Mucsinyi-Horváth Alexandra*:

Mesterséges intelligencia: segítő kéz vagy időzített bomba?

A mesterséges intelligencia (MI) forradalma új korszakot nyitott a pénzügyi szektorban: gyorsabb döntéshozatalt, hatékonyabb kockázatelemzést és személyre szabott ügyfélélményt. Mindezek az előnyök már kézzelfoghatók, elérhetők. Az MI azonban nem csupán lehetőségeket rejt, hanem komoly etikai, compliance és IT biztonsági kockázatokat is hordoz. Hol húzódik a határ az innováció és a felelős működés között? Az Európai Unió szigorúbb keretek közé szorítja az algoritmusok világát, de ez önmagában nem elég. Az MI lehet a segítő kéz, amely a pénzügyi rendszerek stabilitását támogatja vagy az időzített bomba, amely visszaélésre, adatvédelmi incidensekre és akár piaci manipulációra is lehetőséget ad. Milyen kihívásokkal kell szembenézniük a pénzügyi intézményeknek az MI alkalmazása során és hogy lehet elkerülni a technológia rejtett veszélyeit?

Az MI széles körű elterjedése mögött számos etikai aggályt lehet azonosítani, mint például diszkriminációt, munkaerőpiaci kockázatot vagy éppen az úgynevezett „deepfake" technológiát, amely az MI segítségével képes digitális tartalmak valósághű manipulálására. Ez lehetővé teszi arcok cseréjét, hangok szintetizálását vagy akár teljesen fiktív tartalmak generálását. Ezek az etikai aggályok megfigyelhetők mind a felhasználásával, tulajdonlásával, elszámoltathatóságával és az emberiségre gyakorolt hosszútávú hatásaival kapcsolatban egyaránt. Joggal merülnek fel bennünk olyan kérdések például, hogy vajon az MI elfogulatlan? Vajon azonosan ítéli meg a különböző emberek különböző értékrendjét? Meg tudjuk védeni tőle vagy általa a magánéletünket?

Az MI rendszerek használata során gyakran jelentkező diszkrimináció, elfogultság abban az esetben érhető tetten, ha azok a tanítóadatokban vagy a tervezés során jelen vannak, vagy éppen hiányoznak. Számos oka lehet annak, hogy az MI rendszerek nem működnek jól bizonyos populációk esetében. Mivel az önálló tanulásra képes algoritmusok az általuk hozzáférhető adatokból dolgoznak, így előfordulhat például, hogy egy adott demográfia alulreprezentált és ezért az MI döntéseiben ez nem tükröződik megfelelően. Ha például egy arcfelismerő rendszer több kaukázusi karaktereken tanulta meg felismerni az emberi arcvonásokat, akkor jobban fog teljesíteni a többséghez tartozók esetében. De hasonló példa említhető meg egy nyelvi szoftver esetében is, amikor is a fordítóprogram az orvos szót automatikusan hímneműnek, míg az ápoló kifejezést nőneműnek fordította.

Az öntanulás vagy önfejlesztés nyomán kialakított MI-k minden szándékosság nélkül is mutathatnak elfogultságot, ugyanis attól függetlenül, hogy a válasz statisztikailag korrekt, még nem biztos, hogy ténylegesen is az. Ezért szükséges lehet humán beavatkozás a fejlesztési folyamatokba, így biztosítva jobban a szegregáció kiküszöbölését.

Manapság már nem kell bemutatni a deepfake technológiát, amivel egy személy valamely ismertetőjegyét az MI segítségével úgy változtatják meg, hogy az illető valaki másnak tűnjön. E technológia képes befolyásolni és félrevezetni embereket, mivel hajlamosak elhinni, amit a saját szemükkel látnak. Azáltal, hogy látjuk, bizonyítékot szolgáltatunk az agy több területének, amely hatására meggyőzővé teszi számunkra a látottakat. Első ránézésre nem feltétlenül gondolnánk, hogy a látottaknak nincs valóságalapja. Az arcfelismerő modelleknek nem könnyű feladat az emberek

azonosítása, ugyanis az emberi arc arcmimikája változik, az arc részeit takarhatják különböző tárgyak, mint például szemüveg vagy nők esetében előfordulhat változatos frizura vagy más egyéb ideiglenes elváltozás is, illetve bizonyos mértékben a telefonok kamerája is torzít. Mivel az arc egyedi, amelyet annak bizonyos részeinek apró eltérései okoznak, lehetőséget biztosít az MI megzavarására. Például az is előfordulhat, hogy egy alanyt smink nélkül felismer az alkalmazás, azonban annak használatát követően már nem.

Amikor az MI és a (jog)szabályoknak való megfelelés (compliance) kapcsolatáról beszélünk, fontos azonosítani, majd pedig kezelni az adatvédelmi és szabályozási hiányosságokat, valamint biztosítani az átláthatóságot és a meglévő szabályok betartását. Az MI alkalmazása során szükséges szavatolni az adatok védelmét és az érintettek jogainak tiszteletben tartását. Ennek szabályozása tekintetében az egyik legnagyobb mérföldkőnek számító előrelépés 2021. áprilisában kezdődött, amikor is az Európai Unió Bizottsága közzétette az MI-ről szóló rendelettervezetét az „Artificial Intelligence Act"-et (AI Act), ami 2024. augusztus 1-jén hatályba is lépett, azonban a különböző rendelkezések fokozatosan válnak alkalmazandóvá. A tiltott MI gyakorlatokra vonatkozó szabályok például már 2025. február 2-tól érvényesek.

A pénzügyi intézmények esetében számos magas kockázatú MI megoldás működik, így az AI Act kiemelten érinti az ágazat szereplőit. A hitelbírálati algoritmusok, a csalásmegelőző rendszerek vagy a piaci kockázatelemző modellek mind olyan területek, ahol a megfelelés elengedhetetlen. Az érintett pénzügyi intézményeknek gondoskodniuk kell arról, hogy ezek az MI rendszerek átláthatóak, auditálhatóak és torzításmentesek legyenek. Egy, további – talán a legnagyobb – kihívás a döntések reprodukálhatósága.

Az, hogy az MI rengeteg adatot használ fel a működéséhez, a veszélyeit is tovább növeli. Alkalmazása során kulcsfontosságú kérdés az adatvédelem és a magánélet védelme. Egyrészt aggályokat vet fel, hogy a nagy mennyiségű adatot hogyan dolgozzák fel, használják fel, illetve tárolják, valamint az is, hogy ezekhez férhet hozzá. Másfelől pedig nem átlátható és nem egyértelmű a kezelt adatok köre (a legtöbb esetben személyes, érzékeny adatokról is szó lehet).

A kockázat értékeléséhez első körben fontos megnézni, hogy milyen jelenleg is hatályos szabályozás vehető figyelembe. Amikor adatvédelmi szabályozásról beszélünk, elsőre az uniós adatvédelmi rendelet, a GDPR juthat eszünkbe. A tudatosság és oktatás növelése fontos minden érintett számára, azonban egységes szabályrendszerrel már keretet biztosíthatunk a kockázatok csökkentése érdekében.

Míg a compliance és az etikai megfelelés elsősorban az MI rendszerek jogszerű és tisztességes működésére összpontosít, nem mehetünk el az IT biztonsági kockázatok mellett sem. Ezek a technológia sérülékenységeire és a potenciális fenyegetésekre hívja fel a figyelmet. Egy MI-t alkalmazó rendszer nemcsak az adatvédelmi előírások betartásától válik megbízhatóvá, hanem attól is, hogy ellenáll a kibertámadásoknak, manipulációnak és visszaélésnek.

Ahogy említettük, az MI mögött álló programok „tudása" az adatokkal való tanításból ered. Ezeket az adatokat védeni szükséges, hogy az integritás ne sérüljön. Minél több adatot ismer meg azonban az MI, annál jobban tanul és annál pontosabban tud előre jelezni, elemezni. Ennek a kettőségnek

a legnagyobb hátulütője, hogy ezek az adatok manipulálhatók vagy megmérgezhetők. Emiatt jelent meg egy új típusú fenyegetés, az úgynevezett „data poisoning” vagyis az adatmérgezés.

Ennek is több fajtája lehet, közülük a legismertebb az adatkészlet szándékos, célzott megváltoztatása. Ebben az esetben a támadók úgy mérgezik meg az adatkészletet, hogy pontatlan vagy helytelenül minősített adatokat adnak a tényleges adatkészlethez. Például az e-mail spamszűrője is gépi tanulást használ a biztonságos és a levélszemét elektronikus üzenetek megkülönböztetésére. A tanító adatok mindössze 1%-ának megváltoztatásával az algoritmus már hatástalanná válhat. A támadók ezt követően az adathalász e-maileket jelszavak, felhasználónevek ellopására vagy vállalati számítógépes rendszerbe való behatolásra használhatják fel.

Egy másik hatékony módszer az MI rendszer mérgezésére az, hogy egy egészséges modellt egy mérgezettre cserélnek. A betanítás után a modell csak egy képre vagy dokumentumra hasonlító fájl lesz. A hagyományos kibertámadási módszereket alkalmazó támadók így hozzáférhetnek az ellenőrzött modelleket tároló rendszerhez és megváltoztathatják vagy lecserélhetik azt mérgező modellekre.

Az elmúlt években rengeteg új szolgáltató jelent meg a pénzügyi piacon, ami új lehetőségeket, de új kockázatokat is hordoz. A kibercsalók azonban eközben ráadásul egyre professzionálisabbak: együttműködnek a többi bűnözővel és folyamatosan új átverési típusokat fejlesztenek, sokszor az MI segítségével. A mesterséges intelligenciával jelentősen kiszélesedett a csalók repertoárja, ugyanis már nemcsak szöveggel, hanem hanggal, képekkel, chatbotokkal, videóval is tudnak támadni, s kódokat is jóval könnyebben tudnak írni.

Az MI tehát forradalmasítja a pénzügyi szektort, de komoly etikai, compliance és IT biztonsági kihívásokat is felvet. Az AI Act és a GDPR szigorú kereteket szab, így a pénzügyi intézményeknek időben és proaktívan kell felkészülniük a megfelelésre. Kulcsfontosságú az MI rendszerek folyamatos auditálása és átláthatóságuk biztosítása. Kiemelt figyelmet kell fordítani az adatvédelmi és informatikai biztonsági intézkedésekre is, például az adatmérgezés elleni védelmi mechanizmusokra és az adatok titkosítására. Az MI alkalmazásának hatékony és felelős működtetése érdekében elengedhetetlen az oktatás, valamint az emberi felügyelet biztosítása.

„A technológia csak egy eszköz. Az számít, hogyan használjuk.”

- Karen Bartleson amerikai digitális szakértő–

A kérdés már nem az, hogy használjuk-e az MI-t, hanem az, hogy biztonságosan és felelősen teszszük-e.

** A szerző a Magyar Nemzeti Bank vezető felügyelője*

„Szerkesztett formában megjelent 2025. augusztus 12-én az Economx.hu oldalon.”