

A GENERATÍV MESTERSÉGES INTELLIGENCIÁVAL KAPCSOLATOS EURÓPAI UNIÓS ETIKAI ÉS JOGI SZABÁLYOK

THE ETHICAL AND LEGAL RULES OF THE EUROPEAN UNION ON GENERATIVE ARTIFICIAL INTELLIGENCE

Csík Aurél*

Absztrakt

A mesterséges intelligencia korunk egyik legfelkapottabb témája. A MI alkalmazásával kapcsolatban az idő előrehaladtával egyre több aggály és probléma felmerült, amelyek miatt a jogalkotóban is felmerült a szabályozás kérdése. A cikkemben elsősorban ennek a jogalkotásnak az uniós eredményeit szeretném körbejárni, nevezetesen a frissen elfogadott AI Act és a soft law Etikai Irányelve által lefektetett rendszert és követelményeket, majd igyekszem válaszolni arra a kérdésre, hogy ezek a jogforrások képesek-e lépést tartani a technológia fejlődésével.

Kulcsszavak: mesterséges intelligencia, Etikai Irányelvek, AI Act, Európai Unió, ChatGPT

Abstract

Artificial intelligence is one of the most trending topics of our time. As time has gone by, concerns and problems with the use of AI have grown, leading to a growing number of regulatory issues in the legislature. In this article, I would like to focus on the achievements of this legislation in the EU, namely the system and requirements laid down by the recently adopted AI Act and the soft law Ethical Guidelines, and then try to answer the question of whether these legal sources are able to keep pace with the development of the technology.

Keywords: artificial intelligence, Ethical Guidelines, AI Act, European Union, ChatGPT

1. Bevezetés

A mesterséges intelligencia és az azzal való visszaélés a legmeghatározóbb beszédtemák közé tartoznak, ennek ellenére viszont az is megállapítható, hogy ezek a kérdések nem kifejezetten újkeletűnek. A mesterséges intelligencia, mint eszköz már régóta foglalkoztatja az emberiséget, az olyan problémák, mint például a dezinformáció és az ellenfelek hiteltelenítése pedig a történelem kezdete óta része volt a politikának.

A cikkben elsősorban a mesterséges intelligenciával felmerülő aggályokat és kockázatokat szeretném bemutatni, kiemelve a generatív mesterséges intelligenciát, amely az MI kategóriájába tartozó technológiák közül leghamarabb válik a mindennapi életünk részévé.

Az első fejezetben éppen ezért a mesterséges intelligenciát, mint XXI. századi témát szeretném körüljárni, amelyben kitérek a ChatGPT-re és a hozzá hasonló programokra, valamint azok társadalomra gyakorolt hatására és a lehetséges veszélyeire is.

* Csík Aurél, harmadéves nappali tagozatos PhD hallgató, Miskolci Egyetem, Deák Ferenc Állam- és Jogtudományi Doktori Iskola, acsik96@gmail.com. Témavezető: Prof. Dr. Paulovics Anita, egyetemi tanár, Miskolci Egyetem Állam- és Jogtudományi Kar, Államtudományi Intézet, Alkotmányjogi Intézeti Tanszék.

A következő részben bemutatom az *AI Act* által felállított mesterséges intelligencia kategóriákat, amelynek az alapja az úgynevezett kockázatalapú megközelítés. Ez a rendszer a társadalomra való veszélyességtől függően négy nagyobb kategóriába osztja a mesterséges intelligenciát alkalmazó technológiákat.

Az utolsó részben az Európai Unió által kidolgozott Etikai Irányelvek legfontosabb pontjait igyekszem majd összefoglalni, amelyek az ismertebb problémákra hatékonyak megoldásul szolgálhatnak. Ezt különösen fontos kérdésnek tartom, mivel véleményem szerint az etika és az erkölcs a mesterséges intelligencia szemszögéből értelmezhetetlen fogalmak, így ezeknek a megvalósítása elsősorban az emberek feladata.

Cikkemben tömör és adekvát válaszokat keresek azokra a kérdésekre, amelyek az embereket ma leginkább foglalkoztatják a mesterséges intelligenciával kapcsolatban, és összegzésként értékelem a vizsgált szabályozást.

2. A generatív mesterséges intelligencia meghatározása

A mesterséges intelligencia, mint kifejezés 2022 novemberében valósággal berobbant a köztudatba¹, amelynek az oka elsősorban az a felismerés volt, hogy a koncepció a korábban elképzelt science fiction történetekből átcsúszni látszik a kézzelfogható valóságunkba. Az esemény, amely előidézte ezt a változást, a *ChatGPT* nevű chatbot prototípusának megjelenése volt, amelyet az *OpenAI* nevű mesterséges intelligenciával foglalkozó kutatólabor fejlesztett². A chatbot megrengette a tudományos világot, pillanatok alatt népszerűvé vált, 2023 januárjára pedig a leggyorsabb ütemben növekvő felhasználói bázist is magáénak tudhatta³. Az előbbiekhöz indokul elsősorban az szolgálhat, hogy az alkalmazás képes olyan szövegeket létrehozni, amelyek első olvasatra akár valódi emberektől is származhatnának. A *ChatGPT* felkapottsága így a mesterséges intelligenciát és annak a társadalmunkra gyakorolt hatásait a fősodratú témák közé emelte⁴.

A pozitív fogadtatás mellett azonban többen aggályukat is kifejezték az etikai és egzisztenciális kérdéseket is felvető technológiai innovációval szemben. Elon Musk – aki az *OpenAI* létrehozásában is közreműködött⁵ – például úgy fogalmazott az *X* (korábbi nevén *Twitter*) nevű platformon⁶, hogy a *ChatGPT* ijesztően jó, amivel közelebb kerültünk a veszélyesen erős mesterséges intelligencia létrehozásához. Kiemelném továbbá a program korábban megemlített „emberszerűségét”, mivel a mesterséges intelligenciából eredő legtöbb jogi és etikai kérdés ebből az attribútumból ered. Az elmúlt időszakban többen is aggodalommal

¹ A Google Trends – amely egy Google-keresési adatokat feldolgozó szolgáltatás – számain látható, hogy az „artificial intelligence” kifejezésre és az „AI” mozaikszóra történő keresések száma 2022 novembere óta látványosan megemelkedett, a legtöbb keresésre – cikk megírása pillanatáig – 2023 márciusában és 2024 februárjában került sor.

² What is AI chatbot phenomenon ChatGPT and could it replace humans? (letöltés ideje: 2024.04.03.)

³ Krystal HU: ChatGPT sets record for fastest-growing user base - analyst note <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> (2024.04.04.)

⁴ James VINCENT: ChatGPT proves AI is finally mainstream — and things are only going to get weirder <https://www.theverge.com/2022/12/8/23499728/ai-capability-accessibility-chatgpt-stable-diffusion-commercialization> (letöltés ideje: 2024.04.04.)

⁵ Grace KAY: The history of ChatGPT creator OpenAI, which Elon Musk helped found before parting ways and criticizing <https://www.businessinsider.com/history-of-openai-company-chatgpt-elon-musk-founded-2022-12> (letöltés ideje: 2024.04.04.)

⁶ Elon Musk X postja a ChatGPT-ről, <https://twitter.com/elonmusk/status/1599128577068650498> (letöltés ideje: 2024.05.10.)

tekintettek arra az eshetőségre, hogy a mesterséges intelligencia idővel kiválthatja az emberi munkaerőt több szektorban is. Elég csak a forgatókönyvírók és színészek sztrájkjára gondolni⁷, amely hónapokra megbénította Hollywoodot és az amerikai filmgyártást. A sztrájk legfontosabb követelései között az is ott volt, hogy a forgatókönyvírás során – illetve magában a filmgyártásban is – kerüljék el a mesterséges intelligencia alkalmazását, mivel az veszélyezteti az alkotói integritást.

A *ChatGPT* elsődleges funkciója az, hogy különböző utasítások alapján – jellemzően a felhasználóval folytatott interakció során – szövegeket „generál” az interneten fellelhető információk alapján. Ez önmagában nem tűnhet nagy mérföldkőnek, mivel a *ChatGPT* előtt is rendelkezésre álltak különböző chatalkalmazások és digitális személyi asszisztensek, elég ha csak az *Apple Siri*re vagy az *Amazon Alexa*jára gondolunk. A *ChatGPT* innovációja leginkább az, hogyan és milyen minőségben éri el azt, amit az elődjei elértek. Ez pedig az, hogy szinte bármit – szöveget, képet vagy hangot – képes előállítani, amelynek a minősége alkalmanként az ember alkotta produktumokat is képes túlszárnyalni.

Mivel ezek a chatbotok az elmúlt időszakban páratlan népszerűsége tettek szert, ezért a laikusok a mesterséges intelligenciát, mint technológiát is ezekkel a vívmányokkal azonosítják. Ez viszont téves irány, ugyanis a fenti működési elvre épülő alkalmazások a mesterséges intelligencia kategóriájába tartoznak ugyan, viszont nem fedik le teljes mértékben annak alkalmazási körét. A *ChatGPT*-t és az ahhoz hasonló programok különböző tartalmakat állítanak elő, ezért a mesterséges intelligencia ezen alfaját generatív mesterséges intelligenciának nevezzük.

A generatív mesterséges intelligencia technológiájának az előképe az *ELIZA* névre hallgató program volt, amelyet Joseph Weizenbaum tervezett a hatvanas években⁸. A program elsődleges célja annak bizonyítása volt, hogy az emberi kommunikációhoz képest a számítógép által generált kommunikáció kevésbé cizellált. Ezt támasztja alá a program neve is, mivel Weizenbaum az *ELIZA*-t Eliza Doolittle-ről nevezte el⁹, aki George Bernard Shaw *Pygmalion* című drámájának központi szereplője. A dráma arról szól, hogy Doolittle-ből, aki egy tájszólassal beszélő virágáruslány, hogyan csinálnak előkelő úrihölgyet. Weizenbaum azért választotta ezt a nevet, mert lehet, hogy Eliza Doolittle a történet végén azt az illúziót kelti, hogy ő egy arisztokrata, azonban a felszín alatt ugyanaz a vidéki lány marad, aki az átalakulás előtt volt. Eképp a mesterséges intelligencia – viselkedjen bármennyire is emberhez hasonlóan – sem tudja önmagát megtagadni. Weizenbaum a mesterséges intelligencia J. Robert Oppenheimerének is tekinthető, mivel az elsők között volt, aki megfogalmazta a félelmét azzal a technikai vívmánnyal kapcsolatban, amelynek a megteremtésében ő is közreműködött. Weizenbaum legnagyobb hatású művében, a *Computer Power and Human Reason: From Judgement to Calculation* című könyvében felvetette azt a kérdést, hogy szabad-e hagyni a gépeket gondolkodni, amely miatt őt a tudományos világ eretnekké nyilvánította¹⁰.

⁷ Dani ANGUIANO – Lois BECKETT: How Hollywood writers triumphed over AI – and why it matters, <https://www.theguardian.com/culture/2023/oct/01/hollywood-writers-strike-artificial-intelligence> (letöltés ideje: 2024.05.10.)

⁸ *ELIZA*: a very basic Rogerian psychotherapist chatbot, <https://web.njit.edu/~ronkowitz/eliza.html> (letöltés ideje: 2024.04.08.)

⁹ Joseph Weizenbaum: Writes *ELIZA*: A Pioneering Experiment in Artificial Intelligence Programming, <https://www.historyofinformation.com/detail.php?id=4137> (letöltés ideje: 2024. 04.08.)

¹⁰ Ben TARNOFF: Weizenbaum’s nightmares: how the inventor of the first chatbot turned against AI, <https://www.theguardian.com/technology/2023/jul/25/joseph-weizenbaum-inventor-eliza-chatbot-turned-against-artificial-intelligence-ai> (letöltés ideje: 2024. 05. 10.)

3. Mesterséges intelligenciát működtető rendszerek osztályozása

A generatív mesterséges intelligenciával kapcsolatban már számos etikai, jogi, sőt egzisztenciális kérdés is felmerült, mióta az emberek azokat széleskörűen alkalmazzák. Mivel az egzisztenciális aggályok elsősorban a filozófia és a pszichológia területére tartoznak, ezért a cikkemben nem ezekre kitérni, csak az etikai és jogi kérdéseket fogom részletesebben kifejteni. Az Európai Unió mesterséges intelligenciát szabályozó rendelete, az *AI Act* alapját is alapját is ezek a kérdések képezik. A rendelet az Európai Unió jelenlegi jogalkotási ciklusának egyik legfontosabb eredménye, amelyet Ursula von der Leyen „Ambíciózusabb Unió” című politikai szerepvállalásában jelentett be először¹¹. A Bizottság a mesterséges intelligencia technológiáját használó különböző programok szabályozásának meghatározása során kockázatalapú megközelítést alkalmazott. A kockázatalapú megközelítés lényege nagyjából úgy írható le, hogy a szabályozás mértéke nagyjából arányos azzal, hogy a program mennyire veszélyes a társadalom prudens működésére nézve. A tervezet a végleges változatában négy ilyen kategóriát foglal magába, mindegyikhez egy-egy szabályozási formát rendelve.

A fentebb említettek közül a legenyhébb kategória a minimális kockázattal rendelkező mesterséges intelligencia¹², amelyekkel szemben a jelenleg is hatályban lévő, meglévő szabályozásokon kívül nem állít fel további kötelezettségeket.

A következő kategóriát az átláthatósági szempontból aggályos technológiák csoportja jelenti¹³. Ebbe a kategóriába elsősorban a természetes személyekkel kapcsolatba lépő vagy tartalmakat generáló programok tartoznak ide. A generatív mesterséges intelligencia elsősorban ebbe a csoportba sorolható. Ennek a kategóriának a lényege, hogy az ide sorolható technológiákról nem dönthető el egyértelműen, hogy magas kockázatot jelentenek-e vagy sem. Elég, ha csak a generatív mesterséges intelligenciával kapcsolatban a hírekbe kerülő problémákra gondolunk. Az egyik legelterjedtebb ilyen technológia a deepfake, amely az *AI Act* alapján olyan mesterséges intelligencia által generált vagy manipulált kép, audio- vagy videotartalom, amely érzékelhetően hasonlít létező személyekre, tárgyakra, helyekre vagy más entitásokra vagy eseményekre, és amely egy személy számára megtévesztő módon eredetinek vagy valóságosnak tűnhet¹⁴. A deepfake jól szemlélteti ennek a kategóriának a lényegét, ugyanis azontúl, hogy bizonyos szektorokban a használata jelentősen megkönnyíti a munkafolyamatokat¹⁵, sajnos emberi életet is tönkre tehet. A deepfake segítségével – akár egy darab kép felhasználásával – bárkiről kompromittáló hamis vagy hamisított felvételeket lehet előállítani, amely ellen már állami szinten is megfigyelhetőek szabályozási törekvések¹⁶. Az ilyen technológiákat alkalmazóknak egyértelművé kell tenniük, hogy a produktum mesterséges intelligencia közreműködésével jött létre, ha pedig munkahelyen használják, akkor erről értesíteni kell munkavállalókat és képviselőiket.

A következő kategória a magas kockázatú mesterséges intelligenciáknak a köre¹⁷. A tervezet szerint ebbe a csoportba tartoznak azok a rendszerek, amelyek nagy valószínűséggel

¹¹ *AI Act* preambulum (javaslat indokai és céljai)

¹² Briefing EU Legislation in Progress Artificial Intelligence Act p.9.

¹³ *AI Act* IV. Cím

¹⁴ *AI Act* 52. cikk (3)

¹⁵ Max KALMYKOV: Positive Applications for Deepfake Technolog, <https://www.dataart.com/blog/positive-applications-for-deepfake-technology-by-max-kalmykov> (letöltés ideje: 2024.04.16.)

¹⁶ Angliában és Walesben bűncselekménynek minősülhet a szexuális tartalmú „deepfake” képek készítése, <https://rtl.hu/kulfold/2024/04/16/anglia-wales-erotikus-deepfake-tartalom-borton-buncselekmeny> (letöltés ideje: 2024.04.16.)

¹⁷ *AI Act* III. cím

negatív hatást gyakorolhatnak az ember egészségére, biztonságára és az alapvető jogaira. Ezt többféle módszerrel lehet megállapítani. Ezek egyike az, hogy a mesterséges intelligenciát a AI Act III. Mellékletében¹⁸ felsorolt szektorokban alkalmazzák. Ezek közé tartozik a biometrikus azonosítás, az oktatás (ideértve a középfokú oktatás minden formáját), az alapvető szükségletekre épülő infrastruktúrák, a foglalkoztatás, az alapvető magánszolgáltatások és közszolgáltatások, a bűnüldözés, a bevándorlás, végül pedig az igazságszolgáltatás és a demokratikus akaratképzés. A magas kockázatú mesterséges intelligencia alkalmazása előtt a tervezet alapján egy olyan eljárást kell lefolytatni, amelynek eredményeképpen megállapítható, hogy a rendszer megfelel a jogszabályoknak és nem okoz hátrányt a működése során.

Az a kör, amellyel kapcsolatban az *AI Act* a leghigorúbban lép fel, az a tiltott mesterséges intelligencia gyakorlatok csoportja¹⁹. Ahogy az elnevezésből is kitűnik, a rendeletben lajstromozott esetekben a mesterséges intelligenciát nem használni. A mesterséges intelligencia nem vehető igénybe megtévesztés, a személy hátrányos tulajdonságainak a kihasználására, (bizonyos kivételekkel) személyazonosításra és az emberek különösen védett adatok menti csoportosítására.

4. A mesterséges intelligenciát érintő etikai irányelvek

Ebben a pontban elsősorban az Európai Unió 2019-es *Mesterséges Intelligenciára vonatkozó Etikai Iránymutatását*²⁰ (továbbiakban: Etikai Iránymutatás) szeretném körbejárni. Az Iránymutatásról mindenek előtt fontos kijelenteni, hogy soft law révén jogi kötőerővel nem rendelkezik, ezért annak a szabályait nem lehet rákényszeríteni a mesterséges intelligenciát fejlesztő vagy éppen azokat használó cégeket és szervezeteket.

Álláspontom szerint az Etikai Iránymutatás legfontosabb megállapítása, hogy a megbízható mesterséges intelligenciának három főbb jellemzővel kell bírnia. Ez a három fő szempont irányításként szolgálhat ahhoz, hogy a jogalkotó miként viszonyul a mesterséges intelligenciához, mint szabályozandó tárgyhoz. A mesterséges intelligenciának egyrészt *jogszerűnek* kell lennie, vagyis a működése és működtetése során tekintettel kell lenni a hatályos az európai, a nemzeti és nemzetközi jogforrásokra²¹. A dokumentum a releváns jogszabályok között nevesíti többek között az Európai Unió elsődleges és másodlagos uniós jog eredményeit, az ENSZ és az Európai Tanács nemzetközi egyezményeit, valamint a tagállami jogalkotást is. Az Etikai Iránymutatás – annak ellenére, hogy a célja nem elsődlegesen jogi tanácsadás, amely a dokumentumban explicit is megjelenik²² – kiemeli továbbá, hogy a szabálykövetésnek egyszerre kell „tevéleges” és tartózkodási” jelleggel megvalósulnia, vagyis nem csak az alapján kell eljárni, hogy mi nem tehető meg, hanem arra is tekintettel kell lenni, hogy mit kell megtenni a szabályok szerint²³. A mesterséges intelligenciára épülő rendszereknek másodsorban *etikusnak* is kell lenniük²⁴, vagyis az adott parancsok végrehajtása során a jogszabályokon túl az etikai elveket és morális szabályokat kell szem előtt tartania. Ez a jog után a megbízhatóság második legfontosabb komponense, viszont azt is be kell látni, hogy a mesterséges intelligencia sok kérdésre nem tud morálisan is adekvát választ adni, mivel

¹⁸ Annex III, <https://www.euaiact.com/annex/3> (letöltés ideje: 2024.04.16.)

¹⁹ AI Act II. cím

²⁰ Az Európai Bizottság által 2018 júniusában létrehozott Mesterséges intelligenciával foglalkozó magas szintű és független technikai csoport Megbízható Mesterséges Intelligenciára vonatkozó Etikai Iránymutatás

²¹ Etikai Irányelvek B. (22)

²² Etikai Irányelvek B. (25)

²³ Etikai Irányelvek B. (23)

²⁴ Etikai Irányelvek B. (26)

természetéből adódóan erre alkalmatlan. Végső, de nem utolsó sorban pedig az ilyen elven működő programoknak – mivel egy technikai vívmányról beszélünk – *stabilnak* és kiszámíthatónak kell lenniük műszaki értelemben is²⁵.

Az Etikai Irányelvek által felállított követelményrendszert három nagyobb csoportra osztották fel a dokumentum szerkesztői. Az első csoportba azok a megbízható mesterséges intelligencia alapját adó kritériumok tartoznak, amelyeknek a célja az alapvető jogokon alapuló etikai elvek érvényesülését biztosítják. A második csoportba azok a követelményeket sorolja a dokumentum, amelyek a megbízható mesterséges intelligencia megvalósításához szükségesek. A harmadik csoportot a megbízható mesterséges intelligencia vonatkozó értékelési lista adja. Ezekre a feltételekre azért is szükség van, mert az ember egyre inkább a számítógépre hagyatkozik, így biztosnak kell lenni abban, hogy a gép által hozott döntések helyesek és összhangban állnak az emberi értékekkel²⁶.

Az első csoportot négy fő kritériumból áll: az emberi autonómia tiszteletben tartásából, a kár megelőzéséből, a méltányosságból és a megmagyarázhatóságból²⁷. Ezeknek a kritériumoknak az alapját az alapvető jogok adják akként, hogy azok iránymutatást adhatnak a mesterséges intelligencia fejlesztése és alkalmazása során. A dokumentum nevesít is²⁸ különböző alapvető jogokat, amelyekből az etikai elvek eredhetnek. Ezek közé az alapvető jogok közé tartozik az emberi méltóság tiszteletben tartása, az egyén szabadsága, a demokrácia, az igazság és a jogállamiság tiszteletben tartása, az egyenlőség, megkülönböztethetőség és szolidaritás és a polgárok jogai.

Az Etikai Irányelvekben lajstromozott követelmények következő csoportja már sokkal méretesebb az előzőnél, amely véleményem szerint az egész dokumentum gerincét is adja. A dokumentum hét követelményt sorol fel a megbízható mesterséges intelligencia megvalósításával kapcsolatban, amelyek egyszerre jelenítenek meg egyéni és társadalmi szempontokat²⁹. A hét követelmény közé tartozik az emberi cselekvőképesség támogatása, a műszaki stabilitás és biztonság, az adatvédelem és adatkezelés, az átláthatóság, a sokféleség, megkülönböztetésmentesség és méltányosság, a társadalmi és környezeti jólét és az elszámoltathatóság. Ezek a kritériumok kölcsönhatásban állnak egymással³⁰ abból a célból, hogy a mesterséges intelligenciára alapuló rendszert a fejlesztésétől egészen annak lekapcsolásáig folyamatosan értékeljék és vizsgálják.

Ebben az aspektusban fontos kiemelni az *átláthatóság követelményét*³¹, amely magában foglalja az MI-rendszer szempontjából releváns elemek – az adatok, a rendszer, az üzleti modellek – átláthatóságát. Az átláthatóság követelménye három fő komponensből áll: a nyomonkövethetőségből, megmagyarázhatóság és a tájékoztatás. A nyomonkövethetőség azt takarja, hogy az MI-rendszer döntéshozatalát elősegítő adatkészleteket és folyamatokat, köztük az adatgyűjtési és adatszámítási folyamatokat, valamint a használt algoritmusokat a lehető leghatékonyabban dokumentálni kell. A megmagyarázhatóság azt takarja, hogy az MI-rendszer műszaki folyamatainak az ahhoz kapcsolódó emberi döntéseknek megmagyarázhatónak kell lenniük, valamint magyarázatoknak kell rendelkezésre állnia a rendszer tervezése során meghozott döntésekről, a rendszer bevezetésének létjogosultságáról, valamint arról, hogy az

²⁵ Etikai Irányelvek B. (27)

²⁶ PUSZTAHELYI Réka: Bizalmunkra méltó MI – a mesterséges intelligencia fejlesztésének és alkalmazásának erkölcsi-etikai vonatkozásairól, in: Publicationes Universitatis Miskolcensis Sectio Juridica et Politica, Tomus XXXVII/2 (2019), pp. 97–120.

²⁷ Etikai Irányelvek B. I. fejezet 2.2. (48) és (50)-(54)

²⁸ Etikai Irányelvek B. I. fejezet 2.1.

²⁹ Etikai Irányelvek B. II. fejezet (57)-(58)

³⁰ Etikai Irányelvek 2. ábra

³¹ Etikai Irányelvek B. II. fejezet 4.

adott MI-rendszer milyen mértékben befolyásolja és alakítja át a szervezeti döntéshozatali eljárást. A tájékoztatás arra irányul, hogy az MI- rendszer nem mutatkozhat be emberként az felhasználóknak, valamint a felhasználók jogosultak tájékoztatást kapni, hogy MI- rendszerrel állnak szemben.

A dokumentum műszaki és nem műszaki eszközöket is felsorol, amelyeknek a célja megbízhatóság eszközölése. Az iránymutatás kiemeli, hogy ennek a célnak a megvalósítása folyamatos, mivel a mesterséges intelligencia, mint technológia folyamatosan fejlődik, ezzel a fejlődéssel pedig lépést kell tartani megbízhatóság biztosítása szempontjából is. A műszaki eszközök közé tartozik a tanulási architektúrák tervezése, az etika és jogállamiság beépítése, magyarázási módszerek használata és a tesztelés és a jóváhagyás, míg a nem műszaki eszközök közé tartozik a szabályozás, a magatartási kódexek létrehozása, a szabványosítás, a tanúsítás, az elszámoltathatóság, az oktatás és a sokféleség biztosítása.

A harmadik és egyben utolsó kérdéskör, amivel a dokumentum foglalkozik a megbízható mesterséges intelligencia értékelésére irányuló eszközök. Ebbe a csoport két részre osztható, egyrészt a megbízható mesterséges intelligencia gyakorlati alkalmazására, másrészt pedig a megbízható mesterséges intelligencia értékelési listájára. Az értékelési lista maga is több komponensből áll. Az ember jelenlétét támogató szempontok között található az alapvető jogokat, az emberi cselekvőképesség támogatását és az emberi felügyeletet. A műszaki stabilitást irányozza elő a támadásokkal szembeni védelem biztosítása, készenléti terv és általános biztonság kidolgozása, a pontosság, a megbízhatóság és reprodukálhatóság. A dokumentum az adatvédelem vonatkozásában a magánélet védelmére, az adatok minőségére és sértetlenségére, és az adatokhoz való hozzáférésre helyezi a hangsúlyt.

Az Etikai Irányelvek által lefektetett kritériumokat megvizsgálva észrevehetjük, hogy az előbbiekben körbejárt kérdéskörök nagyjából kettő-három nagyobb kategóriát fednek le. A mesterséges intelligencia megbízhatóságának az egyik legfontosabb fokmérője az emberi jelenlét az alkalmazása során, mivel a technológia jelenleg a csak emberi parancsra tud végrehajtani feladatokat. Ezen kívül fontos törekvés, hogy a mesterséges intelligencia működése során ne szegje meg a legfontosabb jogszabályokat, amihez olyan apparátust kell létrehozni, amely során a mesterséges intelligencia az ember által adott parancs végrehajtása során a rendelkezésre álló információk között mérlegelje a rendelkezésére álló jogforrások tartalmát. A következő pontban az Etikai Iránymutatás azon rendelkezéseit szeretném részletesebben körüljárni, amelyek hatást gyakorolhatnak mesterséges intelligencia által gerjesztett dezinformáció elleni fellépés tekintetében.

Az egyértelműen körvonalazódik, hogy a mesterséges intelligenciának – főleg, ha olyan problémák esetén beszélünk róla, mint a dezinformáció – ki kell merítenie a megbízható mesterséges intelligenciára vonatkozó követelményeket, vagyis jogszerűnek, etikusnak és stabilnak kell lennie. Ha ez a három feltétel fennállna, elképzelhetetlen lenne, hogy mesterséges intelligenciával álhíreket és deepfake-eket lehessen előállítani.

A dezinformáció szempontjából az alapvető jogokon alapuló kritériumrendszer bír álláspontom szerint a legnagyobb relevanciával. A dokumentum kijelenti, hogy a megbízható mesterséges intelligencia egyik összetevője, hogy működése során tiszteletben tartja az alapvető jogokat. A dezinformáció – különösen a deepfake-vel készülő készülő képes anyagok esetén – tekintetében a legfontosabb alapvető jog, amelyet a mesterséges intelligenciának követnie kell, az az emberi méltóság tiszteletben tartása. A mesterséges intelligenciának mindenkire a neki járó erkölcsi tisztelettel kell tekintenie³², ugyanis minden információ és adat mögött, amelyet a program feldolgoz működése során, egy ember áll, amelynek a testi és szellemi integritását nem sértheti az az eredmény, amire a program a neki adott parancs

³² Etikai Irányelvek (41).

végrehajtását követően jut. Például, ha a mesterséges intelligencia arra kap utasítást, hogy egy bizonyos személyről állítson elő kompromittáló képet, akkor utasítsa el ennek a végrehajtását, mivel ezzel sérülni a célszemély emberi méltósága. A dezinformációval kapcsolatban felmerülhet továbbá a demokrácia, az igazságosság és a jogállamiság tiszteletben tartása is³³. Az Etikai Irányelvek kijelentik, hogy a mesterséges intelligencia nem áthatja alá a demokratikus folyamatokat, amelyek az emberi döntéshozatalra és szavazásra épülnek. A mesterséges intelligencia, mint eszköz nem használható fel arra, hogy a demokratikus közvéleményt egyik vagy másik irányba befolyásolva, ezzel hatást gyakorolva akár arra is, hogy az adott államban milyen politikai erők töltik be a vezető pozíciókat.

A megbízható mesterséges intelligencia megvalósítására irányuló követelmények között is találhatunk olyanokat, amelyek a dezinformáció elleni küzdelem tekintetében nagy relevanciával bírhatnak. Az emberi cselekvőképesség támogatása vonatkozásában is kiemeli a dokumentum az alapvető jogokat. Az emberi cselekvőség és önrendelkezési jog védelme érdekében figyelemmel kell lenni arra, hogy az MI rendszer működése sokszor észrevehetetlen módon befolyásolja az emberi döntést, ezért tisztességtelen manipulációhoz vezethet³⁴. A mesterséges intelligencia rendszerek – elsősorban kapacitásbeli hiányosságok miatt – hátrányosan érinthetik az alapvető jogok érvényesülését. Törekedni kell azoknak a hatásoknak a mérséklésére, amelyet ezek a hiányosságok okoznak. A dokumentum ezen kívül még kiemeli az emberi felügyeletet és az emberi cselekvőképesség támogatását, mivel az ember központi helyen szerepel a mesterséges intelligencia működésének felügyelete során, hiszen az végső soron az ember az, aki fejleszti a mesterséges intelligencia-technológiákat, amelyek az emberi használat során fejlesztik tovább a működésüket.

A kritériumok ezen körében maradván még kiemelném a sokféleség, megkülönböztetésmentesség és méltányossággal kapcsolatos követelményeket, amelyek álláspontom szerint a dezinformáció elleni küzdelemben kulcsfontossággal bírnak. Ezek közül a legfontosabb a méltánytalan torzítás elkerülése³⁵. A dokumentum ugyan elsősorban azt emeli ki ezzel kapcsolatban, hogy mind a tanulás, mind pedig az üzemelés során az adatkezelések hiányossága negatív hatással lehet a történeti adatok miatti nem szándékos torzulás. Ez akár hátrányos megkülönböztetésekhez is vezethet. Álláspontom szerint, azonban ezek a hiányosságok a dezinformációra irányuló tartalmak előállításánál is sérülékennyé teszik a mesterséges intelligenciát.

Végső, de nem utolsó sorban ki szeretném emelni azt, hogy a dokumentum fontosnak tartja, hogy a mesterséges intelligencia a társadalom és a demokrácia érdekeit szem előtt tartva működjön³⁶. A mesterséges intelligenciának a működése során figyelembe kell vennie azt, hogy a társadalom működésére milyen hatással lehet az az eredmény, amelyet a parancs végrehajtása során előállít.

5. Összefoglalás

Cikkem célja elsősorban az volt, hogy bemutassam a mesterséges intelligencia, mint tudományos jelenség jelenlegi szabályozását. Annak ellenére, hogy a mesterséges intelligenciával kapcsolatban, a technológia egyre rohamosabb fejlődése miatt minden ilyen tárgykörben hozott szabályozás el fog avulni, újra és újra elő fognak kerülni azok a kérdések,

³³ Etikai Irányelvek (43)

³⁴ PUSZTAHELYI: i.m. 103.

³⁵ Etikai Irányelvek (80)

³⁶ Etikai Irányelvek (86)

amelyekre a meglévő jogforrások nem vagy máshogy közelítenek meg. Cikkem első fejezetében igyekeztem a mesterséges intelligenciát, mint XXI. jelenséget bemutatni, kiemelve azt, hogy azt széles körűen lehet alkalmazni a mindennapok sor. Mivel jelenleg a mesterséges intelligenciát a generatív mesterséges intelligencia körébe tartozó programokkal azonosítják, ezért az első és legfontosabb célkitűzés, hogy a kettő közti disztingválás társadalmi szinten megvalósuljon. Álláspontom szerint ez kulcsfontosságú jelentőséggel bír ahhoz, hogy a mesterséges intelligenciát átfogóan tudjuk szabályozni.

A következő részben az AI Act által felállított mesterséges intelligencia rangsort mutattam be. Az AI Act által felállított négy kategórián keresztül világosan látszódik, hogy ez a technológia milyen sokféle alkalmazási módot takarhat, és ezek adott esetben milyen veszélyeket jelenthetnek a társadalomra.

Az utolsó részben az Európai Unió által kidolgozott Etikai Irányelvekben foglalt kategória rendszert mutattam be. Mivel a mesterséges intelligencia felhasználói mi magunk vagyunk, a szabályozás pedig alig tud lépést tartani a technológiai fejlődéssel, ezért nekünk és a programok fejlesztőinek kell odafigyelni, hogy azok alkalmazása ne legyen káros se az egyénekre nézve, se a társadalomra nézve.

Álláspontom szerint a mesterséges intelligenciával kapcsolatos, rendelkezésre álló joganyag egy jó irányt vetít elő a technológia szabályozása terén, azonban a szabályok megléte önmagában nem elegendő. A felhasználóknak és a fejlesztőknek is szem előtt kell tartaniuk