

PINTÉR MELINDA⁵

A MESTERSÉGES INTELLIGENCIA ETIKAI ASPEKTUSAI

Absztrakt

Jelen tanulmány a mesterséges intelligenciával (MI) kapcsolatos felvetések és problémák csak egy kis részét, az etikai dilemmákat vizsgálja. A tanulmány elsőként egy nagyon rövid fogalmi áttekintést ad a téma vizsgálatához elengedhetetlenül szükséges fogalmakról, majd röviden kitér annak a kérdésnek a megválaszolására, hogy miért fontos egyáltalán beszélni az MI etikai vonatkozásairól. Ezt követően azt vizsgálja, hogy hogyan lehet jól és rosszul használni az MI-t, mi az MI készítőinek felelőssége mindebben; végezetül pedig nagyon röviden – és nagyon optimistán – bemutatja azt, hogy vajon milyennek kellene lennie az etikus MI-nek.

Kulcsszavak: mesterséges intelligencia, MI, digitális etika

ETHICAL ASPECTS OF ARTIFICIAL INTELLIGENCE

Abstract

Current study examines only a small part of the concerns and problems connected to artificial intelligence (AI), namely the ethical dilemmas. The study will firstly provide a very brief conceptual overview of the concepts essential to the study of the topic, and then briefly address the question of why it is important to speak about the ethical implications of AI at all. Then it examines how can AI be used well and badly, what is the responsibility of the creators of AI in all of this, and finally, very briefly - and very optimistically - presents what an ethical AI should look like.

Keywords: artificial intelligence, AI, digital ethics

⁵ Pázmány Péter Katolikus Egyetem, Bölcsész- és Társadalomtudományi Kar, Nemzetközi és Politikatudományi Intézet, tanárségéd

A Világ gazdasági Fórum szerint a mesterséges intelligenciával (MI) kapcsolatban számos olyan etikai kérdés merülhet fel, amelyre az emberiségnek záros határidőn belül válaszolnia kell (*Bossmann 2016*). Például mi történik, ha már nem lesznek (emberek által) betöltendő állások? Kezelhetjük-e személyként az MI-t, és ha igen, milyen keretek között? Vajon hogyan lehet megalkotni az előítéletmentes MI-t? A számos érintett területben egyetlen dolog a közös: mindegyik egy-egy etikai dilemmát állít a középpontba, jelen tanulmány pedig ehhez kapcsolódóan vet fel néhány konkrét kérdést.

1. Mi a véleményünk az MI-ről?

Azt gondolhatnánk, hogy az „utca embere” semmit nem tud az MI-ről és valójában nem is érdekli – ez azonban egyáltalán nem így van. Érdekes ezért áttekinteni néhány, magyar és amerikai mintán felvett kutatás eredményét arra vonatkozóan, hogy valójában mi a véleményük az MI-ről az embereknek és hogyan viszonyulnak hozzá.

Az amerikaiak 50 százaléka optimistának és tájékozottnak gondolja magát az MI-vel kapcsolatban, míg a „maradék” informálatlannak érzi magát és félelmet kelt benne az MI (*Press 2019*). Az amerikaiak 67 százaléka úgy gondolja, hogy az önvezető autók biztonságosabbak lesznek, mint az ember által üzemeltetettek (*Press 2019*). Az amerikaiak több mint fele (53 százalék) szerint az MI elfogult lehet a döntéshozatal során, továbbá a válaszadók csupán 12 százaléka gondolta úgy, hogy az MI képes különbséget tenni a jó és a rossz között, 56 százalékuk pedig nem hiszi, hogy lehetséges az erkölcsösen viselkedő gépek kifejlesztése (*PEGA Press Releases 2019*).

A hazai adatok kissé mást mutatnak (*Digital Hungary-E-volution 2019*). A magyarok csupán 19 százaléka mondja azt, hogy elfogadja és pozitívan ítéli meg az MI-t, 67 százalék óvatosan áll hozzá, míg 11 százalék idegenkedik tőle. Azok, akik teljesen elfogadják az MI-t, többségében 45 évnél fiatalabb férfiak, akik jóval tájékozottabbak is technológiai kérdésekben és a felhasználás tekintetében is optimista állításokat fogalmaznak meg. A technológiához óvatosan viszonyulók vagy idegenkedők többségében az alacsony iskolai végzettségűek közül kerülnek ki; ráadásul ők olyan potenciális fenyegetésként látják az MI-t, ami adott esetben a munkájukat is elveheti majd.

Láthatjuk, hogy az emberek nem „foghatatlanok” az MI-vel kapcsolatban: tájékozottak, ismerik a technológiát és véleményük is van róla. Éppen ezért (is) kulcsfontosságú az, hogy ennek a technológiának az etikai kérdéseivel foglalkozzunk.

2. Fogalmi tisztázás

Három fogalmat érdemes bővebben bemutatni és magyarázni, ezek pedig az etika, az MI, és az MI, valamint a gépi tanulás fogalmai közötti különbség.

Etika alatt az egyén vagy csoport viselkedését vagy cselekedeteit szabályozó erkölcsi alapelveket értjük. Más szavakkal: „szabályokat” vagy „döntési utakat”, amelyek segítenek meghatározni, mi a jó vagy a helyes. Ebből kiindulva azt mondhatjuk, hogy a szoftverközpontú technológia etikája egyszerűen a „szabályok” vagy „döntési utak” halmaza, amely a technológia (mondjuk egy algoritmus) „viselkedésének” meghatározására szolgál.

De mi az az MI? Jelen tanulmányban olyan adatokat, modelleket és előrejelzéseket értek alatta, amelyek bizonyos input adatokból meghatározott algoritmusok segítségével egyfajta outputot hoznak létre. Mindezek adják meg az MI etikai aspektusait is, amelyekről később fogok szólni. Az MI tehát technológiákat és algoritmusokat használ annak érdekében, hogy az adatokból származó összefüggéseket és kapcsolatokat nyerjen ki automatikusan. Ezen jelentésének „érthetővé tételéhez” pedig az adatmodellekből és az előzetes tapasztalatokból tanul (*Szöllősi 2019*).

Ezzel pedig el is jutottunk az utolsó magyarázatra szoruló fogalompárunkhoz: MI vagy gépi tanulás? Ezt a két fogalmat sokszor használják szinonimaként és valóban nagyon sok közös van jelentéstartalmukban, de nem ugyanazt jelentik: gondoljunk úgy az MI-re mint az autonóm gépi intelligencia szélesebb körére, a gépi tanulásra pedig úgy, mint az MI építésében jelenleg divatos konkrét tudományos módszerekre. Minden gépi tanulás MI, de nem minden MI gépi tanulás.

3. Miért fontos beszélni az MI etikai vonatkozásáról?

Nap mint nap halljuk, hogy az MI már itt jár a sarkon túl és hamarosan teljesen átveszi az uralmat az életünk felett. De vajon mennyire igaz ez? Mennyire beszélhetünk az MI-ről már most is úgy, mint az életünk részéről? Ha például az önvezető autókat vagy például a valós idejű fordítógépek napi szintű használatának az esélyeit, illetve lehetőségeit nézzük, akkor valóban azt mondhatjuk, hogy ez még a jövő zenéje, és egyhamar nem hagyhatunk fel az autóvezetéssel vagy a nyelvtanulással.

Viszont rengeteg olyan terület van, ahol már most is napi rendszerességgel használjuk lakossági szinten is az MI-t, a különféle algoritmusokat, okos technikákat vagy a gépi tanulás egyes formáit (*Narula 2020*). Ilyen például az email postafiókunk spamszűrője, a felsőoktatási intézmények által használt plágiumkereső, a közösségi média oldalak arcfelismerő rendszerei vagy a chatbotok és az okostelefonok beszédfelismerő szoftverei. És ez csak néhány példa az MI lakossági célú felhasználási módjai közül.

De mire használják az MI-t az államok? A rengeteg felhasználási mód közül álljon itt néhány érdekes példa: arcfelismerésre a határon vagy reptereken, a kétes vízumkérelmezők kiszűrésére (*Hendry 2018*) vagy éppen bűnözői hálózatok feltérképezésére a deep/dark weben (*Grierson 2019*).

Bármennyire is hajlamosak tehát sokan a science fiction szintjén gondolkodni az MI-ről, látható, hogy ez már a jelen. És hogy válaszoljunk ezen egység címében feltett kérdésre: ez az, ami miatt fontos beszélni az MI aspektusairól és az etikus használatáról.

4. Mit csinál az MI?

Az MI-t is lehet jó és rossz célokra használni. Míg jó célokra használva könnyebbé teheti az életünket, rossz kezekbe kerülve óriási károkat okozhat. A kérdés, hogy mit tehetünk az utóbbi megelőzése érdekében, hiszen ahhoz, hogy etikus legyen az MI használata, ez kiemelkedő fontosságú kérdés. Alapvetően három olyan területet érdemes meghatározni, amelyek vizsgálata kulcsfontosságú: egyrészt a biztonság, másrészt a magánszféra, az ellenőrzés és a megfigyelés dilemmája, továbbá az ember-MI interakció kérdése.

Biztonság

Mit tehetünk annak érdekében, hogy az MI ne jelentsen potenciális veszélyt az életünkre? Alapvetően és elsődleges módon természetesen minél szélesebb körben garantálni kell a mesterséges intelligencia biztonságos használatát. Ez jelenti például azt, hogy önvezető autók vagy drónok nem okozhatnak baleseteket, de ugyanez a helyzet például a szoftvekekkel, amelyek nem fizikai formában, de képesek kárt okozni. Mindkét esetben szigorú tesztelésnek kell alávetni ezeket az eszközöket.

Ugyanakkor itt ütközünk először abba a problémába, hogy az, hogy mi a helyes, legális vagy jelen példánkból kiindulva biztonságos, helyzettől függően változik. Az emberek döntései nem egy vákuumban születnek meg, hanem az aktuális helyzettől függenek. Ha például éppen egy hozzátartozónkkal száguldozunk a csúcsforgalomban a kórház felé, akkor hajlamosak vagyunk átértékelni a biztonságos vezetéshez kapcsolódó elképzeléseinket. Viszont hogyan tudnak ezekhez a változó, helyzettől függő biztonságértelmezéshez alkalmazkodni a mesterséges intelligencia alapú eszközök? A mesterséges intelligencia „tesztelése” tehát egy soha véget nem érő folyamat.

Az MI-nek képesnek kell lennie arra, hogy reagáljon a felmerülő új helyzetekre, ezeket összefüggéseikben lássa, és annak megfelelően döntsön, hogy mi a „helyes”. Az MI-nek természetesen alkalmazkodnia kell ezekhez a folyamatosan változó helyzetekhez... Újabb etikai dilemma azonban, hogy kinek, kiknek, milyen csoportoknak az értékeihez, normáihoz, biztonságérzetéhez alkalmazkodjanak ezek a rendszerek. A „helyes” célhoz való igazítás, más néven az érték-igazítás a biztonság kulcsfontosságú témája az MI-ben.

Magánszféra, ellenőrzés és megfigyelés

Szintén lehet jóra és rosszra használni az MI-t akkor, ha az emberek közvetlen megfigyeléséről van szó. Mennyire etikus egyáltalán az, ha megfigyelő rendszereket készítünk, még ha a jó cél érdekében is tesszük mindezt? Tudjuk, hogy a repülőtéren a biztonsági ellenőrzés a saját érdekünket szolgálja, ugyanakkor jogosan érezhetjük kiszolgáltatottnak magunkat, amikor zokniban próbálunk áthaladni az átvilágító kapun, miközben a táskánk tartalmát egy gép vizsgálja át.

A magánélet feladása, illetve ennek kockázata feláldozható-e a biztonságért? Rengeteg adatot szolgáltatunk magunkról. Visszaélés esetén viszont csak jóval később szembesülünk a

károkkal. Kérdés, hogy megéri-e a saját biztonságunk érdekében önként kiadni az adatainkat, betekintést engedni különféle mesterséges intelligenciáknak a magánéletünkbe? Ha ugyanis mindezek rossz kezekbe kerülnek, akkor az beláthatatlan károkat okozhat.

Ember – MI interakció

A jó és rossz használatok utolsó példajaként pedig a mesterséges intelligencia és az ember közötti interakciók területét érdemes megemlíteni. Nagyon könnyű ugyanis a mindennapjainkat az okos asszisztensek és egyéb, az életünket könnyebbé tevő, mesterséges intelligenciát használó eszközök köré szervezni, de ennek természetesen ára van. Könnyen kialakulhat egy MI-től való „függőség”, amikor életünk egyre nagyobb és nagyobb részében támaszkodunk rá - olyan helyzetekben is, amelyekben ez nem lenne indokolt. Tudunk azonban pozitívumokat is említeni.

Ennek egyik legjobb példája a MAPS elnevezésű kísérleti projekt (*Servick 2019*), amely a hazai médiában is nagyobb nyilvánosságot kapott (*Molnár 2019*). Az Egyesült Államokban a baleseteket követően az öngyilkosság a második leggyakoribb halálok a tizenévesek és a fiatal felnőttek körében. Kutatók ezért kifejlesztettek egy olyan megfigyelőrendszert, amelynek segítségével a programban résztvevő 50 önkéntes, potenciálisan veszélyeztetett (azaz akinek korábban már volt öngyilkossági kísérlete) fiatal mobiltelefonjuk segítségével vizsgálható tevékenységét monitorozzák: milyen honlapokat keresnek fel, mit posztolnak a közösségi médiában, milyen zenét hallgatnak, vagy éppen a telefon GPS adatai alapján merre járnak. A projekt célja az, hogy olyan mintázatokat találjanak, amelyek arra utalnak, hogy a projekt résztvevője öngyilkosságot készül elkövetni.

Ehhez viszont totális megfigyelés kell. Itt is felmerül a kérdés, hogy megéri-e mindez a jó cél érdekében? Itt jön képbe az MI készítőinek felelőssége.

5. Az MI készítőinek felelőssége

Az MI készítőinek és/vagy megrendelőinek óriási a felelőssége abban, hogy tényleg etikus mesterséges intelligencia készül-e el. Az MI ugyanis olyan lesz, amilyennek megalkotjuk. Az emberi „tökéletlenség” az ember által megalkotott gépekben is testet ölt(het). Számos kérdés merül fel ebben az etikai dimenzióban is (*Nalini 2019*).

Torzítás vagy pártatlanság?

Vegyük például a kognitív torzítás jelenségét. Ezeket a gondolkodási sémákat az MI kiküszöbölheti – ha mindezek figyelembe vételével történik a megalkotása –, ugyanakkor fel is erősítheti őket, ami torzításhoz, diszkriminációhoz, igazságtalansághoz vezethet. Az elfogult algoritmikus rendszerek tisztességtelen eredményekhez, diszkriminációhoz és igazságtalansághoz vezethetnek. Felmerül a kérdés, hogy ezek inkább a modell működési hibáinak tekinthetők, vagy egy lépéssel korábban, a modell alkotójának – akár szándékos, akár nem szándékos – elfogultságában keresendők?

Hajlamosak vagyunk azt hinni, hogy egy algoritmus majd teljesen pártatlanul fog működni, és nem lesznek rá hatással az emberi gondolkodás sajátosságából fakadó torzítások. Ez azonban nem lehetséges akkor, ha már a bemeneti adatok is rosszak. Garbage-in, Garbage-out: önmagában a technológiával nem tudjuk leküzdeni a társadalmi struktúránkban jelen levő torzításokat és előítéleteket. Valójában még rosszabbá tehetjük őket, és ezt gyakran meg is tesszük.

Felelősségre vonhatóság

Szintén nagy a szerepe az MI megalkotóinak a felelősségre vonhatóság területén. Alapvetően szoftverekről, algoritmusokról, robotokról beszélünk, de ha valamilyen jogszabályi előírás lehetővé teszi, hogy adott esetben ezeket az eszközöket tegyük felelőssé hibákért, okozott sérelmekért, károkért, akkor az emberek ezekbe az eszközökbe vetett bizalma is nagyobb lehet. Az igazságszolgáltatás lehetővé tétele ilyen helyzetekben elősegíti a bizalmat; így hosszú távon kialakul egy olyan bizalom ezekkel a technikákkal szemben, amely lehetővé teszi a minél szélesebb körű felhasználást, ugyanis az emberek már nem fognak félni attól, hogy tisztességtelen vagy diszkriminatív lépések esetén tehetetlenek a „gépekkel” szemben

De mi történik, ha a kár bekövetkezik? Van-e lehetőség a negatív hatások orvoslására? Járhat-e kártérítés? Ha igen, hogyan? Ez még gyerekcipőben jár. Kérdés egyrészt, hogy kit terhel ilyen esetben a felelősség: a gépet vagy az alkotót? Másrészt vannak már olyan szervezetek, amelyek azon dolgoznak, hogy mindez, a felelősségre vonhatóság megvalósulhasson.

Transzparencia, értelmezhetőség és magyarázhatóság

Az utolsó, de talán legfontosabb alkotói felelősség az MI-vel kapcsolatban, amit említenék, az az átláthatóság kérdése. Az MI-vel szembeni bizalomhiány nagy része abból fakad, hogy az

emberek egyáltalán nem ismerik azt, hogy hogyan működnek ezek az algoritmusok vagy eszközök. Nem azért, mert ne értenék meg, hanem a „fekete doboz” technika miatt, azaz azért, mert a vállalatok nem akarják nyílt forráskódúvá tenni a fejlesztéseket, nem akarják elárulni a „titkos szász” összetételét – e mögött pedig természetesen saját gazdasági érdekeik húzódnak meg.

És ismét eljutottunk a kérdéshez: mennyire etikus ez a titkolózás? Mi van akkor, ha az MI károkat okoz? Elegendő csak ezt tudni és cselekedni, vagy értenünk is kell, hogy hogyan hozta a döntést a gép? Az, hogy ezek a fejlesztések közkincként legyenek kezelve, mindenki érdeke lenne. Ez egyúttal növelné például ezen cégek szervezeti átláthatóságát és a bizalmat is a legújabb MI fejlesztések iránt, de ahogyan az előbb is említettem: a gazdasági érdekek sokat nyomnak a latban. Amíg viszont nem teljesen transzparenszek ezek a fejlesztések, számos etikai dilemmával kell szembenéznük továbbra is.

6. Milyen tehát az etikus MI?

És végül, de nem utolsó sorban: mindazon dilemmák, tényezők, illetve kívánalmak alapján, amelyeket jelen tanulmányban említettem, milyennek is kellene lennie az etikus mesterséges intelligenciának (*Fan 2019*)?

Az etikus MI transzparen. Tudjuk, hogy hogyan működik, tudjuk, hogy milyen célt, és hogyan kinek, minek a céljait és érdekeit szolgálja. Az etikus MI igazságos és pártatlan. Objektív, és semmilyen irányba nem részrehajló vagy torzító. Az etikus MI nem árt: soha nem okozhat károkat, baleseteket, nem veszélyeztetheti akár az egyéneket, akár az egész társadalom biztonságát. Az etikus MI felelősséggel rendelkezik, vagy úgy is mondhatnánk, hogy felelősségre vonható. Ha bárkit bármilyen kár ér az MI működése miatt, akkor a jogi felelősség ténye legyen kimondható. Végül pedig az etikus MI olyan, ami tiszteli a magánszférát, és amelyet nem használnak fel olyan tevékenységekre, amely az ehhez való jogot sérti – legalábbis nem úgy, hogy arról a megfigyelt személy ne tudna. Az etikus MI a magánélet védelmét és a védelemhez való jogot egyaránt jelenti.

Irodalomjegyzék

Artificial Intelligence in 2019: Getting Past the Adoption Tipping Point - *A report from Blumberg Capital exploring consumers' comfort level, knowledge of and sentiment towards artificial intelligence.* Blumberg Capital.

<https://www.blumbergcapital.com/cnt/uploads/2019/07/Artificial-Intelligence-in-2019-Blumberg-Capital.pdf> [Letöltve: 2020. február 28.].

BOSSMANN, J. (2016) Top 9 ethical issues in artificial intelligence. *World Economic Forum.*

<https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/>

[Letöltve: 2020. január 9.].

Consumers Failing to Embrace AI Benefits, Says Research. *PEGA Press Releases.*

<https://www.pega.com/about/news/press-releases/consumers-failing-embrace-ai-benefits-says-research> [Letöltve: 2020. február 28.].

FAN, S. (2019) What Does Ethical AI Look Like? Here's What the New Global Consensus Says. *Singularity Hub.* <https://singularityhub.com/2019/09/10/what-does-ethical-ai-look-like-theres-now-a-global-consensus/> [Letöltve: 2020. február 28.].

GRIERSON, J. (2019) Home Office to fund use of AI to help catch dark web paedophiles.

The Guardian. <https://www.theguardian.com/technology/2019/sep/17/home-office-artificial-intelligence-ai-dark-web-child-sexual-exploitation> [Letöltve: 2020. február 28.].

HENDRY, J. (2018) Australia's new visa system could use AI to spot dubious applicants.

iTnews. <https://www.itnews.com.au/news/australias-new-visa-system-could-use-ai-to-spot-dubious-applicants-481148> [Letöltve: 2020. február 28.].

Kiderült, hogy a magyarok mit is gondolnak a mesterséges intelligenciáról. *Digital Hungary - E-volution.* <https://www.digitalhungary.hu/e-volution/Kiderult-hogy-a-magyarok-mit-is-gondolnak-a-mesterseges-intelligenciarol/7872/> [Letöltve: 2020. február 28.].

MOLNÁR Cs. (2019) Felismerhető lehet az öngyilkosság veszélye, de ehhez totális megfigyelés kell. *Index.hu.*

https://index.hu/techtud/2019/10/28/ongyilkossag_depresszio_mobilhasznalat_facebook_felismerhető_veszely_totalis_megfigyeles/ [Letöltve: 2020. február 28.].

NALINI, B. (2019) The Hitchhiker's Guide To AI Ethics. *Medium*.

<https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer-1bfd039124b0>
[Letöltve: 2020. február 28.].

NARULA, G. (2020) Everyday Examples of Artificial Intelligence and Machine Learning. *Emerj Artificial Intelligence Research* <https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/> [Letöltve: 2020. február 28.].

PRESS, G. (2019) Artificial Intelligence (AI) Stats News: 50% Of Americans Optimistic And 50% Fearful About AI. *Forbes*. <https://www.forbes.com/sites/gilpress/2019/08/04/artificial-intelligence-ai-stats-news-50-of-americans-optimistic-and-50-fearful-about-ai/#49d10d2926fa>
[Letöltve: 2020. február 28.].

SERVICK, K. (2019) Suicide attempts are hard to anticipate. A study that tracks teens' cellphone use aims to change that. *ScienceMag*.
<https://www.sciencemag.org/news/2019/08/suicide-attempts-are-hard-anticipate-study-tracks-teens-cellphone-use-aims-change> [Letöltve: 2020. február 28.].

SZÖLLŐSI Z. (2019) A Mesterséges Intelligencia rendszerek kihívásai és kockázatai. *Deloitte Magyarország*.
<https://www2.deloitte.com/content/dam/Deloitte/hu/Documents/technology/hu-technology-mesterseges-intelligencia-19-june.pdf> [Letöltve: 2020. február 28.].