



School of Law

San Franciscó-i Egyetem jogi kara

University of San Francisco Law Research Paper No. 2018-05

TÖRVÉNY ÉSZ NÉLKÜL: AI, ETIKA ÉS JOGTUDOMÁNY

JOSHUA P. DAVIS

TÖRVÉNY ÉSZ NÉLKÜL:
AI, ETIKA ÉS JOGTUDOMÁNY

JOSHUA P. DAVIS¹

Úgy tűnik, hogy bármit, amit a számítógépek meg tudnak tenni, a végén meg is teszik, még hozzá jobban, mint mi, és sokkal hamarabb, mint ahogyan azt vártuk. A matematikai számításoktól kezdve a közösségi hálózataink létrehozásán és fenntartásán át a személyes asszisztenseink szerepéig jutottak el. Azt mondják, hogy hamarosan a barátaink lesznek, és élet-halál döntéseket hozhatnak az autónkat vezetve. Talán a törvényeink értelmezését is átveszik. Nem olyan nehéz elképzelni, hogy a számítógépek ezt a feladatot is ellátják, amennyiben a jogértelmezés pusztán leírást vagy előrejelzést foglal magában. Sokkal nehezebb elképzelni, hogy a számítógépek érdemi erkölcsi ítéleteket hozzanak. A végső védőbástya tehát az ellen, hogy a jogértelmezést átadjuk a számítógépeknek - hogy a számítógépek bitorolják az ügyvédek, az állampolgárok, sőt a bírák felelősségét és tekintélyét -, az lehet, hogy az erkölcsi ítélet szerepének elismerése a törvények megmondásában. Ez a lehetőség összeköti az élvonalbeli és a hagyományos jogágot. A jogtudomány központi vitája az elmúlt legalább fél évszázadban az erkölcsnek a jogértelmezésben betöltött szerepéről szólt. Hirtelen ez a vita nagy aktualitást és sürgősséget kapott. A jogtudomány segíthet tisztázni és körülhatárolni a számítógépek szerepét jogrendszerünkben. A mesterséges intelligenciáról való elmélkedés pedig segíthet megoldani azokat a jogtudományi vitákat, amelyek évtizedek óta foglalkoztatnak bennünket.

I. Bevezetés

A VILÁG LÉLEK NÉLKÜL ²Franklin Foer komor képet fest arról, hogy a technológiai fejlődés milyen módon fenyegeti kultúránkat, egyéniségünket és végső soron elménket. Elon Musk úgy véli, hogy a mesterséges intelligencia ("AI") lesz a legvalószínűbb oka a világháborúnak.

¹ Professzor és igazgató, Jogi és Etikai Központ, San Franciscó-i Egyetem Jogi Kara. Nagy köszönettel tartozom Brad Wendelnek a számos beszélgetésért, amelyekből az elemzésemet merítettem, Suzanne Mawhinney-nek a kutatásban nyújtott kiváló segítségéért, valamint a Case Western Law School mesterséges intelligenciáról és etikáról szóló konferenciájának szervezőinek és résztvevőinek, különösen Tim Casey-nek és Brooke Raunignak.

² FRANKLIN FOER, ELME NÉLKÜLI VILÁG: (2017).

Foer könyve a modern technológiát nagy szkepticizmussal és aggodalommal szemlélő legújabb hullám egyik tagja. *Lásd pl.* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); NOAM COHEN, THE KNOW-IT-ALLS: THE RISE OF SILICON VALLEY AS A POLITICAL POWERHOUSE AND SOCIAL WRECKING BALL (2017); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASE INEQUALITY AND THREATS DEMOCRACY (2016); JAMES BARRAT, OUR FINAL INVENTION:

A MESTERSÉGES INTELLIGENCIA ÉS AZ EMBERI KORSZAK VÉGE (2015). A kockázatokra kevésbé koncentrált alternatív megközelítéshez lásd pl. MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE (2017); NICK BOSTROM, SUPERINTELLIGENCE: PATHS,

DANGERS, STRATEGIES (2014).

III.³ Pauline Kim azt sugallja, hogy egy vállalat jogellenes foglalkoztatási megkülönböztetést alkalmazhat anélkül, hogy bárki is tudná - vagy talán soha nem is tudhatná -, hogy ezt teszi, mivel olyan mesterséges intelligenciára támaszkodik, amely folyamatosan fejlődik, és nem hagy feljegyzéseket az ajánlásai alapjáról.⁴ Ahogy Jacob Weisberg írja: "Az algoritmusok gyorsabban fejlesztik az emberek szabályozására szolgáló képességeiket, mint ahogy az emberek kitalálják, hogyan szabályozzák az algoritmusokat".⁵ Úgy tűnik, hogy a technológiai fejlődés komoly veszélyt jelent az emberi etikára. Emellett régóta fennálló jogtudományi vitákra is fényt deríthetnek.

Gondoljunk csak arra, hogy a fejlődő technológiák által felvetett rejtélyek hogyan adnak újszerű hangsúlyt a jogtudomány egyik legrégebbi és legalapvetőbb nézeteltérésének: az erkölcs szerepének a jog meghatározásában. Ez a jogtudományi nézeteltérés a jogi pozitivisták és a nem pozitivisták (néha természetes jogászoknak nevezik őket) között van.⁶ Túlságosan leegyszerűsítve, a jogi pozitivisták azt állítják, hogy meg lehet mondani, mi a jog, anélkül, hogy erkölcsi ítéletet hoznánk arról, hogy milyennek kellene lennie.⁷ Ezzel szemben a nem pozitivisták - legalábbis e megfogalmazás szerint - azt állítják, hogy erkölcsi ítéletekre van szükség annak meghatározásához, hogy mi a jog.⁸ A pozitivisták és a nem pozitivisták közötti vita az elmúlt ötven évben a jogtudomány középpontjában állt.⁹ Miközben megpróbáljuk megoldani a technológiai fejlődés okozta dilemmákat, ősi kérdésekkel találjuk magunkat szemben: milyen szerepet kell játszania a jognak a cselekedeteink irányításában, és ha erkölcsi kötelességünk betartani a törvényt, vajon a jog tartalmának meghatározásakor konzultálnunk kell-e az erkölccsel?

Az autonóm autók meggyőző példát szolgáltatnak. Úgy tűnik, hogy programozásuk nem csak technikai tudást, hanem erkölcsi filozófiát is igényel. Tegyük fel, hogy egy autonóm autó egyetlen utasával éppen összeütközik egy gyalogoscsoporttal. Tegyük fel továbbá, hogy a gyalogosokkal való ütközés minimálisra csökkenti az autóban ülő személy kockázatát, de valószínűleg megöl egy gyalogost.

³ Miután elutasította az észak-koreai fenyegetéssel kapcsolatos aggodalmakat, mint a civilizációra leselkedő "egzisztenciális fenyegetést", így tweetelt: "Az AI fölényért folytatott verseny nemzeti szinten a harmadik világháború valószínű oka, imo". *Lásd:*

<http://money.cnn.com/2017/09/04/technology/culture/elon-musk-ai-world-war/index.html>

⁴ Pauline Kim, *Data-Driven Discrimination at Work*, WILL58. & MARY L. REV. 857, 902-09 (2017).

⁵ Jacob Weisberg, *The Digital Poorhouse*, NYRB (20187., 47június).

⁶ Lásd Joshua P. Davis, *Legality, Morality, Duality*, 2014 UTAH L. REV. 61-63. Pontosabban, sok pozitivisták vallja azt, amit néha "társadalmi ténytézésnek" neveznek: a jogi pozitivismus azt vallja, hogy "minden jogi tényt végső soron kizárólag a társadalmi tények határoznak meg". Id. 61. o. (idézi SCOTT J. SHAPIRO, *LEGALITY* (272011)).

⁷ Id. 61-63.

⁸ Id.

⁹ A vita volt a leghíresebb számok Hart L.A. Hart és Ronald Dworkin. Hart és Dworkin nézeteltérése csírája volt a jogi pozitivismusról folyó vitáknak, Hart valószínűleg az elmúlt fél évszázad legbefolyásosabb pozitivistája, Dworkin pedig a legbefolyásosabb nem pozitivisták, bár kevés jelenlegi teoretikus fogadhatja el bármelyikük nézetét korlátlan formában. Lásd pl. Scott Hershovitz, "The End of Jurisprudence", 124 YALE L.J. 1160, 1162 (2015) ("Több mint negyven éven át a jogtudományt a Hart-Dworkin vita uralta."); Scott Shapiro, "The 'Hart-Dworkin' Debate: A Short Guide for the Perplexed," in RONALD DWORKIN (222007) ("Az

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély
nélkül
elmúlt négy évtizedben az angol-amerikai jogfilozófiát lefoglalta - egyesek szerint
megszállottan - valami, amit 'Hart-Dworkin' vitának neveznek.").

nagyobb számú ember. Másik lehetőség, hogy az autó lecsúszik egy szikláról, megölve utasát, de sok más életet megkímélve.¹⁰

Az autonóm autó számára, amely ezt a döntést hozza, legalább két ijesztő kihívás áll fenn. Az első technológiai. Az autónak fel kell mérnie a különböző lehetséges lépések következményeit. Ez más - és látszólag még nagyobb kihívás -, mint a szokásos közlekedési szabályoknak megfelelő működés. Nem kis teljesítmény olyan autót építeni, amely képes megállni a piros lámpánál, szabályosan kanyarodni, kikerülni más járműveket, a megengedett sebességhatáron belül haladni, a gyalogosoknak a zebránál elsőbbséget adni, és így tovább.¹¹ Az autótól azt kérni, hogy két vagy több különböző rossz megoldás esetén mérje fel a valószínűsíthető életveszteséget, még ennél is lényegesen nehezebbnek tűnik. De tegyük fel, hogy az autó tervezői megbirkóznak ezzel a technikai feladattal - kitalálnak egy olyan módszert, amellyel az önvezető autó valószínűségi alapon meghatározza a különböző nemkívánatos forgatókönyvek lehetséges káros következményeit.

Egy másfajta kihívás még mindig fennáll. Ez a kihívás az előírást jelenti, nem csupán a leírást vagy a jóslatot. Mit *kellene* tennie az autónak? Feláldozza-e az autó a "vezetője" életét, hogy megvédjen másokat? A válasz attól függ, hogy hányan lennének veszélyben? Attól függ-e, hogy ki az erkölcsi felelős, ha egyáltalán felelős valaki azért a szerencsétlen helyzet kialakulásáért, amelyben a kár veszélyének valakit el kell érnie?¹²

Lépjen be a jog és az ügyvédek. Az önvezető autók számítógépének programozói - és azok, akik a programozókat alkalmazzák - valószínűleg figyelembe veszik a jogot. Részben prudenciális megfontolásból. A lehetséges jogi felelősségre vonás fényében akarnak majd változtatni a viselkedésükön. Felelősségre fogják-e vonni az autógyártókat, ha arra utasítják az autókat, hogy bizonyos körülmények között áldozzák fel - vagy ne áldozzák fel - a "sofőröket"? Megengedi-e a jog, hogy az autógyártók

¹⁰ Az ilyen jellegű dilemmákat gyakran nevezik "troliproblémáknak", ezt a kifejezést Judith Thomson filozófus alkotta meg, részben Philippa Foot munkásságára építve. Lásd Bryan Casey, *Amoral Machines, or: How Roboticians Can learn to Stop Worrying and Love the Law*, 111 Nw. U. L. Rev. 1347, 1353 & n. 36 (2017) (Judith Jarvis Thomson, *Killing, Letting Die, and the Trolley Problem*, 99 *Monist* 204, 206 (1976) és Philippa Foot, *The Problem of Abortion and the Doctrine of Double Effect*, *Oxford's Rev.* (1,19673) tárgyalása).

¹¹ A közelmúltban történt eset, amikor az Uber önvezető autója megölt egy gyalogost, jól példázza a feladat nehézségeit. Lásd <https://www.npr.org/sections/thetwo-way/2018/03/27/597331608/arizona-suspends-ubers-self-driving-vehicle-testing-after-fatal-crash>. Természetesen egyetlen halálos áldozat nem bizonyítja, hogy az önvezető autók kevésbé biztonságosak, mint az emberi sofőrök. Megjegyzendő az is, hogy az Uber önvezető autóinak is több nehézsége lehetett, mint más gyártók önvezető autóinak. Lásd: <https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html>.

¹² Amint fentebb említettük, ezeket a problémákat gyakran "troliproblémáknak" nevezik. Lásd a 6. lábjegyzetet. Nagy szakirodalom elemzi őket, és empirikus kutatásokat végeznek arra vonatkozóan, hogy az emberek hogyan reagálnak rájuk, és számos kommentátor felismerte alkalmazásukat az önvezető autókra és más új technológiai fejlesztésekre. Lásd például: W. Bradley Wendel, *Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles*, *Case W. L. Rev.* (2018); Bryan Casey, *Amoral Machines, or: How Roboticians Can learn to Stop Worrying and Love the Law*, 111 Nw. U. L. Rev. 1347, 1353 & n. 36 (2017) és az ott hivatkozott források.

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély
nélkül.

3

a fogyasztók lehetőségei? Ha igen, a fogyasztókat felelősségre vonják-e, ha "önző" autókat vásárolnak, szemben az "önzetlen" autókkal? Fel kell-e tárnuk a gyártóknak ezt a kockázatot? Kétségtelenül számos egyéb fontos prudenciális kérdés merül fel, amelyekkel foglalkozni kell ahhoz, hogy a gyártók és vásárlók megalapozott, önérdékű döntéseket hozhassanak.

Itt is szerepet játszhat a technológia. Eljöhet az a nap - talán hamarabb, mint gondolnánk vagy szeretnénk -, amikor a technológia részt vehet a jogi értelmezésben.¹³ Lehet, hogy ugyanolyan jól, majd jobban, mint az emberek. Sokan úgy gondolták, hogy egy számítógép soha nem győzheti le a sakkvilágbajnokot. A játék egyszerűen túl bonyolult. Aztán az egyik mégis megtette.¹⁴ Ugyanez történt nemrégiben a Go esetében is, amely egy ősi játék, amely sokkal bonyolultabb, mint a sakk, és amelyben az embereknek még nagyobb relatív előnyük volt a számítógépekkel szemben.¹⁵ Talán csak idő kérdése, hogy a jogértelmezéssel - vagy legalábbis a jog értelmezésének megjósolásával - is így legyen.

Tegyük fel, hogy a számítógépes programozók kifejlesztenek egy mesterséges intelligenciával rendelkező programot - nevezzük Herkulesnek¹⁶ -, amely még a legtapasztaltabb és letehetségesebb ügyvédeknel is hatékonyabban jelzi előre a lehetséges jogi kimeneteleket. Vajon a számítógépek akkor a mi hipotetikus esetünkben feleslegessé tették volna az ügyvédeket? Nem feltétlenül. A jogi értelmezés ugyanis többről szólhat, mint a prudensségről. Egy másik lényeges szempont lehet az erkölcsi. Lehet, hogy *erkölcsi* kötelességünk követni a törvényt - vagy legalábbis figyelembe venni azt, amikor arról döntünk, hogyan cselekedjünk. Ha igen, akkor hogyan kellene a jogértelmezőknek - feltehetően jogászoknak - megmondaniuk az autonóm autógyártóknak, hogy mit ír elő (vagy tilt vagy engedélyez) a törvény? És kiszoríthatja-e őket a Hercules? Megszüntetheti-e a Hercules az ügyvédek - vagy más emberi lények¹⁷ - szerepét a szóban forgó élet-halál döntések meghozatalában? Vajon egy

¹³ Lásd pl. Richard Susskind és Daniel Susskind, *A szakmák jövője: How Technology Will Transform the Work of Human Experts* 283 (2015) ("Ahogy a gépek egyre inkább képessé válnak, arra a kérdésre, hogy "Mi marad az emberi szakemberek feladata?", szintén nehéz ellenállni annak a következtetésnek, hogy a válasznak az kell lennie, hogy "egyre kevesebb"). Lásd még id. 66-71. o. (összefoglalva többek között azt, hogy a technológia ma már milyen módon képes elvégezni olyan feladatokat, amelyek történelmileg az ügyvédek számára voltak fenntartva).

¹⁴ Tegmark, *Life 3.0* at 51.

¹⁵ Lásd <https://www.theguardian.com/technology/2017/may/23/alphago-google-ai-beats-ke-jie-china-go>. A világ legjobb Go játékosa, Ke Jie, a jelek szerint megdöbbent, hogy veszett egy számítógép ellen, ami állítása szerint soha nem fordulhat elő. Id. A válasza áradozó volt: "Úgy érzem, hogy a játéka egyre inkább a 'Go istenéhez' hasonlít. Tényleg, ez zseniális". Id.

¹⁶ Ezt a nevet Ronald Dworkin fiktív idealizált tolmácsától kölcsönözöm, akit először a Ronald Dworkin, *Hard Cases*, 88 HARV. L. REV. 1057, 1083 (1975), majd később továbbfejlesztette a RONALD DWORKIN, *LAW'S EMPIRE* (1986) című könyvében. Dworkin Herkulese, hasonlóan a szövegben szereplő programhoz, olyan módon képes szintetizálni a különböző jogforrásokat, amire egyetlen ember sem képes.

¹⁷ Természetesen a Herkules nem váltaná ki az összes embert, ha nem tudná meghozni a megfelelő üzleti döntéseket. Tegyük fel, hogy képes rá - legalábbis olyan mértékben, amilyen mértékben ezek a döntések a profit maximalizálása érdekében születnek, és nem vonnak maguk után olyan erkölcsi ítéleteket, amelyek a jogi értelmezéshez is szükségesek lehetnek. Lásd pl. Susskind és Susskind, *The Future of the Professions*, 13. lábjegyzet, 78-84. o. (többek között arról, hogy a technológia hogyan szorította ki a vezetési tanácsadókat).

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély
nélkül.

4

számítógép, amely képes programozni az autót, amely viszont eldönti, hogy elkaszálja-e a gyalogosokat, vagy feláldozza a "sofőrjét"? Eljutottunk-e oda, hogy emberi elme nélkül is van jogunk - és jogi etikánk -?

Vegyük észre a lehetséges párhuzamot az autonóm autó előtt álló kihívásokkal, amikor nem lehet elkerülni az emberi élet elvesztését. Az autónak, mint megjegyeztük, először is leíró vagy előrejelző értékelést kell készítenie a különböző cselekvési módok következményeiről, másodsor pedig előíró értékelést kell készítenie arról, hogy mi a helyes választás a különböző lehetőségek között. Ugyanez a megkülönböztetés alkalmazható a jogértelmezésre is. Az első kihívás a jog leírása vagy annak előrejelzése, hogy mások valószínűleg hogyan értelmeznék azt. Tekintettel a jog meghatározatlanságára, azt várnánk, hogy az eredmény különböző lehetséges értelmezések lesznek, amelyeknek különböző valószínűséggel fogadja el őket egy tekintélyes jogértelmező.¹⁸ A második kihívás az, hogy a rendelkezésre álló értelmezések közül ki kell választani egy adott jogértelmezést. Hogyan *kell* értelmezni a jogot? Ez magában foglalhat némi erkölcsi ítéletet - és megőrizheti a jogászok és más emberi lények szerepét, hacsak és amíg Herkules nem képes nemcsak leíró vagy előrejelző ítéleteket hozni a jogról, hanem erkölcsi ítéleteket is.¹⁹

Ez a tanulmány egy újszerű jogtudományi tézist alkalmaz, hogy amellet érveljen, hogy az ügyvédek - és más emberi lények - még Herkules felemelkedése után is relevánsak maradnak. A tézis szerint a jog természetéről szóló legjobb beszámoló az értelmezés céljától függően változik. Pontosabban, amikor a jogértelmezők pusztán a jog leírására törekcsenek - vagy arra, hogy megjósolják, hogyan fogják mások értelmezni -, akkor a jog a legjobban a jogi pozitívizmusnak megfelelően értelmezhető. Amikor azonban a jogértelmezők a jogra mint az erkölcsi útmutatás forrására tekintenek, akkor az erkölcsre kell támaszkodniuk ahhoz, hogy a jogot kellően meghatározóvá tegyék ahhoz, hogy hasznos legyen.²⁰ Ennélfogva nem pozitivistaként (vagy természetes jogászként) kell eljárniuk. Ha tehát az autonóm autók programozásában részt vevő ügyvédeknek erkölcsi kötelességük azt tanácsolni ügyfeleiknek, hogy tartsák be a törvényt - és ez a tanulmány okokat javasol arra, hogy miért tehetik ezt meg -, akkor úgy kell eljárniuk, mint az ún.

¹⁸ Van itt egy párhuzam az erkölcsi nézeteltéréssel. A filozófusok "metanormatív" elméleteket dolgoztak ki abban a reményben, hogy találnak egy olyan mechanizmust, amely az intelligens gépeket az egymásnak ellentmondó erkölcsi elméletek összeegyeztetésére irányítja. Lásd pl. Kyle Bogosian, Implementation of Moral Uncertainty in Intelligent Machines, 27 MINDS & MACHINES 591, 595-603 (2017) (többek között William MacAskill, Normative Uncertainty (2014), elérhető a [https://s3.amazonaws.com/academia-oldalon\).edu.documents/34857095/Normative_Uncertainty_Compl_ete.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1527105349&Signature=WpMsYO%2FztrWnQLhruSpAY3U884E%3D&response-content-disposition=inline%3B%20filename%3DNormative_Uncertainty.pdf](https://s3.amazonaws.com/academia-oldalon).edu.documents/34857095/Normative_Uncertainty_Compl_ete.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1527105349&Signature=WpMsYO%2FztrWnQLhruSpAY3U884E%3D&response-content-disposition=inline%3B%20filename%3DNormative_Uncertainty.pdf)). Amint azt Bogosian helyesen elismeri, az erkölcsről szóló nézeteltérések kiterjednek a metamoralitással kapcsolatos nézeteltérésekre is, és magukban foglalják azokat. Id. at 603-04.

¹⁹ Ha ez a nap elérkezik, a számítógépek - amint azt alább tárgyaljuk - kiszoríthatják az ügyvédeket és a bírakat egyaránt.

²⁰ Ezt a tézist máshol kezdtem el kidolgozni, időnként együttműködve Manuel Vargas filozófussal. Lásd például Joshua P. Davis, Legality, Morality, Duality, 2014 UTAH L. REV. 55; MANUAL VARGAS AND JOSHUA P. DAVIS, AMERICAN LEGAL REALISM AND PRACTICAL GUIDANCE IN REASONS AND INTENTIONS IN LAW AND PRACTICAL AGENCY (2015) (G. Pavlakos &

V.

Rodriguez-Blanco, szerk.); Joshua P. Davis, Legal Dualism, Legal Ethics, and Fidelity to Law, 2016 ABA J. OF PROF. LAWY. 1; Joshua P. Davis & Manuel R. Vargas, Legal Dualism, Naturalism, and the Alleged Impossible of a Theory of Adjudication, in NATURALISTIC JURISPRUDENCE (Giudice, szerk. megjelenés: 2019) (tervezet a szerzőnél).

természetes jogászok. A jog értelmezése során erkölcsi ítéleteket kell hozniuk. És ez indokolhatja, hogy embereknek továbbra is különleges szerepük van abban, hogy megmondják, mi a jog. Ha ez igaz, akkor nem biztos, hogy hamarosan - vagy akár soha - eljutunk oda, hogy emberi elme nélkül is legyen jogunk - és jogi etikánk -.

A II. rész a technológiára összpontosít. A II.A. rész először a mesterséges intelligenciában rejlő lehetőségeket vizsgálja, röviden áttekintve az elmúlt évek néhány fejlesztését, és azt sugallja, hogy a számítógépek nem csak átvehetik az autók vezetését, hanem képesek lehetnek megjósolni, hogyan értelmezik majd az önvezető járművekre vonatkozó jogszabályokat. A II.B. rész tárgyalja a mesterséges intelligencia növekvő szerepét kísérő veszélyt - azt, hogy a számítógépek olyan döntéseket hoznak, amelyek aggályos következtetésekre támaszkodnak, vagy amelyek nemkívánatos hatásokkal járnak. A II.C. rész megjegyzi az e kérdések kezelésével kapcsolatos kihívást - azt, hogy nehéz meghatározni és megérteni, hogy a számítógépek hogyan végzik az általuk készített értékeléseket, ami látszólag szükséges lépés az általuk jelentett potenciális veszélyek elkerülése érdekében. A II.D. rész ezután megállapítja a számítógépek szerepének valószínűsíthető korlátozását: a számítógépek képesek leírni és megjósolni - segítve minket abban, hogy kiválasszuk a céljaink elérésének eszközeit -, de nem biztos, hogy képesek a szükséges erkölcsi ítéleteket meghozni annak meghatározásához, hogy milyen végső célokat kell követnünk.

A III. rész ezután a joggyakorlatra tér rá. A III.A. rész fontos kapcsolatot mutat ki a számítógépek II.D. részben tárgyalt lehetséges szerepe és a jogtudomány között: egyrészt a számítógépek nem képesek végső értékítéleteket, beleértve az erkölcsi ítéleteket is, másrészt a jogtudományban az elmúlt ötven évben a központi vita az erkölcsi ítéletek szerepéről szólt a jog megítélésében.²¹ A III.B. rész a jogtudomány e központi vitájának egy újszerű megoldását tekinti át, amelyet az elmúlt években dolgoztam ki (időnként Manual Vargas filozófussal közösen írt írásokban): "Jogi dualizmus".²² Ez azt sugallja, hogy az erkölcsnek nem kell szerepet játszania annak megmondásában, hogy mi a jog, amikor a jogértelmező csupán a jog leírására törekszik, vagy arra, hogy megjósolja, hogyan fogják mások értelmezni, de az erkölcsnek szükséges szerepe van, amikor a jogértelmező erkölcsi útmutatást keres a jogból. A jogi dualizmus tehát kijelöli a számítógépek jogértelmezésben játszható szerepének lehetséges korlátját: nem helyettesíthetik az emberi lényeket, amikor a jog az erkölcsi útmutatás forrásaként szolgál. A III.C. rész felvet néhány okot arra, hogy ez a pont nem pusztán akadémikus - a jog bizonyos körülmények között valószínűleg az erkölcsi útmutatás forrásaként szolgálhat, többek között az autonóm autók programozásának meghatározásakor. III.D. része végül megjegyzi, hogy a technológia és a jogtudomány közötti kapcsolatra irányuló vizsgálat megoldást kínál a jogtudomány fő vitájára, valamint egy potenciálisan termékeny új kutatási irányt is: a jogi dualizmus elfogadása utat mutathat a jogi pozitivisták és a természetjogászok között régóta tartó vitán túl, amely segíthet körülírni a mesterséges intelligencia megfelelő szerepét a jogértelmezésben; a jogi pozitívizmus és a nem pozitívizmus (vagy természetjog) közötti különbségtétel újradefiniálása pedig olyan módon mozdíthatja elő a jogtudományt, amely nyomon követi, hogy milyen típusú ítéleteket hozhat és milyeneket nem hozhat a mesterséges intelligencia.

A IV. rész rövid következtetést tartalmaz. Azt sugallja, hogy egy olyan világról való elmélkedés, amelyben a számítógépek ügyvédként és bíróként szolgálhatnak - akár reális előrejelzésről, akár csupán provokatív kísérletről van szó -, megtaníthat minket valamit a mesterséges intelligencia természetéről és a jog természetéről.

- ²¹ Lásd a fenti megjegyzést.
²² Lásd a fenti megjegyzést.

II. Számítógépek és jogértelmezés: Lehetőség, veszély, kihívás, korlát

A számítógépek és a mesterséges intelligencia (vagy gépi tanulás) növekvő szerepe társadalmunkban lenyűgöző és összetett. Ez a cikk nem próbál hiteles kijelentéseket tenni ezzel kapcsolatban. A jelen elemzés szempontjából azonban négy korlátozott megfigyelés fontos. Ezek a mesterséges intelligenciában rejlő lehetőségekre, az általa jelentett veszélyre, a veszély kezelésével kapcsolatos kihívásra, valamint a mesterséges intelligenciában rejlő lehetőségek látszólagos korlátaira vonatkoznak. Ez az elemzés keretet biztosít annak feltárásához, hogy a jogtudomány milyen módon tájékoztathat minket arról, hogy mi a megfelelő szerepe a mesterséges intelligenciának a társadalmunkban, és hogy a mesterséges intelligencia milyen módon tájékoztathatja a jogtudományról alkotott elképzeléseinket.

A. Lehetőség: A számítógépek elkerülhetlensége a jogértelmezésben?

Vegyük először a számítógépekben és a mesterséges intelligenciában rejlő lehetőségeket. Hajlamosak felülmúlni a várakozásainkat. Képesek olyan elemzések elvégzésére, amelyeket nemrég még elérhetetlennek tartottunk. A szakértők egykor kételkedtek abban, hogy egy számítógép valaha is legyőzi a sakkvilágbajnokot. Az emberi elme - a maga intuíciójával - egyszerűen jobbnak bizonyult. Aztán mégsem volt az.²³ Ma már a legjobb sakkozók oktatóként használják a számítógépeket - arra, hogy új lehetőségeket ismerjenek fel, és következtetéseket vonjanak le a legjobb sakklépésekről.²⁴ A számítógépek felsőbbrendűek. Az emberek nem tudják legyőzni őket. Legfeljebb tanulhatunk tőlük.

Az autonóm autók egy másik fontos példát szolgáltatnak. Az önvezető autókról szóló viták viszonylag új keletűek - legalábbis azok körében, akik nem a technológiai innovációra szakosodtak. A projekt ijesztőnek tűnt. Az emberi elme hatalmas mennyiségű adatot dolgoz fel és rendszerez, miközben egy gépjárművet működtet. A való világ nem egy sakktábla. A sakktáblák egy rögzített térben - egy nyolcszor nyolc rácsban - mozognak, a mozgásra vonatkozó világosan előírt szabályokkal. Nincsenek szórakozott tizenévesek, akik mobiltelefonon sms-eznek, miközben reggeliznek és rádióállomást váltanak. Nincsenek kisgyerekek, akik váratlanul egy mozgó gyalog vagy huszár elé szaladnak. Az emberi megismerés - beleértve a fizikai világról szóló információk érzékelését és értelmezését, valamint a valós idejű reagálást - hatalmas előnynek tűnik a számítógépekkel szemben. Most azonban sok kommentátor szerint az önvezető autók biztonságosabbak, mint az emberi sofőrök, és "a közúti halálesetek legalább 90%-át" képesek megszüntetni.²⁵ Az emberi sofőrök kiszorítása elkerülhetetlen lehet. Az autonóm autók annyira biztonságosabbak lehetnek az emberi vezetőknél, hogy egy napon a nem is olyan távoli jövőben nehéz lesz biztosítást kötni azoknak, akik maguk akarnak vezetni.²⁶

²³ A sakkvilágbajnok 1997-ben kikapott egy számítógéptől. Lásd: <http://www.sciencemag.org/news/2016/03/update-why-week-s-man-versus-machine-go-match-doesn-t-matter-and-what-does>.

²⁴ Id.

²⁵ TEGMARK, LIFE supra3.0, note , at 99.

²⁶ Lásd id., 109. c. (annak tárgyalása, hogy a különösen biztonságos önvezető autók az emberi sofőrökénél olcsóbb biztosításhoz vezethetnek).

Tekintettel erre az előzményekre és összefüggésekre, nem tűnik túlzónak az a lehetőség, hogy a számítógépek értelmezzék a jogot - legalábbis abban az értelemben, hogy megjósolják, hogyan fognak eljárni a bíróságok.²⁷ Az irodalomelméleti szakemberek szeretik azt mondani, hogy nincs önértelmező szöveg.²⁸ Talán nincs is. Vagy talán nem is volt. De talán lesz is. A számítógépek talán képesek lesznek arra, hogy a szöveget más szövegekre fordítsák - felismerjék, felerősítsék és tisztázzák a jelentéseket, hasonlóan az emberi lényekhez. Lehet, hogy képesek lesznek megmondani, hogy mit jelentenek a szövegek, beleértve a jogi szövegeket is, vagy legalábbis képesek lesznek szintetizálni az emberi nyelvhasználatot, és megjósolni, hogy az emberi lények hogyan értelmeznék a szövegeket, beleértve a jogi szövegeket is, és hogyan reagálnának rájuk.²⁹ Herkules talán hamarabb felemelkedik, mint azt váránk.

Ebben az összefüggésben gondoljunk Franklin Foer leírására a Google törekvéseiről:

A Google duzzadó portfóliójának középpontjában egy fő projekt áll: A vállalat olyan gépeket akar létrehozni, amelyek lemásolják az emberi agyat, majd továbblépnek azon. Ez a lényege a globális tudás teljes körű adatbázisának létrehozására irányuló kísérleteinek, valamint az algoritmusok képzése érdekében tett erőfeszítéseinek, hogy azok képessé váljanak a minták felismerésére, megtanítsák őket a képek megkülönböztetésére és a nyelv megértésére.³⁰

Foer szerint a Google arra törekszik, hogy újraterejtse az emberi elmét, és továbbfejlessze azt.

Hasonlóképpen, Foer értelmezésében a Facebook célja, hogy a számítógépek és az őket tervező mérnökök uralmával kiszorítsa a hétköznapi kormányzást, és ³¹még a választási részvételre is hatással legyen.³² Foer idézi Mark Zuckerberget, a Facebook alapítóját, aki elismerte: "A Facebook sok szempontból inkább hasonlít egy kormányra, mint egy hagyományos vállalatra. Van ez a hatalmas közösségünk, és más technológiai cégeknél sokkal inkább mi határozzuk meg a politikát".³³

²⁷ Lásd Susskind & Susskind, *The Future of the Professions*, 69-70. o. és nn. 156-57 ("A Big Data technikák olyan rendszerek alapjául szolgálnak, amelyek a szakértő pereskedőknél is jobban előre jelzik a bírósági döntések eredményét, a szabadalmi vitáktól (a Lex Machina szolgáltatás) az Egyesült Államok Legfelsőbb Bíróságáig") (idézi <<https://lexmachina.com>>; Daniel Katz, Michael Bommarito, and Josh Blackman, "Predicting the Behavior of the Supreme Court of the United States: A General Approach (2014. július 21.), <<https://dx.doi.org/10.2139/ssrn.2463244>>).

²⁸ Lásd pl. Patrick Sullivan, "Reception Moments," *Modern Literary Theory, and the Teaching of Literature*, 45 *J. of Adolescent & Adult Literacy* 568, 568 (2002) ("A jelentés létrehozásának folyamatát ma már egyfajta együttműködésnek tekintjük a szerző, az olvasó, a kultúra vagy "értelmező közösség", amelyben a szerző és az olvasó él, és a nyelv, amellyel a szöveget felépítették.").

²⁹ Nem kívánok állást foglalni az irodalomelméleten belüli nézeteltérésekről. Csak utalni szeretnék, és agnosztikus maradok azzal kapcsolatban, hogy mi az, amit az AI értékelni fog, ha és amikor pontos állításokat tehet a szövegek jelentéséről, legalábbis előrejelzési céllal.

³⁰ FOER, *WORLD WITHOUT MIND*, 33. o. Ha valaki vitatkozni akarna ezzel az összefoglalóval, megkérdőjelezhetné, hogy a Google algoritmusokat képez-e és tanít-e, vagy inkább azt kellene mondani, hogy a számítógépes programok képezik és tanítják magukat.

³¹ Foer azt írja Zuckerbergről, hogy örökölt "egy tartós fantáziát, egy álomsorozatot, amelyben kidobjuk a csöves politikusokat, és mérnökökkel helyettesítjük őket - a

szabályzókkal való uralom". Id. 61. o. ³² Id. 75. o.
³³ Id. 61.

A Google arra törekszik, hogy a mesterséges intelligencia képes legyen mindarra a gondolkodásra, amire az emberek képesek, csak hatékonyabban.

A Facebook átveheti az általában a kormányok számára fenntartott feladatokat. Nemrég pedig arról olvashattunk a hírekben, hogy a mesterséges intelligencia először teljesített jobban, mint az ember egy elismert olvasásértései teszten.³⁴ A számítógépes bírák már nem is tűnnek olyan messzinek.³⁵

Ezután jegyezzük meg Max Tegmark megjegyzéseit a nemrégiben megjelent, LIFE 3.0 című könyvében: AZ EMBERNEK LENNI A MESTERSÉGES INTELLIGENCIA KORÁBAN:

Mivel a jogi folyamat absztrakt módon számításnak tekinthető, amely a bizonyítékokról és a jogszabályokról szóló információk bevitelével és a döntés meghozatalával jár, egyes tudósok arról álmodnak, hogy *robotbírókkal* teljesen automatizálják azt: Olyan mesterséges intelligencia-rendszerek, amelyek fáradhatatlanul ugyanazokat a magas jogi normákat alkalmazzák minden ítéletre anélkül, hogy az emberi hibáknak, például az elfoglaltságnak, a fáradtságnak vagy a legújabb ismeretek hiányának engednének.³⁶

Tegmark azt sugallja, hogy a bíró szerepe tisztán mechanikus, ha összetett is. Az ideális bíró pusztán számításokkal foglalkozik. Ha ez így van, akkor a mesterséges intelligenciával rendelkező bírák - akiket ő robotbíróknak nevez - számos előnnyel járhatnak. Elméletileg nem befolyásolná őket a tudatalatti elfoglaltság.³⁷ Ráadásul megismételhetők és nem fáradnak el, így nem kell aggódnunk amiatt, hogy túl kevés bíró van, vagy hogy a meglévők kimerülnek.³⁸ Továbbá korlátlan memória- és tanulási kapacitással rendelkeznek, így többé nem kell aggódnunk a szakértelem vagy a tudás hiánya miatt.³⁹ Lépjen be Herkules (vagy sok Herkules).

Hogy igazságos legyek, Tegmark elismeri a robotbírák potenciális felelősségét is. Meghackelhetik őket.⁴⁰ Az átláthatóság hiánya is előfordulhat, ami alááshatja a jogrendszer iránti tiszteletet.⁴¹ És előfordulhat, hogy nem orvosolják, hanem inkább megismétlik a diszkriminációs mintákat, amint azt egy nemrégiben készült tanulmány is bizonyítja, amely kimutatta, hogy a visszaesés előrejelzésére tervezett szoftverek az afroamerikaiakkal szembeni elfoglaltságot eredményeznek az ítélelhozatal során.⁴²

³⁴ Lásd Sherisse Pham, "Computers Are Getting Better than Humans at Reading", [cnn.com](http://money.cnn.com/2018/01/15/technology/reading-robot-alibaba-microsoft-stanford/index.html) (2018. január 15.), elérhető a <http://money.cnn.com/2018/01/15/technology/reading-robot-alibaba-microsoft-stanford/index.html> oldalon.

³⁵ Lásd pl. Anna Ronkainen, *From Spelling Checkers to Robot Judges?: Some Implications of Normativity in Language Technology and AI & Law* (elérhető a <http://ssrn.com/abstract=1879426> oldalon (a szoftverek bíráskodásban való lehetséges alkalmazásáról, valamint a gyakorlati lehetőségekről és kihívásokról).

³⁶ TEGMARK, ÉLET 3.0: (1052017).

³⁷ Id. 105-06.

³⁸ Id. 106.

³⁹ Id.

⁴⁰ Id.

⁴¹ Id.

⁴² Id. at 106-07 & n.36 (idézve a Julia Agnwin, Jeff Larson, Surya Mattu és Lauren Kirchner által készített Pro Publica jelentést, "Machine Bias", (2016. május 23.), elérhető a

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély

nélkül
<http://tinyurl.com/robojudge> oldalon). Lásd még CATHY O'NEIL, WEAPONS OF MATH
DESTRUCTION: HOW

Tegmark víziója éppoly lenyűgöző, mint amennyire ijesztő. Elgondolkodtat bennünket, hogy vajon van-e bármi, amit az ember az elméjével meg tud tenni, és amit a számítógépek nem fognak hamarosan jobban csinálni. Arra is rá kell vennie bennünket, hogy elgondolkodjunk azon, hogy a számítógépes "gondolkodás" milyen veszélyeket rejthet magában.

B. Veszély: A számítógépek helytelen okokból cselekszenek?

A számítógépek és a mesterséges intelligencia egyre növekvő szerepe a döntéshozatalban félelemmel kell, hogy eltöltsön bennünket, mind a pozitív - félelmetes -, mind a negatív - borzalmas - értelemben. Gondoljunk csak arra, hogyan reagált az egyik kommentátor - Frank Lautz, a New York-i Egyetem Game Centerének igazgatója - a DeepMind, a Google titkos mesterséges intelligencia leányvállalata által kifejlesztett AlphaZero nevű új sakkprogramra.⁴³ Az AlphaZero lehet a világ legjobb sakkozója. Sőt, más szempontból is figyelemre méltó. Nem az emberi sakkozás évszázados tapasztalataiból merít - mint sok elődje, köztük a GO AI -, hanem a semmiből építi fel algoritmusait.⁴⁴ Mr. Lautz válasza:

Egy ideig, körülbelül két hónapig, azt mondhattuk magunknak: "Nos, a Go AI több ezer évnyi felhalmozott emberi gondolkodást tartalmaz, a heurisztikák, közmondások és híres játékok összes felhalmozott tudását". Ezt a történetet már nem tudjuk elmondani. Ha ezt nem találod legalább egy kicsit ijesztőnek, akkor erősebb fából faragtak, mint én. Félelmetesnek találom, de gyönyörűnek is. Minden meglepő valamilyen módon gyönyörű.⁴⁵

Lautz úr rémülete talán érthetőbb lenne a sakkozáson kívüli kontextusban. Végül is a sakk csak egy játék. De valószínűleg mindannyian ismerjük azt a szerepet, amelyet a számítógépes algoritmusok játszhattak a 2016-os elnökválasztás aláaknázásában. Az oroszok "feltörhették" a közösségi hálózatokat, hogy megváltoztassák sok amerikai politikai preferenciáit, többek között olyan oldalakon keresztül, mint a Facebook és a Twitter.⁴⁶ Nem ez az egyetlen veszély. Gondoljunk Latanya Sweeney harvardi professzor tanulmányára, amely szerint "az afroamerikai neveket gyakran célozták meg olyan Google-hirdetésekkkel, amelyek egyenesen azt sugallták, hogy letartóztatási nyilvántartásuk van, amelyet törölni kellene".⁴⁷

A probléma az, hogy a számítógépek nem zárják ki az elemzés olyan formáit, amelyek erkölcsstelenek, etikátlanok vagy törvénytelenek lehetnek. Képzeljük el például, hogy egy cég - nevezzük AllTooCommon Corp. ("ATCC")-, amely aggasztó szexuális zaklatási mintát mutat. Tegyük fel, hogy az ATCC-nél a nők rendszeresen tapasztalnak helytelen viselkedést, és teljesítményük szenved. Néhányan közülük nem szólnak semmit, és elkedvetlenednek. Mások jelentik a viselkedést, és megtorlást szenvednek el. Bárhogy is legyen, tegyük fel, hogy a nők aránytalanul nagy arányban tapasztalnak jogellenes, hátrányos foglalkoztatási döntéseket. Rosszabbul teljesítenek, mint

BIG DATA INCREASE INEQUALITY AND THREATENS DEMOCRACY (2042016) (megjegyezve, hogy a "Big

Data" állandósítja az egyenlőtlenséget azáltal, hogy "kodifikálja a múltat" ahelyett, hogy "feltalálná a jövőt").⁴³ Lásd Oliver Roeder, "Chess's New Best Player Is A Fearless, Swashbuckling Algorithm," FiveThirtyEight.com (2018. január 3.).

⁴⁴ Id.

⁴⁵ Id.

⁴⁶ Lásd <https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html>.

⁴⁷ FOER, LAW WITHOUT MIND, at 71.

férfiak a látszólag objektív mérések szerint a munkahelyen ellenük irányuló jogellenes magatartás miatt.⁴⁸

Adjon meg egy számítógépet, amelynek feladata a teljesítmény előrejelzése a felvétel, megtartás, kompenzáció és előléptetés céljából. A gép mesterséges intelligenciát használ. Nem egyszerűen egy algoritmust alkalmaz. Mintákat észlel, és ezekre támaszkodva generálja és adaptálja az algoritmusokat. Ilyen körülmények között a számítógép megerősítheti a munkahelyen már meglévő diszkriminációt. Az alkalmazottakról rendelkezésre álló információk és a vállalat által a teljesítmény mérésére használt mérőszámok - beleértve a korábbi értékeléseket, a megtartást, az emeléseket és az előléptetéseket - alapján a számítógép alkalmas lenne arra, hogy megjósolja, hogy a nők - minden más tényezőt figyelembe véve - rosszabbul teljesítenek, mint a férfiak. Ezzel a számítógép tulajdonképpen büntetné a nőket az őket ért hátrányos megkülönböztetésért, látszólag megengedhetetlenül nemi alapon megkülönböztetve őket.⁴⁹

Ez a példa nem pusztán feltételezés. A vállalatok már most is használják a mesterséges intelligenciát alkalmazottaik értékelésére.⁵⁰ A kommentátorok pedig máris aggodalmukat fejezik ki, hogy ezáltal erősíthetik az előítéleteket - és ezt olyan módon tehetik, amely megnehezíti a jogorvoslatot.⁵¹ A jelenlegi jogi doktrínáink nem feltétlenül alkalmasak arra, hogy a vállalatokat ellenőrizzék, amikor a számítógépek olyan elemzések alapján javasolnak döntéseket, amelyek emberi lények által végzett elemzések esetén megengedhetetlenek lennének. Van-e valaha is diszkriminatív szándék, amikor például adatbányászatról és mesterséges következtetésekről van szó? És vajon az ilyen erőfeszítések - eleve - szükségszerűen munkával kapcsolatosak és összhangban vannak-e az üzleti szükségszerűséggel?⁵²

Hasonló kockázatokkal járhatnak az önvezető autókkal kapcsolatos döntések, ha kevésbé nyilvánvalóan is. Gondoljunk csak arra, mi történne, ha egy autót úgy programoznának, hogy minimalizálja az általa okozott jogi felelősséget egy

⁴⁸ A közelmúltban a hírekben megjelent történetek sorát tekintve ez a feltételezés fájdalmasan hihetőnek tűnik. Egy példa a sok közül, gondoljunk csak a New York Times által közölt állítólagos zaklatási kultúrára két chicagói Ford-gyárban. Susan Chira és Catrin Einhorn, "How Tough Is It To Change a Culture of Harassment? Ask Women at Ford", NYT (2017. december). A Ford a túl sok példa közül egyet szolgáltat. Lásd még Emily Steel, "At Vice, Cutting-Edge Media and Allegations of Old-School Sexual Harassment," NYT (2017. dec. 23.).

⁴⁹ Hasonló potenciál és veszély kíséri a számítógépes algoritmusok és a mesterséges intelligencia más felhasználási módjait is. Gondoljunk csak arra, hogy a gyermekvédelmi szolgálatok pittsburghi tisztviselői a számítógépekre támaszkodnak, hogy megjósolják, mely gyermekek vannak veszélyben. Lásd Dan Hurley, "Can an Algorithm Tell When Kids Are in Danger?" (Egy algoritmus meg tudja mondani, mikor vannak veszélyben a gyerekek?) NYT (1/2/18). A gyermekek védelme a családon belüli erőszaktól körülbelül olyan kényszerítő ok, mint amelyet csak találni lehet a technológiára való támaszkodásra. Ennek ellenére csak idő kérdése volt, hogy a Pittsburghben - vagy máshol - hozott döntések a helytelen megkülönböztetés észlelt mintáihoz vezessenek, beleértve a faji alapú megkülönböztetést is. Lásd VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (AZ egyenlőtlenség automatizálása: hogyan profilozzák, rendőrködnék és büntetik a szegényeket).

(2018); Jacob Weisberg, The Digital Poorhouse, NYRB 45-47 (2018. június 7.) (EUBANKS,

nélkül
AUTOMATING INEQUALITY; SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH
ENGINES REINFORCE RACISM (2018)).

⁵⁰ Lásd Pauline Kim, *Data-Driven Discrimination at Work*, WILL48. & MARY L. REV. 857, 902-
09 (2017).

⁵¹ Id.

⁵² Id.

baleset, beleértve a más járműveknek okozott károkat, valamint a megélhetés és az élet elvesztését.⁵³ Természetesen ennek valós idejű értékelése ijesztő technikai feladat lenne. De tegyük fel, hogy az autó számítógépe képes erre. Fel tudja mérni az úton lévő többi jármű értékét, a járművezető, utasok és gyalogosok sérülésének valószínűségét, valamint az ebből eredő potenciális felelősséget, beleértve a jövedelemkiesést is. Tekintettel a társadalomban a tőke és a jövedelem elosztására, nem lenne nehéz megjósolni, hogy az autó kinek az életét értékelné jobban, és kinek kevésbé. Minden más esetben a drágább járművek vezetői jobban járnának, mint a kevésbé drága járművek vezetői. A gazdagok védve lennének. A gazdagok pedig valószínűleg fehérek lennének. Továbbá - megint csak, minden más esetben egyenlő feltételek mellett - a legjobb kereseti éveikben járó emberek kedvezőbb bánásmódban részesülnének, mint a fiatalok és az idősek, a férfiak kedvezőbb bánásmódban, mint a nők, és a fehérek kedvezőbb bánásmódban, mint a színesbőrűek. Az autó életre-halálra vonatkozó döntéseket hozna olyan módon, amely megerősíti az osztály, kor, nem és faj szerinti egyenlőtlenségeket.⁵⁴

Ugyanez igaz lehet a jogértelmezésre is. Valójában a törvény az, amely megteremti azokat az ösztönzőket, amelyek egy önvezető autót arra készítetnek, hogy potenciálisan kifogásolható módon cselekedjen. Kapcsolódó és általánosabb aggályok merülhetnek fel abból, ha egy számítógép értelmezi a törvényt, esetleg megjósolja, hogyan értelmeznék azt a bírák. A bírák aránytalanul nagy arányban fehér férfiak. Vannak-e olyan elfogultságaik, amelyek befolyásolják a döntéshozatali mintáikat? Vajon a más csoportokhoz tartozó bírák is elfogultak, esetleg más irányba? Előre jelzik-e ezek a minták a bírói döntéseket, talán még olyan módon is, amelyet a bírák nem ismernek fel, és amelyet a jog emberi értelmezői nem vesznek észre vagy nem vesznek figyelembe? Ha igen, akkor a Hercules valószínűleg felismeri ezeket a tendenciákat - felismeri a diszkrimináció mintáit az esetjogban -, és olyan jogi értelmezéseket kínálhat, amelyek megerősítik ezeket. Hercules talán még nagyobb mértékben teszi ezt, mint a gyakorló jogászok, akik legalábbis megpróbálhatják megtisztítani a jogot a szennyeződésektől.⁵⁵ Talán Lautz úrnak igaza van - mindannyiunknak legalább egy kicsit meg kellene rémülnünk.

C. Kihívás: A mesterséges intelligencia megkerülhetetlenül megkérdőjelezhetetlen?

Ennél is ijesztőbb, hogy talán csak korlátozottan vagyunk képesek érzékelni, még kevésbé korrigálni a mesterséges intelligenciából eredő egyenlőtlenségeket. Ennek egyik oka, hogy a mesterséges intelligencia gyakran nem átlátható. Ha az lenne, talán képesek lennének a számítógépes döntéshozataltól a nemkívánatos előítéleteket kiszűrni. Ha egy számítógép jelezné, hogy a nők nemét egy bizonyos mértékben hátrányosan veszi figyelembe a nők jövőbeli munkateljesítményének előrejelzésében, akkor lehetséges lenne az elemzésnek ezt a részét kivenni - és korrigálni. Ez azonban a gyakorlatban több okból sem lehetséges.

⁵³ Lásd Bryan Casey, *Amorális gépek*, avagy: NW111. U. L. REV. 1347,1350 (2017) (azzal érvelve, hogy a profitmaximalizáló cégek úgy fogják megtervezni a mesterséges intelligenciát, hogy minimalizálják a jogi felelősséget, többek között az önvezető autók tervezésében).

⁵⁴ Lásd pl. MARTHA CHAMALLAS AND JENNIFER B. WRIGGINS, *THE MEASURE OF INJURY: RACE, NEMEK ÉS A KÁRTÉRÍTÉSI JOG* (2010). Megjegyzendő, hogy Jack Weinstein bíró olyan ítéleteket hozott, amelyek ellenállnak a kártérítési jog általános tendenciájának, megakadályozva, hogy az ügyvédek faji vagy etnikai alapon érveljenek a felelősség csökkentésével kapcsolatban. Ashley Southall, "Award in Lead Paint Lawsuit Can't Be Tied to Ethnicity, Judge Rules," NYT (2015. július 29.).

⁵⁵ Lásd Joshua P. Davis & Manuel R. Vargas, Legal Dualism, Naturalism, and the Alleged Impossible of a Theory of Adjudication, in NATURALISTIC JURISPRUDENCE (Giudice, ed. forthcoming 2019) (draft on file with author).

A számítógépek érvelési módjának javításával kapcsolatos egyik nehézség az lehet, hogy nem áll rendelkezésre elegendő idő. Az önvezető autókra például már most is rendkívüli terhet róttunk, ha azt kérjük tőlük, hogy emberi felügyelet nélkül, biztonságosan működjenek, és válsághelyzetekben valamilyen működőképes kritériumrendszer alapján minimalizálják a károkat. Lehet, hogy nem lehet még egy újabb elemzési réteget ráhúzni - olyat, amely megköveteli az autóktól, hogy értékeljék, hogy az általuk folyamatosan fejlesztett és adaptált biztonsági algoritmusok tartalmazznak-e implicit elfogultságokat. Ha ez a folyamat emberi felügyeletet igényelne, akkor nem lenne lehetséges valós időben. Sőt, úgy tűnik, ez megghiúsítaná az önvezető autók célját. Ha az autóknak saját magukat kellene felügyelniük, akkor még egy újabb mesterséges intelligencia rétegre lenne szükség. Ez pedig nem biztos, hogy megvalósítható. És a mesterséges intelligencia további rétegét is feltehetően felügyelni kellene.

A mesterséges intelligencia átláthatatlanságának másik oka az, hogy folyamatosan hozza létre és adaptálja saját algoritmusait, így még azok sem tudják megjósolni, hogy hogyan fogja csinálni, amit csinál, vagy utólagosan értékelni, hogy mit csinált. Gondoljunk csak Michal Kosinski, a Stanford Graduate School of Business professzorának helyzetére. Egy nyílt forráskódú arcfelismerő algoritmust alkalmazott nyilvánosan közzétett társkereső profilokra.⁵⁶ Egy kényes témát akart feltárni - azt, hogy a fizikai jellemzők korrelálnak-e a személyiségjegyekkel. Először semmi érdekeset nem talált. Amikor azonban arra kérte az algoritmust, hogy a fényképek alapján azonosítsa az alanyok szexuális irányultságát, az 91 százalékos pontossággal tette ezt a férfiak és 83 százalékos pontossággal a nők esetében.⁵⁷ Ez az eredmény megdöbbentő volt. És nyugtalanító. Amikor publikálta tanulmányát, az előre látható felháborodást váltott ki.⁵⁸ Mi mást ismerhetne még fel a mesterséges intelligencia a fényképekből? Mit jelenthet ez a jövőben a magánéletre nézve?

Ami kevesebb figyelmet kapott - ahogy Cliff Kuang a The New York Times Magazine-ban írta -, az "egy valódi rejtély volt, amely szinte figyelmen kívül maradt a média visszhangja közepette: *Hogyan* csinálta a számítógép azt, amit csinált? Mit látott, amit az emberek nem tudtak?"⁵⁹ Még Kosinski sem tudta. Az algoritmust nem úgy tervezték, hogy felfedje az általa észlelt mintákat. Kosinskinek különböző kísérleteket kellett végeznie, hogy a lehető legjobban *kikövetkeztesse*, hogyan csinálja az algoritmus, amit csinál.⁶⁰

Végül, és ez a legspekulatívabb, a számítógép "gondolkodásmódja" nem biztos, hogy az emberi megértésnek megfelel. Legalábbis lehetséges, hogy egy számítógép olyan módon szervezné az információt, amelyet az ember nem tud megérteni - talán azért, mert nem vagyunk elég okosak, talán azért, mert bizonyos minták nem értelmesek számunkra, talán azért, mert vannak bizonyos dolgok, amelyeket nem tudunk érzékelni, és talán más okok miatt.

⁵⁶ Cliff Huang, "Can A.I. Be Taught to Explain Itself?", The New York Times Magazine (2017. november 21.).

⁵⁷ Id.

⁵⁸ Id.

⁵⁹ Id. (kiemelés az eredetiben)

⁶⁰ Id. Lásd még Jacob Weisberg, The Digital Poorhouse, NYRB 45 (2018. június 7.) (amely Kosinski munkáját, valamint egy hasonló rejtélyt tárgyal, amely arról szól, hogy egy német

nélkül
"kézírás-felismerő algoritmus 80 százalékos pontossággal meg tudja jósolni, hogy egy
mintát férfi vagy nő írt-e").

Ez az utolsó pont mélyreható és csúszós lehet, és gondos munkát igényel az ismeretelmélet és az ontológia terén. Ez nem a megfelelő hely arra, hogy ezt az erőfeszítést megtegyük. A jelen célokra elegendő annak felismerése, hogy az idő és az átláthatóság nem biztos, hogy elegendő az emberi lények számára a mesterséges intelligencia ellenőrzéséhez. Valami többre lehet szükség. Ez a pont elérhetetlenül elvontnak tűnhet. De nem feltétlenül az. Próbálja meg elmagyarázni a legtöbb nyolcévesnek a valószínűségszámítást, a számítást vagy a többszörös regresszióelemzést. Egyszerűen nem tudják megérteni, függetlenül attól, hogy a magyarázat világos és teljes, az elmélet pedig megalapozott. A fogalmak egyszerűen túl bonyolultak. Ugyanez igaz a kvantummechanikára és a relativitáselméletre - sok felnőttre, nem csak a gyerekekre. És ugyanez igaz lehet néhány olyan fogalomra is, amelyet a mesterséges intelligencia kifejleszthet és használhat. Talán egyetlen emberi lény - még a legokosabbak és legtájékozottabbak sem - sem lesz képes megérteni őket.

A fenti három pont utal néhány olyan nehézségre, amelyek az Európai Unió általános adatvédelmi rendelete ("GDPR") alapján felmerülhetnek.⁶¹ A GDPR tartalmazza a "magyarázathoz való jogot" - a jogot arra, hogy magyarázatot követeljen arra vonatkozóan, hogy egy személyt érintő algoritmus hogyan jutott a következtetéseihez.⁶² Nem világos, hogy ez mit fog jelenteni a gyakorlatban. A kormánytisztviselőknek nehéz dolguk lesz annak meghatározásával, hogy mi számít elegendő magyarázatnak. Érthetőnek kell lennie? Ha igen, akkor kinek? Az átlagembernek az utcán? Egy szakértőnek? Az adott terület néhány szakembere? Egy hipotetikus személynek, aki elegendő tudással és intelligenciával rendelkezik ahhoz, hogy megértse, még akkor is, ha valójában senki sem rendelkezik mindkettővel? Vajon csak technikailag helyesnek és teljesnek kell lennie, még ha absztrúz - az emberi számításokon túlmutató - is?

Összefoglalva, nehéz megmondani, hogy a "magyarázathoz való jog" mennyire lesz gyakorlatias. Egyes döntések - például az önzetű autók által hozott döntések - gyorsak és visszavonhatatlanok lehetnek. A számítógép "gondolkodásának" utólagos összerakása csak korlátozottan lehet hasznos. Ráadásul nem biztos, hogy megvalósítható az a teher, hogy olyan programokat fejlesszünk ki, amelyekből kiderülne, mit tett a számítógép. Ez a feladat olyan költséges és nehézkes lehet, hogy túl nagy árat jelentene az innováció számára. És mindenesetre előfordulhat, hogy az emberek nem képesek megérteni, mit tett a számítógép, még akkor sem, ha a számítógép bizonyos értelemben képes megmagyarázni az "érvelését". Ez utóbbi különösen a jogértelmezésre vonatkozhat. Lehet, hogy a Herkules bizonyos szempontból hatékonyabban tudja értelmezni a jogot, mint az ember, de nem biztos, hogy képes felvilágosítást adni arról, hogyan jutott a következtetéseihez.⁶³

⁶¹ Májusban lépett hatályba Lásd25,2018.: <https://www.eugdpr.org/>

⁶² <https://www.csoonline.com/article/3254130/compliance/what-does-the-gdpr-and-the-right-to-explanation-mean-for-ai.html>. A GDPR-t kísérő, jogilag nem végrehajtható (71) preambulumbekzdés említi a "magyarázathoz való jogot". Maga a GDPR pedig kimondja, hogy az adatkezelőknek tájékoztatniuk kell a fogyasztókat az adataik felhasználásának módjáról, beleértve "az automatizált döntéshozatal meglétét", és legalább bizonyos körülmények között "érdemi tájékoztatást kell adniuk az érintett logikájáról, valamint az adatkezelés jelentőségéről és az érintettre gyakorolt várható következményeiről". GDPR, 15. cikk, elérhető a <http://www.privacy-regulation.eu/en/article-15-right-of-access-by-the-data-subject-GDPR.htm> oldalon.

⁶³ Ez még akkor is igaz lehet, ha a Hercules olyan nyelven tudja megindokolni döntéseit, amelyet az emberi jogértelmezők használnak. Figyeljük meg a különbséget - amelyet például Brian Leiter gyakran tesz - egy jogi értelmezés - beleértve a bírósági döntést is - tényleges

motivációja és a jogi tolmács, például egy bíró által felajánlott indoklás között. Lásd pl. Brian
Leiter, Explaining Theoretical Disagreement, 76 U. CHI. L. REV. 1215, 1225 (2009)
(magyarázza a "Disingenuity
nélkül

D. Limit: Hiányzik az autonóm autókból az autonómia?

A mesterséges intelligencia lehetőségei tehát fantasztikusak lehetnek, de legalább egy kicsit ijesztőek is, olyan módon, amivel nem könnyű megbirkózni. Ez nem jelenti azt, hogy a bennük rejlő lehetőségek korlátlanok. A számítógépek és a mesterséges intelligencia eddig csak az eszközökkel tudott foglalkozni, a végső célokkal nem. Másképpen fogalmazva, a tények, nem pedig az értékek birodalmában működtek.⁶⁴ Meg kell őket tanítani arra, hogy milyen végső célokat kövessenek.⁶⁵ Amint ez megtörtént, már használhatók a minták felismerésére és a viselkedés előrejelzésére, gyakran sokkal hatékonyabban, mint az emberek. Még arra is képesek lehetnek, hogy közbenső vagy instrumentális célokat fogadjanak el - olyanokat, amelyek segíthetnek nekik elérni a rájuk bízott végső célokat. Egyelőre azonban még nem tudják kiválasztani azokat a célokat, amelyek elérésére törekedniük kell. Bizonyos értelemben tehát az "autonóm autó" kifejezés félrevezető lehet. Az autonómiát gyakran úgy képzeljük el, mint ami a célok, és nem csak az eszközök megválasztását jelenti. Még nem jutottunk el arra a pontra, hogy a számítógépek célokat válasszanak. Úgy tűnik, hogy az embereknek kell megadniuk azokat.⁶⁶

Az a kilátás, hogy a számítógépeket felhatalmazzuk arra, hogy erkölcsi ítéleteket hozzanak, nem túl ígéretes. Végül is az erkölcsfilozófiában kevés az egyetértés az erkölcsi problémák helyes megfogalmazásának módját illetően. A filozófusok vitatkoznak az erkölcsi tételek természetéről, arról, hogyan lehet vagy kell azokat tesztelni, hogyan jutnak az emberek erkölcsi felismerésekre, lehetnek-e az erkölcsi állítások igazak vagy hamisak, mik az erkölcs szervező elvei, és még számos más kapcsolódó kérdéstről. Kétségtelenül ironikus, de az emberi lényeknek az a nagyon korlátozott képessége, hogy az erkölcsfilozófiában a konszenzus felé haladjanak, és magabiztos erkölcsi ítéleteket hozzanak, mindkét tekintetben a számítógépek fölé helyezheti őket. Lehet, hogy nem tudunk jobbat tenni annál, mint hogy magunktól - technológiai segítség nélkül - elkeveredjünk ezeken a zavaros vizeken, részben azért, mert nem rendelkezünk elegendő megértéssel ahhoz, hogy a számítógépeknek elég világos útmutatást adjunk ahhoz, hogy segítsenek nekünk.⁶⁷

Theory", hogy a jogértelmezők, beleértve a bírákat is, álságosan megmagyarázhatják következtetéseik alapját, különösen azt, hogy azt állítják, hogy a törvényt követik, miközben valójában új jogot alkotnak).

⁶⁴ E különbségtétel általános tárgyalását lásd pl.: RONALD DWORKIN, RELIGION WITHOUT ISTEN (2013).

⁶⁵ Azért mondom, hogy "végső" célok, mert a mesterséges intelligencia feltehetően képes lenne olyan instrumentális célokat azonosítani, amelyek segítik a végső célok meghatározott halmazának elérését. Lásd pl. TEGMARK, Supra Note 2, 264. o. (a részcélok és a végső célok kapcsolatának tárgyalása); BOSTROM, Supra Note.

, 132. o. (az instrumentális konvergencia tézisének tárgyalása).

⁶⁶ Így például Max Tegmark négy olyan alapelvet javasol, amely reményei szerint az emberi erkölcsöt egy működőképes magra egyszerűsíti: utilitarizmus, sokféleség, autonómia és örökség. TEGMARK, 2. lábjegyzet, 271. o.

⁶⁷ Az e kihívás leküzdésére irányuló erőfeszítésekről lásd pl. Kyle Bogosian, Implementation of Moral Uncertainty in Intelligent Machines, 27 MINDS & MACHINES 591, 595-603 (2017) (többek között William MacAskill, Normative Uncertainty (2014), elérhető a https://s3.amazonaws.com/academia.edu.documents/34857095/Normative_Uncertainty_Complete.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1527105349&Signature=

nélkül
WpMsYO%2FzuWnQLhruSpAY3U884E%3D&response-content-
disposition=inline%3B%20filename%3DNormative_Uncertainty.pdf).

Ez a cikk nem kívánja megoldani azt a bonyolult kérdést, hogy a számítógép képes-e vagy képes lesz-e erkölcsi ítéleteket hozni. De néhány kihívás előzetes vázlatos bemutatása helyénvalónak tűnik. Három alapvető lehetőség áll rendelkezésre: (1) a felülről lefelé irányuló megközelítés,⁶⁸ (2) az alulról felfelé irányuló megközelítés⁶⁹ és (3) a prediktív megközelítés. A felülről lefelé irányuló megközelítés egy általános elv vagy általános elvek összessége megadásával irányítaná a mesterséges intelligencia döntéseit.⁷⁰ Az egyik probléma az, hogy a filozófusok, a hétköznapi polgárok, a választott tisztviselők vagy látszólag bármely más releváns csoport között közel sincs konszenzus arról, hogy melyek a helyes erkölcsi elvek.⁷¹ Sőt, amikor egy adott esetben megpróbálunk választani a releváns erkölcsi elvek közül, valószínűleg szükség lesz olyan ítéletekre, amelyeket "helyi" erkölcsi ítéleteknek nevezhetünk - olyan ítéletekre, amelyek érzékenyek a kontextusra vagy a környezetre. A tisztán absztrakt erkölcsi ítéletek valószínűleg elégtelennek bizonyulnak az emberi magatartás vagy a mesterséges intelligencia irányításához.⁷² A helyi erkölcsi ítéletek szükségessége miatt az erkölcs felülről lefelé történő megközelítése valószínűleg nem lesz sikeres.

A második megközelítés - az alulról felfelé történő megközelítés - hasonló kihívásokkal néz szembe. Az alulról felfelé irányuló megközelítés inkább indukciót, mint dedukciót jelentene - a mesterséges intelligencia a tapasztalatokon keresztül fejleszti ki saját erkölcsi elkötelezettségeit.⁷³ A mesterséges intelligenciát nem úgy programoznák, hogy általános erkölcsi elveket építsen be, hanem inkább felismerné azokat - vagy legalábbis felismerné az erkölcsi cselekedeteket vagy eredményeket - konkrét kontextusokban, talán az emberi erkölcsi fejlődéshez hasonló folyamaton menne keresztül.⁷⁴ Ahhoz, hogy ez a megközelítés működjön, a mesterséges intelligenciának inputot kellene kapnia arról, hogy adott esetben mi a helyes erkölcsi eredmény vagy cselekvés, vagy ki kellene fejlesztenie valamilyen képességet arra, hogy maga hozza meg a megfelelő erkölcsi ítéleteket. Ez a második megközelítés számtalan problémát vetne fel. Az egyik az erkölcsi (vagy érték-) ítéletek és a cselekvés egyéb indokai - beleértve az óvatosságot, a heurisztikákat és az előítéleteket - összefonódása. Még elvileg sem világos, hogy egy számítógép hogyan tudná szétválasztani a különböző okokat, amelyek egy adott eredmény vagy magatartásmód támogatására vagy elutasítására készítetnek. Továbbá, ha egy számítógép kialakítaná saját erkölcsi ítélőképességét, talán szubjektív tapasztalatot vagy valami ahhoz nagyon közel álló dolgot kellene szereznie - tehát

⁶⁸ A felülről lefelé irányuló rendszerrel kapcsolatos néhány lehetőségről - és kihívásról - lásd Wendell Wallach és Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* 83-97 (2009).

⁶⁹ Az alulról felfelé építkező megközelítés hasonló tárgyalását lásd id. 99-115. o. Wallach és Allen a kettő keverékét is tárgyalja. Id. 117-24. o.

⁷⁰ Id. 83.

⁷¹ Lásd pl. Kyle Bogosian, *Implementation of Moral Uncertainty in Intelligent Machines*, 27 *Minds & Machines* (591,2017592) (a felmérés szerint a filozófia szakosok 200926%-a elfogadta vagy hajlott a deontológia, 24%-a a konzekvencializmus és 18%-a az erényetika felé, a fennmaradó 32% pedig más megközelítéseket részesített előnyben).

Természetesen az erkölcsi nézeteltérések is hasonló problémát jelentenek az emberek számára. Részben a demokratikus folyamatokra támaszkodunk, hogy leküzdjük ezt a problémát - vagy a választott képviselőkre hagyatkozunk, hogy törvényt alkossanak, és bírákat neveznek ki, hogy megfelelő ítéleteket hozzanak a törvény értelmezése és alkalmazása során, vagy pedig bírákat választunk, hogy meghozzák ezeket a döntéseket.

⁷² Emiatt sok filozófus támogatta a reflexív egyensúly fogalmát - az elvek vagy szabályok

nélkül
iteratív folyamat, amely során az egyes esetekre vonatkozó ítéleteket az elvek vagy szabályokhoz rendeljük, és fordítva. Lásd például JOHN RAWLS, A THEORY OF JUSTICE (1971). A témával kapcsolatos szakirodalom nagy.⁷³ Wallach and Allen, Moral Machines, supra note , 99-101. o.

⁷⁴ Id.

hogy érezzék a szükséges érzelmeket és empátiát.⁷⁵ De ez a szubjektív vagy kvázi-szubjektív tapasztalat olyan erkölcsi és egyéb értékítéleteket eredményezhet, amelyeket elfogadhatatlannak tartunk - talán egy olyan technocentrikus világgépet, amely nem felel meg a mi antropocentrizmusunknak.⁷⁶

A harmadik megközelítés - az előrejelzés - azt jelenti, hogy a számítógép nem hoz érdemi erkölcsi vagy más értékítéleteket, hanem inkább megjósolja - és így utánozza -, hogy az emberek hogyan hoznák meg ezeket az ítéleteket. Ez a megközelítés Holmes Bad Man elméletére és híres kijelentésére emlékeztet: "A bíróságok tényleges cselekedeteinek jóslatait, és semmi más nagyképűséget nem értek a jog alatt".⁷⁷ Holmes-t parafrázálva azt mondhatnánk, hogy az emberek által ténylegesen meghozandó erkölcsi ítéletek jóslatai, és semmi más nagyképűség, az, amit AI erkölcs alatt ért. Tekintettel a Holmes Rossz ember elméletével való párhuzamra, a jósló megközelítés alávethető annak az erőteljes kritikának, amelyet H. L. A. Hart fogalmazott meg "A jog fogalma" című művében. Ahogy Hart érvelt - többek között - egy bíró valószínűleg nem járna jól, ha megjósolná, hogy ő maga hogyan döntene egy ügyben, mivel ez a vizsgálat reménytelenül körkörösnek tűnik.⁷⁸ De vegyük észre, hogy a mesterséges intelligenciát nem egyszerűen arra kérik, hogy jósolja meg, hogyan fog dönteni egy bíróság. Arra kérik, hogy jósolja meg, hogy a bírák - vagy más jogértelmezők - milyen *erkölcsi* ítéleteket hoznának (a jogértelmezéshez szükséges mértékben). Ez legalább részben olyan választ ad Hartnak, amely Holmes számára nem állt rendelkezésre. Mégis, a fent tárgyalt két fő szempont a harmadik megközelítésre is vonatkozik. Az erkölcsi ítéletek helyi jellege és az a mód, ahogyan az erkölcsi okok elválaszthatatlanul összefonódnak a cselekvés egyéb okaival, komoly akadályokat gördít az elé, hogy a mesterséges intelligencia megjósolja és utánozza az emberi értelmezőktől elvárható erkölcsi ítéleteket.

A prediktív megközelítés feltárása során hasznosnak bizonyulhat a bírák és az ügyvédek nézőpontjának megkülönböztetése. Vegyük először a robotbíró. Mivel az erkölcsi ítéletek helyi jellegűek, a robotbírónak valószínűleg jelentős mennyiségű bírósági ítéletre lenne szüksége - olyan adatpontokra, amelyek a robotbíró előtt álló döntéshez szorosan kapcsolódó döntéseket tükröznek -, hogy utánozhassa az emberi bírót. A hétköznapi emberi életből vagy más intézményi környezetből származó általános erkölcsi elvek valószínűleg nem lennének elegendőek. Ezek az egyéb kontextusok csak korlátozott útmutatást nyújtanának, mivel a köztük és a bírósági kontextus között fennálló kapcsolat gyengébb lenne. Úgy tűnik tehát, hogy a robotbírák nem lennének képesek teljesen kiszorítani az emberi bírót. Ha mégis, akkor az adatok, amelyekre támaszkodnak, egyre kevésbé lennének és

⁷⁵ Az érzelmeknek az erkölcsi érvelésben játszott lehetséges szerepéről szóló nézetekről lásd pl, PAUL BLOOM, AGAINST EMPATHY: THE CASE FOR RATIONAL COMPASSION (2016); JEFFREY GREENE, MORAL TRIBES: EMOTION, REASON, AND THE GAP BETWEEN US AND THEM (2014); CARLA BAGNOLI, ED., MORALITY AND THE EMOTIONS (2012); MARTHA C. NUSSBAUM, UPHEAVALS OF THOUGHT: THE INTELLIGENCE OF EMOTIONS (2003).

⁷⁶ Az igazsághoz tartozik, hogy antropocentrizmusunknak lehet, hogy nincs szilárd erkölcsi alapja. Lásd pl. PETER SINGER, ANIMAL LIBERATION (1975); PETER SINGER, "Equality for Animals?", in PRACTICAL ETHICS (1979).

⁷⁷ Oliver Wendell Holmes, HARV. L. REV. 457,459 (1897).

⁷⁸ H.L.A. Hart, A jog fogalma 124-54 (1961).

Továbbá, a törvény meghatározatlansága miatt egy előrejelzési elmélet a lehetséges kimeneteleket és az egyes kimenetek valószínűségét fogja eredményezni. Nem világos, hogy a

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély
mesterséges intelligenciának hogyan kellene választania ezek közül. Amint azt alább
tárgyaljuk, a legvalószínűbb kimenetel több okból sem feltétlenül a legvonzóbb.

az idő múlásával egyre kevésbé lennének relevánsak az eléjük kerülő ügyekben, és erkölcsi ítéleteik minősége romlana.

Továbbá, egy robotbírónak nehézséget okozhatna a bírói véleményekben rejlő erkölcsi értékítéleteket elválasztani az azokat motiváló egyéb okoktól - óvatosság, heurisztika, elfogultság *stb.* -, amelyek a bírói véleményeket motiválják. Amennyiben a robotbírónak meg kell jósolnia az erkölcsi ítéleteket - és nem csak azt, hogy a bíróságok hogyan fognak dönteni -, fontos, hogy a mesterséges intelligencia csak az erkölcs szerepét emelje ki a bírósági döntéshozatal mintázatából, és figyelmen kívül hagyja az egyéb okok vagy motivációk szerepét. Nem világos azonban, hogy az AI hogyan tudná ezt a feladatot elvégezni. Bár az AI nagyon jó lehet az emberi döntéshozatal mintáinak azonosításában, nem világos, hogyan tudná megkülönböztetni az emberi érvelés erkölcsi alapjait más alapoktól anélkül, hogy független erkölcsi ítéleteket hozna - éppen ezt a feladatot hivatott elkerülni az előrejelzési megközelítés.

A fenti elemzés azt sugallja, hogy a prediktív megközelítés csak a második legjobb közelítést nyújtja a bírói jogértelmezéshez szükséges erkölcsi ítéletekről. A robotbíró irányító adatok tekintetében emberi döntésekre támaszkodik, és az erkölcsi ítéletekre vonatkozó adatok összefonódnak a bírói döntéshozatal befolyásoló egyéb, nem erkölcsi hatásokkal. Mindkét jelenség miatt a robotbírók előrejelzései valószínűleg eltérnének az emberi erkölcsi ítéletektől.⁷⁹

Vegyük most a *robo*lawyers-t. Az emberi ügyvédek mesterséges intelligenciával való helyettesítése önmagában nem jelentené ~~az~~ kockázatot, hogy a robotügyvédek megfosztanák őket a bírósági döntéshozatalban megjelenő minták felismeréséhez szükséges adatoktól. Mindaddig, amíg emberek szolgálnak bíróként, ezek az adatok továbbra is rendelkezésre állnának. De felmerül egy másik probléma. A bírák más intézményi struktúrákban működnek, mint az ügyvédek. A bíró által levont jogi következtetés nem feltétlenül ugyanaz, mint amit egy ügyvédnek kellene levonnia. Egy bíró például bizonyítékok hiányában elutasíthat egy büntetőügyet. De ahogy Holmes rossz ember elméletének kritikusai megjegyzik, helytelennek tűnik azt mondani, hogy legális a gyilkosság elkövetése, amíg valaki nem hagy hátra elegendő bizonyítékot az elítéléshez. És etikátlan lenne, ha egy ügyvéd úgy segítene az ügyfélnek a gyilkosság elkövetésében, hogy nem hagy bizonyítékokat, és megússza a dolgot.⁸⁰ Tehát a bírósági eljárások kimenetele nem

⁷⁹ Elismerem azt a kényes problémát, hogy az emberek jobb erkölcsi ítéleteket hoznának-e, mint a mesterséges intelligencia. Az erkölcsi ítéletekkel kapcsolatos nézeteltérések miatt még azt is el lehetne képzelni, hogy a mesterséges intelligencia véletlenszerű eltérései az emberi ítéletekhez képest inkább jobban megközelítik a helyes erkölcsi ítéleteket - ha vannak helyes erkölcsi ítéletek -, mint kevésbé. Másrészt egy ilyen szkeptikus nézet arra ösztönözne minket, hogy teljesen lemondjunk az erkölcsi ítéletek meghozatalának kísérletéről.

⁸⁰ Lásd pl. az 1.2(d) Model Rule of Professional Conduct ("Az ügyvéd nem adhat tanácsot az ügyfélnek, hogy olyan magatartást tanúsítson, vagy segítsen az ügyfélnek, amelyről az ügyvéd tudja, hogy bűncselekmény vagy

..... család s.").

Bryan Casey úgy tűnik, hogy az erkölcs és a mesterséges intelligencia kérdésének helyes megközelítése a Holmes-féle Rossz ember nézőpontjának elfogadása. Ahhoz, hogy ne maradjunk "reménytelenül elmerülve", azt sugallja, hogy fel kell hagynunk azzal, hogy erkölcsi ítéleteket próbáljunk hozni arról, hogy mit kellene tennie az AI-nak. Nem világos, hogy állítása pozitív

nélkül.
vagy normatív. Ha csupán azt jósolja meg, hogyan fog fejlődni a mesterséges intelligencia, akkor valószínűleg igaza van abban, hogy a pénzügyi ösztönzőknek - beleértve a törvény által létrehozottakat is - nagy szerepük lesz. Ha azonban az álláspontja előíró jellegű - ha az az érve, hogy az AI vállalatoknak csak arra *kellene* törekedniük, hogy minimalizálják jogi felelősségüket -, akkor az álláspontját számos erőteljes kritika éri, amelyeket már

nem feltétlenül adnak közvetlen útmutatást az ügyvédek számára, például arról, hogy milyen tanácsot adhatnak ügyfeleiknek, vagy milyen magatartást segíthetnek etikailag. Ez problémát okoz az erkölcsi ítéletek helyi jellegének fényében. A bírói kontextus túlságosan távol állhat az ügyvédi kontextustól ahhoz, hogy az egyikből a másikba átvihető erkölcsi ítéletekre vonatkozó következtetéseket lehessen levonni.

Eközben ugyanazok a nehézségek, amelyek a mesterséges intelligenciát az erkölcsi ítéletek és más ítéletek szétválasztása során a robotbíróként való fellépés során sújtják, a robotügyvédre is vonatkoznak.⁸¹

Mindenesetre a számítógépek erkölcsi ítéletalkotási képessége jóval elmaradt a viselkedésminták leírására és előrejelzésére irányuló technikai képességeitől. A mesterséges intelligencia legyőzte a legnagyobb emberi zsenit a Go játékban, de nem tanult meg olyan meggyőző érveket felhozni, amelyek meggyőzhetnének egy morálfilozófust. Vegyük ezzel kapcsolatban Tegmark LIFE 3.0 című könyvének "Célok" című fejezetét.⁸² Éleslátó és elgondolkodtató elemzést nyújt az emberi célok és a mesterséges intelligencia összehangolásának különböző kihívásairól, és ezt a feladatot három összetevőre bontja: *megtanítani* a mesterséges intelligenciát a céljainkra; rávenni a mesterséges intelligenciát, hogy *elfogadja* a céljainkat; és biztosítani, hogy a mesterséges intelligencia *megtartsa* a céljainkat.⁸³ Ebben a vitában olyan módszereket javasol, amelyekkel az AI magától is tanulhat. Megjegyzi például, hogy egy időseket gondozó robot képes lehet arra, hogy egy nyugdíjas ember tevékenységeiből kikövetkeztesse, hogy mit értékel.⁸⁴ Az etikát illetően azonban Tegmark nem javasolja, hogy az AI öntanulásnak szerepe legyen. Ezzel kapcsolatban megjegyzi, hogy a viselkedés pusztá megfigyelése csak korlátozottan lenne hasznos. Legtöbbünk nem mindig cselekszik erkölcsösen. A nyugdíjas ember lehet képmutató, hazug, csaló és tolvaj. Ha ez így van, a robot feltehetően megtanulná, hogy segítsen a férfinak a képmutatásban, a hazugságban, a csalásban és a lopásban. Ez még akkor is így lehet, ha az ember elismerné, hogy ezek a tevékenységek erkölcsileg helytelenek.

Hogyan azonosítaná tehát Tegmark, hogy milyen etikai célokat kellene követnie a mesterséges intelligenciának? A válasza, legalábbis implicit módon, az, hogy az emberekre támaszkodik. Mint kifejti, "az etikai kérdésekkel való foglalkozást a célokhoz igazodó szuperintelligencia megalkotása utánra halasztani felelőtlenség lenne, és potenciálisan

Holmes Bad Man-elmélete és az általa feltételezett jogi pozitivizmus durva formája ellen. Ezek közül kiemelkedik annak a szövegben tárgyalt elképzelésnek az elutasítása, hogy az a személy, aki megússza a gyilkosságot, még mindig megsértette a törvényt (és az erkölcsöt), helyesen értelmezve.

⁸¹ A szövegben szereplő elemzés arra összpontosít, hogy a mesterséges intelligencia képes-e pontos erkölcsi ítéleteket hozni. Van egy másik lehetséges ok is, amiért a számítógépek - legalábbis bizonyos körülmények között - nem lehetnek képesek az emberi lények helyett bíróként vagy ügyvédként szolgálni. Ez az alternatív ok nem arra összpontosítana, hogy a jogértelmező *milyen* következtetésekre jut, hanem arra, hogy *ki* a jogértelmező. Ez a megkülönböztetés nagyjából megegyezik azzal, amelyet Paul Kahn az ész és az akarat között tett. Lásd Paul Kahn, Reason and Will in the Origins of American Constitutionalism, 98 YALE L.J. 449 (1989). Lehet, hogy nem engednénk meg, hogy mesterséges intelligenciák bíróként szolgáljanak, és pedig ugyanabból az okból, amiért nem engednénk meg, hogy szavazzanak - nem (csak) az általuk hozott döntések minősége miatt, hanem azért, mert nem olyan lények, akiknek ilyen szerepet kellene kapniuk egy demokratikus társadalomban. Lásd JERRY KAPLAN,

nélkül
MŰVES INTELLIGENCIA: WHAT EVERYONE NEEDS TO KNOW 98-101 (2016) (egy gondolatmenetet tárgyalva.

kísérlet, amelyben egy polgár fokozatosan egyre nagyobb hatalmat ruház át a mesterséges intelligenciára, hogy az szavazzon a nevében). Hálás vagyok Brad Wendelnek egy vitáért, amely ezt a pontot felvetette, és amelyet egy külön tanulmányban közösen vizsgálunk meg. Ennek a pontnak a kidolgozása meghaladja e cikk kereteit.

⁸² Id. at249 - 280.

⁸³ Id. 260.

⁸⁴ Id. 261.

katasztrofális."⁸⁵ De nem javasolja, hogy a *mesterséges intelligencia* hogyan azonosíthatná helyettünk az erkölcs tartalmát. Inkább négy alapelvet javasol, amelyeket az évek során olvasott műveiből desztillál: utilitarizmus; sokféleség; autonómia; és örökség.⁸⁶ Eltekintve ezeknek az elveknek az érdemeitől - mindegyik elv külön-külön, és a kombináció minden bizonnyal ellentmondásos lenne az erkölcsfilozófusok, a kormányzati tisztviselők és a hétköznapi polgárok körében -⁸⁷, a jelen célokból az a figyelemre méltó, hogy egy, az AI-ról és a gépi tanulásról szóló könyvben Tegmark nem javasolja, hogy az általa felvetett alapvető etikai kérdések megoldását bízzuk az AI-ra. Úgy tűnik, ezt nekünk magunknak kell megtennünk.⁸⁸

A mesterséges intelligencia jelenlegi és talán jövőbeli képtelensége arra, hogy a végső célok meghatározásához szükséges érdemi erkölcsi ítéleteket meghozza, hatással van az eddig tárgyalt különböző kérdésekre. Gondoljunk csak a munkahelyi nemi diszkriminációra. Mint már említettük, egy számítógépes program, amelynek feladata a munkahelyi siker előrejelzése, megerősítheti és állandósíthatja a szexizmust. De a hiba nem feltétlenül a számítógépben van - vagy legalábbis nem csak a számítógépben. Inkább az emberek legalább részben vétkesek - feltéve, hogy van vétkekesség -, mivel szennyezett adatokra támaszkodnak, és a számítógép számára a célokat nem kellő teljességgel határozzák meg. A mesterséges intelligencia jelenleg nem képes

⁸⁵ Id. 269.

⁸⁶ Id. az Ő271. összefoglalása az elvekről:

- Utilitarizmus: A pozitív tudatos élményeket maximalizálni kell, a szenvedést pedig minimalizálni.
- Sokszínűség: a pozitív élmények sokszínűsége jobb, mint ugyanazon élmény sokszoros ismétlése, még akkor is, ha ez utóbbit a lehető legpozitívabb élményként azonosították.
- Autonómia: Az öntudatos entitásoknak/társadalmaknak rendelkezniük kell a szabadsággal, hogy saját céljaikat kövessék, kivéve, ha ez ellentétben áll valamely kényszerítő elvvel.
- Hagyaték: Kompatibilitás olyan forgatókönyvekkel, amelyeket a legtöbb ember *ma* boldognak tartana, inkompatibilitás olyan forgatókönyvekkel, amelyeket lényegében minden ember *ma* szörnyűnek tartana.

Id.

⁸⁷ A nyilvánvaló vitás pontok közé tartozik Tegmark döntése, hogy minden tudatos tapasztalatot értékel, beleértve a mesterséges intelligencia, és nem csak a szerves életformák tapasztalatait, és az, hogy nem jelzi, hogyan kellene a pozitív tudatos tapasztalatokat egymáshoz képest mérlegelni, ami többek között a tapasztalatok fajtáitól (minden öröm egyformán értékes, a perverz, a profán és a tiszta?) és az azokat megtapasztaló entításoktól függően változhat (egy ember öröme vagy fájdalma egyenértékű egy halállal vagy egy számítógépes programéval, feltételezve, hogy a mesterséges intelligencia szubjektív tapasztalatot fejleszt ki?). Tegmark négy alapelve úgy tűnik, hogy egyszerre testesíti meg az utilitarista és a deontológiai elkötelezettséget, ami kritikát vált ki az utilitaristák és a deontológusok részéről egyaránt. Az e morálfilozófiák mindegyikét feltáró, támogató és kritizáló szakirodalom hatalmas, és számtalan kérdést és aggályt vet fel Tegmark megközelítésével kapcsolatban. E kérdések és aggályok elemzése meghaladja e tanulmány kereteit.

⁸⁸ Másképpen úgy is fogalmazhatnánk, hogy magunkra vagyunk utalva, de ez a körülmények között zavaros metaforának tűnik.

nélkül:
Tegyük elismeri annak lehetőségét, hogy a mesterséges intelligencia tudatosságot fog kifejleszteni, és az emberi lényekhez hasonló jogok és felelőségek teljes körét el fogja sajátítani. Id. 281-315276,.

önmagában meghatározza, hogy az egyik végső cél a hátrányos megkülönböztetés
elkerülése - és annak korrekciója -.

Ugyanez igaz azokra is, akik a vezető nélküli autókat irányító kritériumokat biztosítják.
Ha egy autót arra kérünk, hogy minimalizálja a jogi felelősséget, az megkérdőjelezhető
értékitételeket ágyaz be a fejlődő algoritmusába. A kiszámítható eredmény az lehet, hogy a
társadalomban a diszkrimináció mintái állandósulnak. Ami különösen trükkös, hogy az
ártalmatlannak tűnő célok - amelyek első látásra nem tűnnek vitathatónak - a gyakorlatban
megkérdőjelezhetőnek vagy akár elfogadhatatlannak is bizonyulhatnak.

Már megkérdőjeleztük, hogy bölcs dolog-e arra kérni egy önvezető autót, hogy
minimalizálja a jogi felelősséget. Gondoljunk arra, hogy mi lehet a vonzóbb alternatíva: a
várható életveszteségek minimalizálása. Egy ilyen megközelítés minden életet egyformán
kezelne, ellentétben a jelenlegi kártérítési rendszerünk által létrehozott ösztönzőkkel. De vajon
tényleg ezt akarjuk-e elérni? Képzeld el, hogy egy önvezető autó észleli, hogy a baleset
elkerülhetetlen, mert egy részeg sofőr a szembejövő forgalomba kanyarodott. Az autó két
lehetőséggel szembesül:

(1) *szinte biztosan* megöli az ittas vezetőt, és *valószínűleg* megmenti egy ártatlan gyermek életét,
aki átkel az úton; vagy (2) *szinte biztosan* megmenti az ittas vezető életét, és *szinte biztosan*
megöli az ártatlan gyermeket. Ha az autó minimalizálja a várható életveszteséget, akkor
feltehetően megkíméli az ittas vezetőt és megöli az ártatlan gyermeket. Végére is, ez a stratégia
átlagosan a legtöbb életet megóvná.⁸⁹ Az ittas vezető bűnösségét - és a gyermek ártatlanságát -
figyelembe véve azonban ez az eredmény nem biztos, hogy helyes.

Hasonló elemzés alkalmazható a jogértelmezésben részt vevő számítógépre is. Amint
fentebb említettük, ha a bírósági döntéshozatalban elfogultsági mintázatot találunk, a
számítógép ezt a mintázatot beágyazhatja - és várhatóan be is ágyazza - az általa készített
előrejelzésekbe. Gondoljunk arra, hogyan kellene egy számítógépnek kezelnie azt az átható
meghatározatlanságot, amely a legtöbb modern jogtudós szerint létezik a jogban.⁹⁰ A pusztá
leírás vagy előrejelzés - a meghatározatlanság körülményei között - nem fog egyetlen
értelmezést eredményezni. Számos értelmezési lehetőséget fog eredményezni, amelyek közül a
jogértelmező talán különböző eséllyel választja ki őket. Hogyan válasszon közülük a
számítógép?

⁸⁹ Hogy ezt a példát egy kicsit konkrétabbá tegyük - ha téves pontossággal is -, tegyük fel, hogy
az (1) esetben az ittas vezető túlélési esélye 5%, a gyermek túlélési esélye pedig 55%, a (2)
esetben pedig az ittas vezető túlélési esélye 95%, a gyermek túlélési esélye pedig 5%. Az (1)
választás átlagosan 0,6 életet mentene meg, a (2) választás pedig átlagosan 1,0 életet. A várható
életveszteség minimalizálása szempontjából - minden életet egyformán kezelve - a (2) jobb, mint
az (1).

Ez a feltételezés bonyolulttá válhat, ha figyelembe vesszük a várható élettartamot, és a
célú értelmességet, hogy az emberi életből várhatóan hátralévő éveket maximalizáljuk. De ez
nem változtat a lényegen: az ittas és vétkes sofőr életét ugyanolyan értékkel kezeljük, mint egy
ártatlan ember életét.

⁹⁰ Vitathatóan ebben az értelemben vagyunk "most már mindannyian realisták". Lásd Michael S.
Green, Legal Realism as Theory of Law, 46 WM. & MARY L. REV. 1915, 1917 (2000) ("[I]t
gyakran mondják - sőt, olyan gyakran mondják, hogy már közhelyesnek számít, hogy

'közhelynek' nélkül nevezük -, hogy most mindannyian realisták vagyunk"). A legtöbb tudós inkább a jogbizonytalanság mértékéről, mintsem annak létezéséről nem ért egyet.

Az egyik lehetőség az lehet, hogy a számítógép a legvalószínűbb vagy legáltalánosabb jogi értelmezést választja. Ennek van arculati vonzereje. De az, hogy ez a legjobb megközelítés-e, erkölcsi megítélést igényel. Nem nyilvánvaló, hogy a legnépszerűbb választás lesz a "helyes" - attól függően, hogy mi határozzuk meg a "helyes" fogalmát -, különösen, ha aggódhatunk amiatt, hogy a népszerűség gyanús okokból eredhet. A legnépszerűbb eredmény kiválasztása absztrakt módon megfogalmazva kifogástalannak tűnhet, a gyakorlatban azonban kifogásolható eredményekhez vezethet. Lehet, hogy a bírák vagy más jogértelmezők hajlamosak szisztematikus hibákat elkövetni, gyakori félreértésekben szenvednek, kiszámítható kognitív elfogultságoknak esnek áldozatul, vagy akár gyakori, nem kívánatos, tudattalan előítéletek alapján cselekszenek. A jogról alkotott népszerű nézet - vagy az a nézet, amelyik a legnagyobb valószínűséggel népszerű lesz - fontos értelemben téves lehet, még hozzá kiszámíthatóan. Az egyik ok, amiért elszigeteljük szövetségi bíráinkat a közvetlen választási elszámoltathatóságtól, azon a nézeten alapulhat, hogy ami népszerű, az nem feltétlenül helyes, ha a jogról van szó.⁹¹

A konkurens megközelítés úgy nézhetne ki, mint a Ronald Dworkin által híressé vált megközelítés. Azt állította, hogy a jogértelmezés két ítéletet foglal magában, az egyiket "illeszkedésnek", a másikat pedig "igazolásnak" nevezte.⁹² Dworkin szerint az értelmezés magában foglalja annak értékelését, hogy egy valószínűsíthető eredmény mennyire "illeszkedik" a törvényhez - amit úgy is értelmezhetünk, hogy az összeegyeztethető egy számítógép előrejelzésével, hogy a bírák nagy valószínűséggel hogyan értelmeznék azt. Az értelmezés azonban annak értékelését is magában foglalja, hogy egy eredmény mennyire "igazolja" a törvényt - vagyis, hogy erkölcsileg mennyire teszi vonzóvá a törvényt.

Dworkin megközelítése szerint a jogértelmezőknek egyensúlyt kellene teremteniük az illeszkedés és az indokoltság között.⁹³ Ezt a feladatot Dworkin a modellbíró, Herkules elé állította.⁹⁴ De lehet, hogy erre a Herkules nevű számítógépes program nem képes. Ha igazunk van abban, hogy a mesterséges intelligencia nem képes érdemi erkölcsi ítéleteket hozni, akkor Herkules, mint mesterséges intelligencia, nem tudja elvégezni az összes olyan feladatot, amelyet a versengő jogértelmezések közötti választáshoz kell elvégeznie, legalábbis Dworkin szerint. Végző soron a Hercules számára a lényegi *erkölcsi* ítéletek szükségesek ahhoz, hogy elvégezze a munkáját. Az erkölcsi ítéletek pedig egyedülállóan *emberi* ítéletek lehetnek.

Elképzelhető, hogy a nem is olyan távoli jövőben a számítógépek gyorsabban és pontosabban tudják majd felmérni az "alkalmasságot", mint az emberek, ahogyan sakkban vagy Go-ban is meg tudnak verni minket. De megvitattuk azokat az okokat, amelyek miatt kételkedhetünk abban, hogy a számítógépek képesek lesznek pontosabban megítélni az "alkalmasságot", mint az emberek, sőt, abban is kételkedhetünk, hogy képesek lesznek megítélni az "alkalmasságot".

⁹¹ Ez az álláspont nem feltétlenül olyan elitista, mint amilyennek elsőre tűnik. A bírák talán nem jobbak az erkölcsi ítéletalkotásban, mint az átlagpolgárok. De érzékenyek lehetnek olyan intézményi aggályokra, amelyek jogi képzés és tapasztalat nélkül mások számára nem lennének nyilvánvalóak. Továbbá a bírák az adott helyzetekben a kérdéses tágabb elvekre összpontosíthatnak, nem pedig az adott ügy körülményeire, és olyan normákat alkalmazhatnak, amelyek széles körű és tartós támogatást élveznek - és élveznek is - a lakosság széles körében. Lásd Ronald Dworkin, *The Forum of Principle*, in *A MATTER OF PRINCIPLE* (1986); Christopher Eisgruber, *Constitutional Self-Government* (2007).

⁹² Dworkin megközelítésének rövid összefoglalását lásd Davis, *Legality, Morality, Duality*, supra note.

, 94-95. o.,^{nélkül} Joshua P. Davis, Cardozo's Judicial Craft and What Cases Come to Mean,
68 Y.U. L. REV. 777, 809-10 (1993)).

⁹³ Id.

⁹⁴ Id.

egyáltalán.⁹⁵ A fenti érvelés két tételhez vezet: (1) az érdemi erkölcsi ítéletek néha szükségesek lehetnek ahhoz, hogy megmondjuk, mi a törvény; és (2) egyedül az ember lehet képes arra, hogy ezeket az érdemi erkölcsi ítéleteket meghozza. Ezek a tételek együttesen alakíthatják a mesterséges intelligenciának a jogértelmezésben betöltött szerepét. Ha meg tudjuk határozni, hogy a jogértelmezés mikor igényel érdemi erkölcsi ítéletet, akkor tudni fogjuk, hogy az embernek mikor van különleges szerepe a jogrendszerünkben. Van egy érdekes lehetséges csavar is. Ha igazunk van a mesterséges intelligenciában rejlő rendkívüli potenciállal kapcsolatban, akkor egy harmadik pont is igaz lehet: (3) A számítógépek képesek lehetnek arra, hogy esélyeket rendeljenek a különböző erkölcsi ítéletekhez, amelyeket az emberi lények hozhatnak, még akkor is, ha nem képesek érdemi erkölcsi ítéleteket hozni. Ez a három pont elvezet bennünket a mesterséges intelligencia és a jogtudomány metszéspontjához, amelyhez most fordulunk.

III. Jogtudomány: Erkölcs és jogértelmezés.

A számítógépekről és a jogértelmezésről szóló fenti vita kétféleképpen motiválhatja a jogtudomány vizsgálatát. Először is, további aktualitást és sürgősséget ad a jogtudománynak. Új okunk van arra, hogy megoldjuk azt a régóta tartó vitát, amely arról szól, hogy az erkölcs milyen szerepet játszik a jogértelmezésben. Ez segíthet megérteni és meghatározni azokat a funkciókat, amelyeket a mesterséges intelligencia a jogértelmezésben betölthet - és nem tölthet be -.

A mesterséges intelligenciáról való elmélkedés egy másik módon is motiválhatja a jogtudományi kutatást. Lehet, hogy erős intuíciónk van arra, hogy a számítógépek jogértelmezésben betöltött szerepét korlátozni kell - hogy helyet kell fenntartanunk az embereknek. Az erkölcsnek a jogértelmezés szempontjából való jelentőségére való odafigyelés segíthet megérteni és igazolni ezt az intuíciót. Érezzük, hogy bizonyos körülmények között azt szeretnénk, ha a jogértelmezők érdemi erkölcsi ítéletet hoznának - és úgy gondoljuk, hogy erre a számítógépek nem képesek. A számítógépeknek a jogértelmezésben játszott megfelelő szerepéről alkotott nézeteink tehát befolyásolhatják az erkölcsnek a jogértelmezésben betöltött szerepéről alkotott nézeteinket.

Ezek a pontok egy újszerű jogtudományi álláspont feltárásához nyújtanak kontextust. Különböző publikációkban - amelyek egy részét Manuel Vargas filozófussal közösen írtam - elkezdtem kifejleszteni egy érvet, amelyet "jogi dualizmusnak" nevezünk.⁹⁶ Eszerint a jog természetének legjobb leírása a jogértelmezés céljától függően változik.⁹⁷ Amikor a jogértelmező csupán a jog leírására törekszik, vagy arra, hogy megjósolja, hogyan fogják mások értelmezni azt, a jogi dualizmus azt sugallja, hogy nincs szükség erkölcsi ítéletekre annak kimondásához, hogy mi a jog.⁹⁸ Más szóval, a jogi pozitívizmus nyújtja a legjobb magyarázatot a jog természetéről ilyen körülmények között.⁹⁹ Máskor azonban a jogértelmezők erkölcsi útmutatást keresnek a jogban. Ilyenkor a jogi dualizmus szerint a jogértelmezőknek gyakran - talán mindig - érdemi erkölcsi ítéleteket kell hozniuk.¹⁰⁰ A természetjog tehát a legjobb magyarázatot adja a jog természetéről ilyen körülmények között.¹⁰¹

⁹⁵ A számítógépek azonban képesek lehetnek megjósolni, hogy az emberek hogyan értékelnék az indokoltságot. Lásd:

Az alábbi III.D. pont tárgyalja ennek a lehetőségnek néhány lehetséges következményét.

⁹⁶ Lásd a fenti megjegyzést.

⁹⁷ Id.

⁹⁸ Id.

⁹⁹ Id.

¹⁰⁰ Id.

¹⁰¹ Id.

A jogi dualizmus megoldást kínál a jogtudomány elsődleges vitájára. Egyúttal módot nyújt a mesterséges intelligencia jogértelmezésben rejlő lehetőségeinek és korlátainak feltérképezésére. Ha a számítógépek - ahogyan azt fentebb javasoltuk - hamarosan felülmúlják az embereket a jog leírásában és a jog értelmezésének megjósolásában, akkor kiszoríthatják az e feladatok elvégzésével megbízott jogértelmezőket.

Másrészt, ha a számítógépek nem tudnak érdemi erkölcsi ítéleteket hozni - ha meg kell mondani nekik, hogy milyen végső célokat kell követniük, és nem tudják ezeket a célokat saját maguk kiválasztani -, akkor a jogi dualizmus azt javasolja, hogy az embereknek továbbra is meg kell tartaniuk a felelősséget a jogértelmezésért.

Természetesen ez az érvelés csak akkor számít, ha a jogértelmezőknek időnként kötelességük a jogra mint erkölcsi útmutatás forrására tekinteni. Van okunk azt hinni, hogy ez így van. Egyes jogtudósok úgy vélik, hogy a jognak általában erkölcsi ereje van. Még azok számára is, akik kételkednek ebben az általános állításban, lehetnek olyan körülmények, amelyek szerintük kötelezik a jogértelmezőket a jog követésére.¹⁰² Az egyik valószínű jelölt a bírák lennének, amikor jogkérdésekben döntenek. Az előttük fekvő ügy eldöntésekor a lehetséges jogértelmezések közül kell választaniuk.¹⁰³ Egy másik jelölt bizonyos körülmények között az ügyvédek lennének - beleértve azt is, amikor tanácsot adnak a joggal kapcsolatban azoknak, akiket azzal bíztak meg, hogy fejlesszék ki az önvezető autók szoftverét, hogy élet-halál kérdésekben döntsenek.¹⁰⁴

Egy további jogtudományi kérdést is érdemes megemlíteni. Ha úgy gondoljuk, hogy a mesterséges intelligencia végül képes lesz felülmúlni az embereket minden leírási vagy előrejelzési feladatban, akkor ez a fölény az erkölcs bizonyos kérdéseire is kiterjedhet. Különösen, a mesterséges intelligencia talán jobban le tudja írni az emberek által vallott - vagy vallott, vagy cselekedett (ami nem feltétlenül ugyanaz) - erkölcsi meggyőződések halmazát, legalábbis bizonyos körülmények között.¹⁰⁵ Ez azonban nem jelenti azt, hogy a számítógépek képesek lennének érdemi erkölcsi ítéletet hozni. Ez a különbségtétel a jogtudományi kutatás egy lehetséges területét jelzi, amely valójában már régen felmerült, amikor Lon Fuller és H.L.A. Hart híres vitát folytatott az erkölcs szerepéről a jogértelmezésben.¹⁰⁶ Más szavakkal, különbséget tehetünk egyrészt az *erkölcsre vonatkozó* leíró és előrejelző állítások, másrészt az *érdemi* erkölcsi ítéletek között. Bizonyos célokra - talán a mesterséges intelligencia megfelelő hatókörének meghatározásával összefüggésben - megkülönböztethetnénk a jog természetéről szóló elméleteket, amelyek a jogértelmezésben lehetővé teszik az erkölcsről szóló leíró vagy előrejelző ítéleteket, azoktól, amelyek lehetővé teszik az érdemi erkölcsi ítéleteket. Ez a megközelítés kiegészíthetné azokat az egyéb módszereket, amelyekkel a tudósok megkülönböztetik a jogi pozitívizmust a természetjogtól.

A III. rész ezeket a pontokat vizsgálja. A III.A. rész keretbe foglalja a jogtudomány központi történeti vitáját, azaz a természetjog és a jogi pozitívizmus közötti vitát. Ez a központi vita a

¹⁰² Ennek a lehetőségnek a megvitatásához lásd Joshua P. Davis, *Legal Dualism, Legal Ethics, and Fidelity to Law*, 2016 J. OF PROF. LAWY. 1, 21-26.

¹⁰³ Id. 22. pont.

¹⁰⁴ Id. 23-25.

¹⁰⁵ De lásd fentebb, hogy milyen nehézségekbe ütközhet a mesterséges intelligencia, ha ilyen jellegű előrejelzéseket tesz a jogértelmezésben.

¹⁰⁶ H.L.A. Hart, Positivism and the Separation of Law and Morals, HARV71. L. REV. 593
(1958); Lon L. Fuller, Positivism and Fidelity to Law-A Reply to Professor Hart, HARV71. L.
REV.
630 (1958).

az erkölcs szerepe a jogértelmezésben - egy olyan vita, amely hatással van a mesterséges intelligencia által betölthető funkciók lehetséges külső hatáira.

A III.B. rész elmagyarázza, hogy a jogi dualizmus hogyan kínál lehetséges megoldást erre a vitára, amelyben a természetjog és a jogi pozitívizmus egy-egy megfelelően meghatározott keretek között egyaránt megfelelő magyarázatot ad a jog természetéről. A jogi dualizmus azt is megmagyarázhatja, hogy miért játszik az ember alapvető szerepet a jogértelmezésben, amikor a jog az erkölcsi útmutatás forrásaként szolgál. Ezáltal módot adhat a mesterséges intelligencia jogértelmezésben betöltött szerepének korlátozására, beleértve a vezető nélküli autók irányítását is.

A III.C. rész ezután néhány olyan körülményt javasol, amelyekben a jognak az erkölcsi útmutatás forrásaként kell szolgálnia. Az egyik valószínű példa az, amikor a bírák (vagy esküdtek) jogi vitákat döntenek el. Más példák valószínűleg akkor merülnek fel, amikor egy önvezető autó eldönti, hogy mely életet kímélje meg és melyeket áldozza fel, és amikor egy vállalat eldönti, hogy egy autonóm autót hogyan kell programozni, hogy élet-halál kérdésekben döntsön.

A III.D. rész azt javasolja, hogy a technológiai fejlődésre való reflektálás miként javíthatja a jogtudomány megértését. Megjegyzi, hogy a mesterséges intelligenciával kapcsolatos aggodalmaink - és azok lehetséges szerepe a jogértelmezésben - segíthetnek igazolni a jogi dualizmust. Ha a jogi dualizmus képes megmagyarázni, hogy miért akarjuk korlátozni a mesterséges intelligencia szerepét a jogrendszerünkben - és mindezt úgy, hogy az intuíciónknak is legyen értelme -, akkor ez okot ad a jogi dualizmus támogatására. A III.D. rész párhuzamot von a jog számítógépes értelmezésének és az erkölcs számítógépes értékelésének lehetőségei között: a számítógépek képesek lehetnek az erkölcs leírására és az emberek erkölcsi ítéleteinek előrejelzésére, de nem képesek érdemi erkölcsi ítéleteket hozni. Ez a jogtudományi gondolkodás különböző iskolájának meghatározására egy érdekes alternatív módot javasol, amely a mesterséges intelligencia jogértelmezésben betöltött szerepének mérlegeléséből ered. Bizonyos célokra a jogi pozitívizmus és a természetjog megkülönböztetésének legjobb módja talán nem az erkölcsnek a jogértelmezésben betöltött szerepe, hanem inkább az *érdemi* erkölcsi ítéleteknek a jogértelmezésben betöltött szerepe lehet. Ily módon nemcsak azt láthatjuk, hogy a jogtudomány hogyan tájékoztathatja a jogértelmezés élvonalbeli kérdéseit, hanem fordítva is. A jogértelmezéssel foglalkozó AI-ról szóló gondolat kísérlet - vagy ha úgy tetszik, gondolat kísérlet - új megvilágításba helyezheti a régóta folyó jogtudományi vitákat.

A. Jogtudomány: A jog természetének meghatározása az erkölcs szerepe alapján.

Ha a Herkulesről szóló fenti fejtegetés helytálló, akkor a mesterséges intelligencia szerepét a jogértelmezésben korlátozhatja az, hogy a törvény megmondása során érdemi erkölcsi ítéleteket kell hoznia. Ez szépen illeszkedne a jogtudomány legalább az elmúlt fél évszázad központi kérdéséhez: az erkölcs szerepéhez a jogértelmezésben.¹⁰⁷ A jogtudomány főbb irányzatait az határozza meg, hogy szerintük milyen szerepet játszik az erkölcs a jogban. A központi szakadék a jogi pozitívizmus és a természetjog között húzódik.¹⁰⁸

¹⁰⁷ Lásd a fenti megjegyzést.

¹⁰⁸ Id.

A jogi pozitivizmus különbséget tesz a jog és az erkölcs között.¹⁰⁹ A jogi pozitivizmusnak különböző ágai vannak. A kizárólagos (vagy kemény) jogpozitivisták azt állítják, hogy az erkölcs nem játszik szerepet abban, hogy megmondjuk, mi a jog.¹¹⁰ Az biztos, hogy a kizárólagos jogpozitivisták elismerik, hogy az erkölcsnek lehet és kell is szerepelnie a jog megalkotásában - például amikor a jogalkotó törvényt hoz, vagy amikor egy bíró új törvényt alkot. De ez szerintük külön kérdés attól, hogy egy adott időpontban mi a jog.¹¹¹

Az inkluzív (vagy puha) jogi pozitivizmus ezzel szemben elfogadja az erkölcs lehetséges - és esetleges - szerepét a jogban.¹¹² Az inkluzív jogi pozitivizmus egyik tipikus meghatározási módja a társadalmi tények tézise. Eszerint a jog tartalma végső soron csak a társadalmi tényektől függ.¹¹³ Ezek a végső társadalmi tények azonban az erkölcsöt relevánssá tehetik az egyes jogi ítéletek szempontjából, így az erkölcs beépülhet a jogba.¹¹⁴

A harmadik megközelítést normatív (vagy etikai) jogi pozitivizmusnak nevezik.¹¹⁵ A pozitivizmusnak ez a változata bizonyos értelemben az inkluzív jogi pozitivizmus fordítottja. A normatív pozitivizmus lehetővé teszi, hogy az erkölcs szerepet játsszon a joggal kapcsolatos végső (vagy alapvető) ítéletekben, de csak akkor, ha ezek az erkölcsi ítéletek arra a következtetésre vezetnek, hogy a jogértelmezőknek nem szabad további erkölcsi ítéleteket hozniuk a jog megítélése során.¹¹⁶ Antonin Scalia bíró vitathatóan ehhez a gondolkodási irányzathoz tartozott.¹¹⁷

A nem pozitivisták - akiket néha természetes jogászoknak is neveznek - hajlamosak ellenállni a jog és az erkölcs közötti különbségtételnek. Úgy vélik, hogy az erkölcsi ítéletek időnként szükségesek a jogi ítéletek meghozatalához. Ronald Dworkin egy "interpretivista" nézetet képviselt, amely a jogértelmezésben az érdemi erkölcsi ítéleteket a leíró ítéletekkel integrálta.¹¹⁸ Lon Fuller hasonló nézetet vallott a

¹⁰⁹ SHAPIRO, LEGALITY, supra note , at 273.

¹¹⁰ Id. 269. o.

¹¹¹ Id.

¹¹² Id. 270-71.

¹¹³ Id. 273.

¹¹⁴ Id. 270-71.

¹¹⁵ Jeremy Waldron, Normatív (vagy etikai) pozitivizmus, in: HART POSZTKRIPT: ESSAYS ON THE UTÓSZÓ A JOG FOGALMÁHOZ (4112001).

¹¹⁶ Joshua P. Davis, Legal Dualism, Legal Ethics, and Fidelity to Law, 2016 J. OF PROF. LAWY. 1, 8.

¹¹⁷ Id. 12-14. o.

A nem-pozitivisták (vagy természetjogászok) is hasonló nézetkülönbségekkel rendelkeznek, bár ezek általában nem ilyen egyértelműen meghatározottak. Lon Fuller például viszonylag hűvösen ölelte magához az erkölcsöt. Ő inkább a "jog belső erkölcsiségére", vagyis a jogi szférára jellemző erkölcsi értékekre összpontosított. Lásd Lon L. Fuller, Positivism and Fidelity to Law-A Reply to Professor Hart, HARV71. L. REV. 630,650, 659-60 (1958); LON L. FULLER, THE MORALITY OF JOG (1964). Ronald Dworkin ezzel szemben sokkal szélesebb körben támaszkodott az erkölcsre. Lásd például: RONALD DWORKIN, JUSTICE FOR HEDGEHOGS (2013).

¹¹⁸ Lásd pl. RONALD DWORKIN, JUSTICE FOR HEDGEHOGS (2013).

bizonyos tekintetben, bár ő inkább az erkölcsi értékek egy speciális osztályára összpontosított, amely a jogrendszerekben sajátosan jelenik meg - amit ő a jog "belső erkölcsiségének" nevezett.¹¹⁹

Az ezekkel a jogtudományi kérdésekkel kapcsolatos álláspontunk hatással lehet a mesterséges intelligenciának a jogértelmezésben betöltött szerepére, különösen akkor, ha elfogadjuk, hogy a mesterséges intelligencia képes megtanulni az összes olyan leíró és előrejelző ítéletet, amelyet az emberek a joggal kapcsolatban hoznak, és jobb munkát végezni, de úgy gondoljuk, hogy nem képes érdemi erkölcsi ítéleteket hozni. Ha ez így van, akkor az embernek különleges szerepe lehet a jogértelmezésben, amennyiben - talán csak annyiban - a természetjog bizonyos körülmények között a legjobb magyarázatot adja a jog természetéről, vagyis amennyiben a jogértelmezés érdemi erkölcsi ítéleteket igényel.

A legnyilvánvalóbb példa a kizárólagos jogi pozitívizmus. A kizárólagos pozitivisták úgy vélik, hogy nincs szükség érdemi erkölcsi ítéletekre ahhoz, hogy megmondjuk, mi a jog.¹²⁰ Ha a számítógépek az embernél jobban - és talán sokkal jobban - képesek lesznek tisztán pozitív ítéleteket hozni a jogról, és ha a kizárólagos jogpozitivistáknak igazuk van a jog természetét illetően, akkor a mesterséges intelligencia képesnek tűnik arra, hogy kiszorítsa az embert a jogértelmezők közül. Az biztos, hogy még sok jogpozitivista is fontolgatja az erkölcs bizonyos szerepét az ítélkezésben. Scott Shapiro például azt állította, hogy a jogértelmezésnek csak a társadalmi tények értékelését kellene megkövetelnie, de az új jog megalkotása - amit a bírácoknak néha meg kell tenniük - és a jog alkalmazása - amit a bírácoknak gyakran meg kell tenniük - időnként erkölcsi ítéleteket igényel.¹²¹

Hasonlóképpen, az inkluzív jogpozitivisták elfogadják, hogy a jog néha erkölcsi ítéleteket tartalmaz, bár a jog tartalma - beleértve azt is, hogy megköveteli vagy megengedi-e az erkölcsi ítéleteket - *végső soron* csak társadalmi ténykérdés.¹²²

Még ha a jogpozitivistáknak igazuk is van a jog természetét illetően, akkor is lehet, hogy az embernek szerepe van a jogértelmezésben. De ez a szerep valószínűleg korlátozott lenne. Eljön majd az a nap, amikor a mesterséges intelligencia képes lesz - és feltehetően képesnek is kell lennie - a jog értelmezéséhez szükséges összes tisztán leíró és előrejelző ítéletet meghozni.¹²³ Azonosítaná és keretbe foglalná az erkölcsi ítéleteket is.

¹¹⁹ Lásd pl. Lon L. Fuller, Positivism and Fidelity to Law-A Reply to Professor Hart, 71 HARV. L. REV. 630,650, 659-60 (1958); LON L. FULLER, THE MORALITY OF LAW (1964).

¹²⁰ SHAPIRO, LEGALITY, supra note , at 269.

¹²¹ Lásd Joshua P. Davis, Legality, Morality, Duality, UTAH2014 L. REV. 55, 76-77 (a SHAPIRO, LEGALITY-t tárgyalva, 274-76. o.).

¹²² Más szóval, a kizárólagos és az inkluzív jogpozitivisták egyetértenek a társadalmi tényállás tézisében. SHAPIRO, LEGALITY, supra note , at 273.

¹²³ A jognak lehetnek olyan részei, amelyek nem alkalmasak a mesterséges intelligencia általi elemzésre, és nem tartalmaznak érdemi erkölcsi (vagy más értékítéleteket). Brian Leiter egy különösen jól kidolgozott nézetet kínált ezen a vonalon. Lásd Brian Leiter, Heidegger and the Theory of Adjudication, 106 YALE L.J. 253, 253 n.3 (1996). Lieter e ponton kifejtett nézetére adott választ lásd Joshua P. Davis & Manuel R. Vargas, Legal Dualism, Naturalism, and the Alleged Impossible of a Theory of Adjudication, in NATURALISTIC JURISPRUDENCE (Giudice, szerk. megjelenés: 2019) (tervezet a szerzőnél).

A jogi "kézművességről" szóló vitákhoz kapcsolódóan lásd pl. STANLEY FISH, DOING

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély

nélkül.
WHAT COMES NATURALLY: CHANGE, RHETORIC, AND THE PRACTICE OF THEORY IN LITERARY
AND LEGAL

a jogvita rendezéséhez szükséges. Az embereket ezután arra kérnék fel, hogy pótolják a hiányzó darabokat. Elképzelhető tehát, hogy az emberi bírák olyan szűk körű szerepet játszanának, mint amelyet ma az esküdtek a mi jogrendszerünkben, akik meghatározott és körülhatárolt kérdéseket oldanak meg egy olyan jogrendszerben, amelyet nagyrészt a mesterséges intelligencia irányít. Az ügyvédek pedig hasonló szerepet játszhatnak más környezetben.

Ezek a jogelméleti kérdések nem kizárólag a tudósok hatáskörébe tartoznak. Tekintsük meg e tekintetben John Roberts főbíró nyitóbeszédét a megerősítő meghallgatásán. Híresen a bírakat a hazai pálya bíróihoz hasonlította: "És nem fogom elfelejteni, hogy az én feladatom a labdák és ütések megítélése, nem pedig a dobás vagy az ütés".¹²⁴ A metaforákat sokféleképpen lehet értelmezni. Ezért veszélyes túl sokat olvasni ebből az állításból. De az egyik módja annak, hogy úgy értelmezzük, mintha a főbíró jogpozitivistának nyilvánítanánk, valószínűleg kizárólagos jogpozitivistának vagy normatív jogpozitivistának. A hazai pálya bírói végül is vitathatatlanul pusztán leíró értékelést végeznek: labda vagy ütés volt-e a dobás? Ideális esetben nem saját értékítéletet hoznak az ütőterület meghatározásakor, hanem mechanikusan követik a Major League Baseball szabálykönyvében foglalt szabályt.¹²⁵

Az a nézet, hogy a főbíró a jogi pozitívizmus híve, némi megerősítést nyer abban a nyilatkozatában, hogy "a jogállamiság" alapján fog dönteni az ügyekben, és kijelentette, hogy "nincs napirendje".¹²⁶ Ezek a szavak ismét kissé homályosak. De az egyik hihető módja annak, hogy úgy értelmezzük őket, hogy elkerülik az erkölcsi ítéleteket. Ha Roberts bírósággal számol, hogy a jogértelmezés során egyáltalán nem hoz erkölcsi ítéleteket, akkor törekvő, kizárólagos jogpozitivistának minősülne.¹²⁷

STUDIES (1989); Stanley Fish, *Dennis Martinez and the Uses of Theory*, 96 YALE L.J. 1773 (1987); Dan M. Kahan, *The Supreme Court 2010 Term: Foreword: Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law*, 125 HARV. L. REV. 1, 27-28 (2010) (a "kézműves normák" tárgyalása). Leiterhez hasonlóan Jack Balkin is kifejezetten Heidegger munkásságára támaszkodik. Lásd például: Jack M. Balkin, *Understanding Legal Understanding: The Legal Subject and the Problem of Coherence*, 103 YALE L.J. 105, n159. (1101993).

¹²⁴ Lásd <http://www.cnn.com/2005/POLITICS/09/12/roberts.statement/>

¹²⁵ A Major League Baseball szabálykönyv 2.00 szabálya a strike zónát úgy határozza meg, mint "az a terület a hazai pálya felett, amelynek felső határa egy vízszintes vonal a váll felső része és az egyenruhanadrág felső része közötti középpontban, az alsó szint pedig a térdkalács alatti üregben lévő lián", és amelyet "az ütő állása határoz meg, amikor az ütő felkészül a dobott labda megütésére". A főbíró megjegyzéseinek hangneme azt sugallja, hogy e szabály tartalmának meghatározásához nincs szükség a bírósági érdemi értékítéletére. Természetesen az ellenkezőjével is lehet érvelni. De ez inkább a főbíró jogtudományi szemléletének kritikáját jelentené, mintsem annak meghatározására tett erőfeszítést, hogy maga a főbíró hogyan látja a jog természetét.

¹²⁶ Lásd <http://www.cnn.com/2005/POLITICS/09/12/roberts.statement/>

¹²⁷ Ha helyett csak végső vagy alapvető erkölcsi ítéleteket akart hozni - például arról, hogy a demokráciában a nem választott bíráknak tartózkodniuk kell attól, hogy a jogértelmezés során érdemi erkölcsi ítéleteket hozzanak, talán a demokráciaelméletben rejlő okokból, ahogyan Scalia bírósági ítélete, Joshua P. Davis, *Legal Dualism, Legal Ethics, and Fidelity to Law*, 2016 ABA J. OF PROF. LAWY. 1, 12-13. -, akkor ő normatív pozitivista. Ez a megkülönböztetés a szövegben szereplő vita szempontjából nem számít.

Úgy tűnik, hogy ez is egy olyan feladat, amelyet a számítógépek hamarosan jobban tudnak majd elvégezni, mint az emberek. Itt bővítsük ki a főbíró metaforáját. Elgondolkodhatunk azon, hogy miért hagyatkozunk még mindig emberi bírókra a labdák és ütések megadásában. Miért nem bizzuk *ezt a* feladatot egy számítógépre? A baseballrajongók már látták a televízióban, hogy a számítógépek a jelek szerint pontosabban és megbízhatóbban tudják megkülönböztetni a labdákat és az ütéseket, mint az emberek valós időben. Feltehetően ezért használják a sporthálózatok a számítógépeket a bírók döntésének utólagos elemzésére. És ha a számítógépek még nem is pontosabbak és megbízhatóbbak, a közeljövőben biztosan azok lehetnek. Egyes sportágak már most is számítógépekre támaszkodnak a legfontosabb döntések meghozatalában vagy az emberi döntések felülvizsgálatában. Gondoljunk csak a teniszben a vonalbírói döntésekre.¹²⁸ Úgy tűnik, hogy a fő ok, amiért nem támaszkodunk a számítógépekre, a hagyományok iránti nosztalgikus nosztalgia. A számítógépekre támaszkodó bírók vagy játékvezetők, ahogy a britek mondanák, egyszerűen nem krikett.¹²⁹

A szórakoztatás összefüggésében - és bármennyire is komolyan vesszük, a sport mégiscsak a szórakoztatás egyik formája - érthető, hogy továbbra is emberi bírókra vagy játékvezetőkre támaszkodunk, még akkor is, ha ezek a bírók hajlamosabbak az elfogultságra, és kevésbé megbízhatóak és pontosak, mint a gépek. Végül is a gépek is jobbak lehetnek az emberi *játékosoknál*, de ha egy robot dobna vagy ütne, az tönkretenné a játékot. Ha mi is jobban szeretjük az emberi bírókat és játékvezetőket, akkor nyugodtan engedhetünk ennek a hajlamnak, még akkor is, ha ennek következtében rendszeresen szükségtelen hibákat tapasztalunk.

Ugyanez nem biztos, hogy igaz a jogra. Ha a számítógépek felülmúlják az embert a jogértelmezés tisztán leíró aspektusaiban - ha végül jobbnak bizonyulnak a "labdák és ütések" megadásában -, és ha a jogértelmezés csak ennyivel jár, talán engednünk kellene, hogy a számítógépek átvegyék a hatalmat. Ők lehetnek az ügyvédek - és akár a bíránk is. A nosztalgiának van némi értéke, de a gyorsaságnak, a pontosságnak és végső soron a "jogállamiságnak" is, hogy a főbíró idézzem. Ezért nem meglepő, hogy - amint azt fentebb tárgyaltuk - Max Tegmark javaslata, miszerint "a jogi folyamatot absztrakt módon számításnak tekinthetjük" - ami alatt, úgy tűnik, egy mechanikus, bár összetett folyamatot ért - arra a következtetésre vezetett, hogy a robotbírák sok tekintetben jobbnak tűnnek az emberi bíráknál.¹³⁰ Elméletileg a robotbírák kevésbé lehetnek elfogultak, nem hajlamosak a fáradtságra, és hatalmas tudás felhalmozására képesek.¹³¹ Ha a jogi értelmezés nem más, mint számítás, akkor erkölcsi kötelességünk lehet, hogy túllépünk a nosztalgiánkon, és jogrendszerünket (nagy részt) átadjuk a mesterséges intelligenciának. Hagyjuk, hogy a Herkules jobban csinálja, amit csinál, mint az emberi bírák. A Ludditizmus a jogállamiság gyenge versenytársa lenne.

De mi van akkor, ha a jogértelmezés nem pusztán számításról szól - még csak nem is bonyolult és finom számításról -, hanem valami többet igényel? Mi van, ha érdemi erkölcsi ítéleteket igényel, legalábbis sok esetben? És mi van, ha nem tudjuk úgy meghatározni az erkölcsi céljait, hogy a mesterséges intelligencia érdemi erkölcsi ítéleteket hozhasson? Ilyen körülmények között lehet okunk arra, hogy megőrizzük az emberi lények szerepét a jog kimondásában.

¹²⁸ <https://en.wikipedia.org/wiki/Hawk-Eye>

¹²⁹ Kivéve, hogy ironikus módon az. A krikett egyike azon sportágaknak, amelyek a pontosságot biztosító technológiát alkalmazzák. Id. Az, hogy ellentmondás alakult ki, nem meglepő, sőt elkerülhetetlen. Lásd például: <http://www.news.com.au/sport/cricket/the->

nélkül
ashes/talking-points-from-day-two-of-the-second-ashes-test-in-adelaide/news-
story/55a4a5164fc406b334cb59af186cc4c4.

¹³⁰ TEGMARK, LIFE supra3.0, note , at 105.

¹³¹ ID.

- B. Jogi dualizmus: A központi jogtudományi vita feloldása és a számítógépek szerepének körülírása.

A mesterséges intelligenciának a jogértelmezésben játszott lehetséges szerepével kapcsolatos vizsgálatunk tehát új jelentőséget és időszerűséget kölcsönöz a jogi pozitívizmus és a természetjog között régóta folyó vitának. A vita egyik lehetséges megoldása a jogi dualizmus. Ez azt sugallja, hogy a jog természetének legjobb bemutatása a jogértelmező céljaitól függ.

A jogi dualizmus különösen arra összpontosít, hogy a jogértelmezők milyen különböző módon tudnak a legjobban megküzdeni a jogi meghatározatlansággal. Legalábbis a jogi realisták XIX. század végi és XX. század eleji munkássága óta a legtöbb jogelméletíró elfogadja, hogy a jog jelentős mértékben határozatlan. Egy jogi kérdésre vonatkozóan gyakran többféle jogértelmezés áll összhangban a mérvadó jogforrásokkal.¹³² A jogi dualizmus abból a belátásból indul ki, hogy a jogértelmezőnek a jogértelmezés céljaitól függ, hogyan kell kezelnie ezt a meghatározatlanságot.

Egyes jogértelmezők megtehetik - és meg is kell hagyniuk - a jogi meghatározatlanságot. Aki egyszerűen csak leírja a jogot, az azonosíthatja a különböző lehetséges jogértelmezéseket, és elmagyarázhatja, hogyan egyeztethetők össze a hiteles jogforrásokkal.¹³³ Ha ezt a meghatározatlanságot kiiktatná a leírásából - ha csak azt írná le, amit ő a legjobbnak tart, elismerve, hogy vannak őszinte és ésszerű nézeteltérések -, akkor a leírása valószínűleg hiányos lenne, és ennek következtében szenvedne.

Ugyanez igaz arra is, ha valaki megjósolja, hogyan lehet értelmezni a törvényt. Amennyiben a jog elemzése tisztán prudenciális okokból indokolt - például a potenciális jogi felelősség felmérése egy vállalkozás megvásárlásának eldöntése során -, a legjobb jogértelmezési stratégia az lehet, ha a különböző lehetséges jogértelmezésekhez esélyeket rendelünk, és felmérjük azok pénzügyi következményeit. Ha egy fontos kérdésben a több lehetséges jogértelmezés közül csak egyet választunk, az rossz döntéshez vezethet. A cél nem az, hogy meghatározzuk a legjobb jogértelmezést - bármelyik értelmező véleménye szerint -, hanem inkább az, hogy ésszerű pénzügyi befektetést tegyünk, felismerve, hogy van egy bizonyos fokú bizonytalanság, elismerve, hogy különböző jogi kimenetek lehetségesek, és mérlegelve az esélyeket és az egyes kimenetek következményeit.

Azok a jogértelmezők, akik a törvényt tekintik az erkölcsi útmutatás forrásának, más helyzetben vannak. Nekik gyakran fel kell oldaniuk a jogban rejlő meghatározatlanságot, ha figyelembe akarják venni a jogot annak eldöntésében, hogy mit kell tenniük.¹³⁴ Gondoljunk csak egy bíróra. Tegyük fel, hogy a bírónak van némi erkölcsi kötelessége, hogy megpróbálja követni a törvényt a döntés meghozatalakor. A bíró felismerheti, hogy a vonatkozó jognak többféle értelmezése is létezik.¹³⁵ A döntés meghozatalakor azonban lehet, hogy egyetlenegy kell választania. A bíró nem tehet egyszerűen jóslatot arra vonatkozóan, hogy valószínűsíti, hogy

¹³² Lásd a 89. lábjegyzetet.

¹³³ Davis, Legal, Morality, Duality, supra note , 92. o.; Davis, Legal Dualism, Legal Ethics, and Fidelity to Law, supra note , at. 18.

¹³⁴ Davis, Legal, Morality, Duality, supra note , 92-23. o.

¹³⁵ Id.

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély
nélkül.

uralkodni.¹³⁶ Úgy tűnik, hogy akkor sem teljesítené kötelességét, ha egy egyszerű előrejelzés alapján döntene arról, hogy más bírák - esetleg fellebbviteli bírák - valószínűleg hogyan értelmezik a jogot. Ehelyett lehet, hogy a legjobb ítélőképessége alapján kell feloldania a jogban rejlő bizonytalanságokat, és legalább egy hihető elmélet szerint ezt úgy teheti meg, hogy figyelembe veszi az egyes lehetséges értelmezések viszonylagos illeszkedését a hiteles jogforrásokhoz és azok viszonylagos indokoltságát.¹³⁷ Más szavakkal, előfordulhat, hogy érdemi erkölcsi ítéleteket kell hoznia - valószínűleg a vonatkozó jogi szabályok és normák által keretezetteket - ahhoz, hogy a jog kellően meghatározó legyen ahhoz, hogy egy adott ügyben következtetésre jusson.

Ez nem a megfelelő hely arra, hogy a jogi dualizmus melletti teljes érvelést kifejtjük, hogy megvizsgáljuk az ellene felhozható kritikákat, vagy hogy értékeljük, hogy a jogi dualizmus megfelelő válaszokat ad-e rájuk. Ezek a kérdések inkább általánosak lesznek, és nem lesz különösebb jelentőségük a mesterséges intelligenciának az etikában betöltött szerepe szempontjából. De, mint fentebb említettük, van egy pont a jogi dualizmussal kapcsolatban, amely különösen fontos a jelenlegi vita szempontjából. Segíthet megmagyarázni és indokolni a mesterséges intelligencia jogértelmezésben betöltött szerepének korlátozását, és talán a robotbírák fogalmának elutasítását.

Tegyük fel, hogy négy állítás igaz. Először is, a jog eléggé határozatlan ahhoz, hogy számos jogi kérdésben többféle jogi értelmezés áll rendelkezésre. Másodszor, a bírácoknak - és néhány más jogértelmezőnek - a jogra mint az erkölcsi útmutatás forrására kell tekintenie. Harmadszor, amikor a jogértelmezők erkölcsi útmutatást keresnek a jogban, és a jog erkölcsi ítélet nélkül meghatározhatatlan, az értelmezőknek érdemi erkölcsi ítéletet kell gyakorolniuk, hogy a jog eléggé meghatározhatóvá váljon ahhoz, hogy segítse őket abban, hogy mit kell tenniük. Negyedszer, a mesterséges intelligencia nem képes érdemi erkölcsi ítéleteket hozni; emberi lényekre van szüksége, hogy meghozzák ezeket az ítéleteket, kiválasztva a végső célokat, amelyeket a mesterséges intelligencia követni fog. Ha elfogadjuk ezeket a tételeket, akkor a jogi dualizmus módot ad arra, hogy megértsük és igazolni tudjuk a mesterséges intelligencia szerepének korlátozását a jogértelmezésben. Az emberi jogértelmezőkre - és nem a mesterséges intelligenciára - kell támaszkodnunk, amikor a jog erkölcsi ítélet hiányában meghatározhatatlan, és az erkölcsi útmutatás forrásaként szolgál. Mikor valószínű, hogy ez bekövetkezik?

C. A jogi dualizmus alkalmazása: A jog mint az erkölcsi útmutatás lehetséges forrása.

A jogi dualizmus szerint a jogértelmezéshez csak akkor van szükség erkölcsi ítéletre, ha a jog az erkölcsi útmutatás forrásaként szolgál. A kérdés tehát az, hogy a jog valaha is betölti-e ezt a szerepet. Egyesek úgy gondolhatják, hogy nem. Oliver Wendell Holmes, Jr. például híres módon úgy határozta meg a jogot, mint "próféciaikat arról, hogy a bíróságok mit fognak tenni a valóságban, és semmi más mesterkélt dolog".¹³⁸ Ő dolgozta ki a jogról szóló "rossz ember" elméletét is, amely szerint a jog csak prudenciális, nem pedig erkölcsi aggályokra ad okot. Amint azt fentebb tárgyaltuk, a számítógépek hamarosan sokkal hatékonyabbak lehetnek a jogi döntések előrejelzésében - "jóslatok" felajánlásában -, mint az emberek. Ha ez így van, és ha valaki a Holmes-féle jogi pozitívizmusnak a fentiekben ismertetett formáját vallja, akkor lehet, hogy az embernek nincs különleges szerepe a jog értelmezésében, még akkor sem, ha elfogadja a jogi dualizmust.

¹³⁶ H.L.A. Hart, A jog fogalma 124-54 (1961).
nélkül:

¹³⁷ Lásd a fenti megjegyzések - és a kísérő szöveg.

¹³⁸ Oliver Wendell Holmes, Jr., The Path of the Law, HARV10. L. REV. 457 (1897).

De van okunk megkérdőjelezni Holmes pozitívizmusát. Ahogy például H. L. A. Hart rámutatott, a bírói döntések előrejelzése nem nyújt különösebben hasznos útmutatást a bírák számára.¹³⁹ Egy bírónak úgy kellene döntenie, hogy megjósolja, hogyan fog dönteni? Ez reménytelenül körkörösnek tűnik. Valójában a bírák különösen hihető - és különösen releváns - jogértelmezők, hogy valamilyen erkölcsi kötelességük van a törvény betartására. Erre esküt (vagy megerősítést) tesznek.¹⁴⁰ Ráadásul különleges hatáskörrel és vitathatóan különleges erkölcsi felmentésekkel rendelkeznek, amelyek ezen a lehetséges erkölcsi kötelezettségen alapulnak. Valószínűleg nem hatalmaznánk fel bírákat arra, hogy jelentős döntéseket hozzanak, ha nem gondolnánk, hogy erkölcsileg köteleznék őket a törvény. És talán szigorúan értékelnék azt, ahogyan a hatalmukat gyakorolják - évtizedekre börtönbe vetve embereket, megfosztva őket dollármillióktól -, ha úgy gondolnánk, hogy egyénileg erkölcsileg felelősek a tetteikért, és nem pedig a szerepükből adódóan megengednénk nekik, hogy olyan módon cselekedjenek, ami egyébként erkölcsileg gyanús lenne.¹⁴¹ És úgy gondoljuk, hogy a peres feleknek különleges okuk van arra, hogy panaszkodjanak, ha egy bíró döntése nemcsak számukra kedvezőtlen, hanem "törvénytelen" is. A bírának tehát valószínűleg legalábbis erkölcsi kötelességük, hogy figyelembe vegyék a törvényt, amikor arról döntenek, hogy mit tegyenek. Ennyiben, ha a jogi dualizmusnak igaza van, ez segíthet megmagyarázni, hogy a robotbírák miért nem válthatják ki az emberi bírákat.

Az ügyvédek esetében az elemzés bonyolultabb. A jogi etika egyik központi vitája az, hogy az ügyvédek lényegében jogi pozitivistaként működhetnek-e - bármilyen lehetőséget kínálva és bármilyen jogi érveket hozva fel, amelyek a legjobban megfelelnek ügyfeleik igényeinek vagy kívánságainak -, vagy pedig az ügyvédeknek kötelességük, hogy a jog értelmezése során független ítéletet hozzanak. Ez a vita gyakran úgy zajlik, hogy nézeteltérés van arról, hogy jogrendszerünket úgy kell-e értelmezni, mint amely a kontradiktórius rendszer egy erős formáját testesíti meg.¹⁴² Ez nem a megfelelő hely arra, hogy állást foglaljunk ebben a vitában.¹⁴³ Megjegyzendő azonban, hogy számos elismert tudós szkeptikusan áll az agresszív ügyvédi magatartás igazolásához a kontradiktórius rendszer nevében - legalábbis bizonyos körülmények között. Ha valaki osztja ezt a szkepticizmust, és elfogadja a jogi dualizmust is, akkor könnyen arra a következtetésre juthat, hogy az ügyvédeknek független véleményt kell nyilvánítaniuk arról, hogyan lehet a legjobban értelmezni a törvényt és a jogi kötelezettségeket, és nem csak megjósolni a lehetséges kimeneteleket, mint ahogyan azt Holmes Rossz embere tenné. A jogi dualizmus szerint ez azt jelenti, hogy az ügyvédeknek időnként kötelességük erkölcsi ítélőképességet gyakorolni a

¹³⁹ Lásd H.L.A. HART, A JOG FOGALMA 124-54 (1961).

¹⁴⁰ 28 U.S.C. § 453.

¹⁴¹ A kölcsönösség e fogalmának - a jog megvédi az embereket a tetteikért való felelősségre vonástól, amikor a jogrendszerben betöltött intézményi szerepükön belül működnek - megvitatását lásd Davis, Legal Dualism, Legal Ethics, and Fidelity to Law (Jogi dualizmus, jogi etika és joghűség), fentebb, 20-22. o.). E kérdés szempontjából releváns Hart megkülönböztetése a törvénynek való engedelmisség és a fegyvert viselő ember parancsainak való engedelmisség indoka között. Lásd H. L. A. HART, A JOG FOGALMA 6, 20, 90 (1961).

¹⁴² Lásd pl. W. BRADLEY WENDEL, LAWYERS AND FIDELITY TO LAW (2010); TIM DARE, THE COUNSEL OF ROGUES?: AZ ÜGYVÉD SZEREPÉRŐL ALKOTOTT ÁLTALÁNOS FELFOGÁS VÉDELME (2009); RICHARD MARKOVITZ, A MODERN JOGI ETIKA: ADVERSARY ETHICS IN A DEMOCRATIC AGE (2008); DAVID LUBAN, LEGAL ETHICS AND HUMAN DIGNITY (2007); DEBORAH RHODE, IN AZ IGAZSÁGOSSÁG ÉRDEKEI (2000); WILLIAM H. SIMON, AZ IGAZSÁGOSSÁG GYAKORLATA: AZ

nélkül
IGAZSÁGOSÁG ELMÉLETE (2000); WILLIAM H. SIMON, THE PRACTICE OF JUSTICE: A THEORY
OF
ÜGYVÉDI ETIKA (2000).
143 Lásd pl. DAVID LUBAN, LEGAL ETHICS AND HUMAN DIGNITY (2007); DEBORAH
RHODE, IN THE INTERESTS OF JUSTICE (2000); WILLIAM H. SIMON, THE PRACTICE OF
JUSTICE: A THEORY OF
ÜGYVÉDI ETIKA (2000).

a jogi szolgáltatások nyújtásának részeként a jogszabályok értelmezése. Ilyen körülmények között a robotbírókra vonatkozó korlátozások a robotügvyédekre is kiterjedhetnek. A mesterséges intelligencia egyik formája sem lehet képes az etikus cselekvéshez szükséges érdemi erkölcsi ítéletek meghozatalára. A Herkules - bármennyire is zseniális - csak korlátozottan képes ügyvédi szerepet játszani, ahogyan arra is csak korlátozottan képes, hogy bíróként elnököljön.

Ha az ügyvédeknek bizonyos körülmények között kötelességük független jogi ítéletet hozni - és ennek során *erkölcsi* ítéletet is kell hozniuk -, akkor ez a kötelezettség valószínűleg az önvezető autók programozásának jogi normáira vonatkozó jogi normák értékelése során merül fel. Vegyünk egy kapcsolódó történelmi precedenst: a Ford Pinto ügyét. A Ford úgy tervezte meg a Pintót - beleértve a benzintartály elhelyezését és a kapcsolódó biztonsági intézkedéseket -, hogy a hátulról történő ütközés esetén szokatlanul magas volt a személyi sérülés kockázata. A *Grimshaw kontra Ford Motor Company* ügyben egy ebből eredő perre került sor, amelyben a sérült autós nemcsak a sérülésekért járó kártérítést, hanem büntető kártérítést is követelt.¹⁴⁴ A Ford azzal érvelt, hogy a kaliforniai jog szerint a büntető kártérítés kiszabásához rosszhiszeműségre van szükség, és mivel a Ford nem akart senkinek sem ártani - legfeljebb üzleti döntésként kockáztatta a kárt -, nem rendelkezhetett a szükséges rosszhiszeműséggel. A kaliforniai fellebbviteli bíróság nem értett egyet ezzel, ami némi fényt vet a vállalati szereplő azon kötelezettségeire, hogy a jogi következményeken túlra tekintsen, mint ahogyan azt Holmes Rossz embere tenné.

A kaliforniai fellebbviteli bíróság elutasította azt a felfogást, hogy a Fordnak "szándékában kell állnia egy vagy több személynek kárt okozni" ahhoz, hogy rosszhiszeműen cselekedjen.¹⁴⁵ Ehelyett úgy ítélte meg, hogy a "rosszindulat" szó magában foglalja "a közbiztonság érzéketlen és tudatos figyelmen kívül hagyását tanúsító magatartást azok részéről, akik tömegcikkeket gyártanak és forgalmaznak".¹⁴⁶ Továbbá a "közbiztonság érzéketlen és tudatos figyelmen kívül hagyása" fogalmát úgy határozta meg, hogy az magában foglalja egy autógyártó azon döntését, hogy "a kártérítési kártérítést az üzleti tevékenység költségeinek részeként kezeli, ahelyett, hogy a hibát orvosolná".¹⁴⁷ A bíróság érvelésének egyik értelmezési módja az, hogy a Fordot arra kötelezi, hogy tartsa be a közbiztonság védelmét szolgáló törvényt, és ne kezelje a jogi felelősség kockázatát pusztán pénzügyi megfontolásként a profitmaximalizálás során, a la Holmes rossz embere. E kötelezettség szándékos megszegése - mivel a kompenzációs kártérítés megfizetése nettó módon nyereséges lenne - elegendő volt a büntető kártérítés alátámasztásához.¹⁴⁸

Természetesen a kaliforniai fellebbviteli bíróság véleménye nem dönti el, hogy a Fordnak - és az azt tanácsadó ügyvédeknek - erkölcsi kötelessége volt-e figyelembe venni a törvényt annak eldöntése során, hogy hogyan

¹⁴⁴ Grimshaw v. Ford Motor Co., 119 Cal. App. 3d 757 (4th App. Distr. 1981).

¹⁴⁵ Id. 809.

¹⁴⁶ Id. 810.

¹⁴⁷ Id.

¹⁴⁸ Megjegyzendő, hogy a büntető jellegű kártérítési igényhez jogsérelemre van szükség; a büntető jellegű kártérítésre nincs önálló keresetindítási ok. Tehát a bíróság érveléséhez szükséges volt az a tény, hogy a Ford magatartása törvényt sértett - és nem csak az, hogy vitathatóan erkölcsileg elítélendő volt -. Ha egy vállalati szereplő olyan módon veszélyezteti a nyilvánosság tagjait, amely nem sérti a jogi normát - például ha a magatartása nem minősül

deliktumnak^{nélkül}, akkor nem lenne igény a büntető kártérítésre. A büntető kártérítéshez az alapjául szolgáló jogi jogsértés szükséges. Úgy tűnik tehát, hogy ebben a példában egy jogi norma - nem csak egy erkölcsi norma - megsértése végez némi munkát.

az autók tervezéséhez. Egy olyan bírói vélemény, amely azt jelzi, hogy a törvények betartására kötelesség van - és a jogi szankciókat nem szabad pusztán ösztönzőként kezelni -, nem oldja meg azt a kérdést, hogy a törvény valóban erkölcsi kötelezettségeket teremt-e. A kérdés egyszerűen megismétli önmagát. Van-e a Fordnak erkölcsi kötelessége követni a büntető kártérítés kiszabásában kifejezett bírói megbízást, vagy erkölcsileg megengedhető, hogy a Ford magát a büntető kártérítést - a kártérítés mellett - csupán "az üzletmenet költségeként" kezelje, ahogyan azt Holmes Rossz embere tenné?

De *Grimshaw* legalább szuggesztív. Azt jelzi, hogy maga a jog - ahogyan azt végre egy jelentős bírói vélemény is megtestesíti - úgy véli, hogy egy jogi szabály relevanciája több mint prudenciális és kötelezettségeket keletkeztető. És arra utal, hogy a Ford valami erkölcsileg helytelen dolgot tett - hogy "rosszindulatúan" cselekedett, ami egy erkölcsi és nem csak jogi jelentéssel terhelt kifejezés -, nemcsak azzal, hogy veszélybe sodorta az embereket, hanem azzal is, hogy ezzel megsértette a törvényes jogait.

Továbbá az autógyártók döntéseit meghatározó jogi értelmezéseket nem vetik alá a kontradiktórius eljárás szigorúságának, mielőtt azok súlyos következményekkel járnának. Amikor az ügyvédek a nyílt bíróságon adják elő érveiket - és a bírónak lehetősége van arra, hogy megvizsgálja azokat, és válogasson közülük -, az ügyvédek vitathatóan viszonylag szabadon foglalhatnak el agresszív jogi álláspontokat.¹⁴⁹ A bírósági rendszer, ha úgy tetszik, biztonsági hálóként szolgál. A bírónak vitathatatlanul az a feladata, hogy független jogi - és talán erkölcsi - ítéletet hozzon. Ez felszabadíthatja az ügyvédeket.¹⁵⁰ De ugyanez nem igaz egy autógyártóra, akinek a döntései sok emberéletbe kerülhetnek - és sokakat megóvhatnak -, mielőtt a bírósági rendszernek reális esélye lenne arra, hogy értékelje az autógyártó elemzését, és esetleg felszólítsa az autógyártót magatartásának módosítására. Ilyen körülmények között - amikor a világ egy szereplője és annak ügyvédei olyan döntéseket hoznak, amelyeknek jóval azelőtt vannak következményei, hogy azokat egy pártatlan bíróság értékelné - különösen erős az az érv, hogy az ügyvédek kötelesek független ítéletet hozni, amikor megmondják, mi a törvény.¹⁵¹ Nem támaszkodhatnak ugyanolyan mértékben a kontradiktórius rendszer korrekciós hatásaira, mint egy bíróságon érvelő ügyvéd.

A Ford Pinto ügyben tárgyalt azon kötelezettség, hogy a vonatkozó jogi normát figyelembe kell venni az erkölcsi magatartás eldöntésekor - ha van ilyen -, természetesen kiterjedne az önvezető autók programozására is. A benzintartály elhelyezéséhez hasonlóan a programozási döntések is hatással lesznek arra, hogy ki él és ki hal meg. Ha egy autógyártó cég úgy hoz döntést, hogy az emberek jogainak megsértésével járó emberi sérüléseket és haláleseteket az üzleti tevékenység pusztán költségeként kezeli, akkor a jelek szerint büntető és kártérítési kártérítés is járhat.

Ráadásul, amint azt fentebb tárgyaltuk, az a mód, ahogyan egy autó meghozza ezeket a döntéseket - legalábbis ha mesterséges intelligenciára támaszkodik -, nagyon nehezen felismerhető, így a halálhoz vezető végső ok-okozati lánc értékelése potenciálisan nehezen követhetővé válik. Az a számítás, amelyet egy önvezető autó a magatartásformák megválasztása során végzett, homályos és elvontabb lehet - és így még utólag is nehezen értékelhető.¹⁵² Gyakorlati valóságként a számítógépes programozók a

¹⁴⁹ Davis, Legal Dualism, Legal Ethics, and Fidelity to Law, *Supra* note , 22-25. o.

¹⁵⁰ Id. 24-25.

¹⁵¹ Id. 24. pont.

¹⁵² Lásd a fenti részt.
nélkül

az önvezető autók gyártójának hátsó szobájában - és az őket tanácsokkal ellátó ügyvédek - korlátozott ellenőrzésnek alávetett, mélyreható következményekkel járó, döntő fontosságú döntéseket hoznak. Pontosán ezek azok a körülmények, amelyek között elvárható, hogy az ügyvédek és ügyfelek különösen nagy mértékben kötelesek legyenek figyelembe venni jogi és egyéb erkölcsi kötelezettségeiket - feltéve, hogy a jogi kötelezettségek erkölcsi kötelezettségeket teremtenek. Bármilyen kontradiktórus eljárás valószínűleg legfeljebb késve lesz hatással a döntő fontosságú döntésekre, és még akkor is csekélyek lehetnek az esélyei annak, hogy kiderüljön, mi történt, és ki a felelős.¹⁵³ Ahogy az ügyvéd, aki azt tanácsolja az ügyfelének, hogyan semmisítse meg a bizonyítékokat, hogy megússza a gyilkosságot, valószínűleg szkeptikusak leszünk azzal kapcsolatban, hogy egy ügyvéd megfelelő tanácsot adhat az önvezető autók gyártójának, hogy a vonatkozó jogi normákat sértő lépéseket tegyen, csak azért, mert a gyártó valószínűleg "megússza". Az ügyfeleknek az önvezető autók programozásával kapcsolatban tanácsot adó ügyvédeknek tehát különösen nagy valószínűséggel kötelességük - ha az ügyvédeknek egyáltalán van ilyen kötelességük -, hogy független ítélőképességgel határozzák meg, hogy mit ír elő a jog, és ne csak a különböző választások lehetséges pénzügyi következményeit ismertessék. Ha a jogi dualizmus meggyőzően mutatja be a jog és a jogértelmezés természetét, akkor ez a független jogi ítélőképesség időnként *erkölcsi* ítélőképességgel járna együtt.

D. Jogtudomány a mesterséges intelligenciában; mesterséges intelligencia a jogtudományban.

i. Az erkölcsi ítéletek mint a mesterséges intelligencia szerepének körülírása a jogértelmezésben.

Láttuk, hogy a jogtudomány tanulságai a jog természetéről és a jogértelmezésről útmutatást adhatnak a mesterséges intelligencia lehetőségeiről és korlátairól. A robotértelmezők hamarosan felülmúlhatják az embereket abban, hogy leíró és előrejelző - talán még meggyőző - beszámolókat adjanak a jogról. Az azonban már kevésbé egyértelmű, hogy még Herkulest is be lehet-e programozni arra, hogy olyan érdemi erkölcsi ítéleteket hozzon, amelyek néha szükségesek lehetnek ahhoz, hogy megmondjuk, mi a jog. Ha Herkules nem programozható így, és ha a jogértelmezés néha erkölcsi ítéleteket igényel, akkor lehet, hogy van egy határ, amelyet a mesterséges intelligencia nem léphet át a belátható jövőben, még akkor sem, ha technikai képességei továbbra is rendkívüli ütemben növekednek. Lehet, hogy nem csak fokozati különbség van a Herkules által meghozható ítéletek és egyes ítéletek - az erkölcsi ítéletek - között, amelyek időnként szükségesek a jogértelmezéshez.

Ennek eredményeképpen a Herkules által játszható szerep részben a jog természetének legjobb megértésétől függhet. Ha a kizárólagos jogi pozitívizmus megragadja a jog természetét, akkor a mesterséges intelligencia hamarosan képes lehet az embernél hatékonyabb jogértelmezésre. Végére is, annak megmondásához, hogy mi a jog, csak a társadalmi tények értékelésére van szükség. Ha pontos, megbízható és hatékony jogértelmezést akarunk, akkor lehet, hogy érdemes a jogértelmezési munka nagy részét a számítógépekre bízni. Lehet, hogy ők jobban fogják csinálni, mint mi, és képesek lesznek a bírúk és ügyvédek feladatainak nagy részét elvégezni. Természetesen az embereknek még mindig maradhat némi szerepük. Lehet, hogy egyedülálló módon képesek vagyunk az új jogszabályok megalkotásához szükséges erkölcsi vagy egyéb értékítéleteket meghozni. És talán ugyanez igaz a törvény *alkalmazására is*, szemben azzal, hogy kimondjuk, mit jelent a törvény.

¹⁵³ Davis, Legal Dualism, Legal Ethics, and Fidelity to Law, Supra note , 22-25. o.; Wendel, Supra note , 187-92. o.

az.¹⁵⁴ A mesterséges intelligenciának azonban képesnek kell lennie a többi feladat elvégzésére, majd az emberi bírák vagy jogászok által elvégzendő értékelések azonosítására és kidolgozására.¹⁵⁵

Az inkluzív jogi pozitívizmus ezzel szemben megőrizhetné az ember nagyobb szerepét. Elfogadja annak lehetőségét, hogy a jogba erkölcsi értékítéleteket is be lehet építeni. Amennyiben a jog ezt teszi, az emberi lényeknek különleges szerepük lehet abban, hogy megmondják, mi a jog. Ennek ellenére, hasonlóan a kizárólagos jogi pozitívizmushoz, a mesterséges intelligencia továbbra is nagymértékben elvégezheti helyettünk a jogértelmezés nagy részét, keretbe foglalva azokat a kérdéseket, amelyeket nyitva kell hagynia, hogy azokat olyan emberi lények oldhassák meg, akik képesek érdemi erkölcsi ítéleteket hozni.

A jogi dualizmus ezzel szemben azt sugallja, hogy az erkölcsi ítéletek szükségesek a jogi értelmezéshez számos helyzetben. A dualizmus szerint a jogértelmezőnek különösen akkor kell erkölcsi ítéleteket hoznia, amikor erkölcsi kötelessége figyelembe venni a törvényt, amikor arról dönt, hogyan cselekedjen. A jogi dualizmus így segíthet felismerni, hogy mikor nem szabadna Herkulesnek felhatalmazást adni a jogértelmezésre. Általánosságban azt a következtetést vonhatjuk le, hogy nem szabad emberi bírák helyett robotbírakat alkalmaznunk - a robotbírák bizonyos szempontból vélt előnyei ellenére -, és¹⁵⁶ legalábbis bizonyos körülmények között nem szabad emberi ügyvédek robotügyvédekkel kiszorítanunk.¹⁵⁷ Még általánosabban felismerhetjük, hogy a jogtudomány képes lehet tájékoztatni a számítógépes tanulással kapcsolatos legfrissebb etikai kérdésekről.

Ez fordítva is igaz. A technológia fejlődésének és az ebből adódó etikai kérdéseknek a szemlélése a jogtudomány számára is hasznos lehet. Ha például kételkedünk abban, hogy a robotbírák - vagy akár a robotügyvédek - jó ötlet, és ha e kételyeink egyik oka az, hogy gyanítjuk, hogy a bírácoknak és az ügyvédeknek erkölcsi ítéletet kellene hozniuk, akkor ez egyfajta

¹⁵⁴ Lásd pl. SHAPIRO, LEGALITY, supra note , 141-47. o. (elismeri, hogy az erkölcsi ítéletek időnként szükségesek lehetnek az ítélezésben, még ha nem is a jogértelmezésben).

¹⁵⁵ Nagyjából ugyanez az elemzés látszik érvényesnek a normatív (vagy etikai) pozitivistákra is. Emlékezzünk vissza, hogy a normatív (vagy etikai pozitivisták) szerint a pozitívizmus elfogadásához csak egy alapvető (vagy végső) erkölcsi ítéletre van szükség. A mesterséges intelligencia programozói meghozhatják ezt az ítéletet. E célból úgy tűnik, hogy a normatív (vagy etikai) pozitívizmus a kizárólagos jogi pozitívizmusba omlik bele. Természetesen lehetséges, hogy bizonyos körülmények között az alapvető (vagy végső) ítélet arra készítetne bennünket, hogy elhagyjuk a pozitívizmust, hogy a normatív (vagy etikai) pozitívizmus életben tartja ezt a lehetőséget, és hogy a normatív (vagy etikai) pozitívizmus ezért bizonyos körülmények között eltérhet a kizárólagos jogi pozitívizmustól. Scalia bíró például egyszer híresen azt mondta, hogy ő egy "gyengébb szívű" originalista lenne, ha a korbácsolás eredetileg nem minősülne kegyetlen és szokatlan büntetésnek. Antonin Scalia, Originalizmus: The Lesser of Two Evils, 57 U. Cin. L. Rev. 849, 864 (1989). Később visszavonta ezt az álláspontját. <http://volokh.com/2013/10/07/justice-scalia-repudiates-fainthearted-originalism/>

¹⁵⁶ TEGMARK, LIFE supra 3.0, note , 105-06. o.

¹⁵⁷ Az, hogy az emberi ügyvédek valóban független ítélőképességgel járnak-e el, ahelyett, hogy egyszerűen azt mondanák ügyfeleiknek, amit azok hallani akarnak, egy kissé különálló, bár kapcsolódó kérdés. Ugyanígy az a kérdés is, hogy a jövőben megkívánhatjuk-e - vagy ösztönözhetjük-e - az ügyfeleket arra, hogy a jelentős jogi következményekkel járó döntések

nélkül
meghozatalakor ne robotügyvédekre, hanem független, emberi jogi tanácsadókra
hagyatkozzanak. Kérdezhetnénk, hogy egy számítógép folytathat-e jogosulatlan ügyvédi
tevékenységet? Ezek a kérdések túlmutatnak e cikk keretein.

a jogi dualizmus vonzónak találásának oka, a mellette szóló egyéb érvek mellett. Ha a jogi dualizmus meg tudja magyarázni és igazolni az ellenállásunkat, hogy Herkulesnek túl nagy szerepet adjunk a jogrendszerünkben, az a dualizmus mellett szól.

ii. Az erkölcsi ítéletek különböző fajtái

A technológiai fejlődésre való odafigyelésnek más módjai is lehetnek a joggyakorlatban. Vegyünk egy fentebb rövideen említett kérdést. A mesterséges intelligencia nem csak a jogról, hanem az erkölcsről is képes lehet bizonyos ítéleteket hozni. Az erkölcs, akárcsak a jog, részben társadalmi gyakorlat. Vannak érdekes társadalmi tények az erkölcsről. Még ha a mesterséges intelligencia nem is tud érdemi erkölcsi ítéleteket hozni, képes lehet arra, hogy társadalmi tényként leírja az emberek erkölcsi meggyőződéseit, megjósolja, hogy az emberek milyen erkölcsi ítéleteket fognak hozni, sőt még arra is, hogy azonosítsa azokat az erkölcsi érveket, amelyeket az emberek meggyőzőnek találnak.¹⁵⁸

Vegyük észre azt a párhuzamos szerepet, amelyet Herkules játszhatna a jogi elemzésben. A jogi értelmezők - beleértve a mesterséges értelmezőket is - különböző típusú állításokat tehetnek: leíró, előrejelző, meggyőzőnek nevezhető és érdemi állításokat. Más szóval, egy jogértelmező törekedhet arra, hogy leírja a jogot, megjósolja, hogy mások hogyan fogják értelmezni a jogot, vagy hogy a jog elemzését úgy alakítsa ki, hogy meggyőzzön valakit (ami egyfajta előrejelzéssel jár azzal kapcsolatban, hogy az illető hogyan fog reagálni). Vagy a jogértelmező hozhat úgymond érdemi jogi ítéletet, megkísérelve a lehető legjobb jogértelmezést nyújtani, beleértve - legalábbis a jogi dualizmus szerint - bizonyos esetekben a jog meghatározóvá tételéhez szükséges erkölcsi ítéleteket is, és azt, hogy a magatartás irányítása céljából megmondja, mi a jog. Ugyanilyen típusú ítéleteket lehet hozni az erkölcsről is. Ez pedig a jogtudományi táj felosztásának a hagyományostól eltérő módját sugallja - legalábbis bizonyos célokra másképp.

Az új divízió egyik lehetséges útja a következő. Lehet, hogy a jog bizonyos fajta erkölcsi ítéleteit bizonyos célokból összeegyeztethetőnek tekinthetjük a jogi pozitivizmus (bizonyos változataival). Különösen azt mondhatnánk, hogy az olyan erkölcsi ítéletek, amelyek csak a társadalmi tények értékelését foglalják magukban - amelyek csak leíró, előíró vagy meggyőző értékelést igényelnek -, nem feltétlenül teszik a jogértelmezést csak a természetjoggal összeegyeztethetővé. A természetjogot - legalábbis bizonyos célokból - arra a nézetre korlátozhatjuk, hogy a jogértelmezőknek *érdemi* erkölcsi ítéleteket kell hozniuk - ítéleteket arról, hogy az erkölcs valójában mit jelent, és nem csupán ítéleteket arról, hogy mások hogyan értékelték vagy fogják értékelni az erkölcsöt.¹⁵⁹

¹⁵⁸ Megjegyzendő azonban, hogy a mesterséges intelligenciának nehézséget okozhat az erkölcsi (vagy más értékítéleteket) elkülöníteni az emberek döntéseinek egyéb motivációitól. Lásd fentebb a .

¹⁵⁹ Alternatív megoldásként félretehetnénk a jogtudomány hagyományos felosztását, és új címkéket használhatnánk a jogértelmezésben használt erkölcsi ítéletek e különböző fajtáinak megkülönböztetésére, ahelyett, hogy újraértelmeznénk a jogi pozitivizmust és a természetjogot. Tekintettel a témával kapcsolatos hatalmas szakirodalomra, egy újabb definíció megadása zavart okozhatna.

Egy másik lehetőség az lenne, ha a joggal kapcsolatos tisztán leíró vagy előrejelző állításokat - beleértve az erkölcsi ítéleteket leíró vagy előrejelző állításokat is - "külsőnek", a joggal kapcsolatos, érdemi erkölcsi ítéleteket tartalmazó állításokat pedig "belsőnek"

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély
nevezni. Ha is vagyok Brad Wendelnek, amiért felvette ezt a lehetőséget.

Akárhogy is fogalmazzuk meg szemantikailag a kérdést, az erkölcsről szóló leíró vagy előrejelző ítéletek megkülönböztetése a lényegi erkölcsi ítéletektől hasznosnak bizonyulhat például annak meghatározásában, hogy Herkules mit tehet. Emlékezzünk vissza, hogy feltételeztük, hogy a mesterséges intelligencia hamarosan minden leíró és előrejelző elemzésben jobb lesz, mint az ember. Ez kiterjedhetne az erkölcsről szóló tisztán leíró és előrejelző állításokra is - arra, hogy az emberek társadalmi tényként mit hisznek az erkölcsről, és mit fognak előre megjósolni, hogy bizonyos körülmények között mit fognak hinni.¹⁶⁰ Feltételeztük azonban, hogy a mesterséges intelligencia nem rendelkezik és nem is fog rendelkezni hasonló képességgel ahhoz, hogy érdemi erkölcsi ítéleteket hozzon - ítéleteket arról, hogy valójában mi erkölcsös vagy erkölcstelen. Miközben feltételeztük, hogy a Herkules képes leírni és megjósolni az erkölcsi ítéleteket - és talán még az embernél is jobban képes erre -, azt is feltételeztük, hogy nem tud érdemi erkölcsi ítéleteket hozni.

Ennek a megkülönböztetésnek van visszhangja Lon Fuller írásaiban. A H.L.A. Harttal folytatott híres vitája során Fuller az erkölcs kérdésével, és különösen a katolikus egyház által előírt erkölcsi meggyőződéssel foglalkozott. Úgy tűnt, hogy kényelmetlenül érzi magát attól a gondolattól, hogy egy bíró vallási meggyőződése - és talán különösen a katolicizmusból származó vallási meggyőződések - befolyásolhatják a jogértelmezést. Ez segíthet megmagyarázni, hogy miért hangsúlyozott bizonyos típusú erkölcsi ítéleteket, amelyeket megengedhetőnek tartott a jogértelmezésben.¹⁶¹ E kérdéssel foglalkozva azt sugallta, hogy a pápa kijelentései - különösen a válással kapcsolatban - inkább egyfajta jog, mint egyfajta erkölcs.¹⁶² A valódi erkölcsi ítéleteket - utalt rá - talán úgy lehetne a legjobban értelmezni, mint "a helyes magatartás általánosan osztott nézeteit, amelyek spontán módon, a tapasztalat és a vita révén alakultak ki".¹⁶³

Függetlenül Fuller e kérdésben elfoglalt álláspontjának érdemeitől és pontos tartalmától, elemzése vitathatatlanul tartalmazza a fent tárgyalt gondolat csíráját. Az erkölcsről szóló ítéletek bizonyos fajtáit - leíró és előrejelző állításokat arról, hogy az emberek mit tartanak erkölcsösnek - a jogi pozitivizmussal összeegyeztethetőnek tekinthetjük, legalábbis bizonyos célokra. Egy jogtudós nézetei akkor minősülhetnek jogpozitivistának, ha a jogot úgy határozná meg, mint "jóslatokat arról, hogy a bíróságok valójában mit fognak tenni, és semmi más mesterkéeltséget", még akkor is, ha ezek a jóslatok előrejelzéseket vonnának maguk után arról, hogy a bírák valószínűleg milyen erkölcsi értékeket vallanak, és hogy ezek hogyan fogják tájékoztatni a jogértelmezést.

Hasonlóképpen, ha ezt a megközelítést alkalmazzuk, akkor egy szociológust úgy jellemezhetnénk, mint aki jogi

¹⁶⁰ Mint fentebb már említettük, Hercules képessége, hogy az emberek érdemi erkölcsi ítéleteit elválaszthassa cselekedeteik egyéb alapjaitól, beleértve a jogi értelmezéseket is, korlátozott lehet. Lásd a fenti részt.

¹⁶¹ Fuller úgy tűnt, hogy ezt a kérdést valóban nagyon fontosnak tartja. Lon L. Fuller, Positivism and Fidelity to Law-A Reply to Professor Hart, 71 HARV. L. REV. 630, 638 (1958) ("[I]n sokak gondolkodásában van egy kérdés, amely a jog és az erkölcs kapcsolatáról szóló minden vitában dominál, olyannyira, hogy mindent kiszínez, amit a témáról mondanak vagy hallanak. Arra a fajta kérdésre utalok, amelyet a pápa kijelentése vetett fel a katolikus bírák köteleességéről a válóperekben.").

¹⁶² Id.

¹⁶³ Id.

pozitivistá, ha a szociológus pusztán leíró és előrejelző állításokat tesz egy társadalom törvényeiről, még akkor is, ha ezek az állítások leírásokat vagy előrejelzéseket tartalmaznak arra vonatkozóan, hogy a társadalom tagjainak erkölcsi ítéletei hogyan tájékoztatják a jogi értelmezéseket.¹⁶⁴

A mesterséges intelligenciával összefüggésben a jogi pozitívizmus és a természetjog megkülönböztetésének ez az alternatív módja egy másik előnnyel is járhat. Azt mondhatnánk, hogy a természetjog olyan módon "természetes", ami történelmileg irreleváns volt: elhatárolhatja, hogy hol kell a jogértelmezést a természetes, és hol nem a mesterséges intelligenciának végeznie - amikor nekünk, emberi lényeknek kell felelősséget vállalnunk a törvényeink értelmezéséért. Ha a jogértelmezéshez csupán az erkölcsnek mint társadalmi ténynek a leírására vagy az emberek által vallott erkölcsi értékek előrejelzésére van szükség, akkor az AI hamarosan képes lehet arra, hogy a vonatkozó erkölcsi ítéleteket jobban meghozza, mint mi. Csak ha a jogértelmezés - a feltételezéseink szerint - *érdemi* erkölcsi ítéleteket igényel, akkor az "természetes" abban az értelemben, hogy a mesterséges intelligencia számára nem hozzáférhető. Lehet, hogy a mesterséges intelligencia nem képes a természetes jogot értelmezni, hanem csak azt, amit "természetellenes jognak" nevezhetünk.

IV. Következtetés

Egy tipikus sci-fi filmben minden rosszul sül el, amikor az életünk javítására épített számítógépek elkezdnek saját döntéseket hozni. Gondoljunk csak a Terminátor című filmre. Ott egy mesterséges intelligencia védelmi hálózata öntudatra ébred, képessé válik az önálló cselekvésre, és nukleáris holokausztot indít.¹⁶⁵ Nem egészen ilyen apokaliptikus a kilátás arra, hogy számítógépes programok értelmezik helyettünk a törvényeinket, és talán átveszik a bírói tisztelet, kártérítést ítélnék meg, sőt büntetőjogi szankciókat is kiszabnak. De ettől függetlenül ijesztő. Pontosan miért nem tudnak a számítógépek jobban értelmezni a törvényeket, mint az emberek? Ha a jogértelmezés csupán egyfajta számítás - ahogyan azt legalább egy nagy tudású és megfontolt futurista felvetette¹⁶⁶ -, akkor a számítógépek hamarosan valószínűleg jobbak lesznek benne, mint mi. A számítógépek jól számolnak. Erre tervezték őket. Miért ne tehetnék meg helyettünk? Miért ne hagyhatnánk, hogy a jog emberi elme nélkül működjön?

Ez a cikk egy lehetséges okot javasol arra, hogy ne tegyünk ezt. A jogértelmezés megkövetelheti a végső célok kiválasztását - és a kiválasztásuk során az érdemi erkölcsi ítélet gyakorlását - legalábbis bizonyos körülmények között. A számítógépek pedig nem biztos, hogy képesek kiválasztani a végső célokat, még akkor sem, ha

¹⁶⁴ Érdekes a kapcsolat e megközelítés és a jogi pozitívizmus hagyományos meghatározásai között. Tekintsük a társadalmi tények (vagy társadalmi) tézist, a jogi pozitívizmus meghatározásának gyakori mechanizmusát. Davis, Legality, Morality, Duality, supra note , 61-62. o. A társadalmi tények tézisének egyik megfogalmazása az, hogy a jog tartalma végső soron csak a társadalmi tényektől függ, az erkölcsi tényektől nem. Id. A jogi pozitívizmus tehát úgy értelmezhető, mint a társadalmi tények tézisének helyeslése. Megjegyzendő azonban, hogy az erkölcsről szóló egyes állítások pusztán társadalmi tényként kezelik azt. Egy pusztán leíró jellegű állítás arról, hogy egy társadalom - vagy egy társadalom tagja - valójában mit hisz az erkölcsről, társadalmi tény, nem pedig erkölcsi tény. Leíró vagy előrejelző állítást tehetünk az erkölcsről anélkül, hogy állást foglalnánk az erkölcs tényleges tartalmáról. Csak az érdemi erkölcsi ítéletekhez lehet többre szükség, mint a társadalmi tényekre vonatkozó ítéletekhez.

¹⁶⁵ nélkül
https://en.wikipedia.org/wiki/The_Terminator

¹⁶⁶ TEGMARK, LIFE supra3.0, note , at (1052017).

hamarosan túlszárnyalják az embereket a végső célok elérésében, amelyek elérésével megbízzák őket.

A mai napig, bármennyire is lenyűgözőek voltak a technológiai fejlesztések, úgy tűnik, hogy a leírás és az előrejelzés - az előírt célok szolgálatában álló minták azonosítása - mentén haladnak, de nem a követésre érdemes célok azonosítása mentén. Ahhoz, hogy a számítógépek érdemi erkölcsi ítéleteket hozhassanak, a számítógépek által elvégezhető elemzések *típusának* megváltoztatására lehet szükség - nem csak az általuk elvégezhető feladatok nehézségi fokának megváltoztatására.

Ha ez igaz, akkor egy régóta tartó jogtudományi vita megoldása hamarosan nagy gyakorlati jelentőséggel bírhat. A vita az erkölcs szerepéről szól, amikor megmondjuk, mi a jog. A jogi dualizmus e vita lehetséges megoldását javasolja. Eszerint a természetjog a legjobb magyarázatot adja a jog természetéről, amikor a jogértelmező erkölcsi útmutatást keres a jogban, és a jogi pozitivizmus a legjobb magyarázatot adja, amikor a jogértelmező csupán a jog leírására vagy annak megjósolására törekszik, hogy mások hogyan fogják azt értelmezni. Ha a jogi dualizmusnak igaza van - kissé nyersen fogalmazva -, akkor ez lehetővé teheti számunkra, hogy meghatározzuk a mesterséges intelligencia jogértelmezésben játszható szerepének külső határát. Kizárólag az emberek lehetnek képesek értelmezni a jogot, amikor az arról tájékoztat, hogy erkölcsileg mit kellene és mit nem kellene tennünk. A jogtudomány tehát hozzájárulhat ahhoz, hogy a társadalom hogyan viszonyuljon a technológiához. Azt sugallhatja, hogy az embereknek szerepet kell játszaniuk a vonatkozó jogszabályok értékelésében, például az önvezető autók programozásában.

A technológia a jogtudomány megértését is előmozdíthatja. Ha arra összpontosítunk, hogy a számítógépek bíróként szolgálnak, az segíthet élesíteni intuíciónkat azzal kapcsolatban, hogy milyen típusú ítéletekre van szükség a jogértelmezéshez, és hogy milyen környezetben van ezekre szükség. Az, hogy a számítógépek hamarosan jobbak lehetnek az embereknél a jogértelmezés tisztán leíró és előrejelző (vagy meggyőző) aspektusaiban, néhány olvasónak a gondolkodás során hihető előrejelzésnek tűnhet, még ha megdöbbentő és nyugtalanító is. Mások talán szkeptikusabbak lesznek azzal kapcsolatban, hogy mire lesz képes a mesterséges intelligencia. Akár jóslat, akár gondolatkísérlet, értékes felismerésekre tehetünk szert, ha megvizsgáljuk azokat az intuíciókat, amelyek miatt ellenállunk annak a gondolatnak, hogy a Herkules - egy számítógépes program, amely tisztán leíró és előrejelző jogi ítéleteket tud helyettünk hozni, és ezt jobban teszi, mint mi - ügyvédként vagy bíróként szolgáljon. Ez a gyakorlat segíthet megtanulni valamit a jog természetéről. Miközben mérlegeljük, hogy milyen típusú ítéleteket hozhat az ember, és milyeneket nem hozhat a számítógép, talán felülvizsgáljuk, hogyan különböztetjük meg a jogi pozitivizmust a természetjogtól, vagy úgy dönthetünk, hogy más módon változtatjuk meg a jogtudományi álláspontok felosztását, legalábbis bizonyos célokra. Talán új okot találunk arra is, hogy a jogi dualizmust mint a jogtudósokat az elmúlt legalább fél évszázadban foglalkoztató vita megoldását elfogadjuk: az erkölcs szerepét a jogértelmezésben. Legalább ilyen szempontból a technológiai fejlődés a javunkra válhat.

LAW WITHOUT MIND © Joshua Paul Davis - Minden Jog Fenntartva
VÁLASZTERVEZET - Kérjük, ne idézze vagy terjessze engedély
nélkül.