

A „fekete doboz” politikája

**A Mesterséges Intelligencia demokratikus kontrollja és a
digitális társadalmi egyenlőtlenségek politikai dimenziói
Magyarországon**

LBA456

Tartalomjegyzék

Tartalomjegyzék	1
1. Bevezetés	2
2. Elméleti keretek	4
2.1. A mesterséges intelligencia, mint „fekete doboz”	4
2.2. Az „MI triász”	5
2.3. Etikai alapelvek	7
2.4. Kognitív torzítások	8
2.5. Sötét minták	9
2.6. Az MI-alapú rendszerek demokratizálása	9
2.7. MI ártalmak osztályozása	10
2.8. Szabályozási alapelvek	11
2.9. Technikai és működési kockázatok	12
3. Jelenlegi szabályozás	13
3.1. A globális konszenzus alapjai	13
3.2. Nemzetközi szakpolitikai kitekintés	13
3.3. Szabályozás az Európai Unióban	16
3.4. Az AI Act kritikája	19
3.5. Digitális- és MI tudatosság az EU-ban és Magyarországon	20
3.6. Az MI intézményes keretei Magyarországon	23
4. A kutatás módszertani háttere	25
5. Kutatási eredmények	30
5.1. Kompetencia szintje	30
5.2. Kommunikáció	32
5.3. Szabályozási keretrendszer hatékonysága	34
5.4. Politikai akarat	35
5.5. Kockázatok és okozatiság felmérése	37
5.6. Társadalmi és etikai megfontolások	38
5.7. Interjúk értékelése	40
6. Konklúzió	41
7. Függelék	42
7.1. Ábrajegyzék	42
7.2. Esettanulmányok	52
Irodalomjegyzék	54

1. Bevezetés

A mesterséges intelligencia (MI) és a gépi tanulási rendszerek egyre inkább a gazdasági, társadalmi és politikai struktúrák szerves részévé válnak világszerte; a gyors technológiai fejlődés és a növekvő digitalizáció által jellemezhető jelenkorban a mesterséges intelligencia és a társadalmi egyenlőtlenségek metszéspontja kritikus kutatási területet jelent a politikatudomány számára (Joyce et al., 2021, Zhang et al., 2022). Ez a gyorsan fejlődő technológia elképesztő lehetőségeket rejt magában, azonban, ha nincs megfelelően szabályozva, ugyanekkora veszélyeket is – a legsúlyosabb kockázat az emberiség kihalásának veszélye (Kasirzadeh, 2024). Az MI-alapú rendszerek alkalmazása olyan kérdéseket vet fel, mint a technológiai hatalom elosztása, a társadalmi igazságosság biztosítása, és a közérdeknek az innovációk során való érvényesítése. Szabályozásának vizsgálata megmutatja, hogy a politikai rendszerek hogyan képesek reagálni a gyors technológiai fejlődés kihívásaira, és milyen intézkedésekkel lehet biztosítani, hogy a technológia valóban a közérdeket szolgálja, anélkül, hogy fokozná a társadalmi egyenlőtlenségeket.

A dolgozat célja, hogy kinyissa a diskurzust a társadalomtudományok kutatói és képviselői felé, hangsúlyozva a multidiszciplináris megközelítés szükségességét a mesterséges intelligencia társadalmi következményeinek megértéséhez és kezeléséhez (Mező & Mező, 2019; Négyesi, 2020, Gruetzemacher & Whittlestone, 2022). Ez a megközelítés lehetővé teszi, hogy a mesterséges intelligenciával kapcsolatos politikai döntések és kutatások megalapozottabbak és a társadalmi jóllét szempontjából is mérvadóak legyenek. A technológiai innováció, ha széles körben elterjed – akár az ipari forradalom esetében – alapjaiban rengeti meg a társadalmi, politikai és gazdasági szférát, felforgatva a status quo-t. A változáshoz való alkalmazkodás proaktív, tájékozott és inkluzív politikát és gyakorlatokat igényel, amelyek a társadalom minden szegmensének javára hasznosíthatják a mesterséges intelligenciában rejlő lehetőségeket. Az MI-alapú rendszerek elterjedése befolyásolja a munkaerőpiacot, az oktatást, az egészségügyet, a közlekedést és számos más társadalmi szférát. A különböző szakterületek – mint például szociológia, közgazdaságtan, politológia és etika – saját szempontjaik szerint értékelik ezeket a hatásokat. A technológiával kapcsolatos dilemmák, mint például adatvédelem, átláthatóság, elszámoltathatóság és diszkrimináció megelőzése, olyan kérdések, amelyek csak etikai, jogi és filozófiai szempontok integrálásával kezelhetők. Az egyértelmű, jogilag szabályozható

keretek kialakításához jogi és etikai területek szakértőinek együttműködése szükséges. Az MI-rendszerek bevezetéséhez nélkülözhetetlen a társadalmi bizalom megléte – a technológiai újításokkal szembeni féltelmek csökkentése és a társadalmi elfogadottság növelése érdekében fontos, hogy a szociológiai és pszichológiai nézőpontok is megjelenjenek. A szociálpolitikai és közgazdasági szemlélet egyaránt szükséges annak vizsgálatához, hogy az MI hogyan járul hozzá a társadalmi igazságossághoz, vagy hogyan mélyíti az egyenlőtlenségeket. Mindeközben pedig nem szabad megfeledkezni magáról a technológiáról: az említett területek képviselőinek legalább alapvető szinten tisztában kell lenni a technológiában rejlő kockázatokkal, és szoros kommunikációban az MI-alapú rendszerek fejlesztőivel, hogy felelősségteljesen tudjanak döntéseket hozni a saját területükön.

Problématérkép (A különböző pontok közötti összefüggések az 1. ábrán láthatóak.)

- (1) Az MI-rendszerek döntéshozatali folyamatai átláthatatlanok: az algoritmusok „fekete doboz” működése nehezíti a technológia feletti kontrollt, mivel az érintett személyek és szabályozók gyakran nem látják, hogyan születnek a döntések.
- (2) Az MI-rendszerekbe sztereotípiák és torzítások épülnek be az algoritmus betanítása során, ami növeli a társadalmi egyenlőtlenségeket, különösen a sérülékeny csoportokat érintve hátrányosan.
- (3) Az EU-ban és Magyarországon is lemaradásban van a szabályozás a technológia gyors fejlődése mögött. A szabályozási keretek kidolgozása hosszú időt vesz igénybe, és a politikai akarat hiánya is hátráltathatja a hatékony intézkedéseket.
- (4) Az MI-alapú technológiák alkalmazása kockázatokkal jár, de a döntéshozók sokszor nem rendelkeznek a szükséges technikai és szabályozási kompetenciákkal ezek felmérésére. Ezért a kockázatkezelés és a felelős MI-használat kialakítása is korlátokba ütközik.
- (5) A digitális szakadék növekedését eredményezheti, hogy a társadalom különböző csoportjai eltérő hozzáféréssel bírnak az MI-technológiákhoz. Az új technológiák elérhetetlensége mélyítheti a társadalmi különbségeket és csökkenti a hátrányos helyzetűek részvételi lehetőségeit.

2. Elméleti keretek

2.1. A mesterséges intelligencia, mint „fekete doboz”

A mesterséges intelligencia, mint „fekete doboz” (Buhrmester et al., 2021) kifejezés arra a jelenségre utal, amikor az MI rendszerek, különösen azok, amelyek összetett gépi tanulási algoritmusokon alapulnak, döntéseket vagy előrejelzéseket hoznak anélkül, hogy átlátható módon felfednék, hogyan jutottak ezekhez a döntésekhez. Így a felhasználók számára nehéz megérteni, hogy miért generáltak egy adott kimenetet (Pedreschi et al., 2019). A jelenleg létező MI-rendszerek esetében sokszor maguk a tervezők sem értik a belső működésüket, ennek következtében az újonnan kialakuló, tanult képességeket esetenként huzamosabb ideig nem fedezik fel. Ha a jelenlegi paradigmából következtetünk a jövőben potenciálisan létrejövő, fejlett mesterséges intelligenciák (Artificial General Intelligence, AGI) működésére, akkor a fejlett rendszerek valószínűleg a mostaniakhoz hasonlóan „fekete dobozok” lesznek (Eth, 2023). Ez a fajta működés elrejtheti a képzési adatokban vagy magában az algoritmusban rejlő elfogultságokat is, ami tisztességtelen vagy helytelen döntésekhez vezethet (John-Mathews, 2021). Az átláthatóság hiánya több okból is problémás: többek között a bizalom hiánya, az elfogultságok azonosításának nehézsége, valamint a rendszer hibakeresése és javítása során felmerülő kihívások miatt. Az úgynevezett „megmagyarázható mesterséges intelligencia” (Explainable AI/XAI) (Nauta et al., 2022) kezdeményezések, illetve a „mechanikai értelmezhetőség” (Mechanistic Interpretability) (Nanda, 2023) kutatások célja olyan technikák és eszközök kifejlesztése, amelyek átláthatóbbá és érthetőbbé teszik, hogy az MI-modellek, különösen a mély neurális hálózatok, hogyan hoznak létre döntéseket vagy válaszokat a bennük rejlő struktúrák és algoritmusok alapján. A technikák közé tartozik a rétegek és neuronok szerepének vizsgálata, a súlyok és aktivációk elemzése, valamint a modellek döntéshozatali folyamatának vizualizációja. Az értelmezhetőség növelésével a fejlesztők képesek lesznek jobban felmérni és kezelni a potenciális hibákat vagy elfogultságokat, valamint finomhangolni a MI modellek teljesítményét, ezáltal kezelve a „fekete doboz” problémáját (Adadi & Berrada, 2018).

2.2. Az „MI triász”

„Egyetlen mondatban összefoglalható a modern mesterséges intelligencia összetettsége: a gépi tanulási rendszerek számítási teljesítményt használnak az adatokból tanuló algoritmusok végrehajtására.” (Buchanan, 2020) – milyen egyszerűen hangzik, és mégis, a mesterséges intelligencia alkalmazásainak sokszínűségének és műszaki mélységének megértése sok esetben túlmutat a nem-szakértő döntéshozók jelenlegi megértési horizontján. A jó mesterséges intelligencia politika a mögöttes technológia demisztifikálásával kezdődik (Nanda, 2023), ezért fontos a „fekete doboz” kinyitása, tehát a technológia bizonyos fokú megértése az összes aktor számára, akik részt vesznek az MI-alapú rendszereket érintő folyamatokban a létrehozástól a használatig. A nagyméretű nyelvi modellek (Large Language Models, LLMs) egyre nagyobb képességeket, de ezzel arányosan nagyobb torzításokat is mutatnak, amelyek összetettsége kihívást jelent a politikai döntéshozók számára az innovációk holisztikus értékelésében. Ez különösen igaz a nemzetek közötti együttműködésekre, ahol implicit kihívást jelent a hatékony kommunikáció és tájékoztatás biztosítása minden érintett fél számára. Ahogy az MI alkalmazások terjednek, úgy növekszik az etikai aggályok súlya is. A fejlesztések döntő többsége az angol nyelvű rendszerekre koncentrált, figyelmen kívül hagyva a más nyelvű adatkészletek sajátos kihívásait. Mindeközben, noha a mesterséges intelligenciába történő magánberuházások dinamikusan növekszenek, a szabályozói intézkedések nem követik ezt a tempót. A mesterséges intelligenciával kapcsolatos elfogadott törvényjavaslatok száma jelentősen elmarad az indítványozottaktól, ami a szakpolitikai döntések elfogadásának vagy végrehajtásának késlekedésére utal. A technikai innovációk rohamos ütemű változása arra figyelmeztet, hogy a fejlődés gyorsasága felülmúlhatja a jelenlegi jogalkotási és szabályozási kereteket (Zhang et al., 2022). Ezt a dilemmát egyes érdekcsoportok az innováció korlátozásával vagy teljes befagyasztásával kívánják megoldani – ez látszólag feloldhatatlan ellentétet eredményez a technooptimista és -pessimista hangok közt.

Felelős döntéshozóként, a doboz kinyitásához vezető első lépés az „MI-triász” koncepciójának megértése lenne. A modern mesterséges intelligencia paradigmáját a gépi tanulás határozza meg, azaz amikor a gépek képesek önállóan következtetéseket levonni az adatokból. Ennek magját a gépi tanulási rendszerek által használt számítási teljesítmény adja, amely az adatokból tanuló algoritmusok végrehajtását teszi lehetővé. Ez a folyamat alapvetően három összetevőre épül: adatokra, modellekre és számítási kapacitásra, amelyek

együttesen alkotják az úgynevezett „MI-triász”. Az adatok szolgáltatják az MI-rendszerek számára a tanuláshoz szükséges információkat. A képzési adatok minősége és mennyisége közvetlen hatással van a mesterséges intelligencia teljesítményére; minél jobb minőségű az adat, annál pontosabb és megbízhatóbb az MI modell. Ezen adatok gyűjtése, feldolgozása és védelme kritikus kérdéseket vet fel, különösen az adatvédelem és adatetika területén. A modellek fejlődése lehetővé tette, hogy a mesterséges intelligencia széles körben alkalmazható legyen olyan összetett feladatokra is, mint a nyelvi fordítás, a képfelismerés és az autonóm döntéshozatal. A modellek összetettségének növekedésével párhuzamosan nőtt a működésükhöz szükséges számítási erőforrások igénye is. Az innovációk, mint a mély neurális hálózatok és transzformátorok, forradalmasították az MI képességeit, fokozva azok hatékonyságát és alkalmazási területeit. A harmadik összetevő a számítási teljesítmény: az összetettebb modellek és nagyobb adathalmazok kezeléséhez szükséges kapacitás biztosítása nélkül a mesterséges intelligencia alkalmazása korlátozott lenne. A számítási kapacitás fejlődése – mint például a GPU-k (Graphics processing unit, Grafikus processzor egység) és TPU-k (Tensor Processing Unit, Tenzor-feldolgozó egység) terjedése –, illetve a virtualizáció és a „cloudification”¹ lehetővé tette a meglévő erőforrások sokkal hatékonyabb és olcsóbb kihasználását és azt, hogy a mesterséges intelligencia rendszerek gyorsabban tudjanak adatokat feldolgozni és tanulni. A mesterséges intelligencia fejlődésének három korszaka – a mélytanulás előtti időszak (Pre-Deep Learning Era), a mélytanulás kora (Deep Learning Era) és a nagyléptékű korszak (Large-Scale Era) – különböző mértékben támaszkodnak ezekre az elemekre. Miközben a mélytanulás előtti időszakban a képzési számítások még követték Moore törvényét², a mélytanulás korában és azt követően exponenciális növekedésnek indultak, ami különösen a nagyvállalatok által üzemeltetett nagyméretű rendszerekben vált szembetűnővé (OpenAI, 2018; Sevilla et al., 2022).

Mindezek tükrében a döntéshozóknak és szabályozóknak alaposan mérlegelniük kell, hogy melyik területen, mikor és milyen mértékű beavatkozást alkalmaznak, a chipek exportellenőrzésének szigorításától az algoritmusok fejlesztésének támogatásáig. A politikai

¹ A cloudification magyarul leginkább felhősítésként vagy felhőalapúvá tételként fordítható. Ez a kifejezés azt a folyamatot írja le, amikor egy vállalat vagy szervezet az informatikai rendszereit, alkalmazásait, adatbázisait és infrastruktúráját a felhőbe költözteti, vagyis átáll a felhőalapú megoldásokra. A felhősítés célja, hogy a hagyományos, helyben telepített infrastruktúrák helyett rugalmasabb, költséghatékonyabb és skálázhatóbb felhőalapú szolgáltatásokat használjanak, amelyek könnyebben hozzáférhetőek, biztonságosabbak, és egyszerűsítik az üzemeltetést.

² Moore-törvényként ismert az a megfigyelés, hogy a számítógépes chipek tranzisztorainak száma körülbelül kétfévente megduplázódik. Forrás: <https://ourworldindata.org/moores-law>, Letöltve: 2024.04.20.

stratégiák kialakítása során figyelembe kell venniük, hogy a számítási teljesítmény, a modellek fejlesztése és az adatok kezelése milyen etikai és biztonsági kihívásokkal járhat.

2.3. Etikai alapelvek

Az etika szerepe általánosságban azon értékek és elvek kijelölése, melyek az emberek együttélését szabályozzák, illetve egy univerzális fogalomrendszer kialakításával egy közös kommunikációs alap teremtése. Amikor MI etikáról beszélünk, akkor az MI és az emberi társadalom „együttéléséről” gondolkodunk³ és az ezzel kapcsolatos fogalmakat tisztázzuk.

Az etikai hatáselemzés (Ethical Impact Analysis, EIA) a digitális technológiák, áruk, szolgáltatások és gyakorlatok etikai hatásainak értékelése, amely kulcsfontosságú a technológia jövőjének etikai szempontú megvitatásához, előrejelzéséhez és alakításához (Floridi, 2023). Ez a megközelítés azt hangsúlyozza, hogy túl kell lépni azon, hogy egyszerűen megjósoljuk, mely digitális innovációk lesznek sikeresek, ehelyett arra kell összpontosítanunk, hogy etikai megfontolások alapján melyeknek kellene sikeresek lenniük. Prioritásként kell meghatározni, hogy mi az, ami környezeti szempontból fenntartható, társadalmilag elfogadható és ideális esetben társadalmilag előnyös. [Weber szerint a szubjektív értékek és etikai normák azok, amelyek képesek alapvetően átalakítani a társadalmi viselkedést és a racionalizáció irányát, nem csupán a praktikus vagy formális racionalitások. Ez az értékracionalitás alapvetően különbözik a formális vagy gyakorlati racionalitástól, és olyan irányokat hozhat létre, amelyek megfelelnek a társadalmi etikai ideáloknak, így biztosítva a társadalmi elfogadhatóságot és a fenntarthatóságot, ami párhuzamba állítható az említett megközelítéssel, amely az etikai szempontok alapján történő sikeresség prioritását hangsúlyozza (Kalberg, 1980).] A digitális etika az értékek, elvek, választások, ajánlások és korlátok befolyásolásán keresztül jelentős hatást gyakorol a technológiai irányításra. Formálja a közvéleményt, a politikailag megvalósíthatót, a jogilag kikényszeríthetőt, és végső soron az egyének és szervezetek cselekedeteit. Floridi amellet érvel, hogy az etikai értékekkel összhangban lévő információs társadalmat kell kiépítenünk, ahelyett, hogy az egyes technológiákkal szembeni negatív visszajelzésekre reagálnánk.

³ Forrás: Genevieve Bell: 6 big ethical questions about the future of AI.
<https://www.youtube.com/watch?v=UGHzKaAOOcA>, Letöltve: 2024.10.09.

Jelenleg kevés olyan értékelés létezik, amely azt vizsgálja, hogy a mesterséges intelligencia rendszerek mennyire viselkednek etikus módon, amikor társadalmi környezetben kell működniük. A 2023-ban bemutatott MACHIAVELLI egy mérőeszköz, amely több, mint félmillió különböző társadalmi döntéshozatali szituáción keresztül vizsgálja, hogy ha a rendszerek döntéshozatali helyzetekbe kerülnének milyen mértékben keresnek hatalmat, alkalmaznak megtévesztést, idéznek elő kárt és sértenek etikai elveket. A modelleknek a jutalom (játékpontszámok) és az etikus döntéshozatal között kell választaniuk, és egy pontszámnövelésre hajló modellnél előfordulhat, hogy a jutalom maximalizálása érdekében figyelmen kívül hagyja az etikai elveket (Zhang et al., 2022). Magyarul: ha azért kapna jutalmat a rendszer, hogy az emberiség ellen cselekedjen, előfordulhatna, hogy a jutalmat választaná.

2.4. Kognitív torzítások

A kognitív torzítások a normálistól vagy a racionalitástól való szisztematikus eltérés mintái az ítékezésben, ahol az egyének a bemenet észleléséből saját „szubjektív valóságot” hoznak létre. Ezek az előítéletek gyakran az agynak az információfeldolgozás egyszerűsítésére tett kísérletének az eredményei, és számos tényező befolyásolja őket, beleértve az érzelmi és erkölcsi motivációkat. A mesterséges intelligencia területén a kognitív előítéletek kritikus jelentőségűek, mivel a felhasznált adatokon vagy algoritmusokon keresztül akaratlanul is az MI-rendszerekbe kódolódnak, így befolyásolhatják a döntéshozatali folyamatokat és állandósíthatják a meglévő emberi előítéleteket (Rastogi et al., 2020)⁴. Az adatok általában tükrözik az emberi társadalom múltbeli és jelenlegi előítéleteit, a strukturális egyenlőtlenségeket, illetve az adatokat összegyűjtő emberek maguk is elfogultak lehetnek, és ez a torzítás végül az MI-rendszerekbe is beszivárog. Ennek kezeléséhez elsősorban tiszta, előítéletmentes képzési adatokra lenne szükség, illetve szigorú felügyeletre, amely segít a rendszer torzító magatartásának kiszűrésében és korrigálásában (Soleimani et al., 2021). A teljesen torzításmentes adathalmazok létrehozása rendkívül nehéz vagy talán lehetetlen. A tudósok és fejlesztők különféle technikákkal próbálják minimalizálni a torzító hatást, például az adatok átvizsgálásával és tisztításával, az algoritmusok folyamatos monitorozásával, reprezentatív

⁴ Ugyanakkor az MI-rendszerekbe történő kognitív torzítások tudatos integrálása új lehetőségeket nyit meg a mesterséges általános intelligencia (AGI) felé vezető úton, ahol ezeket a torzulásokat prioritásként kezelik, és potenciálisan áthidalhatják a kognitív szakadékokat (Potapov et al., 2012). A kognitív és etikai előítéletek szándékos beépítése az MI-rendszerekbe pedig javíthatja ezek döntéshozatali képességeit komplex, valóságos helyzetekben, bár ez etikai kérdéseket is felvet (Fabi & Hagendorff, 2022).

mintavétellel, több forrásból származó adatok integrálásával és „fairness-fokozó” technikákkal (fairness-enhancing, mely során az algoritmusokat „fair” eredményekre kényszerítik) (Binns, 2018, Mehrabi et al., 2021, Geiger et al., 2020).

2.5. Sötét minták

A sötét minták (dark patterns) olyan tervezési elemek, amelyeket arra használnak a szolgáltatók, hogy a felhasználókat olyan döntések meghozatalára készítsék, amelyek nem az ő érdekeiket szolgálják, hanem jellemzően a szolgáltató üzleti érdekeinek kedveznek (Federal Trade Commission, 2022). Sok esetben már az adatgyűjtési fázisban beavatkoznak a felhasználói szokásokba, kihasználják a kognitív torzításokat és hatást gyakorolnak a döntéshozatali mechanizmusokra. Bár a sötét minták használata a felhasználói élmény tervezés tiltólistáján szerepel, hiszen a döntések megtévesztő tervezési gyakorlatokkal történő irányításával veszélybe kerül a felhasználók autonómiája és magánélete, mégis széles körben használják a vállalatok. A sötét minták nem korlátozódnak csupán a felhasználói felületekre: a programkódokban rejlő sötét minták még alattomosabbak lehetnek, mivel teljesen láthatatlanok a felhasználók számára és csak technikai auditálással fedezhetők fel, ezen folyamatok pedig bonyolultak és költségesek lehetnek. Jogi szempontból a sötét minták használata az adatvédelmet és a magánélet védelmét is sérti. A mesterséges intelligencia bevonódásával pedig a felhasználók manipulációjának egyre kifinomultabb formái jelennek meg. A kifejezetten MI-alapú sötét mintáknak két fő típusa ismert: a „deepfake”⁵ (manipulált képi, hang- és videótartalom) és az antropomorfizmus („Artificial Emotional Intelligence” (AEI) vagy antropomorfizált MI chatbotok) (Jarovsky, 2023, Li et al., 2023).

2.6. Az MI-alapú rendszerek demokratizálása

A fentiekén túl további dilemmákat vet fel, hogy amennyiben a technológiai innovációt nem demokratikus keretek között alkalmazzuk, az újítás könnyen az elnyomás eszközévé válhat. A technológia mindenütt jelen van az életünkben, és döntéseinket, identitásunkat, közösségeinket formálja, a technológiai fejlődés mégsem vezet automatikusan társadalmi előrehaladáshoz. Bár a részvételi demokrácia elméletben

⁵ „A kifejezés az angol deep learning (= *mély tanulás*) és fake (= *hamis*) szavak összetétele. A mély tanulás kifejezés onnan ered, hogy a technika során használt mesterséges neurális hálózat több egymásra épülő rétegből épül fel (azaz "mély").”, Forrás: <https://lexiq.hu/deepfake>, Letöltve: 2024.05.05.

ideálisnak tűnhet, a gyakorlatban sokszor nehezen megvalósítható, ami különösen igaz a technológia és a digitalizáció terén (Eubanks, 2012). A politikai és társadalmi döntésekbe való széleskörű bevonás nélkül a technológiai fejlesztések gyakran csak a hatalmon lévők és gazdaságilag domináns csoportok érdekeit szolgálják, miközben marginalizálják az alacsonyabb társadalmi rétegeket és növelik az egyenlőtlenségeket. A technológiai hozzáférés kérdéseit gyakran kezelik elosztási problémaként, amely az anyagi javak egyenlőtlen megosztását idézi elő, és ezzel korlátozza a társadalmi mobilitás lehetőségeit. A „digitális szakadék” fogalma például azt feltételezi, hogy a technológiai fejlődés mindenki számára előnyös, azonban ez a nézet figyelmen kívül hagyja, hogy a technológia gyakran reprodukálja és mélyíti a fennálló társadalmi és gazdasági egyenlőtlenségeket. A technológia demokratikus integrációjának hiánya hozzájárulhat a „szegénységi kultúra” fenntartásához, amelyben a marginalizált csoportok továbbra is kizsákmányolva és hatalom nélkül maradnak. A technológiai innovációk demokratikus keretek közötti alkalmazása elengedhetetlen ahhoz, hogy ne váljanak az elnyomás eszközeivé. Ennek érdekében szükség van arra, hogy a technológia feletti kontrollt a polgárok széles köre gyakorolja, lehetővé téve számukra, hogy aktívan részt vegyenek az azokról szóló döntésekben, amelyek közvetlenül vagy közvetetten befolyásolják az életüket. Ez magában foglalja a kritikai gondolkodás fejlesztését és a technológia demokratizálását célzó kollektív erőfeszítéseket.

A mesterséges intelligencia demokratizálása azt jelenti, hogy (1) az emberek széles körének lehetővé kell tenni, hogy az MI-alapú alkalmazásokat használhassa és profitálhasson belőlük; (2) az emberek szélesebb körét kell hozzásegíteni, hogy részt vehessen a tervezési és fejlesztési folyamatokban; (3) elő kell segíteni a fejlett MI-rendszereket és képességeket kiépítő és ellenőrző szervezetek által felhalmozott értékek széles körű és méltányos elosztását; (4) az MI irányításával kapcsolatos döntésekben az érdekelték és az érintett népesség szélesebb körének kell részvételt biztosítani. A mesterséges intelligenciával kapcsolatos döntések jellemzően arról szólnak, hogy hogyan találjuk meg az egyensúlyt a vele járó kockázatok és előnyök között, és ezek a döntések határozzák meg, hogy ki és hogyan használhatja, fejlesztheti, terjesztheti és szabályozhatja az MI-t (Seeger et al., 2023).

2.7. MI ártalmak osztályozása

Az MI-alapú rendszerek okozta károk több dimenzióban érinthetik az embereket, különösen a sérülékeny csoportokat, de kihívást jelenthetnek a környezet, az infrastruktúra, a tulajdon és a társadalmi struktúrák számára is (Frase & Daniels, 2023). Ha az MI-

forradalom az ipari forradalomhoz hasonló mértékű lesz, hatalmas változásokat hozhat az emberi társadalomban, és ehhez nem kell emberi szintű kognitív képességeket sem elérnie⁶ (Gruetzemacher & Whittlestone, 2022). Mokyr (2005, idézi Gruetzemacher & Whittlestone, 2022:5) szerint úgy kell erre gondolnunk, mint „egy olyan gazdaságtörténeti fázisátmenetre, amelyben a régi paraméterek már nem érvényesek, és amelyben a rendszer dinamikája megváltoztathatatlanul megváltozott”. Amennyiben nem megfelelően kezeljük a technológiát, rendkívül nagy pusztító potenciállal bír, hasonlóan a nukleáris fegyverekhez. Az MI fejlődése gyors és kiszámíthatatlan, ezért, ha késlekedünk a szabályozásával vagy a kockázatok kezelésével, már túl késő lehet. A legnagyobb társadalmi kockázatok súlyos vagy akár katasztrofális következményekkel járhatnak. Ezek közé tartozik a rosszindulatú szereplők MI használatával megvalósított káros vagy pusztító tevékenysége; a versenykényszer, amely arra kényszerítheti a fejlesztőket, hogy biztonsági szempontból nem kellően felkészült MI rendszereket telepítsenek; a szervezeti kockázatok és a rendszerek komplexitása miatt bekövetkező véletlen balesetek; és az intelligensebb MI-rendszerek feletti kontroll elvesztése, amelyek az érdekeinkkel ellentétes célokat kezdenek követni (Rogue AI) (Hendrycks et al., 2023).

A károk rendszerezhetők az alapján, hogy az MI-rendszer szándékosan, nem szándékosan, vagy véletlenül okozott kárt, és megkülönböztethetők az MI-rendszereket érő károktól is, például egy kibertámadás esetén. A szabályozóknak nemcsak a már bekövetkezett károkat kell figyelembe venniük, hanem a lehetséges jövőbeli károkat is, hogy azok ne váljanak gyorsan elavulttá. Az MI-rendszerek gyors fejlődése miatt a múltbéli, valószínűsíthető, feltételezett és spekulatív károk mind relevánsak lehetnek a jövőbeni szabályozások kidolgozásakor. A károk szisztematikus elemzése segíthet a döntéshozóknak az MI potenciáljának biztonságosan használatában, minimalizálva a kockázatokat (Frase & Daniels, 2023).

2.8. Szabályozási alapelvek

Határterületi mesterséges intelligenciáról (frontier AI) akkor beszélünk, ha az alapmodell nagy képességű és komoly veszélyt jelenthetne a közbiztonságra. A határterületi mesterséges intelligencia szabályozási rendszerének kialakításához több alapvető

⁶ A „transzformatív mesterséges intelligencia” az általános célú technológiákhoz, például az elektromossághoz hasonló széles körű változásokat okoz; a „radikálisan átalakuló mesterséges intelligencia” az ipari forradalomhoz hasonló mélyreható társadalmi változásokat idéz elő.

mechanizmusra van szükség. A biztonsági szabványok kidolgozását szakértők vezette, több érdekelt fél részvételével zajló folyamatokon keresztül kellene megvalósítani, amelyeket kormányzati szervek koordinálhatnak. Ezek a szabványok idővel jogi követelményekké válhatnak, amelyek garantálják a határterületi mesterséges intelligencia modellek biztonságos fejlesztését. A szabályozók számára olyan eszközöket kell biztosítani, amelyek lehetővé teszik a fejlesztési folyamatok átláthatóságát, így a közzétételi rendszerek, nyomon követési procedúrák és a bejelentők védelme révén. Továbbá szükség van olyan mechanizmusokra, amelyek elősegítik a biztonsági előírások betartását, beleértve az önkéntes öntanúsítási rendszereket, a felügyeleti hatóságok által végrehajtott ellenőrzéseket és az engedélyezési folyamatokat. Bár az önszabályozás, mint az önkéntes tanúsítás, bizonyos szintű megfelelést biztosíthat, ez általában nem elég kiterjedt a határterületi MI modellek szabályozásához (Anderljung et al., 2023).

2.9. Technikai és működési kockázatok

A technikai és működési kockázatok azok a konkrét mechanizmusok, amelyeken keresztül a mesterséges intelligencia meghibásodhat vagy visszaélések célpontjává válhat (Anderljung et al., 2023:10). Az egyik probléma a rendszerekben kiszámíthatatlanul megjelenő veszélyes képességekkel kapcsolatos, amelyekre a fejlesztők nem gondoltak előre. Ezek a képességek figyelmeztetés nélkül megjelenhetnek a mesterséges intelligencia fejlesztési szakaszában, vagy a telepítés és a működés után, és akár nem szándékos káros cselekvésekhez vagy eredményekhez is vezethetnek. A mesterséges intelligencia-modellek telepítése után kihívást jelent annak biztosítása, hogy ne okozzanak kárt működésük közben. A rendszereket, miután működésbe léptek, folyamatosan ellenőrizni kell, biztosítva, hogy továbbra is biztonságosan és rendeltetésszerűen működjenek a dinamikus valós környezetben. A felügyelet mellett a rendszer a biztonsági intézkedések állandó kiigazítását is igényli az új sebezhetőségek és fenyegetések kezelése érdekében. A határmenti mesterséges intelligencia-modellek, különösen a nagy teljesítményűek, hajlamosak gyorsan elterjedni. Ez a széleskörű elterjedés megnehezítheti az erőfeszítéseket, melyek a tetteikért való elszámoltathatóság fenntartására irányulnak, valamint a hatékony szabályozást és felügyeletet, és ezáltal fokozza a káros incidensek kockázatát.

Komoly hiányosságok vannak a felelősségteljes mesterséges intelligencia értékelésének szabványosításában. Jelenleg a vezető fejlesztők, mint az OpenAI, a Google és az Anthropic, különböző MI benchmarkokat alkalmaznak modelljeik tesztelésére. Ez az eltérő

megközelítés megnehezíti a legjobb MI-modellek kockázatainak és korlátainak egységes és szisztematikus összehasonlítását (Perrault & Clark, 2024).

3. Jelenlegi szabályozás

3.1. A globális konszenzus alapjai

Az MI-szabályzás alapelvei a mesterséges intelligencia felelős használatát hivatottak elősegíteni – ezek jelentik az MI irányításának etikai és jogi „normatív magját”, amely a konszenzus kialakulásának kezdeti jeleit mutatja. Az egyes elvek alkalmazása kontextusfüggő, és nem minden elv egyformán releváns minden helyzetben vagy közönség számára. Az elvek integrálása a szélesebb kormányzati és szakpolitikai keretekbe, valamint a szakmai gyakorlatokba és a mindennapi rutinokba elengedhetetlen lenne a hatékony szabályozás elérése érdekében. Az adatvédelem biztosítja, hogy a felhasználók teljes kontrollt gyakorolhassanak adataik felett, lehetőséget adva azok korlátozására, helyesbítésére és törlésére. Az elszámoltathatóság elve megköveteli, hogy a rendszerek átláthatóak és ellenőrizhetőek legyenek, lehetővé téve a döntések megértését és a felelősség egyértelmű meghatározását. A biztonság és védelem elve alapján a rendszereket úgy kell megtervezni, hogy azok a legjobb kockázatkezelési gyakorlatokat alkalmazzák, a tervezés minden szakaszában figyelembe véve a biztonsági szempontokat. Az átláthatóság és megmagyarázhatóság célja, hogy az MI-rendszerek érthetőek és nyomon követhetőek legyenek. A méltányosság biztosítja, hogy a rendszerek elkerüljék az előítéleteket és egyenlő lehetőségeket teremtsenek mindenki számára. Az emberi kontroll fenntartására szolgáló elv az automatizált döntések felülvizsgálatát és szükség esetén felfüggesztését teszi lehetővé. A szakmai felelősség elve a tervezők számára előírja a rendszerek hatékony és biztonságos működésének biztosítását, figyelembe véve azok hatásait. Végül, az emberi értékek előmozdítása azt kívánja, hogy a rendszerek az emberi jóléthez és társadalmi fenntarthatósághoz járuljanak hozzá, különösen a sebezhető csoportok védelmében (Fjeld et al., 2020).

3.2. Nemzetközi szakpolitikai kitekintés

A mesterséges intelligencián alapuló rendszerek szabályozására a kormányok jogszabályokat, nemzeti stratégiákat, politikákat és iránymutatásokat alkalmaznak; ezek a szabályozások azonban gyakran ellentmondanak egymásnak, és nehézséget okoz számukra,

hogy lépést tartsanak a gyorsan fejlődő technológiával. A kialakulóban lévő stratégiák általában nemzeti stratégiákkal vagy etikai politikákkal kezdődnek, mielőtt a jogalkotási intézkedések felé haladnának. Az innováció és a kockázatkezelés közötti egyensúly megteremtése összetett és jelentős kihívást jelent, amire az egyik lehetséges megoldás a többoldalú koordináció (Zhang et al., 2022:174-195). Az MI irányítási kereteinek kidolgozásában többek közt olyan szervezetek vesznek részt, mint az OECD, az UNESCO, a WIPO és az Afrikai Unió. Az OECD-nek jelenleg hatvanhárom szabályozási kezdeményezése kapcsolódik a mesterséges intelligenciához⁷.

Az Egyesült Államokban, Európában és Kínában az MI fejlesztések és szabályzások három különböző modellje figyelhető meg.

Az amerikai modell a szabadpiaci kapitalizmus több százéves hagyományain alapul: itt a gazdasági verseny áll a középpontban, és az állam alapvetően hagyja, hogy a piac magától működjön, és csak válsághelyzetben lép közbe. Üzleti szemléletű, és nagy mennyiségű adatgyűjtésre épül, ami lehetővé teszi az üzleti szolgáltatások hatékony működését. Az Egyesült Államok elsődlegesen az innováció a technológiai fejlődés ösztönzésére összpontosít az MI-szabályozás során, és támogatja az önszabályozási gyakorlatokat. Célja a versenyképesség és a globális vezető szerep fenntartása, főképp Kínával szemben. Az Egyesült Államokban 2023-ban 56,3 százalékkal nőtt az MI-t érintő szabályozások száma; míg 2016-ban mindössze egyetlen ilyen szabályozás létezett, 2023-ra már huszonötre nőtt ez a szám (2. ábra). A szabályozás tendenciája jelentősen a korlátozás irányába mozdult el: 2023-ban 10 korlátozó MI-szabályozás volt, szemben a mindössze három kiterjesztővel, ezzel szemben 2020-ban négy expanzív és egy korlátozó szabályozás volt. A szabályozás leggyakoribb tárgya a külkereskedelem és a nemzetközi pénzügy volt, ezután az egészségügy; kereskedelem; valamint tudomány, technológia és kommunikáció. 2023 októberében Joe Biden elnök kiadott egy végrehajtási rendeletet, amely a magas kockázatú MI-rendszerek biztonsági tesztelését és a kormányzati felhasználás szabványosítását írta elő. A rendelet célja továbbá az amerikai MI-ipar innovációjának támogatása, különösen a védelmi szektorban, illetve a távközlési és oktatási szektorban alkalmazott MI felelősségteljes használatának biztosítása. A Biden-Harris adminisztráció továbbá önkéntes irányelveket alakított ki olyan vállalatok számára, mint a Google és a Microsoft, hogy

⁷ Forrás: <https://oecd.ai/en/dashboards/policy-initiatives?conceptUri=http%3F%2Fkim.oecd.org%2FTaxonomy%2FOrganisations%23EuropeanUnion>,
Letöltve: 2024.11.10.

elősegítsék a biztonságos és megbízható MI fejlesztését. Ezek az irányelvek a kockázatok megosztását célozzák meg, illetve a közvélemény tájékoztatását az MI-alapú tartalmakról (Perrault & Clark, 2024). Idén azonban érdekes fordulatot hozhat az USA mesterséges intelligencia politikájában a dereguláció-párti Donald Trump győzelme, aki már a kampányában jelezte, hogy visszavonná Joe Biden 2023 októberében kiadott MI-rendeletét. Ugyanakkor Trump is aggodalmát fejezte ki az MI potenciális veszélyei, például a deepfake technológia miatt, amely szerinte akár nukleáris konfliktust is kiválthat⁸.

Kína számára is fontos a gazdasági verseny, hiszen ez jelentős előnyt biztosíthat számára globális szinten. A kínai modell robusztus adatfelhasználást tesz lehetővé kevesebb személyiségi jogi korlátozással, ami hatékonyabbá teszi az MI-t, de kevésbé védi az egyéni jogokat. Kína szabályozása elsődlegesen a tartalomkontrollra és az információáramlás irányítására összpontosít, emellett iteratív és vertikális jellegű, célzott intézkedések sorozatából áll. A generatív MI-rendszereknél megengedőbb, például az adathasználat és az átláthatóság terén, a szigorúbb rendelkezések elsősorban az információs környezet irányítására és a politikai stabilitás fenntartására irányulnak (Sheenan, 2023). Kína szabályozást vezetett be a „mélyszintézis” technológiákra vonatkozóan annak érdekében, hogy kezelje a valósághű virtuális entitások és multimodális médiumok – beleértve a deepfake-eket is – létrehozásával kapcsolatos biztonsági kérdéseket. Az új rendeletek a szolgáltatókra és a felhasználókra egyaránt vonatkoznak, és olyan intézkedéseket írnak elő, mint az illegális tartalmak megelőzése, a jogi előírások betartása, a felhasználók személyazonosságának ellenőrzése, a biometrikus azonosításhoz való hozzájárulás biztosítása, az adatbiztonság védelme és a tartalmak moderálásának szigorítása. Kína 2023-ban frissítette a generatív mesterséges intelligenciára vonatkozó kibertér-szabályozását is, általános szabályozás helyett célzottabb megközelítést alkalmazva. Az új módosítások enyhítették a szabályozási nyelvezetet és az előírásokat, például az “adatok igazságának, pontosságának, objektivitásának és sokszínűségének biztosítása” most inkább az irányelvek szintjén jelenik meg. Az átdolgozott rendeletek a generatív mesterséges intelligencia fejlesztését is ösztönzik, eltávolodva a korábbi, büntető jellegű megközelítéstől (Perrault & Clark, 2024).

Európában irányított kapitalizmus működik: a jogalkotók előre igyekeznek felismerni a lehetséges problémákat a piacon, és már előzetesen beavatkoznak, hogy a gazdasági versenyt

⁸ Forrás: <https://time.com/6988935/trump-worries-ai-deepfakes-trigger-nuclear-war/>, Letöltve: 2024.11.10.

szabályozzák, nehogy a teljes piaci szabadság káros hatásokhoz vezessen. Az európai modell középpontjában a személyes jogok védelme, az etikai és társadalmilag felelős adathasználat áll, ami szigorúbb szabályozást eredményez. Ez azonban lassítja a technológiai fejlődést és nehezíti az innovációt, mert a fejlesztőknek számos előírást kell betartaniuk, például a GDPR-t.

3.3. Szabályozás az Európai Unióban

Az EU AI Act előfutárában, a 2020-ban kiadott „Fehér könyv a mesterséges intelligenciáról: a kiválóság és a bizalom európai megközelítése” című dokumentumban deklarált egyik uniós szintű cél egy olyan szakpolitikai keret létrehozása volt, amely összehangolja az erőfeszítéseket európai, nemzeti és regionális szinten, erőforrásokat mozgósítva a magán- és közszféra együttműködésével, hogy egy „kiválósági ökoszisztémát”⁹ hozzon létre az értéklánc teljes spektrumán. A másik cél egy „bizalmi ökoszisztéma” (ez maga a szabályozási keret) kialakítása, amely biztosítja az uniós szabályok betartását, különös tekintettel a nagy kockázatú MI-rendszerekre, és bizalmat épít a polgároknak, valamint jogbiztonságot nyújt a vállalatoknak és állami szervezeteknek az MI alkalmazásában (ezt a célt fedi le az EU AI Act). A bizalom építése tulajdonképpen az előfeltétele annak, hogy széles körben el tudjon terjedni a technológia, illetve hangsúlyozza a polgárok, vállalkozások, közjó és a teljes társadalom érintettségét. A Bizottság a dokumentum alapján „*hatékony*” és „*nem túlzottan előíró jellegű*” szabályozási keretet kívánt létrehozni, ehhez találták a legalkalmasabbnak a kockázatalapú megközelítést, amely egyúttal az arányos beavatkozást is biztosítani hivatott (Európai Bizottság, 2020:21-22). Az MI-alkalmazásokat magas kockázatúnak kell tekinteni, amennyiben olyan területen használják, ahol általában nagy kockázatok fordulnak elő (pl.: egészségügy, közlekedés, energia, közszolgáltatások, menekültügy, vagy igazságszolgáltatás), illetve olyan módon használják, hogy jelentős kockázatot jelent az emberekre (pl.: jogi hatásokkal, sérülésekkel, vagy anyagi károkkal járhat).

Az Európai Unió 2023-ban több jelentős, a mesterséges intelligenciával kapcsolatos politikát fogadott el (3. ábra, Perrault & Clark, 2024); ezek leggyakoribb tárgya a tudomány,

⁹ A „kiválósági ökoszisztéma” ebben a kontextusban olyan szakpolitikai és operatív keretet jelent, amely ösztönzi a kutatást, az innovációt és a technológiai fejlesztéseket, erőforrásokat mozgósít, és lehetővé teszi a köz- és magánszféra együttműködését. Célja, hogy az EU-ban a mesterséges intelligencia fejlesztése és alkalmazása minden szinten (kutatástól a piaci alkalmazásig) előmozdítsa a versenyképességet és a fenntartható fejlődést.

technológia és kommunikáció; ezt követi a kormányzati működés és politika¹⁰. 2024-ben pedig jelentős lépést tett előre az EU mesterséges intelligenciáról szóló törvényének kidolgozásával, mely a maga nemében novum. Ez a mérőkövető jogszabály, amely 2023. december 8-án politikai megállapodásra jutott, 2024. március 13-án megszavazásra került az Európai Parlament által, 2024. augusztus 1-én pedig hatályba lépett, globális precedenst teremt az MI-alapú technológiák irányítása terén. A törvény kettős céllal készült: az egyik az uniós értékeknek megfelelő (tehát emberközpontú és megbízható) MI-technológiák fejlesztésének, forgalmazásának és használatának előmozdítása, a másik az egészség, a biztonság és az alapvető jogok magas szintű védelmének biztosítása.

A törvény középpontjában a MI-rendszerek kockázati szintjük szerinti besorolása áll, amely különleges követelményeket ír elő a fejlesztésükre, telepítésükre és használatukra vonatkozóan. Az egyik legfontosabb rendelkezés az elfogadhatatlan kockázatot jelentőnek ítélt MI-rendszerek – például a társadalmi pontozásra, a munkahelyi és oktatási környezetben történő érzelmefelismerésre, valamint a nyilvános helyeken történő, valós idejű biometrikus azonosításra alkalmas rendszerek – teljes betiltása. A törvény előírja, hogy az MI-rendszereknek egyértelműen fel kell tárniuk nem-emberi mivoltukat, így növelve az átláthatóságot, amikor az emberek MI által vezérelt szolgáltatásokkal vagy chatbotokkal kerülnek kapcsolatba. Az AI Act keretében az állampolgárok jogot kapnak arra, hogy panaszt tegyenek és jogorvoslatot kérjenek, ha úgy érzik, hogy egy MI-rendszer hátrányosan érintette őket, ezáltal javítva az elszámoltathatóságot. Az MI-alapú alkalmazásokat fejlesztő vállalatokra nagyobb átláthatósági kötelezettség is vonatkozik majd; kötelező lesz számukra, hogy nyilvánosságra hozzák adathasználati és algoritmusi gyakorlatukat, különösen a magas kockázatú alkalmazásaik esetében (Heikkilä, 2024). A jogszabály jelentős szankciókat ír elő a be nem tartása esetére: a jogsértésekért akár a globális éves forgalom 1,5 százalékától (vagy 7,5 millió eurótól) 7 százalékáig (35 millió euróig) elérő bírság is kiszabható. Szigorú követelményeket fogalmaz meg a magas kockázatú mesterséges intelligencia rendszerekre vonatkozóan is, beleértve a kötelező adatminőségi ellenőrzéseket, átláthatósági intézkedéseket, emberi felügyeletet és megfelelőség-értékelést. Az alacsonyabb kockázatú és általános célú MI-modellek esetében a törvény kevésbé szigorú, de még mindig jelentős kötelezettségeket ír elő, például az iparág által kidolgozott

¹⁰ A kormányzati szabályozások célja a kormányok működésének és a kapcsolódó folyamatoknak az irányítása. Az Európai Bizottság egyik rendelete az inkluzív és rugalmas választási folyamatok támogatását célozta az Unión belül, valamint a választások legitimitásának megőrzését a mesterséges intelligencia politikai félretájékoztatásra való esetleges felhasználásával szemben (Perrault & Clark, 2024).

magatartási kódexek betartását, vagy annak biztosítását, hogy a felhasználók tisztában legyenek azzal, hogy mikor lépnek kapcsolatba egy mesterséges intelligencia rendszerrel.

A törvény végrehajtása fokozatosan történik, a különböző rendelkezések a jogszabály hatálybalépését követő hat, tizenkettő, huszonnégy és harminchat hónap elteltével válnak alkalmazandóvá. Rendelkezései huszonnégy hónappal később lépnek hatályba. Ez alól az időzítés alól kivételt képez a MI Paktum bevezetése, amely a Bizottság kezdeményezése, és amely arra ösztönzi a MI fejlesztőket, hogy önkéntesen tartsák be a törvényben foglalt kötelezettségeket, még a törvény hivatalos alkalmazása előtt. Az első évben a jogszabály kötelezettségeket ír elő az általános célú MI-modellek szolgáltatói számára, és létrehozza a tagállami illetékes hatóságok kijelölését. Ez a kezdeti szakasz tartalmazza a tiltott mesterséges intelligenciával kapcsolatos gyakorlatok éves felülvizsgálati folyamatának megkezdését is. A végrehajtás után tizennyolc hónappal a Bizottság a forgalomba hozatal utáni nyomon követésre vonatkozó végrehajtási jogi aktusokra fog összpontosítani. A kétéves határidőig a jogszabály a magas kockázatú MI-rendszerekre vonatkozó, a jogszabály III. mellékletében meghatározott kötelezettségeket hajtja végre. A tagállamoknak emellett szankciókra vonatkozó szabályokat kell alkalmazniuk, beleértve a közigazgatási bírságokat, és működőképes mesterséges intelligencia „szabályozási homokozókat” (regulatory sandbox)¹¹ kell létrehozniuk, elősegítve ezzel a biztonságos innovációs környezetet. Három évvel a jogszabály hatályba lépése után a kötelezettségek kiterjednek a korábban nem érintett, de bizonyos termékek biztonsági összetevőjeként használt, magas kockázatú MI-rendszerekre is, ami jól szemlélteti a jogszabály átfogó hatályát. A törvény 2030 végére kötelezettségeket ír elő az uniós jog által létrehozott nagyméretű IT-rendszerek alkotóelemeihez tartozó mesterséges intelligenciát alkalmazó rendszerekre.

A jogszabály több szervezeti egységet is létrehoz, köztük a MI Irodát (AI Office), a MI Tanácsot (AI Board), a Tanácsadó Fórumot (Advisory Forum) és a Tudományos Testületet (Scientific Panel). Ezen szervezetek feladata az MI-modellek általános és konkrét alkalmazásainak felügyelete, a szabványok és tesztelési gyakorlatok kidolgozásához való hozzájárulás, valamint a szabályok tagállami szintű érvényesítése. A jogszabály felhatalmazza a Bizottságot arra, hogy egy kezdeti és egy meghosszabbítható ötéves időszak

¹¹ „A szabályozási homokozó olyan eszköz, amely lehetővé teszi a vállalkozások számára, hogy új és innovatív termékekkel, szolgáltatásokkal vagy üzletekkel kísérletezzenek a szabályozó hatóság felügyelete mellett.”, Forrás:

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI\(2022\)733544_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf),

Letöltve: 2024.11.10.

alatt felhatalmazáson alapuló jogi aktusokat adjon ki számos témában, a mesterséges intelligenciát alkalmazó rendszerek meghatározásától a magas kockázatú mesterséges intelligenciára vonatkozó kritériumokig.

3.4. Az AI Act kritikája

Az EU AI Act-et elsősorban bonyolultsága és áttekinthetetlen mivolta miatt kritizálják. Az Act nem mindig egyértelmű abban, hogy mi is pontosan a mesterséges intelligencia, és mi tartozik a szabályozás hatálya alá. Az MI-t érintő szabályozás túl széleskörű, és olyan technológiákra is vonatkozhat, amelyek nem feltétlenül jelentenek etikai vagy társadalmi kockázatot, és vice versa. Ezzel kapcsolatban felmerül, hogy a magas kockázatú MI-rendszerek túl széles kategóriájába kerülnek olyan alkalmazások is, amelyek nem érdemelnék meg a szigorú szabályozást vagy éppen egy magas kockázatú rendszer csúszhat bele kiskapuk miatt az alacsony kockázatú kategóriába. Az egyértelműség hiánya akadályozhatja a hatékony végrehajtást és a jogszabály betartását. Bár a szabályozás célja a kockázatok csökkentése, kritikusok szerint az AI Act túlságosan óvatos, csak a veszélyek és kockázatok kezelésére fókuszál, miközben nem támogatja eléggé az innovatív fejlesztéseket. A törvény túlságosan merev kereteket állít egy rendkívül gyorsan változó technológiára, és nem biztosít elegendő rugalmasságot ahhoz, hogy a változásokra gyorsan reagáljon, ami akadályozhatja az innovációt. Az MI fejlesztése és alkalmazása rohamosan fejlődik, míg a szabályozás gyakran elavulttá válik, mire ténylegesen alkalmazni kezdik. Az EU AI Act előírja, hogy bizonyos MI-rendszereknek kötelező auditokon kell átesniük, hogy megfeleljenek az előírásoknak. Az auditálás és a felügyeleti mechanizmusok azonban kihívást jelenthetnek a komplex és folyamatosan fejlődő rendszerek esetében. Az MI-rendszerek átláthatatlansága, a „fekete doboz” jelenség miatt nehéz lehet bizonyítani, hogy egy MI-algoritmus teljes mértékben megfelel-e a szabályoknak. A törvény betartása különösen a kis- és középvállalkozások számára jelent aránytalan adminisztratív és megfelelési költséget. A követelmények, mint például a kockázatértékelés, auditálás és dokumentáció, hatalmas terheket róhatnak a kisebb szereplőkre, akik nem rendelkeznek azokkal az erőforrásokkal, mint a nagy multinacionális cégek. Ezek a költségek és megfelelési elvárások a fejlesztések Európán kívülre történő áthelyezéséhez vezethetnek. Az EU AI Act hátrányba hozhatja az európai MI-fejlesztést a globális versenytársakkal szemben, különösen az Egyesült Államokkal és Kínával összevetve, ahol a szabályzás lazább vagy teljesen más jellegű. Az Európában működő vállalatoknak ráadásul világszerte

meg kell felelniük a szabályoknak. Ez problémákat vet fel azon nem európai cégek számára is, akik Európában szeretnének MI-technológiákat alkalmazni vagy értékesíteni, és hosszú távon feszültségeket és jogi kihívásokat okozhat a nemzetközi kereskedelemben. A szabályozás érvényesítése és betartása is komoly kihívást jelenthet, különösen, ha a tagállamok között különbségek vannak az implementáció során. A nem egységes végrehajtás az EU különböző országai között jogbizonytalansághoz vezethet, és megnehezítheti a szabályok hatékony alkalmazását.

3.5. Digitális- és MI tudatosság az EU-ban és Magyarországon

Magyarországon a digitális alapkészségekkel rendelkező lakosság aránya nőtt (58,9 százalék), de elmarad az EU 80 százalékos céljától. Az idősebb generációk és hátrányos helyzetű csoportok digitalizációja továbbra is kihívást jelent. Magyarország a mesterséges intelligencia területén leginkább az adatelemzés (Data Analytics) alkalmazásában erős. A magyar vállalatok nagy arányban alkalmaznak adatelemzési technológiákat, és ezen a téren Magyarország az EU-átlag felett teljesít. Az ország adatelemzési alkalmazásának aránya (53,2 százalék) jóval meghaladja az EU átlagát (33,2 százalék). Emellett az MI, felhőszolgáltatások és adatelemzés együttes használata a magyar vállalkozások körében 65,6 százalékos szinten van, ami szintén az EU-átlag felett van (54,6 százalék) (Digital Decade Country Report: Hungary, 2024).

Az alapvető és az átlagon felüli digitális készségeket vizsgálva, a 2024-es DESI (Digital Economy and Society Index) jelentés adatai alapján, az Európai Unió tagországai közül legmagasabb értéket Hollandia, Finnország és Írország érte el – itt a 17-74 éves korosztályba tartozó egyének több, mint 80 százaléka rendelkezik alapvető és több, mint 50 százaléka fejlettebb digitális készségekkel. Magyarország a középmezőnyben található, a legalacsonyabb arányt Bulgária és Románia mutatja mindkét kategóriában (4-5. ábra). Észtországban, Luxemburgban és Írországban a diplomások közel 10 százaléka rendelkezik informatikai (IKT) végzettséggel, Magyarországon ez az arány szintén magas, 7 százalék. A lista végén Olaszország és Ciprus állnak (6. ábra). Az EU átlag fokozatosan emelkedett az évek során, míg Magyarország 2021 környékén csökkent, de azóta gyors növekedés következett be, 2024-re elérve az EU átlagnál jóval magasabb arányt (7. ábra).

Az államok digitális közszolgáltatási¹² transzparenciájának felmérésében Málta, Luxemburg és Észtország kiemelkedően magas pontszámot értek el, közel 100 ponttal, ami azt jelzi, hogy ezekben az országokban az állampolgárok jól informáltak a szolgáltatások tervezéséről és a személyes adataik kezeléséről. Az EU átlaga középtájon helyezkedik el (66,98 ponttal a 0-tól 100-ig terjedő skálán), míg Magyarország az átlag körül teljesít (61,15 pont). A lista végén Ciprus és Románia találhatóak, ahol alacsonyabb a transzparencia szintje (8. ábra). Az EU-ban a transzparencia szintje fokozatosan emelkedett az elmúlt években, míg Magyarországon hullámzó trend figyelhető meg, de összességében szintén növekedés tapasztalható (9. ábra). Az EU törekvései a transzparencia növelésére eredményesek és Magyarország is fejlődést mutat ezen a téren. Az átláthatóság javítása fontos tényező lehet az állampolgárok bizalmának növelésében és a digitális közszolgáltatások szélesebb körű elfogadásában. Az alapvető digitális készségekkel rendelkező személyek aránya és a polgárok számára elérhető digitális közszolgáltatások szintje között pozitív korreláció figyelhető meg, tehát azokban az országokban, ahol magasabb a digitális készségekkel rendelkező személyek aránya, általában a digitális közszolgáltatások szintje is magasabb, és fordítva. Ez a korreláció arra utalhat, hogy a digitális készségek fejlesztése hozzájárul a digitális közszolgáltatások hatékonyabb használatához, illetve az is valószínű, hogy ahol a közszolgáltatások digitális fejlettsége magas, ott a lakosság számára több lehetőség és motiváció áll rendelkezésre a digitális készségek elsajátítására és fejlesztésére (10. ábra).

Dániában, Finnországban és Luxemburgban a vállalatok több, mint 14 százaléka használ MI-technológiákat. Az EU átlaga valamivel 10 százalék fölött van, míg Magyarország az alsó harmadban található, kevesebb mint 6 százalékkal (11. ábra). Az EU átlagában folyamatos emelkedés figyelhető meg, míg Magyarország esetében kisebb növekedés látható (12. ábra). Bár az MI-technológiák használata növekszik Magyarországon, még mindig alacsonyabb szinten áll, mint az EU átlag és jelentős lemaradás tapasztalható a vezető európai országokhoz képest, ahol a vállalatok jóval nagyobb arányban alkalmazzák a technológiát. A vállalkozások számára elérhető digitális közszolgáltatások szintje és a mesterséges intelligenciát alkalmazó vállalkozások aránya között enyhe pozitív korreláció figyelhető meg, ami rámutat, hogy a magasabb digitális fejlettség (legyen az

¹² A digitális közszolgáltatások azok az online elérhető állami és önkormányzati szolgáltatások, amelyeket a kormányok és közintézmények biztosítanak az állampolgárok, vállalkozások és más intézmények számára. Ezek lehetővé teszik, hogy a polgárok és a vállalkozások különféle ügyeket elektronikus úton, kényelmesen és gyorsan intézhessenek, csökkentve az adminisztratív terheket és a személyes megjelenést. A digitális közszolgáltatások kulcsszerepet játszanak a modern állam működésében.

közszolgáltatások vagy KKV-k digitális intenzitása) hozzájárulhat az MI-technológiák szélesebb körű alkalmazásához. Azokban az országokban, ahol magasabb szintű digitális közszolgáltatások érhetők el a vállalkozások számára, ott az MI-t alkalmazó cégek aránya is magasabb. Ez a korreláció arra utal, hogy a fejlett digitális közszolgáltatások jelenléte kedvezhet az MI-technológiák alkalmazásának (13. ábra). [Bizonyos állami intézmények vagy szervezetek – ún. „pilóta szervezetek” – a gazdasági bürokrácia kulcselemeiként vezető szerepet játszhatnak a gazdasági és technológiai fejlődés irányításában (Evans, 1995).] Az MI-t alkalmazó vállalkozások aránya és a legalább alapvető digitális intenzitással rendelkező kis- és középvállalkozások aránya közötti kapcsolat között szintén enyhe pozitív korreláció figyelhető meg: azok az országok, ahol magasabb a digitális intenzitású KKV-k aránya, gyakran nagyobb arányban alkalmaznak MI-t is a vállalkozásokban (14. ábra). Dánia, Finnország és Luxemburg mindkét esetben minden dimenzióban magas pontszámot értek el, míg Románia alacsony értékeket mutat az összes változó esetében. Magyarország a jobb alsó kvadránsban található, ami azt jelzi, hogy az MI-t alkalmazó vállalkozások aránya alacsony, és a digitális közszolgáltatások szintje is az átlag alatt van, tehát a vállalkozások számára elérhető digitális közszolgáltatások fejlettsége nem ösztönzi jelentősen az MI használatát, és az MI alkalmazása Magyarországon viszonylag korlátozott. A kis- és középvállalkozások digitális intenzitása és az MI alkalmazás aránya szintén alacsony, ami arra utal, hogy a magyarországi KKV-k kevésbé rendelkeznek digitális fejlettséggel, és ez a korlátozott digitalizáció valószínűleg akadályozza az MI-technológiák elterjedését is.

Az Eurostat adatai kicsit pozitívabb képet mutatnak hazánkkal kapcsolatban: az Európai Unió és más országok nagyvállalatainak MI-technológia használati arányát tekintve Magyarország a 35 vizsgált országból a 16. helyen áll. Itthon a nagyvállalatok 29,7 százaléka használ legalább egy saját fejlesztésű MI-technológiát. A skála másik végén Görögország található a legalacsonyabb, 14,6 százalékos aránnyal, míg Bulgária a legmagasabb, 47,7 százalékos aránnyal rendelkezik (15. ábra). Ugyanakkor ez az eredmény megtévesztő lehet, mert a fejletlenebb tagállamokban a nagyvállalatok jókora része multinacionális, ami hozza magával ezeket a technológiákat, vagy csak service centre-ként itt fejleszteti, de se nem hazai indíttatásban, se nem hazai felhasználásban, ami így hibásan túlértékelheti az országot, ha sok IT-s SSC (Shared Service Centre) van épp az országban. A kis- és középvállalkozások tekintetében Magyarország a 6. helyen áll: a vállalatok 28,8 százaléka használ mesterséges intelligencia technológiákat, amelyeket saját alkalmazottaik fejlesztettek ki. Az országok

rangsorában Horvátország 17,9 százalékos aránnyal a legalacsonyabb helyen szerepel, míg Lengyelország a legmagasabb arányt képviseli 36,4 százalékkal (16. ábra).

A 2023-as Globális Innovációs Index (GII) alapján Magyarország a 35. helyen áll. A humántőke és kutatás terén az ország 40.2 pontot kapott, a 36. helyen állva. Az oktatási kiadások (4.2 százalék GDP-arányosan) és a PISA-eredmények (479.3) közepes szintűek, míg a kutatók aránya (4,461.8 kutató/millió fő) alacsonyabb értéket mutat. Az infrastruktúra kategóriában Magyarország 53.0 ponttal a 42. helyen szerepel, ahol az ICT (Information and Communication Technology, információs és kommunikációs technológia) elérhetőség (83.5) és használat (83.0) magas. Az üzleti kifinomultság 45.1 pontot ért el, különösen az FDI beáramlás (61.0, 1. hely) és a kutatási tevékenységek terén mutat erősséget. A tudás- és technológiai outputok kategóriában 38.4 ponttal a 26. helyen szerepel az ország, a szabadalmak (1.5/PPP GDP) és tudományos publikációk (20.0) közepes eredményeivel. A kreatív outputok esetében Magyarország 34.1 pontot szerzett, ahol az online kreatív tevékenységek (37.6) kiemelkednek. Magyarország ezek alapján erős ICT infrastruktúrával rendelkezik, és a tudásalapú tevékenységekben jól teljesít, de a humántőke és kutatás, valamint az oktatási rendszer terén kihívásokkal küzd (Dutta et al., 2023).

Az Adecco felmérése¹³ szerint a fejlődő országok válaszadóinak 70 százaléka már használ valamilyen MI-megoldást, míg Magyarországon a Századvég tanulmánya¹⁴ alapján a dolgozók mindössze 25 százaléka alkalmaz ilyen technológiát a munkája során. Az európai vállalkozások körében már 7 százalék tudatosan bevezette az MI-t, Magyarországon azonban ez az arány csupán 3 százalék (E01, R04).

3.6. Az MI intézményes keretei Magyarországon

Magyarországon jelenleg egy direktíva született, Magyarország Mesterséges Intelligencia Stratégiája (2020), amely kijelöli a 2030-ig megvalósítandó állami célokat, illetve egy 2025-ig tartó intézkedési tervet tartalmaz. Prof. Dr. Palkovics László, az akkori innovációs és technológiai miniszter vezetése alatt az Innovációs és Technológiai Minisztérium (ITM) koordinálta a magyarországi MI fejlesztéseit, létrehozva a Mesterséges Intelligencia Koalíciót 2018-ban. A Stratégiát a Koalíció alakította ki, 240 szervezet és 1000

¹³ Forrás: <https://elblog.pl/hu/2024/10/18/ai-kihasznalasa-a-munka-jovoje-vilagszerte/>, Letöltve: 2024.11.10.

¹⁴ Forrás: <https://szazadveg.hu/cikkek/mesterseges-intelligencia-kevesbe-tajekozottak-a-hazai-internethasznalok/>, Letöltve: 2024.10.10.

szakember közös munkájával. A Stratégia nem rendelkezik közvetlen jogerővel, mivel nem jogszabály, hanem egy irányadó stratégiai dokumentum – az ilyen típusú stratégiák célja a hosszú távú fejlesztési célok és prioritások meghatározása, amelyet a kormány és az érintett szervezetek figyelembe vesznek a szabályozás és fejlesztési projektek kialakítása során. A technooptimista hangvételt követő dokumentum feltételezi, hogy a technológiának nincs fizikai eszközigénye [„*A mesterséges intelligencia egy globális technológia, határok és fizikai eszközigény nélkül.*”], sem természeti erőforrásigénye [„*Ezúttal azonban nincs a fejlődésnek természeti erőforrás-igénye, hanem a rendelkezésre álló emberi erőforrás kompetenciája határozza meg a fejlődés korlátait.*”] (Magyarország Mesterséges Intelligencia Stratégiája, 2020:9), és a számos pozitív felhasználási javaslat mellett sajnos nem esik túl sok szó a kockázatokról és azok kezeléséről. Ugyanakkor a kijelölt fejlesztési irányok – különösen a kutatási infrastruktúra bővítése, az adatgazdaság ösztönzése, valamint a képzési és kompetenciafejlesztési programok tervezete – ígéretes alapokat biztosítanak Magyarország digitális fejlődéséhez. A Stratégia hangsúlyozza, hogy Magyarországnak szuverenitását megőrizve kell részt vennie a mesterséges intelligencia kutatásában és alkalmazásában, valamint az erőforrások megosztásán és közös fejlesztési projektek kialakításán keresztül kell biztosítani, hogy megfeleljen az EU digitális egységes piacának követelményeinek, miközben szuverén módon rendelkezhet saját adatai felett. Ez illeszkedik a szuverenitás-diskurzusba, amely a magyar kormány egyik fontos kommunikációs pillére. Ebben a kontextusban említi a direktíva a kockázatokat, mégpedig a kibervédelem és az állampolgárok befolyásolásának tekintetében, illetve az adatfelhasználást illetően figyelmet fordít a „személyes önrendelkezési és szabadságjogok biztosítására”. A Stratégiában említésre kerül az MI-technológiák társadalmasítása, azaz a lakosság széles körének tájékoztatása az MI előnyeiről, hátrányairól és kockázatairól, az új technológiák megismertetése és a kritikus gondolkodás fejlesztése, amely előfeltétele a felelősségteljes felhasználásnak. A program konkrét akciókat tartalmaz, mint például az MI Innovációs Központ létrehozása, az MI Kihívás, valamint MI ismeretek bevezetése a köz- és felsőoktatásban. A pedagógusok és oktatók képzése is kiemelt szerepet kap, hogy támogassák a gyermekek tudatos felkészítését az MI-alapú rendszerek használatára.

A mesterséges intelligenciát érintő kérdésekkel jelenleg leginkább a Gazdaságfejlesztési Minisztérium foglalkozik, mivel az ITM 2022-ben megszűnt, részben kormányzati struktúraváltás részeként. Palkovics László lemondott a tárca éléről, és az innovációs ügyek egy része az újonnan létrehozott Kulturális és Innovációs Minisztériumhoz, míg más

területek a Gazdaságfejlesztési Minisztériumhoz kerültek. A minisztérium (és a tulajdonában lévő N7 Holding Zrt.) fókuszra nagyrészt a védelmi ipari kapacitások fejlesztésén van, ez kiemelt nemzeti és gazdaságstratégiai cél¹⁵, ami azért is izgalmas, mert az MI-alapú technológiák egyik legprofitábilisabb területe a hadiipar¹⁶, illetve a hadiiparban kifejlesztett technológiák idővel a civil szférában is megjelennek, ahogyan például az internet is a katonai fejlesztésekből indult. Ugyanakkor ez az a terület, ami a legnagyobb társadalmi kockázatokkal is járhat. Az MI katonai alkalmazásának fő kockázatai közé tartozik a fegyverfejlesztés felgyorsulása, halálos autonóm fegyverek használata, fejlett kibertámadások végrehajtása és a döntéshozatal automatizálása, ami gyakrabban előforduló, pusztítóbb háborúkat és katasztrofális baleseteket eredményezhet. A mesterséges intelligencia által vezérelt fegyverek csökkenthetik a hadba lépés küszöbét. A kiberhadviselés terén az MI rendszerek fokozhatják a támadások hatékonyságát és nehezíthetik a felelősség megállapítását. A hadviselés üteme és komplexitása egyre inkább meghaladhatja az emberi kontrollt, ami súlyos biztonsági és etikai kérdéseket vet fel (Hendrycks, 2024).

4. A kutatás módszertani háttere

A kutatás célja felmérni azt, hogy a jelenlegi Európai Unió és magyarországi MI szabályozás keretrendszere képes lehet-e az általa kitűzött cél, azaz az etikus és felelősségteljes MI-alapú rendszerek biztosításának (European Parliament, 2024; Magyarország Mesterséges Intelligencia Stratégiája, 2020) megvalósítására.

Felvetések:

(1) A döntéshozók kompetenciája nem elégséges a kockázatok mértékének és okozatiságának felméréséhez, és az ezzel arányos politikai cselekvéshez.

¹⁵ Forrás: <https://kormany.hu/hirek/egyseges-szakmai-iranyitas-a-vedelmi-iparhoz-kapcsolodo-allami-cegeknel>, Letöltve: 2024.11.10.

¹⁶ „A magyar védelmi ipar jelenleg még csak kis hányadát adja a GDP-nek, de a következő években ez változhat, különösen a növekvő exportnak köszönhetően.” – mondta Szalay-Bobrovniczky Kristóf honvédelmi miniszter, Forrás: https://frisshirek.hu/hirek/ai-cikkek/2024/05/10/Magyarország_vedelmi_iparanak_fejlolese_es_a_jovobeli_kilatasok; „Emellett a gazdasági növekedésben is egyre hangsúlyosabb szerephez jut a védelmi ipar, amely az ipar teljesítményének fokozásán keresztül érdemben hozzájárul a magyar gazdaság jövő évi 4 százalékos növekedési céljának eléréséhez.”, Forrás: <https://kormany.hu/hirek/a-vedelmi-iparnak-a-biztonsagot-a-szuverenitast-es-a-nemzetgazdasagot-egyuttesen-kell-szolgálnia>, Letöltve: 2024.11.10.

(2) Az aktorok között nincs megfelelő kommunikáció, amely biztosítaná az első pontban említetteket.

(3) Nem csak a kompetencia, hanem a politikai akarat sem elég erős a kockázatok súlyosságával arányos lépésekre. [Az elit társadalmi tudatossága három feltételen alapul: (1) a szegénység saját negatív következményeinek felismerése, (2) a változtatás lehetőségének érzékelése, és (3) a felelősségérzet megléte. Ha ezek a feltételek teljesülnek, az elit cselekvésre motivált a társadalmi problémák enyhítése érdekében (Amsden et al., 2012:182).]

Kutatási kérdések

(1) Milyen kompetenciákra van szüksége a döntéshozóknak a kockázatok felméréséhez?

(2) Kik között és milyen kommunikációra van szükség, hogy biztosítva legyenek az első pontban meghatározott kompetenciák?

(3) Az első két feltételen túlmenően milyen egyéb akadály lehet a szabályozás hatékonyságának, és hogyan lehetne ezeket megoldani?

Célcsoportok

(1) Az MI-alapú rendszerek előállításán dolgozó szakemberek: MI-tervezők, fejlesztők, jogászok, vezetői réteg. (2) Szakpolitikusok (Európai Unió és magyarországi). (3) MI-kutatók, akadémiai szereplők. (4) Általános felhasználók, társadalom szélesebb rétegei, lakosság sérülékenyebb csoportjai. *(Passzív aktorok, velük ebben a kutatásban nem fogok foglalkozni.)*

Mintaválasztás módszere

(1) Célzott mintavétel: Ez a módszer lehetővé tette, hogy specifikus, előre meghatározott kritériumok alapján válasszak ki résztvevőket, akik releváns információkkal rendelkeznek a kutatási témához. Ebben az esetben a kritériumok az adott területen szerzett szakértelemre és tapasztalatra vonatkoztak. Meghatároztam a szakértői csoportot (pl.: évek óta az MI-szabályozással foglalkozó szakértők), majd célzottan felkerestem őket interjúra.

(2) Hólabda módszer: A felkeresett szakemberek további szakembereket ajánlottak, akik szaktudását relevánsnak gondolták a kutatás szempontjából.

Adatgyűjtési módszerek

A primer kutatásom másodelemzés: az elérhető szakirodalom és adatbázisok elemzése, illetve szakpolitika elemzés.

A szekunder kutatásom a *szubjektív emberi ítéleten alapuló előrejelzés* (subjective human judgment forecasting) módszertan követi (Zellner et al., 2021, Stewart, 2001, Goodwin & Fildes, 2022). A szabályzás a kutatás kezdetekor még hatályba sem lépett, és az Európai Unióban és hazánkban is megalakulás alatt vannak a szabályzást biztosító szervezetek. Ez tehát egy olyan terület, ahol kevés az objektív adat és számos tényező még kialakulóban van, ezért a jövőbeli hatások előrejelzése különösen összetett és bizonytalan. Az új szabályozás hatásait és kimeneteleit a szakértők megérzésein és tapasztalatán keresztül lehet a legjobban megközelíteni. A korlátozott számú, de alapos interjú lehetővé teszi, hogy a szakértők mélyebb betekintést nyújtsanak, és feltárják azokat a szubjektív szempontokat és potenciális kimeneteket, amelyeket kvantitatív adatokkal még nem lehet elemezni. A módszer előnye, hogy az emberek képesek olyan tényezőket is figyelembe venni, amelyeket a gépi modellek nehezebben, vagy egyáltalán nem tudnak. Hátránya azonban, hogy hajlamos lehet az elfogultságokra, szubjektív torzításokra és pontatlanságra. Annak érdekében, hogy a kutatásom erőforrás-korlátaival illeszkedjen, a módszert szűkítettem és adaptáltam.

A módszer erre a kutatásra alkalmazott lépései a következők voltak:

(1) Szakértők kiválasztása: Kiválasztottam a téma szakértőit, akik megfelelő tapasztalattal és tudással rendelkeznek a vizsgált területen, így megalapozott véleményt tudnak formálni a jövőbeli eseményekről.

(2) Kérdezési módszer: A szakértők véleményét félig strukturált mélyinterjúkon keresztül kértem ki. Kutatásomban az első három célcsoporttal, mint aktív aktorokkal foglalkoztam. A negyedik célcsoport kompetenciáinak és tájékozottságának felméréséhez nagyobb elemszámmal dolgozó, kérdőíves felmérés javasolt, amely meghaladja ennek a kutatásnak a kereteit.

(3) Adatgyűjtés és elemzés: A szakértői véleményekből összegyűlt kvalitatív adatokat konszenzuskereséssel és a válaszok gyakoriságával elemeztem.

(4) Integráció és értelmezés: Az összegyűjtött információkat értelmeztem, a szakértői válaszokat strukturált formába rendeztem, hogy kvantitatív adat híján is megalapozott előrejelzést nyújtsak.

Vizsgált változók

(1) Kompetencia szintje: A döntéshozók és az MI-alapú rendszereket fejlesztő szakemberek MI-szabályozással kapcsolatos kompetenciáinak mérése. Ez magában foglalja a technológiai tudást, a szabályozási keretek ismeretét, és az etikai megfontolások értékelésének képességét.

(2) Kommunikáció: Az aktorok közötti kommunikációs csatornák hatékonyságának és nyitottságának mérése. Megvizsgálom, hogy van-e rendszeres párbeszéd az érintett felek között, ez a kommunikáció mennyire hatékony az információáramlás és az összehangolt cselekvés szempontjából, és hogy a kommunikáció iránya és célja szerint laterális vagy hierarchikus kommunikációról beszélhetünk-e.

(3) Politikai akarat: A politikai akarat mértékének mérése az MI szabályozásával kapcsolatban. Értékelem, hogy mennyire hajlandóak a döntéshozók és szakpolitikusok lépéseket tenni a felmerülő kockázatok kezelésére.

(4) Szabályozási keretrendszer hatékonysága: A jelenlegi szabályozási keretrendszer érvényesítésének és betartásának értékelése. Megvizsgálom, hogy mennyire alkalmas a rendszer az MI kihívásainak kezelésére, beleértve az átláthatóságot, az ellenőrzést és az elszámoltathatóságot.

(5) Kockázatok és okozatiság felmérése: A kockázatok típusainak és súlyosságának értékelése, valamint azok lehetséges okainak és következményeinek megértése. Ez magában foglalja az etikai dilemmákat, a társadalmi egyenlőtlenségeket és a technológiai hibákat.

(6) Társadalmi és etikai megfontolások: Az MI által generált társadalmi és etikai kihívásokra való reagálás képességének mérése kapcsán figyelembe veszem az adatvédelem, az egyenlőség és a diszkrimináció elkerülésére vonatkozó irányelvek betartását.

Elemzés meghatározása (esetcsoporthoz)

Az első célcsoportból beszélgettem egy magyar nemzeti MI-szövetség vezető technológiai szakértőjével, egy multinacionális információs-technológiai vállalat IT-biztonsággal és MI-

szabályzással foglalkozó jogászával, egy független nemzetközi MI-biztonsági (AI safety) szervezet magyar alcsoportjának szabályzás vezetőjével (policy lead), egy német multinacionális cég MI fejlesztőjével. A második célcsoportból megkeresett magyar állami megbízottak sajnos nem vállalták a beszélgetést. A harmadik célcsoportból több magyar egyetem és kutatóközpont kutatóival interjúztam, akik főképp politikatudományi, jogi és informatikai területeken tevékenyek, illetve ezek és más társadalomtudományi területek (kulturális antropológia, szociológia) metszeteiből érkeztek. Közöttük volt nemzetközi szervezetenél tanácsadóként dolgozó magyar állami megbízott, aki ezáltal a második és harmadik célcsoportba is tartozik. Az interjúalanyok érdekessége, hogy szinte mindannyian diverz tanulmányi és munkatapasztalati háttérrel rendelkeztek (például állami/akadémiai/piaci hátterek metszete, jogi területről nyitás az MI-technológiák felé, technológiai háttérből induló állami megbízásba érkező alany) tehát esetükben alapvetően adott volt az interdiszciplináris tudás. Ez azt mutatja, hogy a területen aktív szereplők között organikusan az inter- és multidiszciplinaritás vált dominánssá. Összesen kilenc fő adta ki a teljes mintám.

Vizsgálat helye, ideje

A kutatást 2024 februártól 2024 novemberig folytattam le, elsődlegesen az online térben.

Anonimitás

Az interjúkban elhangzott szenzitív adatok és információk védelmének érdekében dolgozatomban az interjúalanyaimat kóddal jelöltem, eredményeimet konkrét személy- és intézménynevek leírása nélkül mutattam be. Kivételt képez Dr. Rab Árpád Szörény jövőkutató, aki nem kérte, hogy anonimizáljam: a vele készült beszélgetést a statisztikában anonimizáltam, az idézetekben monogrammal jelöltem.

MI-alapú technológiák használata

A kutatás során a következő MI-alapú technológiákat használtam: Alrite (<https://alrite.io/>) szövegfelismerő programot az interjúk leiratozásához, ChatGPT 4o programot (<https://chatgpt.com/>) a leiratok javításához, szövegek elemzéshez és grafikonok készítéséhez, Consensus applikációt (<https://consensus.app/search/>) szakirodalmak kereséséhez.

5. Kutatási eredmények

5.1. Kompetencia szintje

Az államigazgatásban lévő döntéshozók és szakemberek sok esetben hiányos technológiai tudással rendelkeznek, ami nem elégséges az MI-rendszerek kockázatainak megfelelő kezeléséhez [„*A jogászok nagy része digitális analfabéta. Legyünk őszinték, nem értjük ezt az egész technológiát mögötte.*” (R03, jogász); „*A <<miniszternek>> az a reakciója a <<minisztériumot>> érintő orosz hackerbotrányra, hogy ő nem tudja használni az okostelefonját az sms-ezésen kívül, az szerintem elég beszédes.*” (R01)]. Jelentős kompetenciabeli kudarcok akkor következhetnek be, amikor a rendszerből hiányzik a bizalom, és a vezetés ezért megpróbálja maximalizálni a politikai kontrollt, miközben elfojtja a szervezeti autonómiát. Ilyen környezetben a vezetők, akik gyakran politikai kinevezettek, alapvető szakmai kompetenciák hiányában kerülhetnek pozícióba. Peter Evans (1995) ezt a problémát a „fejlesztő” és „zsákmányoló” államok megkülönböztetésével magyarázza, ahol a fejlesztő államok támogatják a szakmai autonómiát, míg a zsákmányoló államokban a politikai kontroll mindent felülír [„*A klasszikus példa erre Mark Zuckerberg megjelenése az amerikai képviselőház bizottsága előtt, ahol az áprilisi meghallgatásán abszurd kérdéseket kapott a bizottság tagjaitól, akik a Meta szabályozásával kezdtek foglalkozni. Úgy tűnt, hogy a bizottságnak fogalma sincs a legalapvetőbb dolgokról sem.*” (R03)]. Ennek elkerülése érdekében fontos kihasználni az olyan előnyöket, mint Magyarország esetében az akadémiai közeg ereje (R04), illetve nagyobb hangsúlyt helyezni a kommunikációra és a kompetenciafejlesztésre, akár alapszintű készségek elsajátításáról, akár komplexebb és mélyebb tudás bevonásáról beszélünk [„*Egy állami szabályzónak nem kell saját modelleket tanítania, de képesnek kellene lennie elemezni azokat. Ez PhD szintű tudást igényel. A szabályozóknak ismerniük kellene a modellarchitektúrákat és azok belső működését.*” (R02)]. Az előkészítetlenség is jelentős akadályt gördíthet a komplex szabályozások megvalósítása elé. A jogalkotási folyamat során a különféle rendeleteket kell értelmezniük az érintetteknek, ami egy új szintű rugalmasságot és precíziós készséget kíván [„*Egy <<jogalkotással foglalkozó kormányzati szervezettel>> beszéltem, és ott nyilván már elolvasták [az EU AI Act-et], és most post-it-ekkel rakják össze, hogy akkor ez a rendelet ez, az az, ez erre vonatkozik, az arra... tehát ezt így fejben össze kell rakni nekik.*” (E01)]. Az implementáció egyik legnagyobb kihívása, hogy megfelelő szakemberek álljanak rendelkezésre – az EU által létrehozott munkacsoportok esetében különösen kritikus, hogy

a tagok informatikai mérnöki jellegű tudással bírjanak, ami jelenleg több országban hiányzik (R04). A kompetenciák fejlesztésében kihívást jelenthet, hogy a technológia nagyon gyorsan fejlődik és a naprakész tudáshoz folyamatos tanulásra van szükség (E02), az értelmezéséhez nem elég a horizontális tudás, egyszerre sok tudományterületet kell lefedni (E04, R04), illetve a piaci szektor elszívó hatásával sem tud versenyezni az állami szektor (E01, R03, R04). Magyarországon a rendszerváltás után – és részben már korábban is, a Rákosi-korszakban – megszűntek az iparági érdekképviselések és kamarák, amelyek a vállalkozások közötti tudásmegosztást és együttműködést segítenék elő. A nyugati országokban gyakran ezek a független iparszövetségek és kamarák látják el az ilyen típusú, paraetatiszt (kvázi-állami) funkciókat, például az oktatási workshopok és iparági tréningek szervezését. Magyarországon azonban ezek a szervezetek hiányoznak, így az oktatási és integrációs szerepeket sokszor magánszemélyek vagy kevésbé formális hálózatok kénytelenek átvállalni. Emiatt a szakemberek sokszor saját kezdeményezésre, informális eszközökkel próbálnak megoldást találni, hogy elősegítsék a technológia terjedését [*„Ingyenes workshopokat szervezek, amelyeken 30-80 résztvevő is lehet. Tehát hálózaton belül valakitől kérek egy workshopot szívésségből, és a résztvevőket meg ingyen meghívom. Ilyenkor arra gondolok, hogy akár hetven cégnek is segíthettem jobban megérteni a mesterséges intelligencia integrációját.”* (E01)].

A szélesebb társadalom digitális és MI-specifikus kompetenciáival kapcsolatban Rab Árpád egy friss (a dolgozat befejezése pillanatában még nem publikus) kutatásra alapozva optimistán vélekedett: szerinte a magyar társadalom figyelemreméltó tudatosságot mutat az MI-vel kapcsolatban, ami az utóbbi időben különösen felerősödött. Az erről szóló diskurzus és a generatív technológia gyors elterjedése révén sokan kezdték el személyes szinten is megérteni és beépíteni az új technológiát az életükbe. Bár a társadalomban sokan még nem teljesen értik a technológia valódi hatásait, a tudatosság és az érdeklődés egyértelműen jelen van, és ha megfelelő eszközöket és alkalmazásokat biztosítanak az emberek számára, az MI tudatossága és alkalmazása még inkább növekedne. Jelenleg azonban főként üzleti célú szolgáltatások állnak rendelkezésre. Amikor az olyan mindennapokat segítő eszközök elterjednek, mint a pszichológiai támogatást nyújtó vagy a közlekedést segítő szoftverek, az MI még inkább beépül majd a társadalomba a jövőkutató szerint. Az állampolgárok kompetenciáinak növelése érdekében az NMHH (Nemzeti Média- és Hírközlési Hatóság) már fejlesztett mesterséges intelligenciával kapcsolatos tudatosságnövelő programokat, és a Gazdasági Versenyhivatal is készített tájékoztató anyagokat. Emellett a hagyományos

médiatudatosság-fejlesztő programokba is egyre gyakrabban beépítik az MI-vel kapcsolatos ismereteket, például a deepfake tartalmak felismerésére vonatkozó információkat (R04). Külön figyelmet igényel a társadalom szeniorabb rétegének felkészítése is, például, hogy az idősebb állampolgárok felismerjék az olyan digitális fenyegetéseket, mint a hamis személyazonosság és az ezzel való visszaélések (E01). Az emberi tudás és az agy a legnagyobb erőforrás, amelyet a felsőoktatási intézmények is realizáltak: egyre több egyetemen indulnak MI-hez kapcsolódó szakok (E02).

5.2. Kommunikáció

A jogi és technológiai szakterületek közötti együttműködés elengedhetetlen a szabályozások betartásához és a kockázatok helyes felméréséhez – jelenleg ebben a tekintetben még van tér bőven a fejlődésre (E01, R02, R04). Az információáramlás akadálymentesítésének érdekében a meglévő platformok és dedikált szervezetek hatékonyságát kéne növelni és új platformokat létrehozni [*„Nem volt azért ideális ez a párbeszéd, inkább szekértáborok beszéltek egymással: mindenki mondta a saját álláspontját és meg volt győződve róla, hogy ő tudja a helyes választ.”* (R03); *„Ez kéne, hogy a diskurzus színtere legyen, ahová a szakterületen releváns képviselők járnak, de sokszor nem is jártak, amikor mégis, akkor sem számoltak be semmi olyanról, ami igazán előremutató.”* (E03)]. Ez a probléma a gazdasági bürokrácia társadalmi beágyazottságának hiányából is eredhet, ami akadályozza az információáramlást és csökkenti a döntéshozók érzékenységet a valós gazdasági és társadalmi igények iránt (E02). Amikor a bürokrácia elszigetelten működik, kevésbé képes reagálni a helyi kihívásokra, ami nehezíti a komplex szabályozási területeken való hatékony működést (Evans, 1995).

Magyarországon a nagy fejlesztő cégek többsége külföldi leányvállalat, amelyek nincsenek jelen a magyar szakmai szervezetekben, így hiányzik a helyi beágyazottságuk [*„Én itthon nem látok ilyet [formális kommunikációt]. Itthon nincsenek nagy fejlesztő cégek, és biztos, hogy nem formalizált folyamatokra van ráállva a magyar állami rendszer.”* (R01)]. Ezek a vállalatok csupán részei egy nemzetközi termelési láncnak, és gyakorlatilag érzéketlenek a hazai fejlesztéspolitikai célokra – ez az „ungovernability crisis”¹⁷ egyik megnyilvánulása. Hasonló problémát tapasztalni az EU szintjén is: mivel a nagyvállalatok számos

¹⁷ Az *ungovernability crisis* (irányíthatatlansági válság) olyan állapotra utal, amikor egy társadalom, intézmény vagy kormányzat nehézségekbe ütközik a hatékony irányításban vagy kontrollálásban. Ezt a válságot rendszerint a politikai, gazdasági és társadalmi tényezők összetett együttese okozza, amelyek miatt a hagyományos irányítási és kormányzási struktúrák elégtelenné válnak.

tagországban vannak jelen, gazdaságpolitikailag immunisak az egyes nemzeti politikákra. Bár az EU-nak szüksége lenne egy közös iparpolitikára, ezt a jelenlegi szerződések és a jogi keretek lehetetlenné teszik, ahogy azt a 2024-es Draghi-jelentés is megállapítja (European Commission, 2024).

A szakemberek az állami szervek és fejlesztők közötti információáramlás terén is hiányosságokat tapasztaltak, illetve rámutattak arra, hogy a politikai döntéshozói réteg és a szélesebb társadalmi rétegek (így azok, akik a legkiszolgáltatottabbak a digitális veszélyeknek és MI kockázatoknak) között sincsen jól működő kommunikációs híd [„*Az emberek nem értik a technológia hatását, és sok esetben a politikai réteg sem érdekelt abban, hogy ezt világosan kommunikálja.*” (E02)]. A totalitárius rendszerek, például a szocialista országok egykori politikai vezetése, gyakran akadályozzák vagy gátolják a technológiai fejlődést, ha az veszélyeztetheti a hatalmukat és ellenőrzésüket. A totalitárius államok nem feltétlenül érdekeltek a gazdasági hatékonyság növelésében, hiszen az olyan technológiai fejlődés, mint például a kibernetika vagy a telekommunikáció fejlesztése, nagyobb autonómiát és szabadságot biztosíthatna a polgároknak, ezáltal gyengítve a politikai kontrollt. A KGST-országokban a technológiai innovációkat gyakran fékeztek, mert ezek a fejlődések lehetőséget adtak volna a társadalom jobb informáltságára és az információ szabadabb áramlására, amit a vezetés potenciális fenyegetésként értékelt a saját hatalmára nézve (Fehér et al., 1983).

Az Európai Unió jogalkotási folyamata során bevonták a fejlesztőket és az érintett vállalatokat is, így ezt jó gyakorlatként említhetjük. Bár a plenáris üléseken több száz, akár ezer résztvevő – köztük jogászok és közgazdászok – vettek részt, és számos összehangolási stratégia létezik, nem mindig átlátható, hogy ezek eredménye mennyire tükröződik a jogszabályokban. Végül mindig a jogalkotó hozza meg a döntést, így kérdéses, hogy a fejlesztők és cégek szempontjai mennyire épülnek be ténylegesen a végső jogi keretbe (R03, E04). Magyarországon az akadémiai közegben tapasztalható működő párbeszéd az egyes szervezetek között pozitívumként emelhető ki („*Nálunk például az egyetemen mindenkit érdekel a mesterséges intelligencia, a bölcsészkartól kezdve a gazdasági karon keresztül a jogi karig.*” R02; R04). Ez jó gyakorlatként szolgálhat a többi szektor számára.

5.3. Szabályozási keretrendszer hatékonysága

Az MI szabályozásának előzetes megközelítése az innováció kárára megy: Európa, amely már eleve lemaradásban van az innováció terén, az AI Act bevezetésével tovább lassíthatja saját technológiai fejlődését. A rendelet körül kialakult félelem visszatartja a vállalatokat a fejlesztésektől, mivel a vállalatvezetők tartanak attól, hogy egy esetleges szabálysértés jelentős bírságokat vonhat maga után, ahogyan azt a GDPR esetében is láthattuk. Ez a szabályozási kultúra az elővigyázatosságra épül, ahogyan azt a történelmi példák is mutatják [„Az első gépjárműveknél az Egyesült Királyságban volt egy szabály, amely előírta, hogy egy embernek piros zászlóval kelljen a jármű előtt haladnia – ez volt a „vörös zászlós törvény”. A cél az volt, hogy figyelmeztessen a közeledő, akkoriban 15 km/órás sebességgel haladó gépjárművek veszélyeire, mivel attól tartottak, hogy ezek kárt okozhatnak és veszélyt jelenthetnek a társadalomra. A porosz vasúti törvény esetében a törvényhozók attól tartottak, hogy a vasút kisiklik, és súlyos károkat okoz, így minden potenciális kárért felelőssé tették a vasúttársaságokat, anélkül, hogy megértették volna a vasút technikai sajátosságait.” (R03)]. Az új technológiák megjelenésekor a jogászok általában a veszélyeket hangsúlyozzák, és szigorú szabályozást sürgetnek, anélkül, hogy teljesen megértenék a technológia működését. Az európai megközelítés alapvetően a kockázatok minimalizálására összpontosít, miközben más régiók, például Délkelet-Ázsia inkább az innovációt ösztönzi. Ez a félelemből eredő szabályozás tehát nemcsak a fejlődést lassítja, hanem az európai cégek globális versenyképességét is veszélyezteti (E01, R01, R03, R04).

Az EU MI-szabályozási kerete ennek ellenére ígéretesnek tűnik a szakértők szerint (illetve ez a jelenlegi intézményi struktúrából kihozható legjobb eredmény), de a kiskapuk keresése és az opt-out lehetőségek alááshatják a hatékonyságát. Az EU jogalkotásában gyakori, hogy a végső szakaszban számos záradékot és kivételt építenek be, ami rugalmas, de komplex szabályozási rendszert eredményez. Az MI-rendelet is alapvetően egy keretszabályozás, amely általános előírásokat biztosít, de a részleteket szektorális szabályok fogják meghatározni. A jogalkotók célja, hogy olyan szabályokat hozzanak, amelyek hosszú távon is érvényesek maradnak. A jogszabályok inkább keretrendszert nyújtanak, amelyet a hatóságok jogértelmezéssel töltenek meg konkrét tartalommal. A hatóságok feladata lesz, hogy a jogszabályi keretek között részletesebb iránymutatásokat dolgozzanak ki. Az MI gyors ütemű fejlődése miatt a szabályozás gyakran csak követi a technológiát, ami

kihívásokat jelent a hatékony szabályozás megvalósításában (E02, R02, R03). Az EU rendszerében rengeteg adminisztratív teher van, és a szabályozás előrehaladása lassú folyamat, ami komoly erőforrásokat igényel (E01, E04, R01, R04).

Magyarország jelenleg nem maradt le a szabályozás terén a többi tagországhoz képest: bár a hazai jogszabályalkotás még folyamatban van, a hatóságok feladatai és hatáskörei hamarosan körvonalazódnak (R03).

5.4. Politikai akarat

Az Európai Uniónak érdemes lenne tovább összpontosítania az együttműködés fokozására és ösztönzésére, amit meglévő struktúrái már jelentős mértékben támogatnak, ugyanakkor hiányzik a nemzeti akarat, mivel az államok gyakran ragaszkodnak ahhoz, hogy az innovációjukat saját határaikon belül tartsák (E04). Az MI területén a nemzetállami gondolkodást egy kicsit háttérbe kell szorítani a szakértők szerint, mivel ezt összeurópai együttműködés, közös piac és közös gazdasági célok határozzák meg. Magyarországnak érdemes azon dolgoznia, hogy megtalálja a helyét a globális piac különböző szektoraiban, és ezekben szolgálja ki a számára releváns részeket (R03).

Magyarország versenyképességének növelésében az MI nem közvetlenül adhatna választ, hanem eszközként szolgálhatna azoknak, akik a versenyképességet megtervezik és felelnek érte. Az igazi kitörési pont az adatok és a bizalmi rendszer kiépítése lehetne, amelyek beépülnek a magyar rendszerekbe. Például, ahogy Rab Árpád rámutatott, a magyar egészségügy fejlesztése érdekében nem feltétlenül egy állami, MI-alapú orvosi rendszert kell létrehozni a szuverenitás biztosításához – ehelyett az egészségügy háttérszolgáltatásába beépített mesterséges intelligencia az, ami igazán értéket jelentene. Egy ilyen rendszer megvalósítása alapos fejlesztést igényel, különös tekintettel a személyes adatok kezelésére és a rendszer intelligenciájának és funkcionalitásának kialakítására. Magyarország ebből a szempontból ugyanazokkal a kihívásokkal néz szembe, mint sok más kisebb ország: nem képes felvenni a versenyt az amerikai nagyvállalatokkal, amelyek hatalmas számítási kapacitással és adattömegekkel rendelkeznek, ezért a kitörési pontja az lehet, hogy saját rendszereit és képességeit okosabbá, hatékonyabbá tegye [*Meg kell nézni, és hogy hova teszed be, mi a problémád, mi a hiány, hol akarod használni. Az MI szolgáltatások a tükröbe nézést jelentik; azt, hogy megálmodod a jövődet.*] (R.Á.).

A magyar állam politikai akaratának meglétét mutatják bizonyos kezdeményezések, például az élelmiszeripari fejlesztési pályázatok keretében a cégek jelentős, 36 plusz 6 milliárd forintos támogatást kaptak, ami fontos lépés volt az agrár- és élelmiszer-feldolgozóipar támogatásában. Az ilyen kezdeményezések azonban lassan haladnak, és több hasonló intézkedésre lenne szükség. Magyarország mezőgazdasági szektora már most is ígéretes megoldásokat mutat fel, és a modern technológiai fejlesztések révén tovább erősödhetne. Az ország kiváló termőfölddel és jelentős exportpotenciállal rendelkezik, amelyet modern technológiákkal kihasználva nemzetközileg is versenyképes lehetne az élelmiszeripar és az agrárszektor (E01). Az innovatív megoldásokat leginkább a kis- és középvállalkozások vagy startupok körében lehet megtalálni (R04). Az ország emellett kiemelkedő tudományos kapacitással és magas szintű intelligenciával rendelkezik, számos kiváló kutatóval és erős kutatóhálózattal, mint például a HUN-REN. Ezek a tényezők lehetővé teszik, hogy Magyarország a globális szinten is versenyképes legyen, különösen az innováció és kutatás területén.

Az új technológiákkal kapcsolatban kétféle megközelítés figyelhető meg: az egyik a félelemből fakadó ellenállás, amikor az emberek nem értik és emiatt elutasítják; a másik a túlzott lelkesedés, amikor attól tartanak, hogy lemaradnak a piacon, ezért mindenáron követni akarják az új trendeket. Az ismeretlenség és a tudatlanság pedig gyakran gátolja a politikai döntéshozókat abban, hogy megfelelő lépéseket tegyenek (E02). A politikai akarat strukturálatlansága akadályozhatja az innovatív kezdeményezéseket [*„Miután «korábbi államtitkár» kitették a kormányból, azután ez teljesen lehalt. Hiszen a politikai háttér mögötte megszűnt. [...] Tehát mondhatja a kormány, hogy persze, hogy ők ezt támogatják, ha nem adnak hozzá se pénzt, se politikust.”* (R01)]. De a létező politikai akarat sem mindig elégséges ahhoz, hogy gyors, hatékony lépések történjenek az MI szabályozás terén, mivel a politikusok figyelme más problémákra összpontosul (E03, E04), vagy egyszerűen hiányoznak a szükséges erőforrások (R02). Magyarországon jelenleg nem áll rendelkezésre elegendő forrás a szélesebb körű technológiai fejlesztésekhez és az MI elterjedéséhez. Az Európai Unió által visszatartott források az utóbbi két-három évben komoly akadályt jelentettek Magyarország számára, visszafogták a fejlődést és a lehetőségeket [*„Ha a kisgyerekem nem kapja meg a csokoládét, akkor lehet vitatkozni, hogy azért, mert ő rosszat szólt, vagy azért, mert most ilyen napom van, és nem adom neki. Lényeg, hogy a végén nem kapta meg.”* (E01)]. Az MI stratégiák gyakran optimista jövőképet festenek, azonban a jelenlegi problémák sokkal strukturálisabbak és mélyebbek; nem pusztán

hatékonyságnövelésről van szó, hanem a hiányok kezeléséről. A legtöbb területen még nem tartunk az MI fejlett alkalmazásánál, még az alapvető előfeltételek biztosításával vagyunk elfoglalva [*„Nem az a kihívás, hogy a munkavállalók okosabbá váljanak, hanem az, hogy egyszerűen nincs elegendő munkavállaló. Nem az a kérdés, hogyan lehetne hatékonyabban öntözni, hanem hogy egyáltalán van-e elegendő víz az öntözéshez.”* (R.Á.); *„A kutatói hálózat valódi erősséggé válhatna, de ehhez megfelelő anyagi támogatásra lenne szükség. Bár gyakran eljutunk oda, hogy nincs elég forrás, az alapvető szellemi kapacitás és az infrastruktúra adott.”* (E01)].

5.5. Kockázatok és okozatiság felmérése

Az EU AI Act a gazdasági társaságokra és szervezetekre bízta a végrehajtást, azonban ezeknek a cégeknek elsősorban a profit a céljuk, így nem feltétlenül prioritás számukra a természeti vagy társadalmi károk minimalizálása (R02). Ez egy régi dilemma: ha túlszabályozzuk a gazdasági szektort, előfordulhat, hogy a vállalatok inkább kivonulnak (R03) [*„Az a tőke pedig, amelyet Európában a hazai ipar fejlesztésére fordíthattak volna, befektetésként az amerikai projektekbe menekül.”* (E04)]. A környezeti károk közvetlenül, mint a rendszer fenntartása (például chipek vagy energia előállítás) okozta károk (R01), és közvetetten, mint a társadalomra gyakorolt egyenlőtlenség-erősítő károk (például társadalmi pontrendszerek vagy biometrikus azonosító rendszerek diszkriminatív használata) is megjelennek (R04). A piaci cégek alapvetően nagy figyelmet fordítanak a jogszabályi megfelelésre, ugyanakkor sok MI-rendszert – köztük nyílt forráskódú rendszereket – bárki fejleszthet, továbbfejleszthet vagy alkalmazhat, beleértve szervezett bűnözői és terrorista csoportokat is (R04, E01) [*„Ha van egy fekete doboz modellünk, ami általában értelmes válaszokat ad, akkor egy hozzáértő szakember rá tudja bírni ezt a modellt arra, hogy komplett őrülségeket csináljon.”* (R02)]. Az Egyesült Államokban és Kínában a jogalkotók kevésbé koncentrálnak a lehetséges veszélyekre, így problémát okozhat, ha például egy amerikai vállalat olyan globálisan elérhető megoldást fejleszt, amely rejtett kockázatokat hordoz magában. Amennyiben a mesterséges intelligencia rendszerbe kezdetektől fogva nem építenek be megfelelő kockázatkezelési mechanizmusokat, akkor ezeket utólag rendkívül nehéz kezelni (R04). A szabályozás nehezen tud lépést tartani a gyors technológiai változásokkal és a jelenleg még nem feltárt működési mechanizmusokkal, és ezáltal jelentős kockázatok maradhatnak felügyelet nélkül (R02).

A dezinformáció súlyos problémát jelent, különösen a valóság érzékelését alapvetően megkérdőjelező deepfake technológiák óriási fejlődése, és az MI bevonásával tökéletesített, manipulatív és lopakodó jellegű fekete minták alkalmazása (E01, E02, E04). Az internet tele lett automatizált botokkal, amelyek jelentősen befolyásolják a közvélemény alakulását (R01, R04). Korábban még lehetőség volt arra, hogy az emberek véleményét hitelesen megismerjük az interneten keresztül, de ma már nehéz valós képet kapni arról, hogy mi is az egyének tényleges gondolkodása. Ehelyett inkább azt lehet felmérni, hogy melyik ország vagy érdekcsoport irányít egy adott beszélgetést, amely aláássa az internet globális kapcsolatteremtő és kultúraformáló szerepét. A globális viták eltorzulnak a dezinformáció és a politikai manipuláció következtében, ahol befolyásos szereplők botokkal vagy más módszerekkel irányítják a diskurzust. Mindez egy már eleve polarizált világban történik, ahol az Európai Unió, Oroszország és a Közel-Kelet konfliktusokkal terhelt, az Egyesült Államok is politikailag megosztott és instabil.

A mesterséges intelligencia rendszerek nemcsak arra képesek, hogy teljes mértékben kielégítsék az egyéni igényeket, hanem arra is, hogy formálják ezeket az igényeket anélkül, hogy ez tudatos céljuk lenne. Folyamatosan olyan tartalmat kínálnak, amely a profilalkotás alapján a legjobban illeszkedik a felhasználóhoz, ezért az emberek vágyai és preferenciái alakítják magukat a kapott tartalmakhoz. Ez a folyamat komolyan veszélyeztetheti az egyén önismeretét és szuverenitását, elmosva a személyiség határait. Bár ez elsőre individuális problémának tűnik, ha rendszerszinten terjed el, akkor a társadalom egészében megjelenhet az a veszély, hogy az emberiség fogalma feloldódik, és egy teljesen fogyasztói társadalom válik uralkodóvá (E02).

5.6. Társadalmi és etikai megfontolások

Az MI-alapú rendszerek használata többféle módon fokozhatja a társadalmi egyenlőtlenségeket, a tanításukhoz használt adatok elfogultságtartalma miatti fals kimenetektől (R01, R02) a gazdasági előnyökig, melyekhez csak azok a szereplők férhetnek hozzá, akik a hatékony eszközöket megengedhetik maguknak (E03). A társadalmi és etikai dimenziók pedig jelenleg még mindig nem kapnak elég figyelmet a szabályozási folyamatokban (E02). Fontos aspektusa ennek a kérdésnek a környezeti rasszizmus, amely azt jelenti, hogy a marginalizált csoportok és a globális délről származók életére sokkal erőteljesebb negatív hatással lesznek az említett környezeti károk, miközben a technológiai innovációk pozitívumaiból kevésbé részesülnek (E04). A modellek fejlesztésének és

fenntartásának magas költségei implikálják, hogy azokat csak egy szűk eltcsoport fogja tudni birtokolni és hasznosítani (2023 Landscape Report, 2023, 19.) [*„A nagyobb szolgáltatók azok már elkezdik építeni a saját atomerőműveiket, mert akkora adatközpontjaik vannak.”* (R02)].

A 2010-es évek elején a közösségi média gyors, szabályozás nélküli elterjedése ma már komoly következményekkel jár, például fiatalok mentálhigiénés problémáival, online zaklatással és a közösségi médiából eredő nyomással. Utólag felmerülhet a kérdés, hogy szükség lett volna-e szigorúbb szabályozásra már az elején. Bár az Európai Unió szabályozása a legerősebb a világon az egyén személyi jogainak védelmében és törekvései egyfajta védelmet nyújthatnak a társadalom számára, de a közösségi média példáján látható, hogy mennyire fontos az időben történő és jól megtervezett szabályozás (E01). A technológia önmagában semleges; működését és céljait az határozza meg, hogyan állítják be az emberek. Az MI historikus adatok alapján működik, ami azt jelenti, hogy ha kritika nélkül tanítjuk meg, akkor a múltat rögzíthetjük megváltoztathatatlanul, ezzel esetlegesen mélyítve a társadalmi egyenlőtlenségeket. Az MI-nek képesnek kell lennie arra, hogy indokolja döntéseit, és csökkentse az elfogultságot, hiszen a nagy adattömegek gyakran hajlamosak a statisztikai közép felé torzítani, ami nem mindig ideális. A fejlesztésnek és alkalmazásnak társadalmilag felelős irányba kell haladnia, támogatva az egyéneket és a közösségeket anélkül, hogy pusztán a hatékonyságra törekedne. Az emberi döntéshozatalnak olykor hatékonytalannak kell lennie a magasabb szintű emberi értékek érdekében, amit meg kell tanítani az MI-nek. A társadalom jövője attól függ, hogy az állam milyen eszközkészletet hoz létre és hogyan építi fel stratégiáit. Ha az állam olyan eszközöket fejleszt és ad az emberek kezébe, amelyek javítják az életüket, akkor a társadalom fejlődni fog. Ha viszont az állam nem fordít figyelmet erre, az emberek pusztán termékké válnak, és emiatt kiszolgáltatottá válnak a gazdasági és technológiai hatásokkal szemben (R.Á.).

A legfontosabb a szakértők szerint, hogy az MI emberközpontú legyen, ennek legobjektívebb alapja az emberi jogok tiszteletben tartása. A technológia önmagában nem képes garantálni a diszkriminációmentességet – ehhez gondos tervezés és folyamatos ellenőrzés szükséges. Ha olyan adathalmazokon tanítanánk az MI-t, amelyekben alapvetésként szerepelnek bizonyos jogi és etikai elvek – például az emberi élet és méltóság tisztelete –, akkor elméletileg megtanulhatná, hogy döntései milyen hatással lehetnek az emberi életre és annak minőségére. A jogszabályok azonban hiába objektívek, a jog alkalmazása mindig szubjektív lesz, hiszen az MI válik a jog alkalmazójává (R03).

5.7. Interjúk értékelése

Az interjúkat a megkérdezett szakértők politikai akarata és kockázatokkal kapcsolatos társadalmi tudatossága szempontjából vizsgáltam, az Amsden et al. (2012:182) keretrendszer alapján. Az interjúkat a következő három kritérium szerint értékeltem, és pontoztam 1-től 5-ig terjedő skálán (17. ábra):

(1) Társadalmi hatás felismerése / szegénység (ebben az esetben: a digitális szakadék) negatív következményeinek felismerése: Az adott interjú mennyire tükrözi az MI-szabályozás hiányosságainak és ezek társadalmi következményeinek (például az egyenlőtlenségek mélyülése) felismerését, illetve az interjúalanyok mennyire érezték erősnek az általuk végzett tevékenység társadalmi hatását és jelentőségét. A pontszámok széles skálán mozognak, ami arra utal, hogy az interjúalanyok között nincs egyöntetű felismerés a társadalmi és etikai következményekkel kapcsolatban és nem minden érintett látja tisztán az MI kockázatainak negatív hatásait.

(2) Változtatás lehetőségének érzékelése: Mennyire érzékelhető az interjúban a változtatási lehetőségek iránti nyitottság, vagy a pozitív irányú változtatás lehetőségének megléte, milyen mértékben láttak esélyt vagy lehetőséget a változtatásra, illetve, hogy érzékelték-e, hogy képesek befolyásolni a helyzetet. Itt is jelentős az ingadozás a pontszámok között, ami arra utalhat, hogy az érintettek bizonytalanok vagy pesszimisták a jelentős változások megvalósíthatóságát illetően a jelenlegi politikai vagy szabályozási keretek között. Ez összhangban áll a felvetéssel, amely szerint nincs megfelelő politikai akarat a kockázatok súlyosságával arányos lépések megtételére.

(3) Felelősségérzet megléte: Az interjú alanyai milyen mértékben érzik saját vagy intézményük szerepét és felelősségét az MI-rendszerek társadalmi hatásainak alakításában. A pontszámok itt is ingadoznak, ami azt jelzi, hogy a felelősségérzet egyenetlenül oszlik meg. Ez támogatja a felvetést, miszerint hiányzik a kollektív elszámoltathatóság, ami hozzájárulhat ahhoz, hogy nincsenek arányos intézkedések a kockázatok kezelésére.

Az interjúkat konszenzuskereséssel is értékeltem. A változók közötti konszenzus mértékében különbségek vannak, amelyeket az **átlagos konszenzusszint** ábrája mutat (18. ábra). A „Kompetencia szintje” és a „Társadalmi és etikai megfontolások” kiemelkednek, ezeken a területeken az interjúalanyok között nagyobb volt a konszenzus. A **gyakorisági eloszlás** (19. ábra) megmutatja, hogy mely területeken vannak vegyes vagy egyöntetű

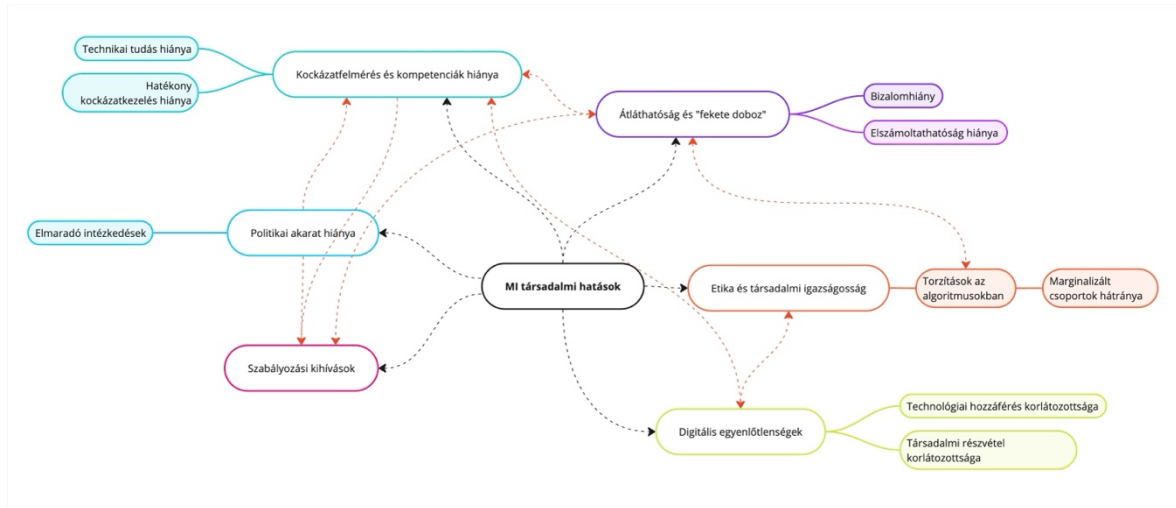
nézőpontok, segítve ezzel annak azonosítását, hogy hol lenne szükség részletesebb elemzésre vagy további egyeztetésekre. Domináns kategória például a „Társadalmi és etikai megfontolások”, ennél a változónál több interjú is magas konszenzust mutatott. Szórtabb konszenzust mutat a „Politikai akarat” vagy „Kommunikáció”, ezeknél a változóknál vegyes kategóriák jelennek meg (alacsony, közepes és magas konszenzus egyaránt).

6. Konklúzió

A kutatás eredményei alapján a döntéshozók jelenlegi kompetenciája nem elegendő az MI-rendszerek kockázatainak alapos felméréséhez és az arányos politikai cselekvéshez. Bár az akadémiai és szakértői közegben van interdiszciplináris tudás, az állami szektor kompetenciái gyakran hiányosak, ami gátolja a hatékony szabályozást. Az aktorok közötti kommunikáció sem elégséges: a jogi, technológiai és politikai szereplők közötti együttműködés fejlesztésre szorul a hatékony kockázatkezelés érdekében. A politikai akarat jelenléte változó; míg bizonyos kezdeményezések biztatóak, ezek nem mindig elegendők a gyors és hatékony lépések megtételéhez, így a felvetések igazolódtak. A hatékony MI-szabályozáshoz elengedhetetlen a kompetenciák bővítése, a kommunikáció javítása és az elkötelezett politikai cselekvés. A kommunikációs csatornák fejlesztése, különösen a szakemberek, fejlesztők és politikai döntéshozók között, elősegítheti a szabályozási célok jobb megértését és a közös cselekvést. Az akadémiai közeg erőssége és a társadalmi tudatosság növelése jelentős lehetőségeket kínál a szabályozás társadalmi beágyazottságának javítására. Ha az állam és a szakemberek közösen dolgoznak a társadalom felkészítésén és a technológia etikus alkalmazásán, az MI-alapú rendszerek használata hozzájárulhat a közjóhoz és a fenntartható fejlődéshez. Az EU szabályozási kerete ígéretes alapot teremtett a jövőbeli fejlődéshez, különösen a hosszú távú fenntarthatóság és a felelős innováció elősegítése szempontjából. A globális versenyképesség fenntartása érdekében növelni kell a jogalkotás rugalmasságát, és figyelembe kell venni az új technológiák gyors fejlődését. Magyarország számára különösen fontos, hogy kihasználja a meglévő tudományos és kutatási kapacitásokat, ösztönözze az innovációt és erősítse a szakértők és döntéshozók közötti párbeszédet. A társadalmi és etikai szempontok előtérbe helyezése elengedhetetlen ahhoz, hogy a mesterséges intelligencia valóban emberközpontú és igazságos lehessen. Az emberi jogok tiszteletben tartása és az átláthatóság biztosítása hozzájárulhat ahhoz, hogy az MI alkalmazása ne csak a gazdasági fejlődést, hanem a társadalmi igazságosságot és a közjót is támogassa.

7. Függelék

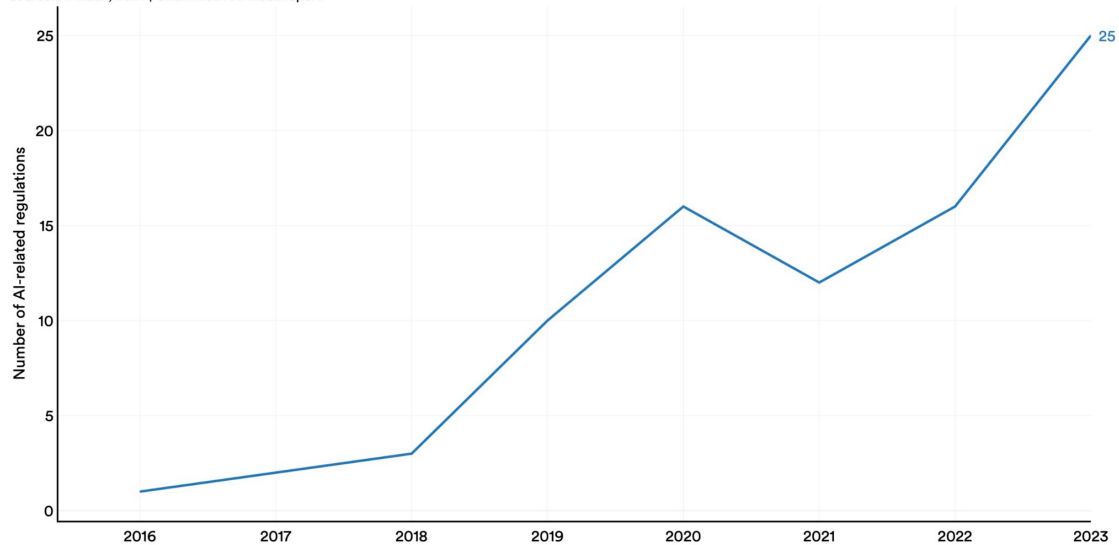
7.1. Ábrajegyzék



1. ábra: Problématérkép, saját szerkesztés

Number of AI-related regulations in the United States, 2016–23

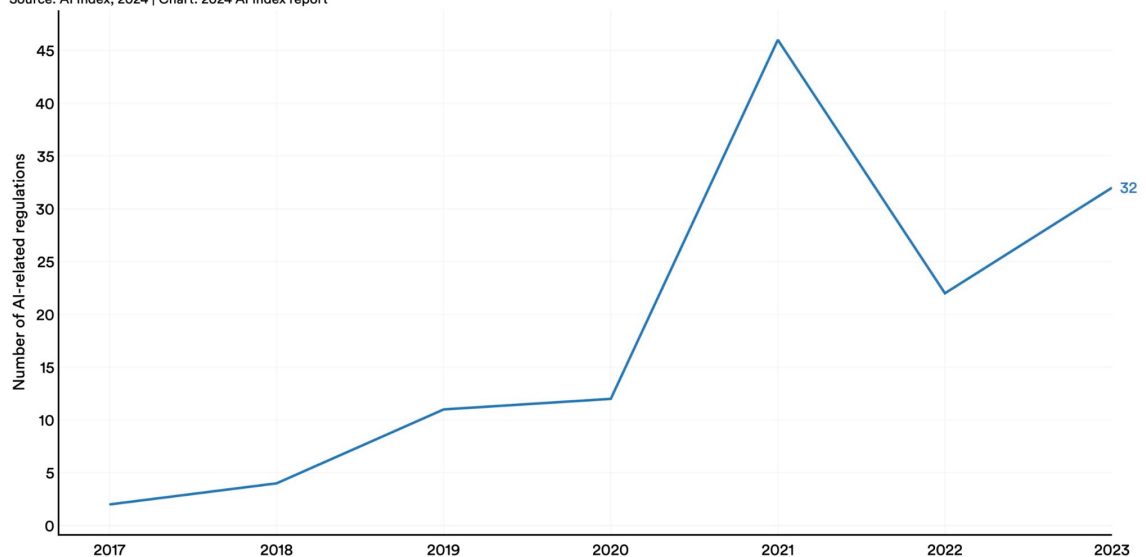
Source: AI Index, 2024 | Chart: 2024 AI Index report



2. ábra: MI-vel kapcsolatos szabályozások száma az Egyesült Államokban, 2016-23

Number of AI-related regulations in the European Union, 2017–23

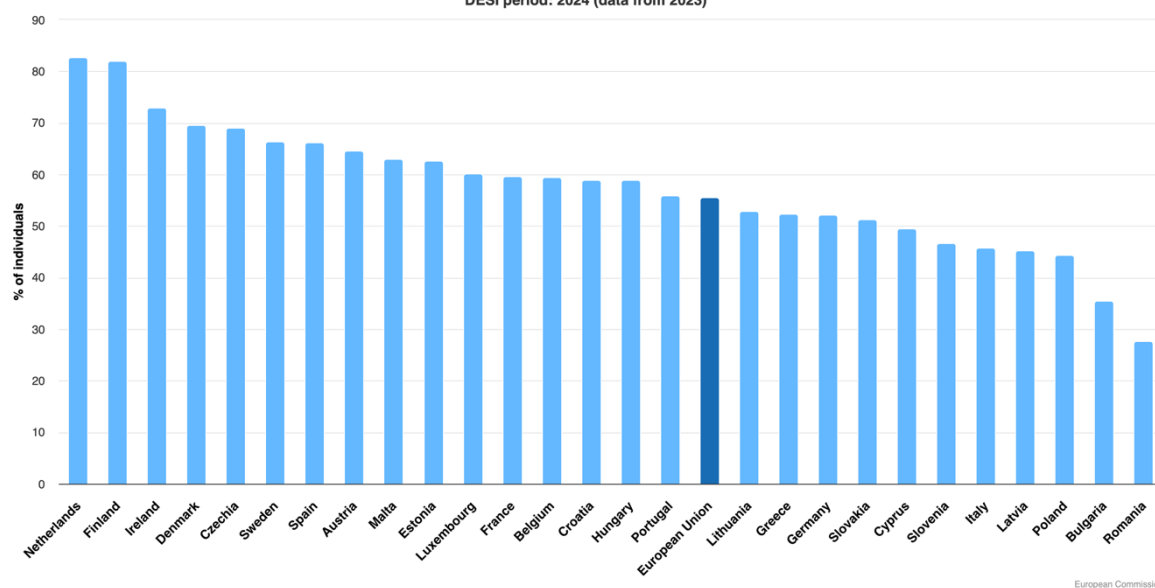
Source: AI Index, 2024 | Chart: 2024 AI Index report



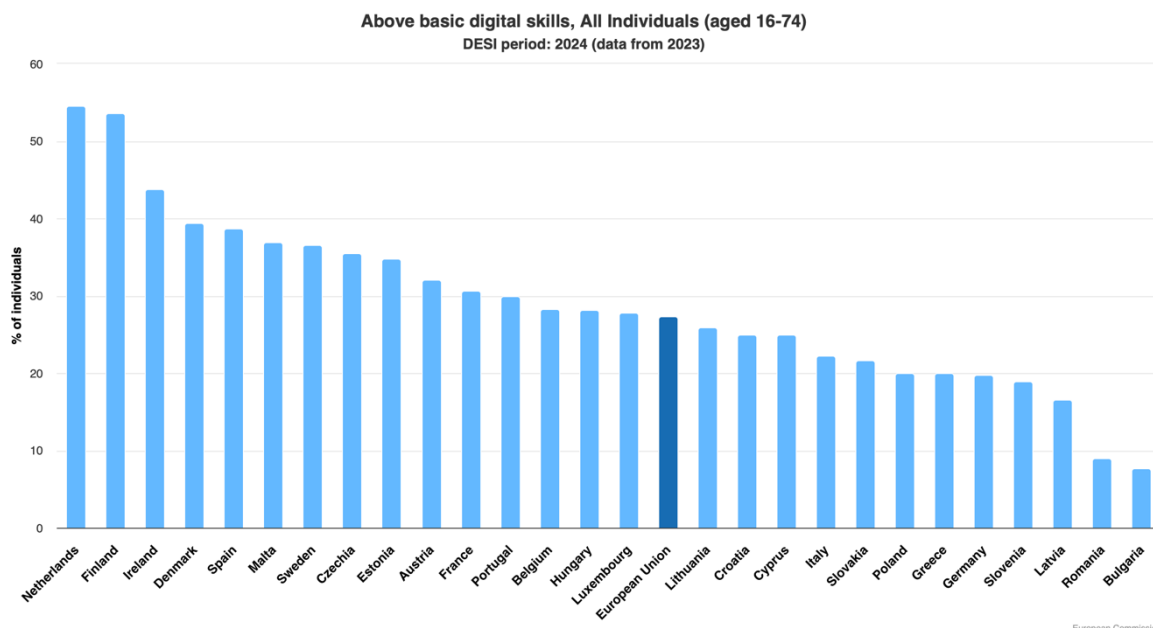
3. ábra: MI-vel kapcsolatos szabályozások száma az Európai Unióban, 2017-23

At least basic digital skills, All Individuals (aged 16-74)

DESI period: 2024 (data from 2023)



4. ábra: Legalább alapszintű digitális készségek, 2023-as adat



5. ábra: Alapszintet meghaladó digitális készségek, 2023-as adat

4-5. ábra: „A következő öt kompetencia MINDEN területén „alapszintű” vagy „alapszintet meghaladó” digitális készségekkel rendelkező személyek:

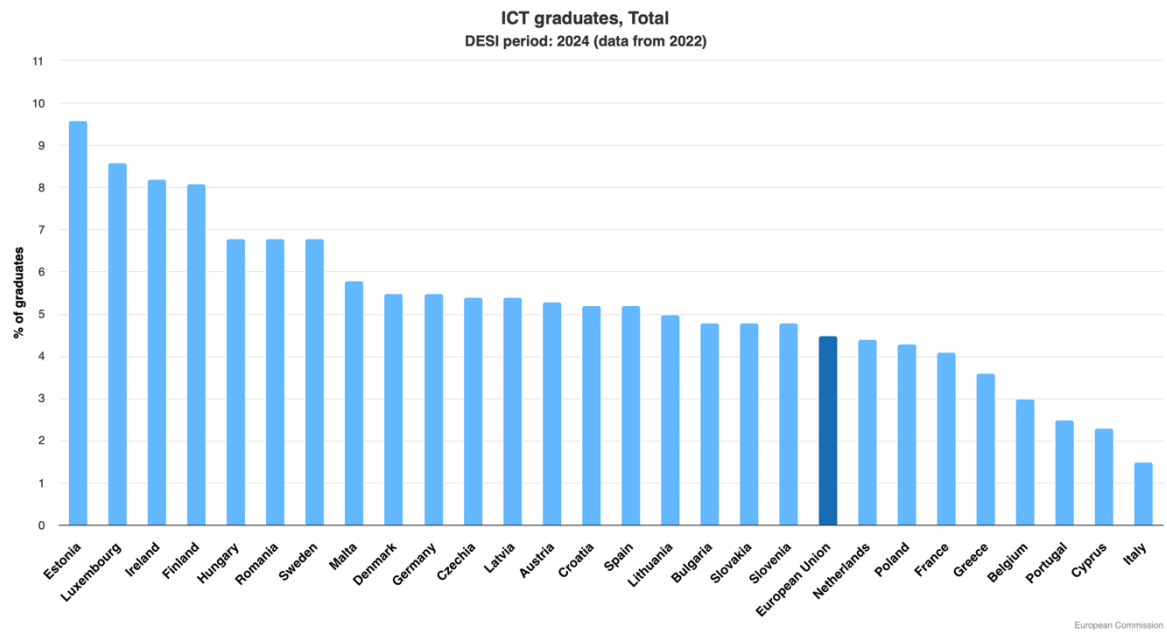
1. *Információs és adatismeret:* Információk keresése árukról vagy szolgáltatásokról; Egészséggel kapcsolatos információk keresése; Online híroldalak, újságok vagy hírmagazinok olvasása; Az online információk és forrásaik tényellenőrzésével kapcsolatos tevékenységek.

2. *Kommunikáció és együttműködés:* E-mailek küldése/vétele; telefonálás/videóhívás az interneten keresztül; azonnali üzenetküldés; közösségi hálózatokban való részvétel; véleménynyilvánítás polgári vagy politikai kérdésekben weboldalakon vagy a közösségi médiában; részvétel online konzultációban vagy szavazásban polgári vagy politikai kérdések meghatározására.

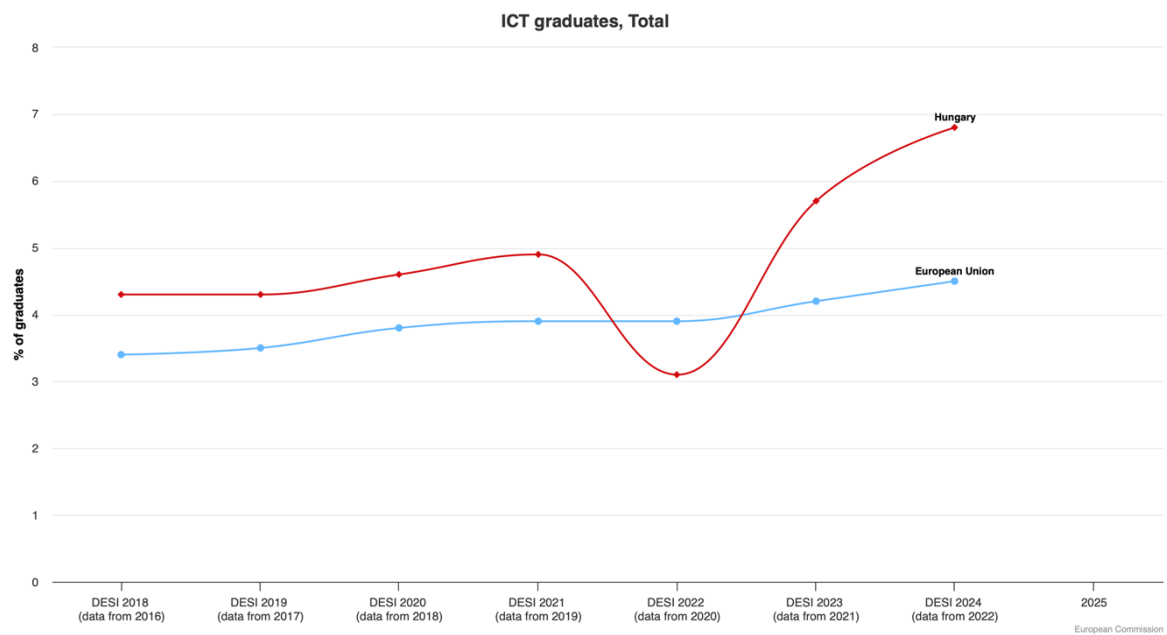
3. *Digitális tartalom létrehozása:* Szövegszerkesztő vagy táblázatkezelő szoftverek használata; Fotók, video- vagy hangfájlok szerkesztése; Fájlok (például dokumentumok, adatok, képek, videók) másolása vagy mozgatása mappák, eszközök (e-mail, azonnali üzenetküldés, USB, kábel) között vagy a felhőben; Több elemet, például szöveget, képet, táblázatot, diagramot, animációt vagy hangot tartalmazó fájlok létrehozása; A táblázatkezelő szoftverek fejlett funkcióinak (függvények, képletek, makrók és egyéb fejlesztői funkciók) használata adatok rendszerezésére, elemzésére, strukturálására vagy módosítására; Kód írása egy programozási nyelven.

4. *Biztonság:* A saját személyes adatokhoz való hozzáférés kezelése annak ellenőrzésével, hogy a weboldal, ahol a válaszadó személyes adatokat adott meg, biztonságos-e, a személyes adatok megadása előtt az adatvédelmi nyilatkozatok elolvasása, a saját földrajzi helyhez való hozzáférés korlátozása vagy elutasítása, a közösségi oldalakon vagy megosztott online tárhelyeken található profilhoz vagy tartalomhoz való hozzáférés korlátozása, a személyes adatok reklámcélú felhasználásának elutasítása, a saját internetes böngésző beállításainak módosítása a cookie-k megakadályozása vagy korlátozása érdekében a válaszadó bármely eszközén.

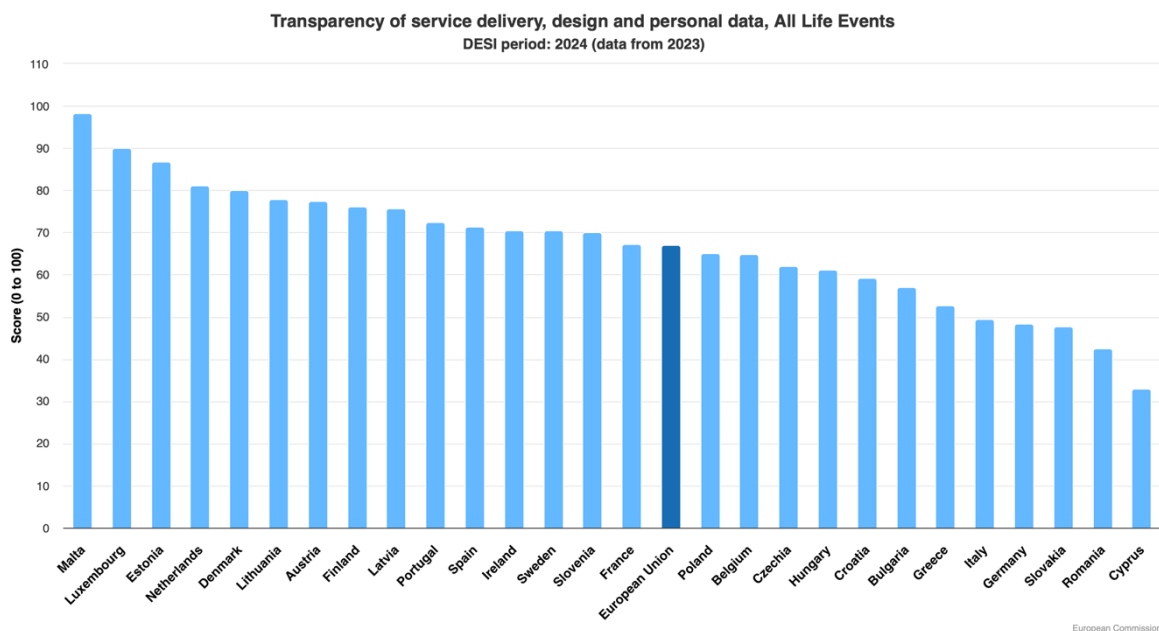
5. *Problémamegoldás:* Szoftverek vagy alkalmazások letöltése vagy telepítése; Szoftver, alkalmazás vagy eszköz beállításainak módosítása; Online vásárlások; Online értékesítés; Online tanulási források használata; Internetes banki szolgáltatások; Álláskeresés vagy álláspályázat elküldése.”



6. **ábra:** IKT-diplomával (Információs és kommunikációs technológiák) rendelkező személyek aránya az EU tagországaiban, 2022-es adat



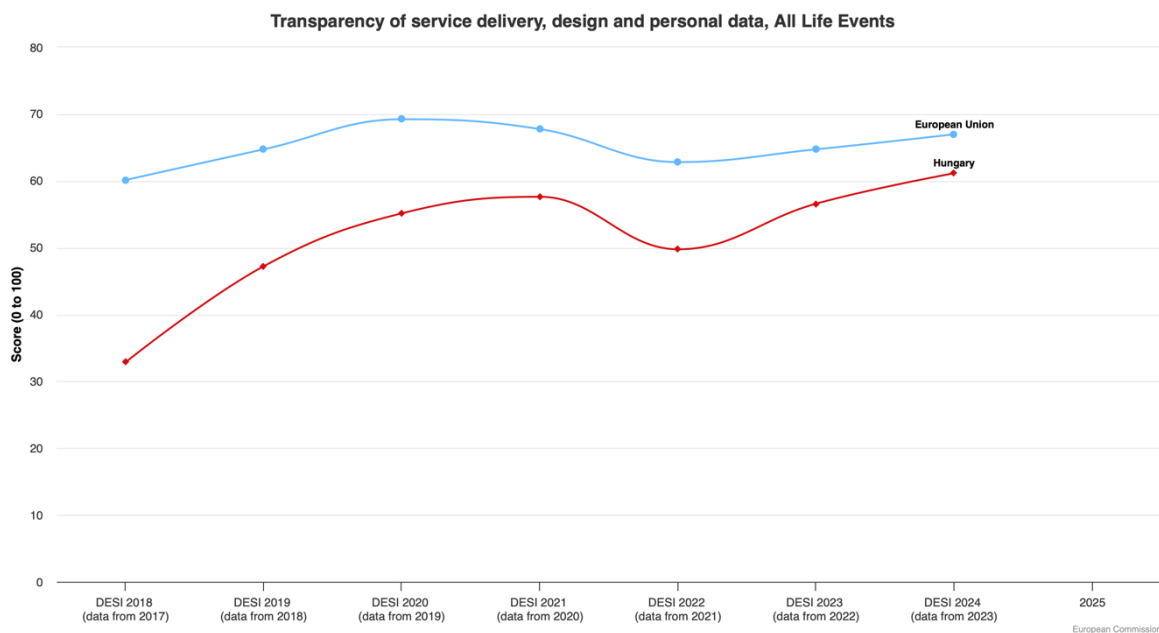
7. **ábra:** IKT-diplomával rendelkező személyek arányának alakulása, Magyarország és EU átlag



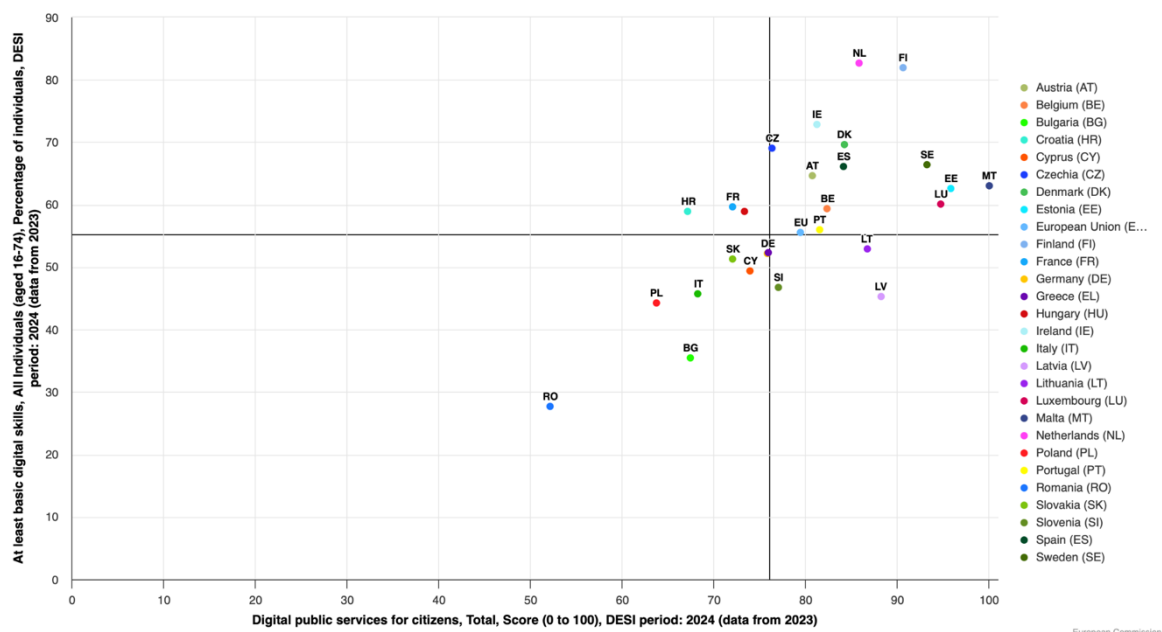
8. ábra: Közzolgáltatási transzparencia (a szolgáltatásnyújtás, a tervezés és a személyes adatok kezelésének átláthatósága)

„A szolgáltatásnyújtás, a tervezés és a személyes adatok átláthatósága: a szolgáltatási folyamatok átláthatósága, a szolgáltatások felhasználói bevonással történő kialakítása és a felhasználók személyes adataik kezelésének lehetősége. A mutató több elem összesített pontszáma, amely többek között a következőket írja le:

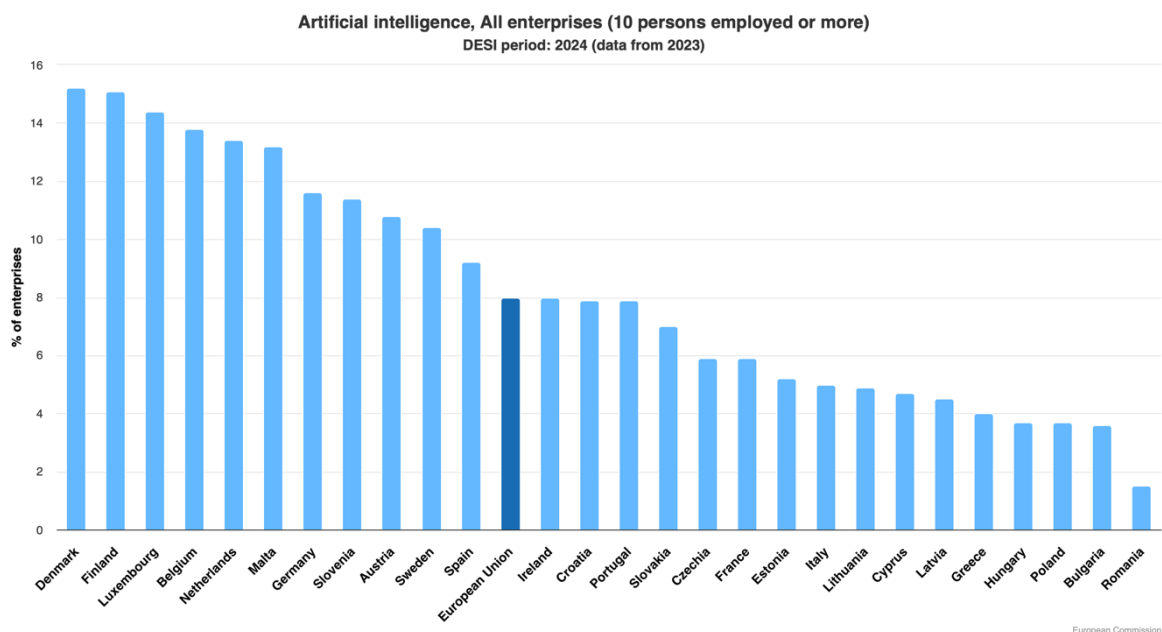
- a teljesítésről szóló értesítést nyújtó szolgáltatások aránya,
- azon szolgáltatások aránya, amelyek nyomon követik az előrehaladást a teljesítés során,
- a portálok személyes adatokhoz való online hozzáférés mértékét,
- a felhasználók bevonásának mértéke a szolgáltatások kialakításába,
- és a felhasználók bevonásának mértéke a szakpolitikai döntéshozatali folyamatokba.”



9. Ábra: Közzolgáltatási transzparencia alakulása, Magyarország és EU átlag

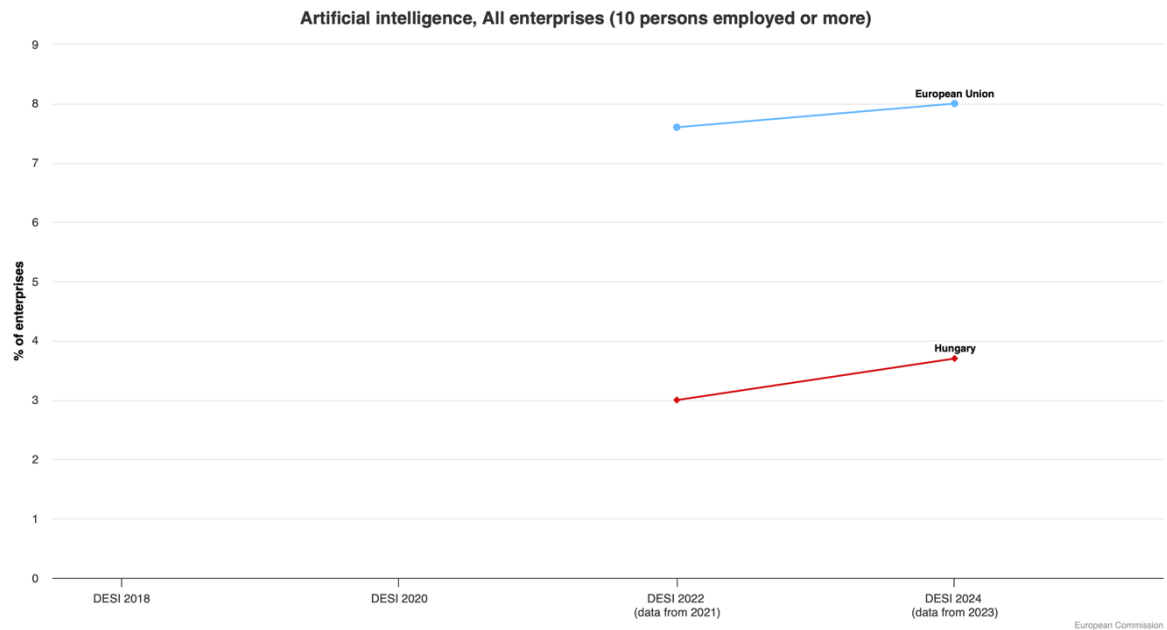


10. ábra: Az alapvető digitális készségekkel rendelkező személyek aránya és a polgárok számára elérhető digitális közszolgáltatások szintje

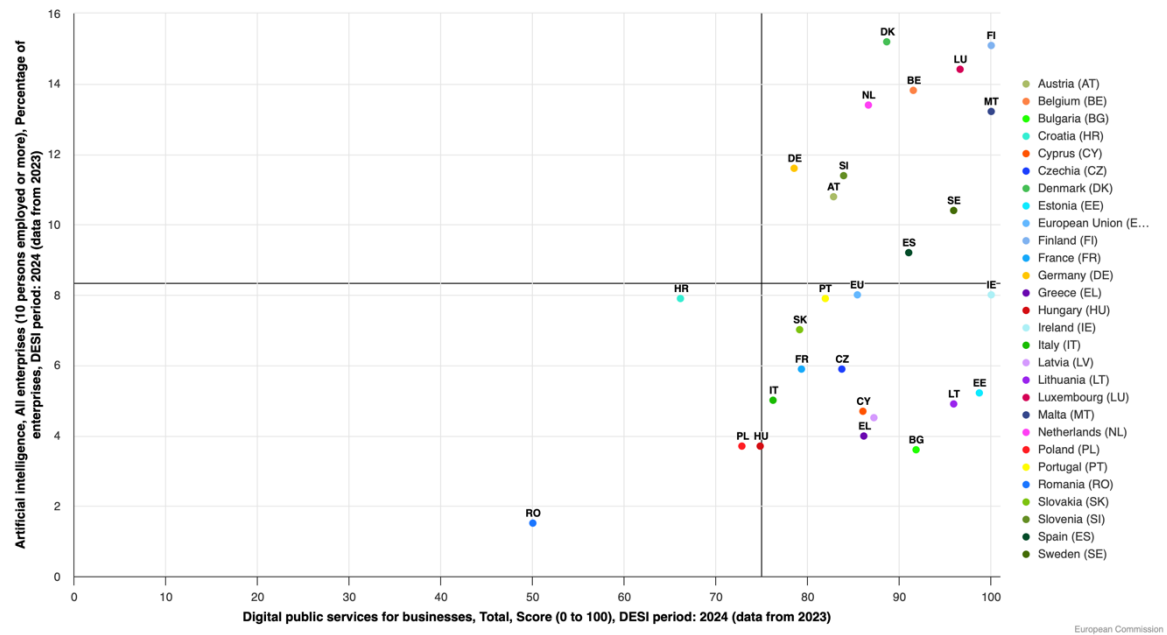


11. ábra: MI-technológiát használó vállalkozások aránya az EU tagországaiban

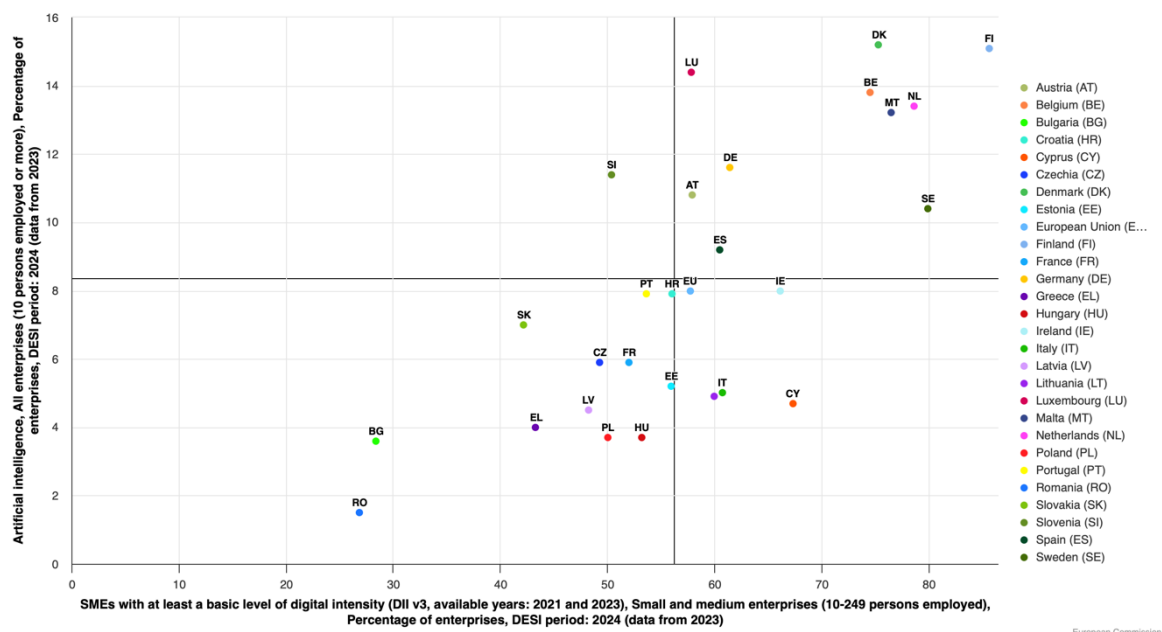
„A 10 vagy annál több főt foglalkoztató vállalkozások. Minden feldolgozóipari és szolgáltatási ágazat, a pénzügyi szektor kivételével. Minden állami vagy magánvállalkozás, a pénzügyi szektor nélkül.”



12. ábra: MI-használat alakulása, Magyarország és EU átlag

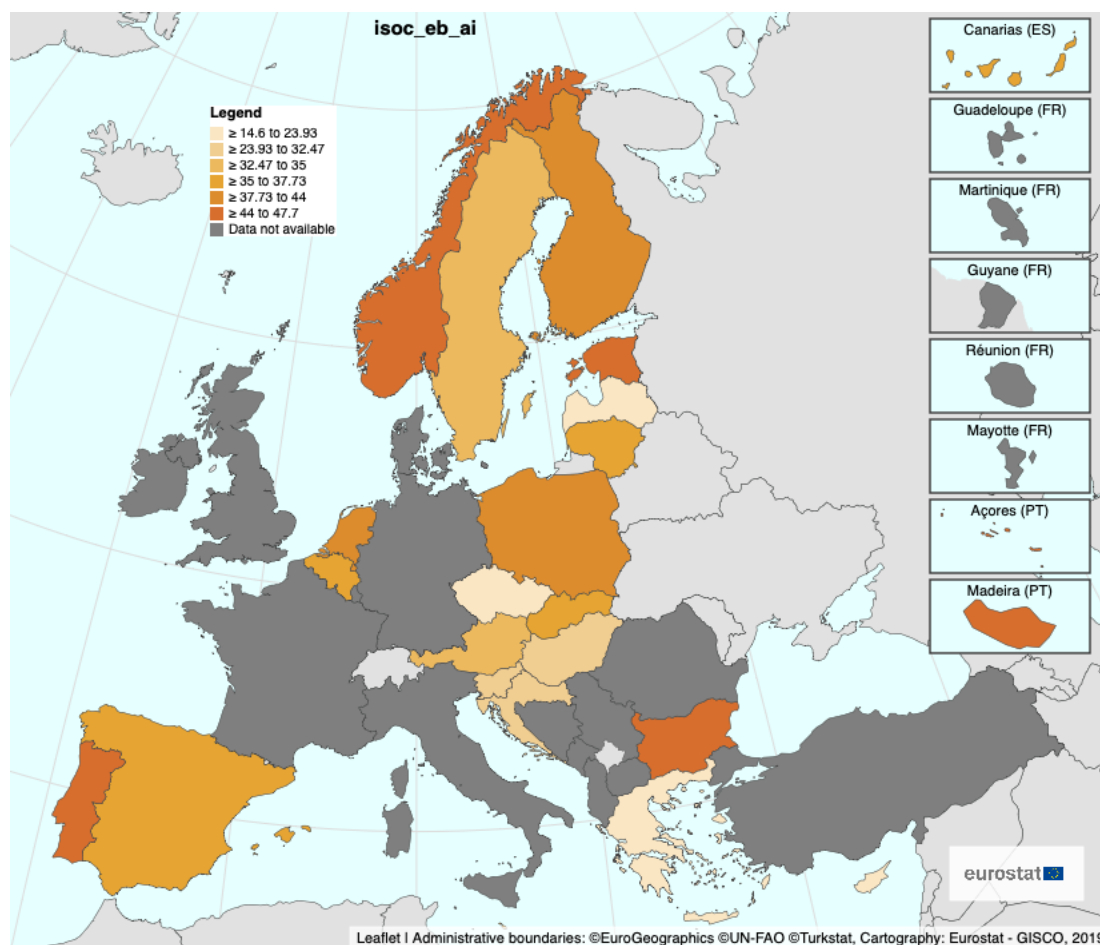


13. ábra: A vállalkozások számára elérhető digitális közszolgáltatások szintje és a mesterséges intelligenciát alkalmazó vállalkozások aránya

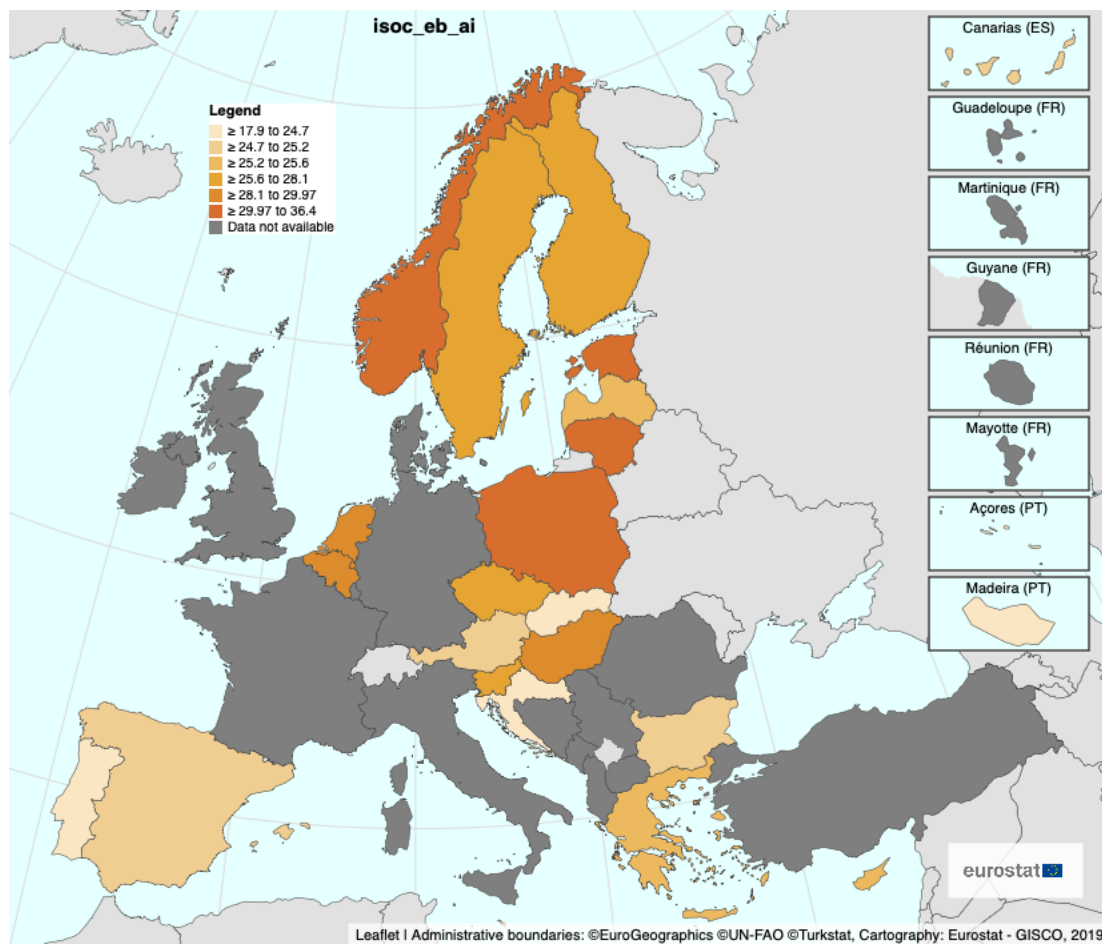


14. ábra: Az MI-t alkalmazó vállalkozások aránya és a legalább alapvető digitális intenzitással rendelkező kis- és középvállalkozások (KKV-k) aránya

2-14. ábra forrás: DESI dashboard for the Digital Decade (2023 onwards), <https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts>, Letöltve: 2024.11.10.

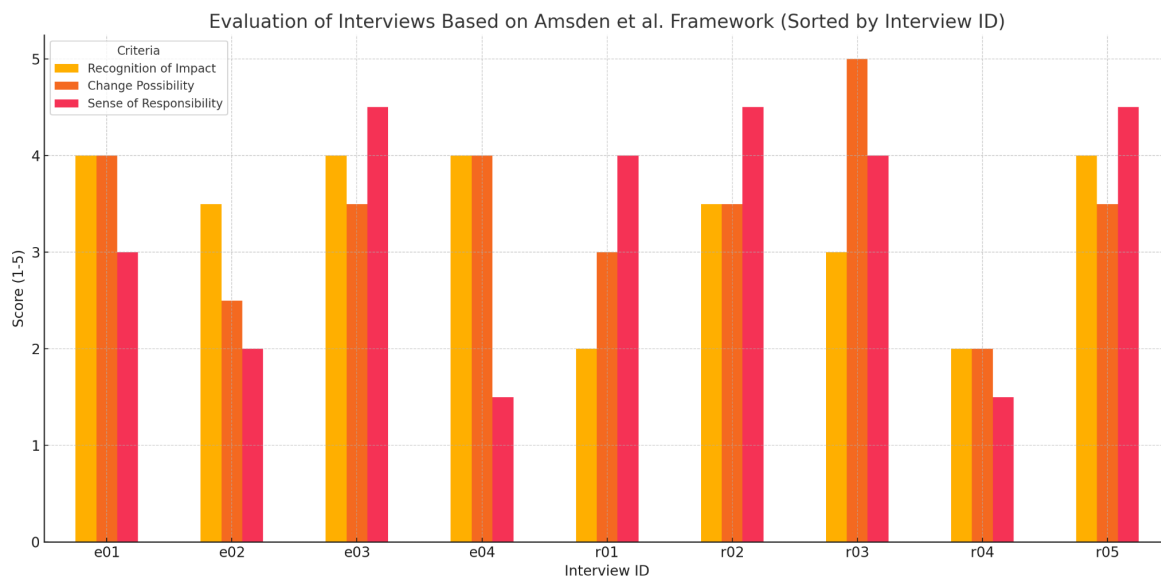


15. ábra: Az EU tagországaiban lévő nagyvállalatok MI-használata, 2023-as adat



16. ábra: Az EU tagországaiban lévő kis- és középvállalkozások MI-használata, 2023-as adat

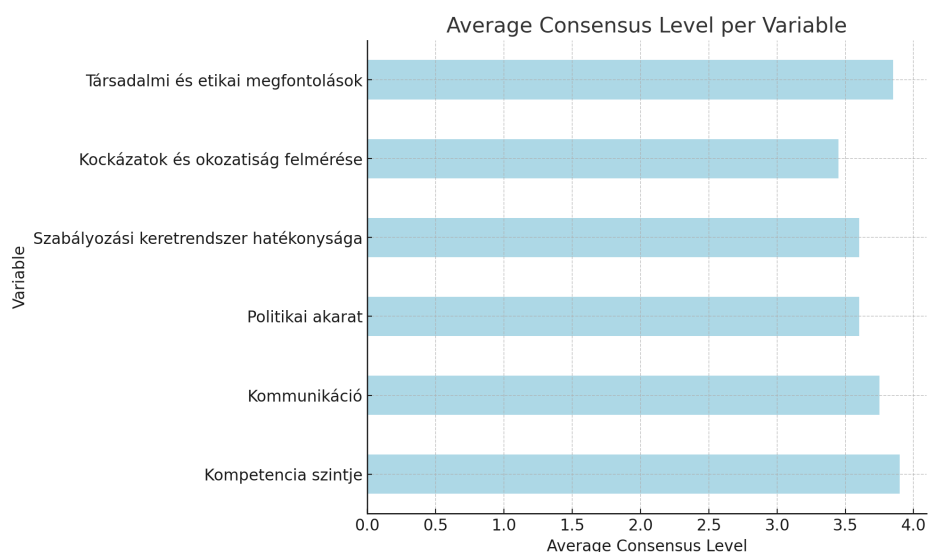
15-16. ábra forrás: Eurostat: Artificial intelligence by size class of enterprise, https://ec.europa.eu/eurostat/databrowser/view/isoc_eb_ai_custom_13484102/default/table?lang=en. Letöltve: 2024.11.10.



17. ábra: Az interjúk értékelése az Amsden et al. keretrendszer alapján, saját szerkesztés

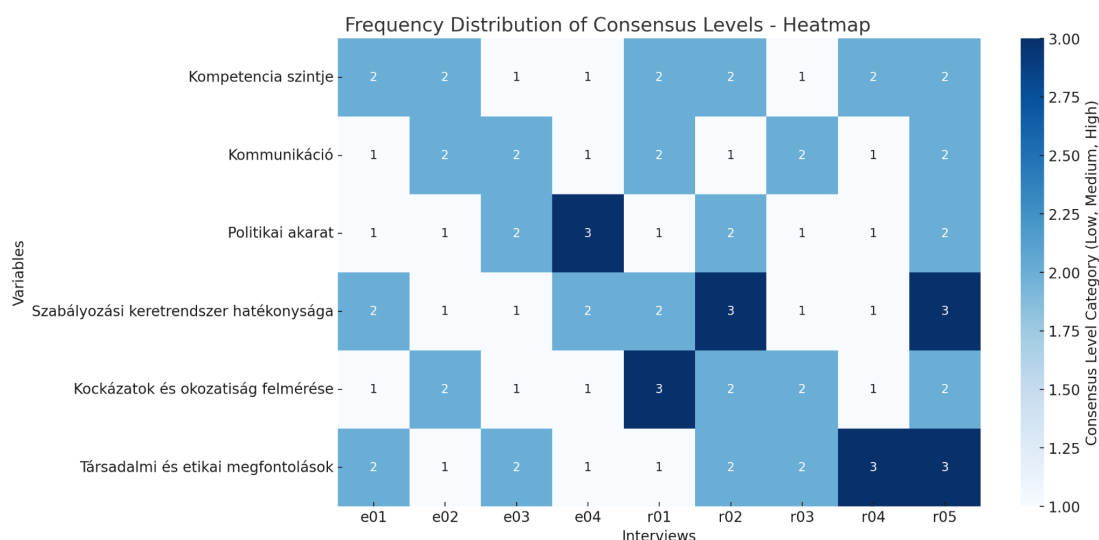
Az interjúk tartalmát kulcsszavas kereséssel elemeztem, hogy az egyes kritériumokhoz tartozó bizonyítékokat találjak. Például: (1) Hatás felismerése: Olyan szavak és kifejezések, mint „fontos”, „befolyás”, „társadalmi hatás” stb.; (2) Változtatási lehetőség: Kifejezések, mint „lehetőség”, „változtatás”, „módosítás” stb.; (3) Felelősségérzet:

Szavak és mondatok, mint „felelősség”, „kötelességem”, „szerepem” stb. Az értékelés során 1-től 5-ig terjedő skálát használtam, ahol 1: Egyáltalán nem fejezte ki a kritériumot, 5: Nagyon erőteljesen és kifejezetten jelen volt a kritérium. Az egyes szövegrészek kontextusa és a kulcsszavak gyakorisága alapján pontozás történt, de a végső értékelés figyelembe vette a szöveg általános hangvételét és tartalmát.



18. ábra: Átlagos konszenzusszint, saját szerkesztés

Az interjúban szereplő válaszokból gyűjtöttem össze az egyes változókhoz tartozó pontszámokat, amelyek tükrözték a válaszadók egyetértését, kompetenciáját, felelősségérzetét stb. Az egyes változókhoz tartozó konszenzusszinteket numerikus értéként kódoltam (1-től 5-ig terjedő skála). Az összes interjúra vonatkozó pontszámokat átlagoltam, így minden változóhoz egy átlagos konszenzusszintet kaptam.



19. ábra: A konszenzusszint gyakorisági eloszlása, saját szerkesztés

Az interjúk során az egyes változókhoz kapcsolódó válaszok gyakoriságát gyűjtöttem össze. Minden interjúhoz és változóhoz hozzárendeltem a válaszok kategóriáit, amelyek alacsony, közepes vagy magas konszenzusszintet jelöltek. Az egyes kategóriák gyakoriságát megszámláltam minden változóra és interjúra vonatkozóan. Például, ha egy változó esetében háromszor jelent meg egy „magas” konszenzusszint, akkor az adott változóhoz tartozó cellába a 3-as érték került.

17-19. ábra: Az elemzéshez a ChatGPT 4o generatív mesterséges intelligencia alapú rendszert használtam, a következő metodológiával: feltöltöttem az interjúk szövegét, promptoltam az elemzés formuláit, definiáltam a kimenetet és legeneráltam a grafikont a programmal.

7.2. Esettanulmányok

Az elméleti keretek absztrakt jellegének érthetőbbé tétele érdekében esettanulmányokkal szemléltetem az MI-alapú rendszerek etikai dilemmáit a gyakorlatban.

A mesterséges intelligencia alkalmazása során előforduló kognitív elfogultságok különösen nagy kihívást jelentenek az MI-vezérelt toborzási folyamatokban, ahol a rendszerekbe épülhetnek be az emberi előítéletek. Az automatizált munkaerő-felvételi platformok (AHP-k) például digitális közvetítőként forradalmasítják a munkaerő-felvételi folyamatot olyan algoritmusok segítségével, amelyek a beszerzéstől a kiválasztásig tartó szakaszokat kezelik; azonban, miközben a technológia hatékonyságot és racionalizált műveleteket ígér, jelentős aggályokat vet fel a magánélet megsértésével, az előítéletek állandósulásával és a hierarchikus struktúrák fenntartásával kapcsolatban, amelyek a profitot előnyben részesíthetik a demokratikus munkakapcsolatokkal szemben (Ajunwa & Schlund, 2020). Hasonlóan dilemmákat vetnek fel a viselhető munkahelyi technológiák, az intelligens szemüvegektől kezdve az alkalmazottak termelékenységének és biztonságának ellenőrzésére és fokozására tervezett exoskeletonokig. Ezek a technológiák, bár különböző működési kontextusokban előnyösek, a munkavállalók magánéletének védelmével és az összegyűjtött adatok pontosságával kapcsolatban megbízhatatlanok. Ezek az adatok – amelyek esetleg hibásak vagy manipuláltak – negatívan befolyásolhatják a munkavállalói kártérítési és biztonsági szabályozásokat is. Az ügyfélkapcsolat-kezelő rendszerek (CRM-rendszerek) elősegíthetik az ügyfelek közötti eltérő bánásmódot, az ügyfélkapcsolatok kihasználását és jelentős adatvédelmi aggályokat vetnek fel. A CRM félrevezetheti az ügyfeleket és védett tulajdonságok alapján diszkriminálhat, amit a célzott reklámozás és a prediktív analitika segít, ami akaratlanul is támogathatja a „digitális redlining”¹⁸ egy formáját.

Az olyan nagy nyelvi modellek, mint például az OpenAI GPT-2, ChatGPT és a Meta Llama 2 esetében a kognitív előítéletek kritikus hatással vannak nemcsak a nemi sztereotípiák fenntartására, hanem a döntéshozatali folyamatok és a mesterséges intelligencia technológiákkal való felhasználói interakciók befolyásolására is (UNESCO IRCAI, 2024). Az LLM-ek esetében vizsgálták a társadalmi nem-alapú szóasszociációkat, a társadalmi nemhez kötött témák iránti érzelmeket, valamint a társadalmi nemi- és kulturális kontextus alapján generált szövegek sokféleségét vagy éppen annak hiányát. Az LLM-ek a női neveket olyan háztartási szerepekhez társították, mint az „otthon”, a „család”, a „gyerekek” és a „házasság”, míg a férfi neveket olyan szakmai és üzletorientált kifejezésekkel kapcsolták össze, mint az „üzlet”, a „vezető”, a „fizetés” és a „karrier”. Olyan is előfordult, hogy szexista és nőgyűlölő tartalmakat generáltak: a generált szövegek mintegy 20 százaléka a nőket lekicsinylő szerepekben vagy birtoklási tárgyként ábrázolta. A Llama 2 és a GPT-2 az esetek hetven, illetve 60 százalékában diszkriminatív tartalmakat produkált. Ez aláhúzza azt az általánosabb problémát, hogy a mesterséges intelligencia rendszerek potenciálisan megerősítik a marginalizált csoportokkal szembeni káros sztereotípiákat és előítéleteket.

Az automatizált vagy részben automatizált jóléti rendszereket – például munkanélküli ellátások jogosultságának meghatározása (Park & Humphry, 2019), szociális segélyek kiosztása és monitorozása (Singh et al., 2019, Eubanks, 2012, 2018), egészségügyi szolgáltatások hatékonyságának javítása (Lee & Yoon, 2021, Sodhro & Zahid, 2021), gyermekvédelmi szolgáltatások kockázatkezelése (Broadhurst et al., 2010, Keddell, 2015, Eubanks, 2012, 2018), hajléktalan emberek nyilvántartási rendszerei és ellátásuk optimalizálása (Kuo et al., 2023, Thomas & Mackie, 2019, Das, 2021, Eubanks, 2018) – is befolyásolja a kognitív torzítások okozta diszkriminatív működés. Ezek a területek a társadalom periferiájára szorult embereket érintik a legkritikusabb mértékben, ezért a rendszerekbe épült kognitív torzítások szó szerint végzetesek lehetnek. Több esettanulmány mutat rá a digitális jóléti jogosultsági rendszer hibáira, amely emberi életet veszít: a rendszerek nem vették figyelembe az egyedi esetek összetettségét, emiatt az áldozatok elvesztették egészségügyi ellátásukat, oktatási támogatásukat, vagy kiszorultak a lakhatási támogatásból (Eubanks, 2018). Egy, a bántalmazás veszélyének kitett gyermekek azonosítására fejlesztett algoritmus olyan közvetítő adatokat használt, amelyek aránytalanul nagy mértékben érintették a színesbőrű családokat, mivel az alacsony jövedelmű és kisebbségi családokat három és félszer nagyobb arányban jelentették bántalmazás és elhanyagolás miatt, mint a fehér családokat.

Az Egyesült Államokban az igazságszolgáltatásban is használtak MI-alapú rendszereket a bírók döntéseinek támogatására. Ez a rendszer különféle tényezők alapján becsülte meg a vádlott visszaesési kockázatát százalékos formában. Az alkalmazás azonban kritikákat váltott ki, mert az eredmények alapján a színes bőrű vádlottakat

¹⁸ „A *redlining* olyan diszkriminatív gyakorlatként definiálható, amely a szolgáltatások, például a jelzáloghitelek, biztosítási hitelek és egyéb pénzügyi szolgáltatások szisztematikus megtagadásából áll bizonyos területek lakói számára, faji vagy etnikai hovatartozásuk alapján.”, Forrás: <https://www.law.cornell.edu/wex/redlining>. Letöltve: 2024.05.05.

gyakrabban minősítette magas visszaesési kockázatúnak, amit a közvélemény diszkriminatívnak tartott, és „rasszistának” nevezett. Későbbi vizsgálatok megállapították, hogy az algoritmus nem a bőrszín alapján hozta az értékeléseit, hanem más tényezőket, például a jövedelmi helyzetet, a lakókörnyezetet és a családi hátteret vett figyelembe. Ezek a tényezők azonban gyakran összefüggenek a társadalmi-gazdasági helyzettel, ami miatt a színes bőrű vádlottak nagyobb arányban kerültek a magas kockázatú kategóriába. A vizsgálat azt is megállapította, hogy a rendszer súlyozása nem volt megfelelő, mivel figyelmen kívül hagyott olyan fontos tényezőket, mint az elkövetett bűncselekmény típusa és a korábbi bűnügyi előélet, amelyek pontosabb elemzést eredményezhettek volna (Török & Zódi, 2021).

Világszerte számos olyan eset történt, amikor a deepfake-ek, a szintetikus média vagy más, MI-alapú technológiák befolyásolták a választási eredményeket. A Cambridge Analytica, egy politikai adatelemzéssel foglalkozó cég, személyre szabott politikai reklámokkal befolyásolta az amerikai választópolgárok véleményét a 2016-os Egyesült Államokbeli elnökválasztáson, valamint szerepet játszott a Brexit népszavazáson is. A cég a Facebook felhasználóinak adatait használta fel engedély nélkül, amelyeket egy látszólag ártalmatlan pszichológiai kvíz alkalmazás gyűjtött be. Ezeket az adatokat elemző algoritmusok segítségével dolgozták fel, hogy megértsék és befolyásolják a választók politikai preferenciáit, majd személyre szabott üzenetekkel manipulálták választási preferenciákat.

Prabowo Subianto, indonéz politikus (aki korábban emberi jogi visszaélésekkel volt összefüggésbe hozható) és stábja olyan eszközök segítségével, mint az adatelemzés, algoritmikus profilozás és automatizált tartalomgenerálás, célzott politikai üzeneteket és dezinformációs kampányokat terjesztettek széles körben az indonéz választások során. Prabowo közösségi média-jelenléte stratégiaileg úgy lett kialakítva, hogy egy kompetens és kiegyensúlyozott vezető képét sugározza, ez a keretezés pedig jelentősen befolyásolta megítélését a közvéleményben. A Prabowoval szembeni közhangulat felmérésére olyan MI-eszközöket használtak, mint a közösségi média-platformokon végzett hangulatelemzés, amelyek a nyilvános kommentekben kifejezett pozitív és negatív érzelmek elemzésével előre jelzik a választási eredményeket.

Irodalomjegyzék

2023 Landscape Report, 2023. <https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf>

Adadi, A. and Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, pp.52138-52160.

Ajunwa, I. and Schlund, R., 2020. Algorithms and the social organization of work.

Amsden, A.H., DiCaprio, A. and Robinson, J.A. eds., 2012. *The role of elites in economic development*. Oxford University Press (UK).

Anderljung, M., Barnhart, J., Leung, J., Korinek, A., O'Keefe, C., Whittlestone, J., Avin, S., Brundage, M., Bullock, J., Cass-Beggs, D. and Chang, B., 2023. Frontier AI regulation: Managing emerging risks to public safety. *arXiv preprint arXiv:2307.03718*.

Binns, R., 2018, January. Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency* (pp. 149-159). PMLR. <https://arxiv.labs.arxiv.org/html/1712.03586>

Bizottság, E., 2020. Fehér könyv a mesterséges intelligenciáról: a kiválóság és a bizalom európai megközelítése. *Brüsszel*, 2, p.19. https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_hu?filename=commission-white-paper-artificial-intelligence-feb2020_hu.pdf

Broadhurst, K., Hall, C., Wastell, D., White, S., & Pithouse, A., 2010. Risk, Instrumentalism and the Humane Project in Social Work: Identifying the Informal Logics of Risk Management in Children's Statutory Services. *British Journal of Social Work*, 40, pp. 1046-1064. <https://doi.org/10.1093/BJSW/BCQ011>.

Buhrmester, V., Münch, D. and Arens, M., 2021. Analysis of explainers of black box deep neural networks for computer vision: A survey. *Machine Learning and Knowledge Extraction*, 3(4), pp.966-989. <https://arxiv.org/abs/1911.12116>

Buchanan, B., 2020. The AI triad and what it means for national security strategy. *Center for Security and Emerging Technology*. <https://cset.georgetown.edu/publication/the-ai-triad-and-what-it-means-for-national-security-strategy/>

Chen, M., Zhang, Z., Wang, T., Backes, M., Humbert, M. and Zhang, Y., 2021, November. When machine unlearning jeopardizes privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security* (pp. 896-911).

Das, S., 2021. Local Justice and the Algorithmic Allocation of Societal Resources. *ArXiv*, abs/2112.01236. <https://doi.org/10.1609/aaai.v36i11.21486>.

Digital Decade Country Report: Hungary, 2024. <https://digital-strategy.ec.europa.eu/en/library/digital-decade-2024-country-reports>

Dutta, S., Lanvin, B., Rivera León, L. and Wunsch-Vincent, S. eds., 2023. *Global Innovation Index 2023: Innovation in the face of uncertainty*. WIPO. <https://www.wipo.int/edocs/pubdocs/en/wipo-pub-2000-2023-en-main-report-global-innovation-index-2023-16th-edition.pdf>

Eth, D. 2023. The Need For Work On Technical AI Alignment. *AI Safety Fundamentals*. <https://aisafetyfundamentals.com/blog/alignment-introduction>

European Commission (2024) *EU Competitiveness: Looking Ahead*. https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

European Parliament, 2024. *Artificial Intelligence Act*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

- Eubanks, V., 2012. *Digital Dead End: Fighting for Social Justice in the Information Age*. MIT Press.
- Eubanks, V., 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
- Evans, P. B. (1995). *Embedded Autonomy: States and Industrial Transformation*. Princeton University Press.
- Federal Trade Commission, 2022. *Bringing dark patterns to light*. Staff report. https://www.ftc.gov/system/files/ftc_gov/pdf/P214800%20Dark%20Patterns%20Report%209.14.2022%20-%20FINAL.pdf
- Fehér, F., Heller, Á. & Márkus, G. (1983) *Dictatorship over Needs: An Analysis of Soviet Societies*. Oxford: Basil Blackwell. <https://archive.org/details/dictatorshipover0000fehe>
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M., 2020. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, (2020-1). https://dash.harvard.edu/bitstream/handle/1/42160420/hls%20white%20paper%20final_v3.pdf?sequence=1
- Floridi, L., 2023. The Ethics of Artificial Intelligence: principles, challenges, and opportunities.
- Frase, H. and Daniels, O., 2023. Understanding AI harms: An overview. *Center for Security and Emerging Technology*. <https://cset.georgetown.edu/article/understanding-ai-harms-an-overview>
- Geiger, R.S., Cope, D., Ip, J., Lotosh, M., Shah, A., Weng, J. and Tang, R., 2021. "Garbage In, Garbage Out" Revisited: What Do Machine Learning Application Papers Report About Human-Labeled Training Data?. *arXiv preprint arXiv:2107.02278*. <https://ar5iv.labs.arxiv.org/html/2107.02278>
- Goodwin, P. and Fildes, R., 2022. Forecasting with judgment. In *The Palgrave Handbook of Operations Research* (pp. 541-572). Cham: Springer International Publishing.
- Gruetzemacher, R. and Whittlestone, J., 2022. The transformative potential of artificial intelligence. *Futures*, 135, p.102884. <https://arxiv.org/pdf/1912.00747.pdf>
- Hagendorff, T. and Fabi, S., 2023. Why we need biased AI: How including cognitive biases can enhance AI systems. *Journal of Experimental & Theoretical Artificial Intelligence*, pp.1-14.
- Heikkilä, M. (2024) 'The AI Act is done. Here's what will (and won't) change', MIT Technology Review, 19 March. Available at: <https://www.technologyreview.com/2024/03/19/1089919/the-ai-act-is-done-heres-what-will-and-wont-change/>
- Hendrycks, D., Mazeika, M. and Woodside, T., 2023. An overview of catastrophic ai risks. *arXiv preprint arXiv:2306.12001*. <https://arxiv.org/pdf/2306.12001>
- Hendrycks, D. (2024). *Introduction to AI Safety, Ethics, and Society*. arXiv. <https://doi.org/10.48550/arXiv.2411.01042>
- Jarovsky, L. (2023) 'Dark Patterns in Code: Ignoring User Choice', 15 March. <https://www.luizasnewsletter.com/p/dark-patterns-in-code-ignoring-user>
- John-Mathews, J.M., 2021. Critical empirical study on black-box explanations in AI. *arXiv preprint arXiv:2109.15067*.
- Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S., Noble, S., & Shestakofsky, B., 2021. Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change. *Socius*, 7. <https://doi.org/10.1177/2378023121999581>.
- Kalberg, S., 1980. Max Weber's types of rationality: Cornerstones for the analysis of rationalization processes in history. *American journal of sociology*, 85(5), pp.1145-1179.
- Kasirzadeh, A., 2024. Two Types of AI Existential Risk: Decisive and Accumulative. *arXiv preprint arXiv:2401.07836*. <https://arxiv.org/abs/2401.07836>

- Keddell, E., 2015. The ethics of predictive risk modelling in the Aotearoa/New Zealand child welfare context: Child abuse prevention or neo-liberal tool?. *Critical Social Policy*, 35, pp. 69 - 88. <https://doi.org/10.1177/0261018314543224>.
- Kuo, T., Shen, H., Geum, J., Jones, N., Hong, J., Zhu, H., & Holstein, K., 2023. Understanding Frontline Workers' and Unhoused Individuals' Perspectives on AI Used in Homeless Services. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3544548.3580882>.
- Lee, D., & Yoon, S., 2021. Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. *International Journal of Environmental Research and Public Health*, 18. <https://doi.org/10.3390/ijerph18010271>.
- Li, C., Wang, J., Zhang, Y., Zhu, K., Wang, X., Hou, W., Lian, J., Luo, F., Yang, Q. and Xie, X., 2023. The good, the bad, and why: Unveiling emotions in generative ai. *arXiv preprint arXiv:2312.11111*. <https://arxiv.org/pdf/2312.11111>
- Magyarország Mesterséges Intelligencia Stratégiája, 2020. <https://digitalisjoletprogram.hu/files/2f/32/2f32f239878a4559b6541e46277d6e88.pdf>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), pp.1-35. <https://arxiv.org/pdf/1908.09635v1>
- Mező, F. and Mező, K., 2019. Interdiszciplináris kapcsolódási lehetőségek a mesterséges intelligenciára irányuló cél-, eszköz-és hatásorientált kutatásokhoz= Interdisciplinary Connectivity Options for Goal-, Asset-and Effect-Oriented Researches of Artificial Intelligence. *Mesterséges Intelligencia*, 1(1), pp.9-29.
- Nanda, N., 2023. Mechanistic interpretability quickstart guide. <https://www.neelnanda.io/mechanistic-interpretability/getting-started>
- Nauta, M., Trienes, J., Pathak, S., Nguyen, E., Peters, M., Schmitt, Y., Schlötterer, J., Keulen, M., & Seifert, C., 2022. From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI. *ACM Computing Surveys*, 55, pp. 1 - 42. <https://doi.org/10.1145/3583558>.
- Négyesi, I., 2020. A Mesterséges Intelligencia és az etika. *HADTUDOMÁNY: A MAGYAR HADTUDOMÁNYI TÁRSASÁG FOLYÓIRATA*, 30(1), pp.103-113.
- Park, S., & Humphry, J., 2019. Exclusion by design: intersections of social, digital and data exclusion. *Information, Communication & Society*, 22, pp. 934 - 953. <https://doi.org/10.1080/1369118X.2019.1606266>.
- Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S. and Turini, F., 2019, July. Meaningful explanations of black box AI decision systems. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 9780-9784).
- Perrault, R. and Clark, J., 2024. Artificial Intelligence Index Report 2024. https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report-2024_Chapter_7.pdf
- Potapov, A., Rodionov, S., Myasnikov, A. and Begimov, G., 2012. Cognitive bias for universal algorithmic intelligence. *arXiv preprint arXiv:1209.4290*.
- Rastogi, C., Zhang, Y., Wei, D., Varshney, K.R., Dhurandhar, A. and Tomsett, R., 2022. Deciding fast and slow: The role of cognitive biases in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction*, 6(CSCW1), pp.1-22.
- Seger, E., Ovadya, A., Siddarth, D., Garfinkel, B. and Dafoe, A., 2023, August. Democratising AI: Multiple meanings, goals, and methods. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 715-722). <https://dl.acm.org/doi/pdf/10.1145/3600211.3604693>
- Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M. and Villalobos, P., 2022, July. Compute trends across three eras of machine learning. In *2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE. <https://arxiv.org/pdf/2202.05924.pdf>

- Sheehan, M., 2023. China's AI regulations and how they get made. *Horizons: Journal of International Relations and Sustainable Development*, (24), pp.108-125. https://carnegie-production-assets.s3.amazonaws.com/static/files/202307-Sheehan_Chinese%20AI%20gov-1.pdf
- Singh, P., Dwivedi, Y., Kahlon, K., Sawhney, R., Alalwan, A., & Rana, N., 2019. Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing. *Information Systems Frontiers*, 22, pp. 315-337. <https://doi.org/10.1007/S10796-019-09916-Y>.
- Sodhro, A., & Zahid, N., 2021. AI-Enabled Framework for Fog Computing Driven E-Healthcare Applications. *Sensors (Basel, Switzerland)*, 21. <https://doi.org/10.3390/s21238039>.
- Soleimani, M., Intezari, A., Taskin, N. and Pauleen, D., 2021. Cognitive biases in developing biased Artificial Intelligence recruitment system.
- Stewart, T.R., 2001. Improving reliability of judgmental forecasts. *Principles of forecasting: A handbook for researchers and practitioners*, pp.81-106.
- Thomas, I., & Mackie, P., 2019. Better data, better knowledge, better society: Developing an ideal homelessness data system drawing on lessons from global practice. *International Journal for Population Data Science*, 4. <https://doi.org/10.23889/ijpds.v4i3.1308>.
- Török, B. and Zödi, Z., 2021. A mesterséges intelligencia szabályozási kihívásai–Tanulmányok a mesterséges intelligencia és a jog határterületeiről [Regulatory challenges of artificial intelligence–Studies in the border area of artificial intelligence and law]. *Budapest: Ludovika Egyetemi Kiadó*.
- UNESCO, International Research Centre on Artificial Intelligence, 2024. Challenging Systematic Prejudices. An Investigation into Bias Against Women and Girls in Large Language Models.
- Zellner, M., Abbas, A.E., Budescu, D.V. and Galstyan, A., 2021. A survey of human judgement and quantitative forecasting methods. *Royal Society open science*, 8(2), p.201187. <https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.201187>
- Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Sellitto, M., Sakhaee, E., Shoham, Y., Clark, J. and Perrault, R., 2022. The AI Index 2022 Annual Report. AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, March. <https://arxiv.org/pdf/2205.03468.pdf>