

Schretnner Attila

## AI Ethics

\* A mesterséges intelligencia etikai, kockázati és irányítási vonatkozásai \*

### I. AI 2021-ben

2021 november 25-én az UNESCO<sup>1</sup> összes tagállama megállapodott egy, a mesterséges intelligencia (a továbbiakban: "AI") technológia egészséges fejlődését lehetővé tévő értékeket és alapelveket rögzítő dokumentum elfogadásáról<sup>2</sup>. Hasonló módon fogadott el ajánlásokat az OECD<sup>3</sup> és a G20<sup>4</sup> országcsoport is. De mi teszi szükségessé, hogy nemzetközi szintű megállapodások születhessenek ebben a témában, amely első érzésre még jövőszerűnek tűnik?

Az ok, hogy a gyakorlatban épp az ellenkezőjét látjuk ennek a megérzésnek. Nemzetközi szinten a vállalatok 86%-a válaszolta, hogy az AI – részben a COVID-nak is köszönhetően – *mainstream* technológiának számít<sup>5</sup>. Továbbá vállaltok 67%-a, 2021-ben tovább erősítette az AI képességeit<sup>6</sup>. Ennek oka, hogy a vállalatok nagy része azt várja az AI technológiától, hogy növelje a folyamatok hatékonyságát (74%), új üzleti modelleket hozzon létre (55%), valamint támogassa új termékek és szolgáltatások piacra vitelét (54%)<sup>7</sup>.

Ez a trend már itthon is mutatkozik, Magyarország is rendelkezik 2020-2030 időszakra szóló Mesterséges Intelligencia Stratégiával<sup>8</sup>. A stratégia becslései szerint Magyarországon az AI 2030-ban 6 400 Mrd Ft-tal fog hozzájárulni a GDP-hez és 900.000 munkavállalót fog érinteni a hatása. Már jelenleg is bevált alkalmazási területei az AI-nak a távközlési, banki, biztosítói, kiskereskedelmi, logisztikai és gyártási iparágak, de az államigazgatás területén is megjelenik a használata.

### II. Etikai kérdések az AI-al kapcsolatban

A fenti számok ismeretében látható, hogy az AI technológia ma is jelentős hatást fejt ki a globális és a magyar gazdaságra, valamint az emberek mindennapjaira. Így érdemes megvizsgálni, a technológia alkalmazásának, milyen etikai kérdései lehetnek, amelyekre az azt alkalmazó szervezetnek érdemes odafigyelni.

---

<sup>1</sup> Egyesült Nemzetek Nevelésügyi, Tudományos és Kulturális Szervezete

<sup>2</sup> UN: 193 countries adopt first-ever global agreement on the Ethics of Artificial Intelligence; Link: <https://news.un.org/en/story/2021/11/1106612>

<sup>3</sup> OECD: Recommendation of the Council on Artificial Intelligence, Link: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

<sup>4</sup> G20: G20 Ministerial Statement on Trade and Digital Economy, Link: <https://www.mofa.go.jp/files/000486596.pdf>

<sup>5</sup> Joe McKendrick (2021): AI Adoption Skyrocketed Over the Last 18 Months, Harvard Business Review

<sup>6</sup> Appen: The 2021 State of AI and Machine Learning Report

<sup>7</sup> The AI Journal: AI in a post-COVID-19 world, Link: <https://aijournal.com/report/ai-in-a-post-covid-19-world/>

<sup>8</sup> Magyarország Mesterséges Intelligencia Stratégiája 2020-2030, Link: <https://ai-hungary.com/api/v1/companies/15/files/137203/view>

Az Európai Parlament (a továbbiakban: „EP”) szakértői egy tizenkét (12) elemből álló listában foglalták össze az AI kapcsán felmerülő etikai kérdéseket<sup>9</sup>. Ezek közül a vállalatok számára releváns lehet:

- Munkaerőpiaci hatás – Hány munkahelyet teremt és veszélyeztet az AI egyre szélesebb körű alkalmazása? Milyen algoritmus és mintázatok mentén szűri az egy adott állásra jelentkező személyeket az AI? Milyen teljesítményértékelési döntéseket készít az AI a munkavállalókról?
- Pénzügyi hatás – Milyen hatása van az AI által hozott döntéseknek a vállalatok eredményességére és más pénzügyi mutatóira?
- Elszámoltathatóság és felelősség – Hogyan határozható meg, hogy ki egy adott AI által hozott döntés felelőse?
- Megfelelés – Hogyan biztosítható, hogy az AI által hozott döntések megfeleljenek a mindenkori jogszabályoknak, szabályozói elvárásoknak, ajánlásoknak, standardokan, és jó gyakorlatoknak? Hogyan garantálható az AI által kezelt személyes adatok megfelelő (GDPR szerinti) kezelése?
- Biztonsági – Hogyan garantálható az AI informatikai környezetének biztonsága? Hogyan védhető az AI rendszer a csaló szándékú külső beavatkozásoktól?

Ezeket az aggodalmakat támasztja alá az EY felmérése is<sup>10</sup> a felhasználók körében. Habár a felhasználók túlnyomó része (~95%) tisztában van az AI fogalmával, de technológia működését már csak töredékük (~30%) érti. Ezen felül a felhasználók aggnak, hogy a vállalatok szabálytalanul használják fel személyes adataikat, a vállalatok által használt technológiák nem transzparens, és manipulálhatják a magatartásukat.

### III. Irányelvek és jó gyakorlatok az etikus AI területén

Ezen etikai kérdések és felhasználói aggodalmak kezelésére számos ajánlás és jó gyakorlat született az utóbbi néhány évben, amelyek gyakorlati útmutatást is nyújtanak az AI technológia alkalmazására és a használatából fakadó kockázatok kezelésére. Ezek közül:

1. A Világgazdasági Fórum („*World Economic Forum*”, a továbbiakban: „*WEF*”) és a szingapúri állam közösen fejlesztette ki a „*Companion to the Model AI Governance Framework*” modellt<sup>11</sup>. A modell általánosan alkalmazható üzleti, állami és civil szervezetek számára is. A modell négy alappillérré épít: irányítási struktúra, emberi beavatkozás az AI döntéshozatalba, működés menedzsment, érintettekkel való interakció és kommunikáció.
2. Szintén a WEF gondozásában jelent meg a „*Empowering AI Leadership: AI C-Suite Toolkit*” módszertan<sup>12</sup>. Ez a módszertan kifejezetten vállalati kontextusra készült. A módszertan az AI-hoz kapcsolódó területek érettségének felmérésére szolgál a kockázatok, irányítási rendszer és működés dimenziói mentén.

---

<sup>9</sup> European Parliamentary Research Service: The ethics of artificial intelligence: Issues and initiatives, Panel for the Future of Science and Technology, Link: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS\\_STU\(2020\)634452\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf)

<sup>10</sup> EY: How focusing on ethics can unleash the full potential of AI, Infusing trust to deliver value in a transparent and accountable way, link: [https://assets.ey.com/content/dam/ey-sites/ey-com/en\\_uk/topics/tmt/ey-ai-and-ethics.pdf](https://assets.ey.com/content/dam/ey-sites/ey-com/en_uk/topics/tmt/ey-ai-and-ethics.pdf)

<sup>11</sup> WEF: Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organizations, Link: <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgisago.pdf>

<sup>12</sup> WEF: Empowering AI Leadership: AI C-Suite Toolkit, Link: [https://www3.weforum.org/docs/WEF\\_Empowering\\_AI\\_Leadership\\_2022.pdf](https://www3.weforum.org/docs/WEF_Empowering_AI_Leadership_2022.pdf)

3. Az AI Ethics Impact Group által készített módszertan<sup>13</sup> is alkalmazható a szervezetek széles körére. A módszertan az átláthatóság, elszámoltathatóság, bizalmasság, igazságosság, megbízhatóság és fenntarthatóság értékelési szempontokon nyugszik.

Az AI etikai, kockázati és irányítási kérdéseiben még nincs teljes szakmai konszenzus. A fent említett példákon túl létezik számos irányelv és iránymutatás, valamint néhány nagyobb technológiai vállalat, mint például a Microsoft<sup>14</sup> saját etikai elveket is kidolgozott, amelyek elérhető nyilvánosan.

## IV. EU rendelettervezet

Ezek az jó gyakorlatok és irányelvek azonban nem kötelező érvényűek. Így a fentebb is tárgyalt felhasználói aggodalmakat a mérlegre helyezve az Európai Biztosság elkezdte egy új rendelet – Javaslat az Európai Parlament és a Tanács Rendelete a Mesterséges Intelligenciára Vonatkozó Harmonizált Szabályok (A Mesterséges Intelligenciáról Szóló Jogszabály) Megállapításáról és Egyes Uniós Jogalkotási Aktusok Módosításáról<sup>15</sup> – kidolgozását. Itt fontos hangsúlyozni, hogy a dokumentum jelenleg csak javaslat státuszban, van így a végleges elvárások pontos meghatározáshoz meg kell várni a végleges szöveg megjelenését. Azonban a szabályozás fő irányvonalai így is kiolvashatók a jelenleg elérhető dokumentumokból.

A rendelet az irányítási rendszerek többségéhez hasonlóan kockázat alapú megközelítést alkalmaz<sup>16</sup>, ezen belül pedig elfogadhatatlan kockázattal, nagy kockázattal, illetve alacsony vagy minimális kockázattal járó kategóriákba sorolja az AI rendszereket. Az elfogadhatatlan kockázatú AI rendszerek közé tartozhatnak majd például:

- Természetes személyek magatartásának szándékos befolyásolására, torzítására használt AI rendszerek
- Természetes személyeket célzó, valós idejű, távoli biometrikus azonosítására használt AI rendszerek nyilvánosság számára hozzáférhető helyeken

---

<sup>13</sup> AI Ethics Impact Group: From Principles to Practice An interdisciplinary framework to operationalise AI ethics, Link: [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO\\_2020\\_final.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf)

<sup>14</sup> Microsoft AI principles, Link: <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimar6>

<sup>15</sup> Link: <https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

<sup>16</sup> European Commission: Regulatory framework proposal on artificial intelligence, Link: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

A nagy kockázatú AI rendszerek közé pedig például az alábbiak lehetnek besorolhatók:

- Természetes személyek valós idejű és nem valós idejű, távoli, biometrikus azonosítására szolgáló AI rendszerek
- Víz-, gáz-, fűtési és villamosenergia-szolgáltatás során használt AI rendszerek
- Természetes személyek hitelképességének értékelésére vagy hitelpontszámuk megállapítására használt AI rendszerek
- Természetes személyek toborzására vagy kiválasztására használt AI rendszerek

Az elfogadhatatlan kockázattal rendelkező AI rendszerek alkalmazása főszabály szerint tilos lehet a rendelet jelenlegi szövegtervezete alapján. A nagy kockázatú AI rendszer esetén elvárás lehet – a teljesség igénye nélkül - egy sor kontroll és működési változás alkalmazása, többek közt:

- Kockázatkezelési rendszer bevezetése
- Minőségirányítási rendszer bevezetése
- AI rendszerek folyamatos fejlesztése
- Átláthatóság és felhasználó tájékoztatása
- Emberi felügyelet garantálása
- Adatkezelési gyakorlatok megfeleltetése (beleértve a személyes adatokat is)
- Cyber-biztonsági elvárások teljesítése
- Naplózás biztosítása
- Bejelentés felügyeleti hatóság felé
- Műszaki dokumentáció összeállítása

A rendelet extraterritoriális hatállyal is bírhat (azaz az EU-n kívüli vállaltokra is vonatkozhat, ha az EU-s állampolgárok is az általuk használt AI rendszerek hatálya alá esnek) majd és büntetési tétele magasabb lehet a GDPR-nál<sup>17</sup>, akár 30 millió EUR vagy az adott cégcsoport globális árbevétele 6%-a is lehet.

## V. AI és belső ellenőrzés

Az AI rendszerek térnyerésével a belső ellenőrzésnek szükséges vizsgálnia azt, hogy a vállalat hogyan végzi a kapcsolódó kockázatok kezelését, valamint, hogy teljesülnek-e a jó gyakorlatok és szabályozások által megfogalmazott elvárások.

A belső ellenőrzésnek be kell vennie az AI rendszerek vizsgálatát az audit tervébe. A belső ellenőrzés, akár már az adott AI rendszer bevezetésének fázisában is bevonódhat a folyamatba tanácsadó vagy informális szerepben annak érdekében, hogy megismerkedhessen az új technológiákkal és felkészült legyen arra a helyzetre, amikor már vizsgálnia kell a működő rendszert.

Az AI rendszeren túl a belső ellenőrzésnek értenie kell az AI-hoz kapcsolódó etikai, kockázati és irányítási dimenziókat is, hogy a vállalati működés más területeihez hasonlóan el tudja látni bizonyosságot nyújtó funkcióját a menedzsment által működtetett kockázatkezelő intézkedések felett.

---

<sup>17</sup> General Data Protection Regulation

## **VI. Összefoglalás**

Látható a trend, hogy közeljövőben egyre több és szofisztikáltabb jó gyakorlat, standard és szabályozás fog megjelenni az AI rendszerek vonatkozásában. Erre pedig célszerű proaktív módon felkészülni a vállalatoknak. Első lépésben célszerű lehet megismerni a jó gyakorlatokat, standardokat és szabályozásokat. Következő lépésben a relevánsnak ítélt elvárásokat javasolt összevetni a vállalt saját képességeivel és felmérni az AI rendszerekhez kapcsolódó érintetteket és kockázatokat. Ezután az azonosított eltérések alapján megfogalmazható egy AI etikai stratégia, egy fejlesztési ütemterv és létre hozható az ezek végrehajtását lehetővé tévő irányítási rendszer kialakítása.