmazás tervezése, megalkotása vagy használata során. Bár ez a második típusú felelősségre vonás bonyolultabb, a legtöbb jogrendszer akkor is felelősnek tartja az embert, ha nem állt szándékában bűncselekményt elkövetni. A hagyományos jogelmélet szerint mindezen esetek állítólagos újdonsága hasonlít egy olyan állat tulajdonosának vagy tartójának felelősségére, "amelyről

vagy ismert, vagy feltételezhető, hogy veszélyes az emberiségre" [Davis, 2011].

A gondatlanság hagyományos bűncselekményéhez hasonlóan azonban van egy probléma: az AI-rendszer helytelen értékelése esetén, amely végül a család csődhez vezet, az AI rendszer

társaság, az embereket csak a mesterséges intelligencia rendszer értékelése által kiváltott csődbűncselekményért lehetne felelősségre vonni, mivel a csalás mentális elemének követelménye az igazgatóság emberi tagjai esetében hiányozna. Ezért a társaság, sőt, az AI-ügynök büntetőjogi felelőssége lenne az egyetlen módja annak, hogy valakit csalárd csődbűncselekmény elkövetésével vádoljanak. Ez a forgatókönyv azonban azt jelenti, hogy a legtöbb jogrendszernek módosítania kellene magát, hogy vagy a robotot, mint a társaság büntetőjogi képviselőjét, vagy a társaságot, mint olyat lehessen felelősségre vonni.

#### 4.2 Polgári elszámoltathatóság

A mesterséges intelligencia-rendszerek viselkedésének jelenlegi hagyományos értelmezése a polgári jog területén az ilyen rendszereket a társadalmi interakció egyszerű eszközeinek tekinti. Míg a büntetőjogban a rossz AI-viselkedésért való felelősségre vonás jellemzően azokat az egyéneket terheli, akik önként követnek el a törvény által tiltott jogellenes cselekményt, addig a polgári jog területén tovább kell különböztetnünk a szerződéseket és a deliktumokat. A szerződésekből a polgári jogi felelősségre vonás többnyire a szerződő fél AI-rendszeren keresztül tanúsított káros magatartása által érintettek kártalanítására vonatkozik; a kártérítési jogban a fizetés a magánfelek közötti, általában az állam által előírt, az AI által okozott károk megtérítésére vonatkozó kötelezettségekből következik. Mindkét esetben vannak problémák.

A kártérítési jogban tekintsük a hibás termékekért való felelősségről szóló 85/374/EGK európai irányelvet. Bár az irányelvet végrehajtó nemzeti jogszabályok az adatokat és a képződményeket is bevonhatják a termék fogalmába, továbbra sem egyértelmű, hogy a mesterséges intelligencia adaptív és dinamikus jellege a gépi tanulási technikák, frissítések vagy ~~hívások~~ révén magában hordozhatja-e, és ha igen, milyen mértékben, a "termék" hibáját vagy hibát okozhat. Emellett figyelembe kell vennünk az olyan AI-rendszerek forgatókönyvét is, amelyek megfelelően működtek, de hibás adatok vagy rossz bemeneti adatok alapján káros eredményt hoztak létre. Itt sem egyértelmű, hogy kit kellene felelősségre vonni, azaz vagy az AI-rendszer gyártóját vagy kivitelezőjét, vagy az adatszolgáltatót, például az internetszolgáltatót, aki nem biztosította a kapcsolatot, vagy mindkettőt. Az adatszolgáltatók azonban nem tekinthetők az európai irányelv értelmében vett "gyártóknak"; ráadásul az AI-tevékenység mögött álló ökoszisztéma olyan összetett lehet, hogy az súlyosan befolyásolhatja a jogászok azon képességét, hogy a jogi okozati összefüggés és a vétkesség fogalmán keresztül elvágják a felelősségi láncot [Karnow, 1996]. Nem meglepő tehát, hogy az EU Bizottsága 2016 szeptemberében megkezdte a fent említett irányelv módosításának folyamatát.

A szerződések területén a mesterséges intelligencia mint eszköz megközelítése azt jelenti, hogy a mesterséges ügynök (AA) által létrehozott jogok és kötelezettségek közvetlenül kötik az emberi megbízót (P), mivel az AA minden cselekménye P cselekményének minősül. Emellett P nem tud kibújni a felelősség alól azzal, hogy vagy nem állt szándékában ilyen szerződést kötni, vagy az AA döntő

hibát követett el. Ez utóbbi esetben, pl. az AA kiszámíthatatlan viselkedése esetén, P kártérítést követelhet az AA tervezőjével és gyártójával szemben. A mechanizmus szerint a



bizonyítási teher, P-nek kell bizonyítania, hogy az AA hibás volt, és hogy ez a hiba akkor állt fenn, amikor az AA a gyártó ellenőrzése alatt állt; továbbá, hogy a hiba volt a P által elszenvedett sérülések közvetlen oka. Mégis, nehéz elfogadni, hogy az AA-k által megállapított jogok és kötelezettségek közvetlenül az emberekre szállnának, mert a megbízó az AA által kötött szerződés konkrét tartalmát, vagy megállapodását akarta. Inkább azért ruháznak át jogokat és kötelezettségeket az emberekre, mert az AA-kra ruházzák azt a jogkört, hogy a nevükben cselekedjenek. Míg a hagyományos megközelítés egy hegeli éjszakába torkollik, ahol a felelősség mindenféle formája szürke, ezért szükségesnek tűnik a mai jogi szabályok módosítása, hogy a robotok üzemeltetőit és felhasználóit a gép különböző hibáinak és az eset körülményeinek megfelelően vonják felelősségre. Például az emberek nem kerülhetik el azt a szokásos konvenciót, hogy a robotok döntő hibát követnek el, azaz a szerződés érvénytelenítését, amikor a szerződő félnek tudnia kellett volna egy olyan hibáról, amely a robot kiszámíthatatlan viselkedése miatt egyértelműen a szerződés kulcsfontosságú elemeit érintette, mint például a termék piaci árát vagy a szerződés tárgyának lényegét. Általánosságban véve a célnak egyensúlyt kell teremtenie az olyan egyének között, akik azt állítják, hogy ne tegyék tönkre őket a társfinanszírozott ügyintézők döntései vagy viselkedése miatt, és az ilyen gépek szerződő felei között, de- mandálva a velük való biztonságos interakció lehetőségét. Ezt az egyensúlyt több kutató is megcélózta [Allen és Widdison, 1996; Weitzenboeck, 2001; Barfield, 2005;

Sartor, 2009; Pagallo, 2013; stb.).

Ez utóbbi megközelítés azonban komoly problémával szembesül, mégpedig az adatok hiányával, amely vagy a jogi felelősség jelenlegi alapértelmezett szabályaitól függ, vagy attól, hogy a mesterséges intelligencia rendszereket nem lehet strukturálatlan környezetben tesztelni. Mivel ezeket a problémákat a technológiai kutatás és fejlesztés területének a jogalkotók által szabályozni kívánt különböző módjai kapcsán kell megvizsgálni, ezeket a kérdéseket a következő szakaszban külön-külön mélyítjük el.

## 5 A szabályozás kihívásai

Az előző szakaszok bemutatnak néhány olyan, a mesterséges intelligencia technológiák által előidézett esetet, amelyek előbb-utóbb arra készítetik majd a nemzeti és nemzetközi jogalkotókat, hogy beavatkozzanak a büntető- és polgári jog területén. Összességében a jogalkotók három különböző típusú kihívással szembesülnek. Ezek a következőkre vonatkoznak: i) a mesterséges intelligencia-technológia sajátosságai; ii) a szabályozási rendszerek közötti verseny; és iii) az ilyen kihívások metasabályozási szintű kezelése.

Ami a mesterséges intelligencia jellemzőit illeti, jogi szempontból e technológia legveszélyesebb aspektusa a mesterséges intelligencia-rendszerek kiszámíthatatlan viselkedéséhez kapcsolódik, amely a gépi tanulási technikákon, a mögöttük álló ökoszisztéma összetettségén stb. múlhat. A jogalkotók ezt a helyzetet szabaleset-ellenőrzés olyan módszereivel kezelték, amelyek vagy a tevékenység mértékét korlátozzák pl. szigorú felelősségi szabályok révén, vagy az elővigyázatosság elvén keresztül igyekeznek

megelőzni az ilyen tevékenységeket. A jogi felelősség jelenlegi alapértelmezett normái

azonban ördögi kört vonnak maguk után, hiszen minél inkább érvényesülnek a szigorú liability szabályok, annál kevésbé tudjuk tesztelni a mesterséges intelligencia rendszereinket, annál inkább akadályozhatják az ilyen szabályok a kutatást és fejlesztést a területen [Pagallo, 2016b]. Az olasz polgári légitörvényhatóság, azaz az "ENAC" által a közelmúltban hozott, a drónok használatára vonatkozó rendkívül részletes szabályozás hulláma jól illusztrálja ezt a zsákutcát. Ennek eredményeképpen gyakran nem áll rendelkezésünkre elegendő adat az események valószínűségéről, következményeiről és költségeiről, hogy meghatározzuk a kockázati szinteket és ezáltal a biztosítási díjak és további mechanizmusok összegét, amelyeken az ilyen gépek viselkedéséért való felelősségre vonás új formái múlhatnak.

De hogy a dolgok még bonyolultabbá váljanak, figyelmet kell fordítanunk a szabályozási rendszerek közötti versenyre, mint például a piaci erők vagy a társadalmi normák. Minden szabályozási rendszer a maga eszközeivel a társadalmi viselkedés szabályozására tart igényt, és akár egy másik szabályozási rendszer igényét is alkalmatlanná vagy feleslegessé teheti. Gondoljunk csak azokra az esetekre, amikor a technológiai innováció folyamatának szabályozására irányuló jogi szándék kudarcot vallott, pl. az EU elektronikus pénzről szóló 46. számú irányelve 2000-ből. Nem sokkal a bevezetése után az online fizetések további formái, mint például a PayPal, arra kényszerítették a brüsszeli jogalkotókat, hogy beavatkozzanak, és 2009-től az új, 110-es irányelvvel módosítsák magukat. Függetlenül azonban attól, hogy milyen forgatókönyvet vizsgálunk, a szabályozási rendszerek közötti ilyen verseny nem normatív ürességben zajlik, hanem értékek és elvek jelenléte strukturálja [Pagallo és Durante, 2016]. A terület technológiai és jogi normái mellett tehát a hangsúlyt az episztemikus normákra kell helyezni, azaz az AI-rendszerek információvalóságának megértési módjaira, valamint azokra a társadalmi álláspontokra, amelyek lehetővé teszik a felhasználók számára, hogy megbízzanak az ilyen AI-rendszerekben, és értékeljék a készségek minőségét, függetlenül attól, hogy ezek a készségek megfelelnek-e a társadalmi igényeknek vagy sem. Ennek megfelelően a politikai döntéshozóknak és a jogalkotóknak szem előtt kell tartaniuk a társadalmi elfogadottság és kohézió mértékét, amely befolyásolja a saját döntéseiket, vagyis azt, hogy létezik-e közös értékrend és elvek vagy sem. Ez az elágazás kritikus, mert valami újat mond nekünk a mesterséges intelligencia normatív kihívásairól metaszabályozási szempontból. Alapvető döntést kell hozni a döntések autonóm AI-rendszerekre történő delegálásával kapcsolatban. A politikai megoldás nem csak az AI döntések kiszámíthatóságának és megbízhatóságának mértékétől függ. A kérdés inkább a társadalmi egyetértés vagy egyet nem értés degressziója körül foroghat, amely a vizsgált normatív kontextust jellemzi. Hogyan kell tehát a jogalkotóknak kezelniük a versengő szabályozási rendszerek közötti kölcsönhatást? Hogyan lehet megelőzni az olyan jogalkotást, amely akadályozhatja a mesterséges intelligencia kutatását? Hogyan kellene kezelnünk az AI egyes rendszereinek peculiaris kiszámíthatatlanságát és

kockázatos viselkedését? Hogyan kellene jogilag szabályozni a jövőt?

Szerencsére többféle jogi technika létezik, amelyekkel megfelelően kezelhetjük ezt a kihíváshalmazt. Elég, ha ezek közül hármat megemlítünk. Először is, a hangsúlyt Brandeis bírónak a *New State Ice Co. kontra Leibmann ügyben* (285 US 262 (1932)) kifejtett kísérleti föderalizmus doktrínájára kell helyezni. Az elképzelés lényege, hogy a szabályok tartalmának konkretizálása, amely

az egyéni viselkedést a jogrendszerek közötti előnyös együttműködés révén szabályozza. Ez történik manapság az önvezető autók területén az Egyesült Államokban, ahol hét állam hozta meg saját törvényeit az ilyen típusú technológiára vonatkozóan. A lehető legjobb fényében ugyanez a politika fog működni az uniós szabályozással az adatvédelem területén [Pagallo, 2017b].

Másodszor, fel kell hívni a figyelmet az implementációs semlegesség elvére, amely szerint a szabályozás definíció szerint az adott technológiára jellemző, de nem részesíti előnyben annak egyik vagy több lehetséges megvalósítását. Az Egyesült Államok Közlekedési Minisztériumának *szövetségi automatizált járművekre vonatkozó politikája* - amelyet a 2.3. szakaszban említettünk - jól szemlélteti ezt a jogi technikát. Bár a szabályozás definíció szerint az adott technológiára, pl. az autonóm járművekre vonatkozik, nem részesíti előnyben annak egyik vagy több lehetséges megvalósítását. Még ha a törvény az adott technológia egy bizonyos tulajdonságát határozza is meg, a jogalkotók úgy is megfogalmazhatják a jogi követelményt, hogy a nem megfelelő megvalósítások is módosíthatók, hogy megfeleljenek a követelményeknek.

Harmadszor, a jogalkotók elfogadhatják a jogi kísérletezés formáit. A japán kormány például az elmúlt 14 évben kidolgozott egy módszert a robotika normatív kihívásainak kezelésére, mégpedig az empirikus tesztelésre és fejlesztésre szolgáló speciális zónák, azaz egyfajta élő laboratórium, a *Tokku* létrehozásával [Weng et al, 2015]. Hasonlóképpen, az autonóm járművek területén több uniós ország is támogatta ezt a fajta megközelítést: Svédország támogatta a világ első nagyszabású autonóm vezetési kísérleti projektjét, amelyben az önvezető autók a közutakat használják a mindennapi vezetési körülmények között; Németország számos, különböző szintű automatizálással végzett tesztet engedélyezett az autópályákon, például az Audi tesztszelei egy autonóm módon közlekedő autóval az A9-es autópályán Ingolstadt és Nürnberg között. Általánosságban elmondható, hogy a kísérletezés ezen formái a törvényileg szabályozatlan különleges zónákon keresztül jelentik azt a jogi alapot, amely alapján empirikus adatokat és elegendő ismeretet lehet gyűjteni ahhoz, hogy racionális döntéseket lehessen hozni számos kritikus kérdésben. Javíthatjuk a megértésünket arról, hogy a mesterséges intelligencia rendszerek hogyan reagálhatnak különböző kontextusokban és hogyan elégíthetik ki az emberi igényeket. Jobban fel tudjuk mérni a mesterséges intelligencia-rendszerek ellenőrzésének esetleges elvesztése által okozott kockázatokat és veszélyeket, hogy kordában tudjuk tartani őket. Továbbfejleszthetünk olyan elméleti kereteket, amelyek lehetővé teszik számunkra, hogy jobban értékeljük a nemkívánatos viselkedést elkerülő potenciális rendszerek terét. Emellett racionálisan foglalkozhatunk e kísérletezés jogi vonatkozásaival, lefedve a következő generációs mesterséges intelligencia-rendszerek által felvetett számos lehetséges kérdést, és kezelhetjük az olyan követelményeket, amelyek gyakran félelmetes akadályt jelentenek az ilyen típusú kutatás számára, mint a biztonsági okokból szükséges nyilvános engedélyek, a személyes adatok feldolgozásához és felhasználásához való hivatalos hozzájárulás, a kockázatok biztosítási modellek és

hitelesítési rendszerek révén történő elosztásának mechanizmusai stb.

Természetesen e jogi technikák némelyike kölcsönhatásba léphet egymással és erősítheti egymást [Pagallo, 2017c]. Ennél is fontosabb, hogy a jogi rugalmasság olyan mechanizmusát képviselik, amely lehetővé teszi számunkra, hogy a szabályozási rendszerek közötti kölcsönhatással foglalkozzunk

bölcsen. Végeredményben jogosan állítható, hogy a technológiai fejlődés folyamatát irányító jog célja nem akadályozhatja a technológiai fejlődés folyamatát, és nem igényelhet túl gyakori felülvizsgálatot e fejlődés irányításához [Pagallo, 2016c].

## 6 Következtetések

A tanulmány egy tömör fenomenológiát kínál az ~~autonóm~~ autonóm mesterséges intelligencia-rendszerekről és a jogról, hogy meghatározhassuk, milyen mértékben érinti már napjaink jogrendszereit a technológiai innováció ezen területe. A 3. szakaszban a sci-fi forgatókönyvek elvetésével a 4. szakasz három olyan módot tárt fel, amelyekkel egyes AI-rendszerek viselkedése a jog kiskapui közé esik. Ennek megfelelően a jogalkotóknak három különböző típusú kihívással kell megküzdeniük, amelyeket az 5. szakaszban vizsgáltunk. Ezek a mesterséges intelligencia-technológia sajátos jellemzőire, a versengő szabályozási rendszerek közötti kölcsönhatásra, valamint arra vonatkoznak, hogy a jogalkotók hogyan kezelhetik az ilyen kihívásokat metaszabályozási szinten. Egyes jogi technikák, mint például a kísérleti föderalizmus szabályai vagy a végrehajtási neutralitás megközelítése olyan jogi rugalmassági mechanizmusokat kínálhatnak, amelyek lehetővé teszik az ilyen kihívások megfelelő kezelését.

Mindent egybevetve, alapvető döntést kell hozni azzal kapcsolatban, hogy a döntéseket egyre inkább az autonóm mesterséges intelligencia-rendszerekre és számtalan intelligens mesterséges ágensre bízjuk. A politikai megoldás nem csak az AI döntések kiszámíthatóságának és megbízhatóságának mértékétől függ. Az issue a társadalmi egyetértés vagy egyet nem értés mértéke körül foroghat, amely a vizsgált normatív kontextust jellemzi. Az AI-rendszerekre, autonóm robotokra vagy szoftverügynökökre bízott feladatok olyan javakat és érdekeket érintenek, amelyek nem csak a "társadalmi elfogadhatóság" mértékét befolyásolják, amely a delegálási folyamatban rejlő kockázatra vonatkozik. Az arra vonatkozó technikai és jogi döntések, hogy a mesterséges intelligenciával működő rendszerekre bízott feladatok milyen hatással lehetnek az eszközökre és az emberi érdekekre, az ezeket az eszközöket és érdekeket megalapozó értékeket és elveket is érinthetik.

A technológiai függőség, valamint a delegálás és az autonómia megfelelő fokának kérdéseit tehát a társadalmi kohézió azon normatív kontextusban meglévő mértékének megfelelően kell megítélni, amelyben az AI-rendszerekre delegált feladatok és döntések következményeit értékelik. A jogi normák meghatározásánál, mint az ágensek kommunikációját és interakcióját lehetővé tevő eszközöknél figyelembe kell venni a további szabályozási rendszerekkel való kölcsönhatást, valamint azt, hogy a társadalmi kohéziót milyen mértékben befolyásolja a technológia, például a robotika által kiváltott munkanélküliség [Floridi, 2017]. Minél erősebb a társadalmi kohézió, annál nagyobb a kockázat a delegálási folyamatban, amely a mesterséges intelligencia-rendszerekre és autonóm mesterséges ágensekre bízott feladatok és döntések nem teljesen előre jelezhető következményeinek normatív értékelésével

társadalmilag elfogadható. Mivel a mesterséges intelligencia-rendszerek itt maradnak, a jognak arra kell törekednie, hogy bölcsen szabályozza kölcsönös kapcsolatainkat. A jogi rugalmasság e dokumentumban bemutatott mechanizmusai megmutatják, hogy ez hogyan lehetséges.

## Hivatkozások

- [Allen és Widdison, 1996] Tom Allen és Robin Widdison. Tudnak-e a számítógépek szerződéseket kötni? *Harvard Journal of Law & Technology* 9(1): 26-52, 1996.
- [Allen et al., 2000] Colin Allen, Gary Varner és Jason Zinser. Prolegomena bármely jövőbeli mesterséges erkölcsi ágenshez. *Journal of Experimental and Theoretical Artificial Intelligence*, 12: 251-261, 2000.
- [Barfield, 2005] Woodrow Barfield. A virtuális környezetekben működő szoftverügynökök jogi kérdései, *Presence* 14(6): 741-748, 2005.
- [Burri, 2016] Thomas Burri. The Politics of Robotic Autonomy. *European Journal of Risk Regulation*, 7(2): 341-360, 2016.
- [Cath et al., 2016] Corinne Cath, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo és Luciano Floridi. *Művészeti intelligencia és a "jó társadalom": az USA, az EU és az Egyesült Királyság megközelítése*.  
Elérhető:  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=290624](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=290624), 2016. december (utolsó hozzáférés: 2017. április 19.).
- [Coudert, 1995] Allison P. Coudert. *Leibniz és a Kabbalah*. Kluwer Academic, Boston-London 1995.
- [Davis, 2011] Jim Davis. The (common) Laws of Man over (civil) Vehicles Unmanned, *Journal of Law, Information and Science*, 21(2): 166-179, 2011.
- [Floridi, 2017] Luciano Floridi. Robotok, munkahelyek, adók és felelőségek, *Filozófia és Technológia*, 30: 1-4, 2017.
- [Floridi és Sanders, 2004] Luciano Floridi és Jeff Sanders. On the Morality of Artificial Agents, *Minds and Machines*, 14(3): 349-379, 2004.
- [Hallevy, 2015] Gabriel Hallevy. *Felelőség a mesterséges intelligencia rendszereket érintő bűncselekményekért*. Springer, Dordrecht, 2015.
- [Hegel, 1892-6] *Előadások a filozófia történetéről*. E.S. Haldane angol fordítása, elérhető a <https://www.marxists.org/reference/archive/hegel/works/hp/hpcontent.htm> oldalon (utolsó hozzáférés: 2017. április 19.).
- [Hildebrandt et al, 2010] Mireille Hildebrandt, Bert-Jaap Koops és David-Olivier Jaquet-Chiffelle. Az elszámoltathatósági szakadék áthidalása: Jogok az új entitások számára az információs társadalomban? *Minnesota Journal of Law, Science & Technology*, 11(2): 497-561, 2010.
- [Karnow, 1996] Curtis E. A. Karnow. Liability for Distributed Artificial Intelligence, *Berkeley Technology and Law Journal*, 11: 147-183, 1996.
- [Minski, 1967] Marvin Minski. *Számítás: Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, N.J., 1965.
- [Pagallo, 2005] Ugo Pagallo. *Introduzione alla filosofia digitale: da Leibniz a Chaitin*. Giappichelli, Torino, 2005.
- [Pagallo, 2013] Ugo Pagallo. *A robotok törvényei: Crimes, Contracts, and Torts*. Springer, Dordrecht, 2013.
- [Pagallo, 2016a] Ugo Pagallo. *Leibniz: una breve biografia intellettuale*. Kluwer, Padova, 2016.
- [Pagallo, 2016b] Ugo Pagallo. Még az angyaloknak is szükségük van a szabályokra: AI, Roboethics, and the Law. In Gal A Kaminka et al (szerk.), *ECAI 2016. Frontiers in Artificial Intelligence and Applications*, pp. 209-215. IOS Press, Amszterdam 2016.
- [Pagallo, 2016c] Ugo Pagallo. Három tanulság a jogkövető intelligens közlekedési rendszerekhez, *JusLetter IT*, 2016. november 24., *JusLetter IT*, 2016. november 24. Elérhető a <http://jusletter-it.weblaw.ch/issues/2016/24-November-2016/three-lessons-learned-9251e5d324.html> címen.
- [Pagallo, 2017a] Ugo Pagallo. A nagy adatok jogi kihívásai: Putting Secondary Rules First in the Field of the EU Data Protection, *European Data Protection Law Review*, 3(1): 36-46, 2017.
- [Pagallo, 2017b] Ugo Pagallo. Amikor az erkölcs nem elég: Robots, Ethics, and the Rules of the Law, *Minds and Machines*, 2017. január.
- [Pagallo és Durante, 2016] Ugo Pagallo és Massimo Durante. The Pros and Cons of Legal Automation and its Governance, *European Journal of Risk Regulation*, 7(2): 323-334, 2016.
- [Prakken és Sartor, 2015] Henry Prakken és Giovanni Sartor. Jog és logika: A Review from an Argumentation Perspective. *Artificial Intelligence* 227: 214-245, 2015.
- [Sartor, 2009] Giovanni Sartor. Kognitív automaták és a jog: Electronic contracting and the intentionality of software agents, *Artificial Intelligence and Law* 17(4): 253-290, 2009.
- [Simon, 1965] Herbert Simon. *Az automatizálás alakja az emberek és a vezetés számára*. Harper & Row, New York, 1965.
- [Solum, 1992] Lawrence B. Solum. A mesterséges intelligencia jogi személyisége. *North Carolina Law Review*, 70: 1231-1287, 1992.
- [Weng et al., 2015] Yueh-Hsuan Weng, Yusuke Sugahara, Kenji Hashimoto és Atsuo Takanishi. A "Tokku" különleges övezet, a robotok és a jog metszéspontja: A Case Study on Legal Impacts to Humanoid Robots, *International Journal of Social Robotics*, 7(5): 841-857, 2015.
- [Weitzenboeck, 2001] Emily Mary Weitzenboeck. Electronic Agents and the Formation of Contracts, *International Journal of Law and Information Technology*, 9(3): 204-234, 2001.