Ernst & Young Kft.
Ernst & Young Ltd.
H-1132 Budapest Váci út 20.
1399 Budapest 62. Pf.632, Hungary

Tel: +36 1 451 8100
Fax: +36 1 451 8199
www.ey.com/hu
Cg. 01-09-267553

Schrettner Attila

AI Ethics - 2. rész

** Data governance szerepe a mesterséges intelligencia etikus használatában**

Ez a cikk az AI Ethics téma feldolgozását folytatja, a sorozat előző, bevezető cikke elérhető a következő linken: https://iia.hu/images/hirlevel/ethics.pdf. A sorozat egyes részei a kapcsolódó EU direktíva három nagyobb elvárás csoportja köré szerveződnek: data governance (jelen cikk), kockázatkezelés és AI specifikus kontroll tevékenységek

I. Adatok és mesterséges intelligencia

A mesterséges intelligencia megoldások rendszerint nagy mennyiségű adathalmazra épülnek[1]. Így az etikus AI megoldások megteremtésének első lépése az őket „tápláló" adatok etikus kezelése. Ez a lépés jól - de nem teljesen - elkülöníthető a már a rendszer által generált outputokból származó kockázatoktól.

Az adatok etikus kezelésére két különböző szempontból is tekinthetünk:

- Az egyik nézőpont folyamati szemléletű, amely az üzleti igény definiálásának, adatok beszerzésének, tárolásának, felhasználásának, archiválásának és törlésének folyamatát, azaz az adatok teljes életciklusát vizsgálja

- A másik nézőpont pedig mesterséges intelligencia „tanítása" közben zajló három fázis szerint - a tréning, validációs és tesztelési (*„training", „validation", „testing"*) adatokat vizsgálja

A fő megfontolás mindkét szempont esetén a teljeskörűség, azaz a *data governance*-nek az adatok teljes életciklusára és mindhárom elkülönült adathalmazra ki kell terjednie. Fontos hangsúlyozni, hogy itt a *data governance* nem csupán a GDPR-nak[2] való megfelelést jelenti. Az AI rendszerek használata során adatvédelmi kockázatokon túl számos más kockázat is felmerülhet azáltal, hogy az adatok rossz inputot jelentenek az AI rendszer és a kapcsolódó algoritmus(ok) számára[3]:

[1] The European Commission's High-Level Expert Group on Artificial Intelligence (2018): A Definition of AI: Main Capabilities and Scientific Disciplines
[2] Az Európai Parlament és a Tanács (EU) 2016/679 Rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet)
[3] University of Pennsylvania, Wharton (2022): Artificial Intelligence/Machine Learning Risk & Security Working Group (AIRS): Artificial Intelligence Risk & Governance

- **Nem megfelelő adatminőség** - technikai szempontból, pl.: az adatok nem megfelelő formátumban vannak

- **Nem megfelelő adatmennyiség** – a rendelkezésre álló adatmennyiség kevés az AI rendszer megfelelő működéséhez

- **„Prekoncepció (*„bias"*) az adatokban"** – az adatok technikai és mennyiségi szempontból is megfelelők, de tartalmilag valamilyen hiba van bennük (lásd: jelen cikk következő fejezete), például a rendelkezésre álló adatok egy adott társadalmi csoportra koncentrálnak, így belőlük előállított modell és a modellt használó AI rendszer kevésbé lesz alkalmas más társadalmi csoportokra vonatkozó döntési helyzetek kezelésére

- **Tiltott adatkategóriák** – valamilyen törvény vagy más szabály által tiltott adat található az adathalmazban

A fenti kockázatok általánosak, az egyes vállalatok iparágának, működési sajátosságainak vagy egyéb paramétereinek megfelelően további kockázatok is azonosíthatók és a *data governance* keretrendszernek ezekre is ki kell terjednie.

II. „Prekoncepció az adatokban"

A fenti kockázatok közül érdemes kiemelni az „prekoncepció az adatokban" kockázatot. Ez két szempontból is érdekes, egyrészt ez a kockázat leginkább AI rendszerek kontextusában jelentkezik, azok tanulásához kapcsolódóan, másrészt ez egy olyan eset, ami látens módon is jelen lehet, nehéz első ránézésre, vagy felületesebb vizsgálatokkal kiszűrni.

Ráadásul az adatokban lévő prekoncepciók akár az adott vállalat ügyfeleinek alapvető jogainak sérüléshez is vezethetnek, mint méltósághoz való jog, egyenlőség és diszkriminációmentességhez való jog, fogyasztóvédelemhez való jog, alapvető szolgáltatásokhoz való hozzáférés, stb[4]. Ezért nagyon fontos, hogy a *data governance* kiemelt figyelmet fordítson erre az aspektusra.

Ezt a kockázatot leginkább példákon keresztül lehet szemléletesen bemutatni[5]:

- **Egy globális technológiai cég AI rendszeren alapuló toborzási rendszere hátrányosan megkülönböztette a női jelentkezőket**[6] - az eltérés oka az volt, hogy az AI rendszer mögött álló algoritmust nagy arányban férfi alkalmazottak önéletrajzaiból álló adatokkal tanították, aminek többek között az is volt az oka, hogy a cég alkalmazottai között eleve magasabb arányban találhatók férfiak. A példa amiatt is érdekes, mert rámutat, hogy még egy, a digitális technológiák terén élenjáró vállalat is eshet ilyen hibába.

[4] European Union Agency for Fundamental Rights (2020): Getting The Future Right Artificial Intelligence and Fundamental Rights
[5] James Manyika, Jake Silberg, and Brittany Presten (2019): What Do We Do About the Biases in AI?, Harvard Business Review
[6] Jeffrey Dastin (2018): Amazon scraps secret AI recruiting tool that showed bias against women, Reuters

- **Két szakszervezet is támadta az egyik személyszállítással foglalkozó vállalatot a sötétebb bőrtónusú sofőröket nem felismerő AI rendszere miatt**[7] - A cég biztonsági okok miatt bevezetett egy AI rendszert, ami bizonyos időközönként ellenőrizte a sofőrök személyazonosságát úgy, hogy az okostelefonjuk élő kamera képét összehasonlította a sofőr arcáról előzetesen eltárolt fotókkal. Két sofőr szakszervezet is azt állítja az AI rendszer alacsonyabb hatékonysággal működik sötétebb bőrtónusú sofőrök esetén, amely szélsőséges esetben ki is zárhatja a sofőröket a rendszerből, így gyakorlatilag elvesztik a munkájukat.

- **Irányítószám szerinti megkülönböztetés hitelképesség elbírálásakor**[8] - Számos pénzügyi szolgáltatónál felmerülő AI rendszer rendellenesség, hogy más paraméterektől függetlenül, a szegényebb irányítószámú településeken vagy kerületekben élő ügyfelek alacsonyabb „hitelképességi" pontszámot kapnak, így magas kamat mellett jutnak hitelhez, vagy el is eshetnek a hitelhez jutás lehetőségétől.

Ezek a prekoncepciók ráadásul üzleti értelemben is hátrányt jelenthetnek egy vállalat számára, a fenti példákra reflektálva:

- A toborzási rendszer hibája miatt tovább lehetnek betöltetlenek a nyitott pozíciók, ami miatt projektek késhetnek, ami bevételkieséshez vezethet

- A belső folyamatok indokolatlan megakasztása költségtöbbletet okozhat

- A hitelkérelmek elbírálásakor kiszűrésre kerülhetnek olyan ügyfelek, akiket egyébként ki tudna szolgálni a pénzügyi szolgáltató, ami bevétel kieséhez vezethet

III. Data governance jó gyakorlatok

A fenti problémák kezelésére azonban léteznek jógyakorlatok. A megfelelő data governance teremti meg az AI rendszerhez kapcsolódó adatok kezelésének szervezeti kereteit, kijelölve azokat a felelősöket és döntési pontokat, valamint felállítva azokat a vállalaton belüli intézményeket, amelyek segítenek kiküszöbölni a fent említett esetek bekövetkezését.

[7] Chris Vallance (2021): Legal action over alleged Uber facial verification bias
[8] Sian Townson (2020): AI Can Make Bank Loans More Fair, Harvard Business Review

Az EY is rendelkezik saját *data governance* módszertannal[9]. Eszerint egy jól átgondolt *data governance* keretrendszer kialakítása során figyelembe kell venni legalább:

- **Stratégiai kérdések:**
  - A vállalat stratégiáját, technológia/IT stratégiáját és mesterséges intelligencia stratégiáját (ha léteznek ezek a dokumentumok)
  - Az AI rendszerekhez kapcsolódó érintettek elvárásait
  - A vállalat által használ AI rendszerek és kapcsolódó felhasználási területek mennyiségét és milyenségét
  - A vállalat által kezelt adatvagyon méretét és milyenségét
  - A vállalat által kezelt adatvagyon monetizációs modelljét
- **Szervezési kérdések:**
  - A vállalat szervezeti struktúráját, bizottságait és más *governance* rendszerek működését
  - A szervezeti modellt, amelyben meg kívánják valósítani a *data governance* keretrendszert
  - A *data governance*-ért felelős szervezeti egység felállítását vagy kijelölését és átalakítását
  - A *data governance* keretrendszer egyéb szereplőinek kijelölését
  - A *data governance* keretrendszerhez kapcsolódó folyamatok kialakítását vagy a meglévő folyamatok átalakítását
- **Működési kérdések:**
  - A *data governance* keretrendszer működtetéséhez szükséges erőforrásokat
  - A *data governance* rendszer működését leíró mutatószámokat (KPI, KRI)
  - A *data governance* keretrendszer működéséhez szükséges informatikai támogatást
  - A kapcsolódó szabályzatok, kódexek, kézikönyvek és egyéb dokumentumok elkészítését vagy frissítését
  - Folyamatos fejlesztést és változásvezetés beépítését

[9] EY Data Governance Playbook

A *data governance* egy remek terület lehet az *ESG rating* javítására is főként azoknak a vállalatoknak, amelyeknél az üzleti modell középpontjában állnak az adatok, pl.: IT, pénzügyi vagy telco szektorokban.

Illetve a *data governance* témakörébe tartozik számos informatika, informatikai biztonsági és adattudomány (*„data science"*) területére vonatkozó kérdés is, amiket jelen cikknek nem célja tárgyalni.

IV. EU direktíva tervezet a data governance-ről

Ezen jó gyakorlatok egy részére reflektál az EU direktíva tervezete is kötelező elvárásként a magas kockázatú AI rendszerek esetén[10]. A direktíva a *data governance* kifejezésre magyarul az „adatkormányzás" kifejezést használja.

Az EU direktíva is kiköti, hogy a *data governance* keretrendszer kialakításánál legalább az alábbi döntési pontok, kompetenciák kijelöléséről és szervezeti keretek kialakításáról rendelkeznie kell a vállalatnak[11]:

- Az AI rendszer megtervezésének főbb döntési pontjai;
- Adatgyűjtés mikéntje;
- Releváns adat-előkészítési műveletek, mint például annotáció, címkézés, tisztítás, gazdagítás és összesítés;
- Vonatkozó hipotézisek, előfeltevések megfogalmazása, különös tekintettel azokra az információkra, amelyeket az adatoknak mérniük kell és meg kell jeleníteniük;
- Szükséges adathalmaz elérhetőségének, mennyiségének és alkalmasságának előzetes értékelése;
- Esetleges torzítások – beleértve a prekoncepciók - meglétének vizsgálata (itt a direktíva eredeti, angol nyelvű szövege a *„bias"* kifejezést használja*)*;
- Esetleges adathiányok vagy hiányosságok azonosítása, valamint e hiányok és hiányosságok kezelésének módja.

Érdemes itt kiemelni, hogy a direktíva tervezet is alkalmazza az adathalmaz hármas bontását a tréning, validációs és tesztelési adatokra. Így a fenti feltételeket mindhárom adattípus esetén biztosítani kell.

[10] European Commission: Regulatory framework proposal on artificial intelligence, Link: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[11] European Commission: Regulatory framework proposal on artificial intelligence, Article 10.

Building a better working world

V. Data governance és belső ellenőrzés

A *data governance* keretrendszer megfelelőségének felmérése a belső ellenőrzés számára is jelenthet kihívásokat. Ilyen területek lehetnek[12, 13]:

- Törzsadatok („*master data*") ellenőrzése
- Metaadatok ellenőrzése
- Back-up adatok ellenőrzése
- Üzleti és IT terület közti megállapodások, SLA-k ellenőrzése
- Adatminőség ellenőrzése

A kihívások áthidalásával azonban jelentős értéket teremthet a belső ellenőrzés, hiszen garanciát tud nyújtani az üzleti vezetőknek, hogy a döntéseket alátámasztó adatok megfelelőek, így bátran használhatók új *insight*-ok előállítására.

VI. Összefoglalás

A *data governance* keretrendszer tehát az egész *AI Ethics* téma alapját, gerincét adja. Ez biztosítja, hogy az AI rendszerek a megfelelő adatokkal legyenek ellátva és így segíteni tudják azt az üzleti célt, amelyre létrehozták őket. Mivel a *data governance* keretrendszer outputja – a megfelelően előkészített adat és az optimalizált algoritmus – az AI rendszer inputját jelenti, így az itteni nemmegfelelőségeknek a teljes AI rendszer szempontjából jelentős tovagyűrűző hatása lehet. Érdemes tehát az adott vállalatnak az *AI Ethics* felkészülését a *data governance* keretrendszer kiépítésével kezdeni.

[12] IIA Belgium (2020): Data Ethics, Where does internal audit fit?, Link: https://iiabelgium.org/wp-content/uploads/2020/08/GKB-Data-Ethics.pdf
[13] IIA Global (2017): Understanding and Auditing Big Data, Link: https://www.iia.nl/SiteFiles/Publicaties/GTAG-Understanding-and-Auditing-Big-Data.pdf