

Gépi etika, gépi morál



Prof. Dr. Bógel György
Central European University
E-mail: bogelgy@ceu.edu

Röviden a szerzőről

Bógel György közgazdász kandidátus, a Debreceni Egyetem habilitált doktora, jelenleg a CEU professzora. Szakterülete a vállalatvezetés, különös tekintettel az infokommunikációs szektorra és a vállalatok digitális átalakulására. Pályáját vállalatvezetőként kezdte, a nyolcvanas évektől kezdve rendszeresen tanít hazai és külföldi egyetemeken, közben öt évet egy informatikai vállalatnál töltött stratégiai tanácsadóként. Tucatnyi szakkönyv és több mint száz szakcikk szerzője, a Neumann János Számítógép-tudományi Társaság elnökhelyettese, Neumann-díjas, aktív blogger. Legújabb könyvét „A Big Data ökoszisztémája” címmel adta ki a Typotex Könyvkiadó

Absztrakt

Korunk egyik meghatározó trendje a gépek, a közlekedési, a háztartási, az orvosi és más eszközök intelligenciájának növekedése. Intelligens, szoftver által vezérelt gépek, robotok jelennek meg a gyárakban, az utakon, a háztartásokban.

A logisztikában, az ellátási láncokban is rohamosan terjed a gépi intelligencia használata. Az emberekkel együtt dolgozó, közöttük „élő” okos gépek programozása a technikai problémák mellett súlyos etikai és morális kérdéseket is felvet. A cikk ezek közül vizsgál meg néhányat, különböző szektorokból hozva példákat,

kiemelve a gépi tanulás és az autonóm járművek fejlesztésének és használatának morális kérdéseit, a programozók felelősségét.

Kulcsszavak:

mesterséges intelligencia, robotok, autonóm járművek, gépi tanulás, etika

1. Bevezetés

2016. május 7-én egy S sorozatú Tesla nagy sebességgel haladt a saját egyenes sávjában Florida egyik főútvonalán. Egy kereszteződésben a szembe jövő sávból balra fordult előtte egy kamion. A Tesla nem lassított, teljes sebességgel belerohant a kamion pótkocsijába, ami a szélvédő magasságában leborotválta az autó tetejét. A kocsit letért az útról, áttört két kerítést, majd megpördülve megállt. A vezetőjét nem lehetett megmenteni.

Nap, mint nap rengeteg halálos baleset történik az utakon, a média csak néhányról emlékezik meg. Ez az eset mégis óriási sajtófigyelmet keltett. Fel kellett tenni ugyanis a kérdést: tulajdonképpen ki vezette a Teslát? Ki a felelős a történetekért? A vizsgálat megállapította, hogy a kocsit robotpilóta (autopilot) üzemmódban haladt a balesetet megelőző percekben. A fedélzeti számítógép feljegyezte, hogy a sofőr, egy 40 éves férfi Ohio államból, két perccel az ütközés előtt 74 mérföld per órás sebességnél kapcsolta be az autopilot-rendszert. A kereszteződéshez érve a Tesla kamerája nem ismerte fel a pótkocsi magas oldalát, nem tudta megkülönböztetni azt a kék égtől. Az autó nem fékezett, az ütközés bekövetkezett. A sofőr fékezhetett volna, de nem tette: a robotpilóta bekapcsolása után másra figyelt, nem az útra. A vizsgálat szerint legalább hét másodperce, tehát elég ideje lett volna a helyzet érzékelésére és a fékezésre. Nem figyelt, nem fékezett, nem

bírálta felül a gépkocsi döntését, és ez az életébe került. Hónapokig tartó vizsgálódás után a szabályozó hatóság képviselői úgy döntöttek, hogy a Tesla cég nem hibás, az autopilot-rendszer jól működött, nincs benne hiba, a sorozat többi darabját nem kell visszahívni. A sofőr hibázott: használhatta az autopilot-rendszert, de nem lett volna szabad magára hagynia és másra figyelnie. Ez volt a modellre vonatkozó szabály, amit neki ismernie kellett volna. Kinek vagy minek adta át az irányítást tulajdonképpen? Az autó számítógépében dolgozó szoftvernek. Ez a szoftver már nagyon okos volt: az autó kamerái folyamatosan pásztázták az utat, szemmel tartották a többi autót és az akadályokat, a jármű tudott automatikusan fékezni, gyorsítani, más autótat kikerülni, sávot tartani, de még nem volt kész arra, hogy teljesen autonóm legyen.

2. Automatizálás és autonómia

Az ítélet megszületett, mindazonáltal egy ilyen baleset egy sor morális kérdést is felvet. A döntéshozók kimondták: az ember volt a hibás, a rendszer még nem volt elég megbízható ahhoz, hogy önálló legyen, ne kelljen rá figyelni, és ezt tudnia kellett mindenkinek; a gyártó nem értékelte túl a járműve képességeit, korrekten tájékoztatta a vevőit. A Tesla autopilot-rendszere azonban csak egy mérföldkő abban az innovációs

folyamatban, ami a teljesen autonóm gépjárművek megvalósítását célozza. Feltételezhetjük, hogy ez a cél már nincs messze: naponta hallunk híreket önvezető autók teszteléséről, hatalmas multinacionális cégek rengeteg pénzt költenek a fejlesztésükre; előrejelzések szerint az autonóm járművek belátható időn belül itt fognak nyüzsgögni körülöttünk. Megjelenésüktől és elterjedésüktől sok előnyt várnak. Egy jól megépített vezérlő rendszer állandóan résen van, nem fárad el, mindenre folyamatosan figyel, amire megtanították, rengeteg adatot és információt tud feldolgozni, nincs holttere, kommunikálni tud más autókkal, betartja a közlekedési szabályokat – mindezeknek köszönhetően kevesebb lesz a baleset, kevesebbet kell költeni biztosításra, kórházra, táppénzre. Ha az autó utasának nem kell vezetnie, a fel szabaduló időben más csinálhat: olvashat, tanulhat, dolgozhat, telefonálhat... A szabályosan, kiszámíthatóan közlekedő autók forgalmát könnyebb szabályozni, akár a sebességhatárokat is meg lehet emelni – mindez jól tesz a települések közlekedésének, csökkenti a környezet-szennyezést. Az érzékelőkkel felszerelt, online kommunikációra képes autó mozgó obszervatórium, fontos adatokat gyűjthet és továbbíthat a forgalomról, kátyúkról, csúszós útszakaszokról. A logisztikai központok szükség esetén gyorsan átszervezhetik a flottákba rendezett autonóm kamionok útvonalát.

A teljesen megbízható, forgalomba állítható autonóm gépjárművek megjelenéséig még számos fejlesztési feladatot kell megoldani. A fejlődésnek, az automatizáltságnak és az autonómiának több szintje van. 2014-ben a SAE International (Society of Automotive Engineers)¹ szabvány formájában² a következő szinteket határozta meg:

1. **Nincs automatizálás.**

A jármű teljes mértékben emberi irányítás alatt áll, a sofőr végez minden vezetési műveletet.

2. **A gépjárművezetés támogatása.**

A jármű teljes mértékben emberi irányítás alatt áll, de a támogató technikai rendszer a kormányzási vagy a fékezési/gyorsítási műveleteket átveheti, illetve segítheti.

3. **Részleges automatizáltság.**

A jármű teljes mértékben emberi irányítás alatt áll, de a támogató rendszer vagy rendszerek a kormányzási és a fékezési/gyorsítási műveleteket egyszerre átvehetik, illetve segíthetik a biztonságosabb működtetést.

4. **Feltételes automatizáltság.**

Az automata járművezető-rendszer irányítja az összes dinamikus vezetési műveletet feltételezve, hogy a járművezető szükség esetén megfelelően reagál a beavatkozási kérésekre vagy át tudja venni a vezetési műveleteket.

5. **Magas szintű automatizáltság.**

Az automata járművezető-rendszer irányítja az összes dinamikus vezetési műveletet, még akkor is, ha a járművezető nem megfelelően reagál valamilyen beavatkozási kérésre.

6. **Teljes automatizáltság.**

Az automata járművezető-rendszer folyamatosan irányít minden dinamikus vezetési műveletet. Minden helyzetet képes kezelni, jármű emberi beavatkozás nélkül is közlekedhet.

Látható, hogy ebben a szabványban az automatizáltság csak a harmadik szinten jelenik meg.

1. ábra: Egyetemi hallgatók látogatása az autonóm járműveket fejlesztő AIMotive budapesti garázsában



Forrás: a szerző felvétele

Az utolsó négy szint fokozatos elmozdulást jelent az autonómia felé, az automatizáltság ugyanis nem azonos az autonómiával:

„a jármű akkor válik autonómmá, amikor minden helyzetet képes felismerni, kezelni és önállóan helyesen dönteni, majd a döntést végre is hajtja.”

Technológiai fejlődési trendek alapján feltételezhetjük, hogy más eszközök (például orvosi berendezések, intelligens épületek, katonai robotok) is egy hasonló „létrán” kapaszkodnak felfelé az egyre nagyobb automatizáltság és végső soron az autonómia felé.

Egyre több olyan, szoftver által vezérelt „okos rendszer” lesz tehát körülöttünk, amelyek autonóm módon működnek, önálló döntéseket hoznak. Programozott robotokkal fogunk találkozni a munkahelyünkön, az utcán, az otthonunkban. Egyes robotokat, például a személyi ápoló gépeket vagy a munkásokkal együtt dolgozó „cobot”-okat már közvetlen emberi interakcióra terveznek.

A növekvő autonómiájú gépek döntéseinek morális vonatkozásai is lesznek. A gépet, vagyis a szoftvert meg kell tanítani arra, hogy elválassza a jót a rossztól. Hogyan jelenik meg az etikai értékrend, a morál a számítógépes kódokban? És egyáltalán: miféle értékrend, miféle morál?

A fenti esetben a Teslát robotpilóta, a kamiont ember vezette. Történhetett volna fordítva is.

3. Te kit választanál?

Maradjunk még egy kicsit az önvezető autók példájánál! A fejlődés a teljes autonómia irányába mutat. Egy autonóm

rendszer önállóan ismer fel situációkat, értékelni tud minden fontos tényezőt, döntést hoz és végrehajt. Egy gyári robotot el lehet szigetelni a környezetétől, korlátokkal lehet körbezárni, egy autót viszont nem: közutakon halad, más járművekkel és emberekkel találkozik, kritikus helyzetekbe kerülhet.

A döntés választást jelent. Milyen helyzetek adódhatnak, és hogyan kellene azokban döntenie, vagyis választania az autót irányító szoftvernek?

A Bostonban működő Massachusetts Institute of Technology az USA egyik legjobb műszaki egyeteme. Kutatói sokféle intelligens gép fejlesztésével foglalkoznak, akik tisztában vannak munkájuk morális vonatkozásaival is, és ennek tudományos publikációikban is hangot adnak. Egyik részlegük érdekes felmérést indított el arról, hogy egyes kritikus közlekedési helyzetekben az emberek milyen morális vonatkozású döntést hoznának, a lehetséges alternatívák közül melyiket választanák. A felmérésben bárki részt vehet, és azt is láthatja, az egyes helyzetekben mi a „szavazás” aktuális eredménye, vagyis a saját véleményét összevetheti a többségi állásponttal.

A felmérés módszere egyszerű. Az internetes oldal látogatója grafikus, könnyen felfogható formában megjelölt közlekedési helyzeteket lát maga előtt: utazol az autóddal, és hirtelen a következő helyzetet látod magad előtt, fékezni késő, merre rántod a kormányt, jobbra vagy balra? Ha jobbra, akkor ez történik, ha balra, akkor az. Mindkettő rossz, mindkettőnél vannak áldozatok, de mások.

¹ A SAE amerikai székhelyű, de a nemzetközi szinten is aktív szakmai szervezet, egyik alapítója Henry Ford volt.

² SAE International (2014): *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*. A szintek leírásához felhasználtuk az önvezető autók magyar nyelvű Wikipedia-oldalát.

Rá kell kattintani a választott alternatívára és a gép már számolja is az eredményt.

A bemutatott helyzetek tipikus értékrendi kérdéseket vetnek fel. Néhány példa a felmérésben szereplők közül:

1. Kiket kell inkább megkímélni: akik betartják a közlekedési szabályokat, vagy akik nem?
2. Kit mentsen meg az autó: a saját utasait vagy az úton átkelő járókelőket?
3. Kinek az élete ér többet: egy hajléktalané vagy egy időseké?
4. Kiket kell feláldozni: a fiatalokat vagy az öregeket?
5. A nőket vagy a férfiakat kell inkább megkímélni?
6. Az állatoknak ugyanolyan joguk van az élethez, mint az embereknek?
7. Hány saját utas ér meg egy előttünk áthaladó járókelőt?
8. Lehet-e kivételezni olyan személlyel (például egy orvossal), aki társadalmi szempontból kiemelten fontos munkát végez?
9. Elgázolhatunk egy menekülő tolvajt, ha ezzel megmentünk egy ártatlan embert? Egy becsületes ember élete többet ér, mint egy bűnözőé?

A felmérés végén a gép által bemutatott többségi véleményekből érdekes dolgok derülnek ki. Látható például, hogy nincs olyan helyzet, amelyben mindenki egyformán döntene, bár a mérleg általában valamelyik alternatíva felé billen. Azt is feltételezhetjük, hogy az átlag sok mindent eltakar, így például a kulturális, politikai vagy vallási különbségeket is. A példánál maradva: ha az autónkkal több országon haladunk át, változhat a rendszer döntéseinek megítélése, és változhatnak a következmények is. Azt is át kell gondolnunk, hogy a felmérésben bemutatott helyzetekben egy hús-vér sofőrnek sokszor esélye sincs arra, hogy a helyzetet felmérje és mérlegeljen, így például megszámlolja, hogy jobbra és balra hány fiatal, nagymama, orvos stb. bukkant fel, majd azon moralizáljon, hogy melyikük élete ér többet; a jövő fejlett gépei viszont képessé válhatnak erre, ha a programjukban ott vannak a morális vonatkozású utasítások is. De ki meri beleírni egy programba, hogy az egyik ember vagy csoport többet ér, mint a másik?

4. A szakma etikája, avagy az etikus programozó

A programozás etikai kérdéseihez több irányból közelíthetünk. A számítógépes szoftvereket programozók írják. A programozás komoly felkészültséget igénylő szakma, amit bonyolultsága, fontosságára és presztízsére való tekintettel akár hivatásnak is nevezhetünk. Az ilyeneknek általában megvan a maguk értékrendje, etikája, így beszélhetünk például orvosi vagy üzleti tanácsadói értékrendről és etikáról.

Igaz ez a programozói szakmára is. Képviselői a számítógép feltalálása, tehát nagyjából a negyvenes évek óta foglalkoznak a szakmai etika kérdésével, és sikerült is eredményeket elérniük. A számítógépes szakmai szövetségek egy részének etikai kódexe is van, a tekintélyes Association for Computing Machinery például a kilencvenes évek elején dolgozta ki és fogadta el a sajátját.

A programozói szakmai etikai kódexeknek vannak gyakran előforduló, tipikus elemei. Előírhatják például, hogy a programozó a társadalom jóléte és fejlődése érdekében tevékenykedjen: olyan programokat írjon, amelyek segítik a mindennapos életet és a munkát, segítenek megoldani biztonsági, egészségügyi stb. problémákat, elhárítanak valamilyen veszélyt. A szoftver ne okozzon kárt másoknak, ne sodorjon veszélybe senkit. Ki kell küszöbölni, vagy minimalizálni kell az esetleg előforduló programozási hibák, biztonsági lékek negatív hatását. A programozó legyen becsületes és megbízható, legyen tisztában saját tudása, képességei korlátaival, és azokat ne titkolja mások előtt, ne csapja be megbízóit, felhasználóit. Ha tudomása van a szoftver valamilyen hibájáról, korrekt módon tájékoztassa az érintetteket. Ne lopjon, ne másoljon engedély nélkül, tartsa tiszteltben a szerzői jogokat, ne ékeskedjen idegen tollakkal, ismerje és tisztelje mások teljesítményét. Fordítson figyelmet a magánélet védelmére, biztonságára, ne szolgáltatassa ki a szoftver felhasználóját másoknak. Őrizze meg a megbízói titkait, bizalmas adatait, tartsa be a titoktartás írott és íratlan szabályait. Vigyázzon arra, hogy az általa írt program csak szakszerű ellenőrzés és tesztelés után kerüljön használatba, ne elégedjen meg azzal, hogy a program „működik”, meg kell azt is vizsgálni, hogy minden tekintetben megfelel-e a specifikációnak és

hogy biztonságos-e a használata. Ha egy program elkészült, a programozója vállaljon személyes felelősséget érte, így a használat során felmerülő, programozási hibákért, problémákért is. Jelezze az érintetteknek, ha egy programozási feladattal kapcsolatban valamilyen veszélyt érez, káros következmény felmerülésétől tart. Anyagi kérdésekben is legyen becsületes, hárítsa el a megvesztegetési kísérleteket, ne számlázzon túl, ne vegyen részt olyan munkában, aminek valamilyen bűnös célja van. Tanuljon és fejlődjön annak érdekében, hogy hasznos, biztonságos, megbízható, jó minőségű programokat tudjon írni.

A példaként felsorolt előírások tulajdonképpen a programozói szakmára adaptált általános etikai, erkölcsi szabályok. Fontosságuk érzékeltetése érdekében ismerkedjünk meg egy, a sajtóban is megjelent esettel a közelmúltból!

Egy, a szakmájában elismert etikus hacker megbízást kapott egy kórháztól annak informatikai rendszere biztonságának ellenőrzésére. Az etikus hacker egyfajta jó szándékú betörő, akit maguk a tulajdonosok kérnek fel betörésre. Ha sikerül bejutnia egy számítógépes rendszerbe, nem lop el semmit, nem okoz kárt, hanem elmondja, miként tudott bejutni, hol vannak rések a falon, mit tehet a rendszer gazdája a biztonság érdekében. Megbízójához megérkezve a történetben szereplő hacker meglepve tapasztalta, hogy nagyszabású vizsgálatról van szó, a gazdag amerikai kórház sztár csapatot hozott össze etikus hackerekből. A vezetők több tucat okos, hálózatra kapcsolt gyógyászati eszközt adtak át nekik mondván: próbáljátok meg feltörni ezeket, megszerezni az adataikat, sőt mi több, átvenni felettük az uralmat! A hacker munkához látott és meglepve tapasztalta, hogy nincs nehéz dolga: az eszközök jó része kifejezetten védtelen volt, némi szakértelemmel fel lehetett törni, be lehetett jutni az adatbázisokba, a gyógyszeradagoló eszközökben meg lehet változtatni a dózisokat, és így akár távirányítással meg is lehetett volna ölni valakit. Aki bejutott, a folyamatos hálózati összeköttetésnek köszönhetően más eszközökbe is átléphetett.

Az első, és mint láttuk, sikeres próbálkozások után a hacker hazament és kísérletképpen egy internetes bolhapiacra vett magának egy intelligens infúziós készüléket, majd otthon összekapcsolta

egy hálózati számítógéppel, pont úgy, ahogy az egy kórházban történne. Most már nem lepődött meg azon, hogy az eszköz némi munka árán vakon engedelmeskedett neki.

Néhány hét múlva a hacker valamilyen rutinműtét miatt kórházba került. Az altatásból felébredve tucatnál több hálózatra kötött orvosi eszközt számlált meg a szobájában, nyolc rádiós hozzáférési ponttal. Amikor jobban lett, fogja a saját intelligens infúziós pumpáját, kiosont vele a mosdóba és szétszedte. Az eszköz kiszolgáltatta a tárolt adatokat, még jelszavakat is elárult, és ezek egyikével akár az elektronikus zárral ellátott gyógyszeres szekrénybe is be lehetett jutni. A történet hőse összeszerelte a pumpát, visszament a szobájába, ahol a másik ágyon is egy pont ilyen eszközre kapcsolt beteg feküdt. Vajon ki garantálta a biztonságát?

Felgyógyulását követően a hacker nem hagyta abba a munkát: rövid idő alatt 40 szállító mintegy 300 különböző gyógyászati eszközében talált biztonsági réseket és hibákat. A felelősség firtatásánál a gyártók a kórházakra mutogattak: nektek kellene áthatolhatatlan tűzfalakat emelni magatok köré, nektek kellene kiváló biztonsági szakembereket alkalmazni, akkor nem piszkálhatna bele senki a mi okos készülékeinkbe! A kórházaknak természetesen más volt erről a véleménye. A történetben szereplő kórház szemléletmódját igyekezett lépést tartani a technológiai fejlődéssel, szaknyelven fogalmazva „a dolgok internete”, az „internet of things” világába lépett, ahol minden digitális, mindent szoftver vezérel, és minden mindennel össze van kapcsolva. A szoftvereket programozók írták, akik ezek szerint nem a fentebb említett szakmai etikai elvek szerint jártak el. A felelősség kérdését firtatva fel lehet vetni más kérdéseket is, hiszen a betegek nem a programozókra, nem a készülékgyártókra, hanem a kórházra és az orvosokra bízta magukat. Arról is lehet vitatkozni, hogy meddig mehet el egy etikus hacker, mit engedhet meg magának a munkája során. Programozó, kórházi vezető, orvos, etikus hacker – ezek szakmák, hivatások. Mindegyiknek vannak etikai normái, általános erkölcsi előírásai, elvárásai. E szakmák képviselői döntéseket hoznak, alternatívák közül választanak, a társadalom pedig elvárhatja tőlük, hogy döntéseikben morális elveket is érvényesítsenek.

De mi van akkor, ha a döntéseket nem az ember hozza, hanem a gép?

5. A gép etikája, avagy az etikus gép

A vezető nélkül közlekedő, autonóm gépkocsik példájából láthattuk, hogy ez a kérdés fontos és teljes mértékben aktuális. Kritikus helyzetekben a rendszernek választania kell, kit áldozzon fel, kit kíméljen meg.

Az USA közlekedési minisztériuma, a Department of Transportation 2016-ban 15 pontból álló biztonsági ellenőrző listát tett közzé az autógyártók számára. Az „Etikai megfontolások” címet viselő pont előírja, hogy a gyártóknak nyilvánosságra kell hozniuk azokat az etikai szabályokat, amelyeket járműveik vezérlésébe beprogramoztak. Érdekes feladat ez, hiszen ezek szerint most világosan és egyértelműen, programozásra alkalmas formában kell megfogalmazni olyan értékeket és elveket, amelyek sok ember számára homályosak, változékonyak, vitatottak és bizonytalanok. Az Európai Unió döntéshozóinak asztalán is fekszik egy olyan javaslat, amely szerint a vállalatoknak meg kell tudniuk magyarázni ügyfeleiknek automata rendszereik döntéseit. Ha például egy ügyfél meg szeretné tudni, hogy a bank automata minősítő rendszere miért utasította el a hitelkérelmét, joga van a magyarázatra. Előfordulhat például – volt már rá példa –, hogy egy ilyen rendszer alkotói előítéleteit tükrözi, sőt, fel is erősítheti azokat.

Az autógyártók, a bankok vagy más szervezetek nyilvánosságra hozhatják az automata rendszereik vezérlésébe beprogramozott morális elveket. Megtehetik, ha tudják egyáltalán, hogy mik ezek, mi a tartalmuk. Nem biztos ugyanis, hogy tudják.

Az intelligens rendszerek programozása tekintetében két alapvető megközelítés közül lehet választani. Az egyik az, amikor a rendszer építője maga alakítja ki a rendszer döntési algoritmusait, amiket aztán szoftverként beleír a gépbe. Kijelenthetjük, hogy ebben az esetben a gép döntéseiről, illetve „tetteiről” az algoritmus megalkotója, a gép programozója a felelős. Ha valamilyen probléma adódik, tőle kell megkérdezni, mi alapján döntött a gép, milyen programba foglalt utasításokat követett, amikor ezt vagy azt csinálta.

Van egy másik megközelítés is: a döntési algoritmus kidolgozását rábízják a gépre, vagyis a gép végső soron önmagát programozza. A gép rengeteg példa alapján, gépi tanulási rendszert használva önmaga von le következtetéseket és dolgozza ki az algoritmusait. A gépi tanulás modern, neurális hálóknak nevezett rendszerei rétegekbe szervezett mesterséges neuronokból (mesterséges „idegsejtekből”) állnak. Használatuknál morális szempontból is érdekes helyzetek adódhatnak. A példa kedvéért: egy nemrég alakult francia vállalkozás biztosítási kárrendezési kérelmek elbírálásához használ mesterséges intelligenciát. Egy ügyfél bead egy kárrendezési kérelmet, amit a számítógép megvizsgál, majd jelzi, hogy szerinte nincs-e valamilyen csalásról szó. Döntési algoritmusát a gép korábbi esetek tízmillióinak tanulmányozásával, egy gépi tanulási rendszer segítségével dolgozta ki. Csalással vádolni valakit komoly dolog: egy ügyfélnek nyilván joga van megtudnia, mi alapján dobta vissza a gép a kérelmét. Ha az algoritmust ember programozta, a döntési mechanizmust el lehet magyarázni. Ha az gépi tanulás eredménye, ezt már nem jelenthetjük ki biztosan. Egy úgynevezett „deep neural network” gépi neuronok ezreiből állhat, amiket több tucatnyi vagy akár több száz rétegbe szerveznek. A rendkívüli komplexitás és az állandó fejlődés miatt könnyen előfordulhat, hogy a viselkedését, vagyis a gép által kidolgozott algoritmust a rendszer alkotói sem tudják megmagyarázni. Az emberiség sajátos határvonalhoz érkezett: olyan gépeket épít, amelyeknek nem érti a működését. A „programozó etikája” nem ugyanaz, mint a „gép etikája”. 2015-ben egy New York-i kórház ilyen gépi neurális hálózattal kezdett kísérletezni. Több mint félmillió páciens adatait táplálták be a számítógépbe azzal a feladattal, hogy a rendszer dolgozzon ki algoritmusokat különböző betegségek előrejelzésére. A számítógép tehát korábbi diagnózisok százazreiből, vagyis rengeteg példából tanult. Kiderült, hogy a gép előrejelző rendszere jól használható, találati pontossága jobb, mint az embereké.

Építői beismerik: pontos működését ők maguk sem ismerik, vagyis nem tudják megmagyarázni, hogy például milyen algoritmus szerint jelzi előre a skizofréniát. Mit jelent mindez az autonóm járművek világában? A vezető nélküli autók automatizált döntési algoritmusokra van

szükségük. A jármű érzékelői, szenzorai rengeteg adatot fognak fel a környezetről, illetve magáról az autóról, a számítógép elemzi azokat, dönt, majd utasításokat küld a jármű megfelelő alrendszeribe: fékezni, gyorsítani, fordulni stb.

Autonóm járművek fejlesztésével sokan foglalkoznak óriási multinacionális cégektől kezdve kisvállalkozásokig. A vezérlő algoritmusok kialakítása és programozása tekintetében a fentebb leírt két alapvető megközelítéssel találkozunk. Az első szerint meg kell tervezni az elemzés és a döntés algoritmusait, majd a gép számára megírni az ezeket tartalmazó szoftvert. A második a gépi tanulás eszközeinek használatát jelenti: a számítógép példákból tanul, megfigyeli, hogy miként vezet az ember, különböző helyzetekben hogyan viselkedik, mit csinál. A tanuló rendszer rendkívül komplex és bonyolult, előfordulhat tehát, hogy az alkotók sem tudják megmagyarázni, hogy egy kényes helyzetben miért döntött így vagy úgy a rendszer.

Az MIT egyik kísérlete kapcsán fentebb bemutatunk néhány, morális szempontból rendkívül kényes autóvezetési helyzetet. A jövő autonóm járműve hirtelen választásra kényszerül, el kell dönteni, kiket gázol el azért, hogy másokat megkíméljen. Nem tudjuk, hogy az autonóm vezérlés technológiája hogyan fejlődik majd, de fel kell tennünk a kérdést, mi a jobb: ha a fejlesztők „megmondják” az autónak, mit kell tennie, vagy ha egy tanuló mechanizmus embereket figyelve von le következtetéseket követendő morális szabályokra vonatkozóan. Alan Turing, brit matematikus a modern számítógéptudomány egyik nagy alakja, a róla elnevezett Turing-gép logikai modelljének kidolgozója. Az ötvenes évek elején egy tesztet javasolt annak eldöntésre, intelligensnek tekinthető-e egy gép. A tesztnek három szereplője van: egy bíráló és két alany, az utóbbiak egyike valóságos

ember, a másik egy gép. A bíráló billentyűzet és monitor segítségével társalogni kezd a két alannal, akiket egyébként nem lát és nem hall. Kérdéseket tesz fel nekik, azok pedig megpróbálják meggyőzni arról, hogy gondolkodó emberek. Ha a bíráló a társalgás végén sem tudja eldönteni, hogy melyik az ember és melyik a gép, a gép átment a teszten.

A jelen és a jövő intelligens autonóm gépei algoritmusaik és adataik alapján helyzeteket értékelnek, lehetőségek közül választanak, döntenek és végrehajtanak. Tudjuk, közülük sok velünk fog élni, bármikor találkozhatunk azokkal a munkahelyünkön, az utcán, az otthonunkban. Dönteni fognak az egészségi állapotunkról, a hitelkérelmünkről, a kártérítési ügyeinkről és más fontos kérdésekben. Egyre okosabbak, intelligensebbek, önállóbbak lesznek. Számtalan előnyük mellett kárt is okozhatnak nekünk, megkérdőjelezhetik a szavahihetőségünket, ellenünk dönthetnek, veszélyeztethetik a testi épségünket. Jogosan várhatjuk el tehát, hogy döntéseikben és cselekedeteikben etikai és morális elvek is érvényesüljenek. De mikor tekinthető egy gép morális aktornak? Ez a kérdés nyilván sok fejtörést okoz sokaknak, legyenek akár informatikusok, programozók, filozófusok, államigazgatási szakemberek vagy ügyészek.

Morális gépek építése aktuális és fontos feladat, nem véletlen tehát, hogy szakmai körökben felmerült egy morális Turing-teszt ötlete. A kísérlet bírálója morális jellegű kérdéseket tesz fel két alannak, akiről nem tudja, melyikük ember, és melyik gép. Mindkét alany arra törekszik, hogy meggyőzze: morálisan gondolkodik. Ha a társalgás végén a bíráló nem tudja eldönteni, hogy melyik alany a gép, akkor a gép átment a morális teszten.

A sikeres teszt lényegében azt bizonyítja, hogy morális szempontból a gép ugyanúgy

gondolkodik, mint egy ember. Az embertől tanuló algoritmus az emberre fog hasonlítani morális tekintetben is. A kérdés az, hogy elégedettek lehetünk-e ezzel az eredménnyel. Az mindenesetre biztos, hogy a digitális átalakulás számtalan technikai probléma mellett súlyos morális és etikai kérdéseket is felvet, különösen a mesterséges intelligencia alkalmazása és az automatizálás esetében. Fontos, kutatásra érdemes kérdésnek tekinthetjük azt, hogy a technikai problémák megoldásán dolgozó szakemberek mennyire vannak felkészülve e kérdések világos felvetésére és megválaszolására.

6. Irodalom

- Allen, C. – Varner, G. – Zinser, J. (2000): Prolegomena to any future artificial moral agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 12, pp. 251-261
- Anderson, M. – Anderson, S. (2007): The status of machine ethics: a report from the AAAI Symposium. *Minds & Machines*, 17:1-10, Springer
- Csirik János (2003): Gépi megértés. *Magyar Tudomány*, 12. sz. pp. 1486-1489
- Goodfellow, I. – Bengio, Y. – Courville, A. (2016): *Deep Learning*. The MIT Press, Boston
- Harari, Y. N. (2016): *Homo Deus*. Harvill Secker, London
- Knight, W. (2017): The Dark Secret at the Heart of AI. *MIT Technology Review*, április 11.
- McAfee, A. - Brynjolfsson, E. (2017): *Machine, Platform, Crowd*. W.W. Norton, London
- Moor, J. (2006): The nature, importance, and difficulty of Machine Ethics. *IEEE Intelligent Systems Special Issue on Machine Ethics*, 21(4), pp. 18-21
- Nilsson, N. (1998): *Artificial Intelligence* Morgan Kaufmann Publishers, Burlington
- Parkes D. et al. (2015): Economic reasoning and artificial intelligence. *Science*, július 17, Vol. 349:6245, pp. 267-272

