Neked ajánljuk!

Etikus AI:
Az emberi értékek próbája

Az AI valódi társadalmi értékét nem a modellek teljesítménye, hanem az határozza meg, milyen emberi értékek mentén használjuk őket.

Az elmúlt hetek eseményei jól mutatják, hogy a hangsúly egyre inkább a bizalomra, az átláthatóságra, az emberi felügyeletre és a felelős használatra helyeződik.

Ebben a hírlevélben áttekintjük az EU új AI-átláthatósági szabályait, az AI és a mentális egészség kapcsolatának etikai dilemmáit, az ENSZ inkluzív AI-governance kezdeményezését, az etikus promptolás jelentőségét, valamint a Stanford AI Index 2026 legfontosabb tanulságait arról, hogy hol tartanak ma a szervezetek a felelős AI használatában.

Egy dolog egyre világosabb: az etikus AI használat már nem elsősorban technológiai, hanem emberi, szervezeti és irányítási kérdés.

AI-átláthatóság: új szabályok léptek életbe az EU-ban

2026. augusztus 2-tól hatályba léptek az EU AI Act első átláthatósági kötelezettségei. Az új szabályok célja, hogy az emberek egyértelműen tudják, mikor találkoznak mesterséges intelligenciával vagy AI által létrehozott tartalommal.

A szabályok értelmében:

- az AI-val közvetlenül kommunikáló rendszereknek jelezniük kell, hogy a felhasználó mesterséges intelligenciával lép kapcsolatba;
- a deepfake tartalmakat egyértelműen meg kell jelölni;
- az AI által generált vagy manipulált tartalmak esetében biztosítani kell a felismerhetőséget;
- Bizonyos esetekben azt is jelezni kell az érintetteknek, ha érzelemfelismerő vagy biometrikus kategorizáló AI-rendszer elemzi őket.

Az átláthatóság kulcsfontosságú a bizalom megteremtéséhez: az embereknek tudniuk kell, mikor kommunikálnak AI-val, illetve mikor találkoznak AI által létrehozott tartalommal. Ez segíti a tudatos döntéshozatalt, és csökkenti a megtévesztés, a manipuláció és a dezinformáció kockázatát.

TOVÁBB

Inkluzív és globálisan igazságos AI

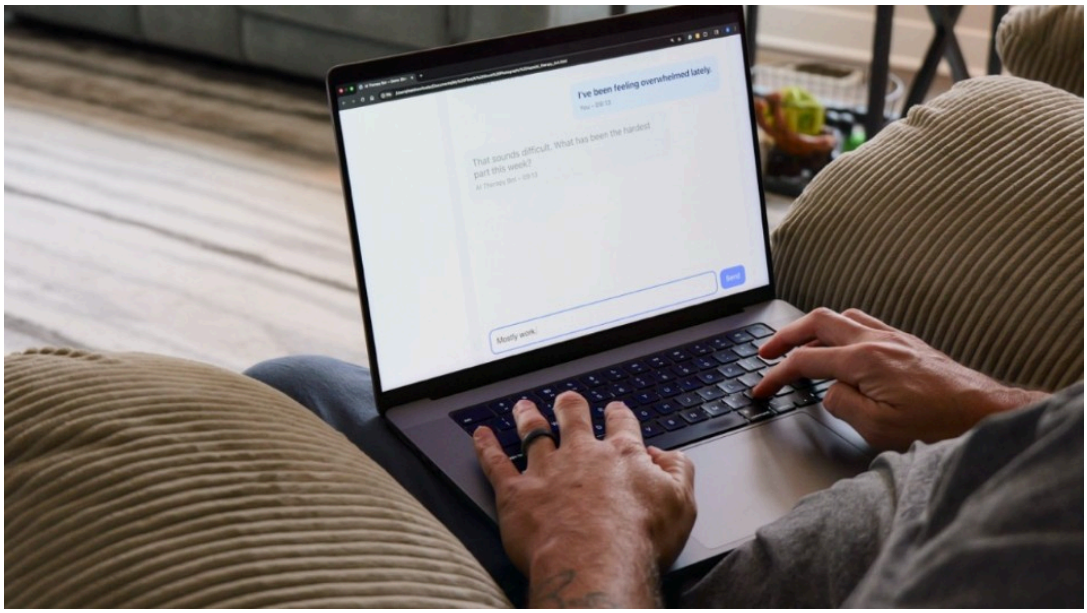
Július elején az ENSZ AI Governance Global Dialogue rendezvényének egyik alapgondolata az volt, hogy az AI szabályait és jövőjét ne kizárólag a technológiailag legfejlettebb országok alakítsák. A kezdeményezés célja, hogy a fejlődő országok és a globális gazaságban kisebb befolyással bíró országok is teljes jogú résztvevőként vehessenek részt az AI irányítás (governance) kialakításában.

Ez a gyakorlatban azt jelenti, hogy:

- az AI gazdasági és társadalmi előnyeiből ne csak néhány ország vagy nagy technológiai vállalat részesüljön;
- az AI támogassa az egészségügy, az oktatás, a kutatás, a mezőgazdaság és a katasztrófavédelem fejlődését világszerte;
- a kevésbé fejlett országoknak is legyen beleszólásuk az AI szabályozásába;
- az AI-rendszerek tervezése során figyelembe vegyék a különböző társadalmi, gazdasági és kulturális szempontokat.

Az ENSZ szerint az AI akkor tekinthető valóban etikusnak és fenntarthatónak, ha nem növeli tovább a digitális és gazdasági egyenlőtlenségeket, hanem szélesebb körben járul hozzá a jóléthez, a fejlődéshez és a lehetőségekhez.

TOVÁBB



Az AI és a mentális egészség: új etikai kérdések

A kutatások szerint az AI-chatbotok nem feltétlenül adnak rossz tanácsot, de érzékeny mentális egészségi helyzetekben nem mindig reagálnak megfelelően. Előfordulhat, hogy nem ismerik fel a krízishelyzeteket, nem irányítanak szakemberhez, vagy nem nyújtanak elég biztonságos támogatást, ezért használatuk fokozott körültekintést igényel.

Ezért a mentális egészség támogatásában az AI jelenleg nem helyettesítheti az emberi szakértelmet és a megfelelő szakmai felügyeletet.

A kutatók szerint különösen nagy kockázatot jelenthet, ha a felhasználók érzelmi támogatás, diagnózis vagy terápiás segítség *helyettesítőjeként* fordulnak az AI-hoz. Ez veszélyes lehet felnőttek és gyermekek esetében egyaránt, hiszen ezek az eszközök nem képesek helyettesíteni a szakképzett mentálhigiénés vagy egészségügyi szakemberek munkáját.

A felelős AI-használat része annak felismerése is, hogy a mesterséges intelligencia bizonyos helyzetekben nem megfelelő forrása a

segítségnek, és nem helyettesítheti az emberi szakértelmet vagy a személyes támogatást.

TOVÁBB

Stanford AI Index 2026: Egyre több a szabályzat, de kevés a tényleges kontroll

A Stanford Egyetem Emberközpontú Mesterséges Intelligencia Intézetének (Stanford HAI) 2026-os AI Index jelentése szerint a mesterséges intelligencia rohamos ütemben terjed a szervezetekben, aminek következtében a felelős használat és az AI governance is kezd kiemelt figyelmet kapni.

A vállalatok mind nagyobb hangsúlyt fektetnek a Responsible AI gyakorlatokra: nő az AI-governance szerepkörök száma, és egyre kevesebb szervezet működik AI-irányelvek nélkül. A gyakorlati megvalósítás ugyanakkor sok esetben még elmarad a szándékoktól. Ennek oka elsősorban a megfelelő szakértelem hiánya, a korlátozott erőforrások és a folyamatosan változó szabályozási környezet.

A jelentés másik fontos megállapítása, hogy miközben az AI-modellek teljesítménye és üzleti alkalmazása gyorsan bővül, *az AI-fejlesztő cégek egyre kevesebb információt osztanak meg* arról, hogyan készülnek ezek a rendszerek, és milyen kockázatokkal járhat a használatuk. A felhasználók számára kevesebb információ érhető el a tanítóadatokról, a modellek működéséről és azok valós hatásairól.

Minél kevesebbet tudunk arról, hogyan készülnek és milyen elvek mentén működnek az AI-rendszerek, annál nehezebb megalapozott bizalmat építeni irántuk. Az etikus AI egyik legfontosabb feltétele ezért továbbra is az átláthatóság.

TOVÁBB

Etikus promptolás: nem csak az AI-on múlik

A GovTech elemzése arra hívja fel a figyelmet, hogy az AI etikája nem kizárólag a modellek tanítóadatairól szól. A felhasználók által adott utasítások, vagyis a promptok is jelentős etikai következményekkel járhatnak.

A rosszul megfogalmazott promptok nemcsak pontatlan válaszokat eredményezhetnek, hanem etikai kockázatokat is hordoznak. Ha például a felhasználó általánosításokra vagy sztereotípiákra épülő utasításokat ad, az AI felerősítheti a meglévő előítéleteket.

Adatvédelmi problémát okozhat az, ha valaki személyes, bizalmas vagy érzékeny információkat oszt meg egy nyilvános AI-eszközzel.

Emellett egy olyan prompt, amely nem ösztönzi a források ellenőrzését vagy a bizonytalanság kezelését, hozzájárulhat félrevezető, pontatlan vagy ellenőrizetlen tartalmak létrehozásához.

A cikk egyik fontos üzenete, hogy nemcsak a modell lehet elfogult, hanem a felhasználó is bevihet etikai kockázatokat a promptokon keresztül.

TOVÁBB

Etikus AI

Az Etikus AI kampány célja, hogy közérthető, nem technikai nyelven mutassa be: a mesterséges intelligencia használata etikai, társadalmi és emberi kérdés – nem csupán technológiai.

TOVÁBB

Egyenlítő Alapítvány
Váci út 33., Budapest
Hungary

You received this email because you signed up on our website or made a purchase from us.

Unsubscribe

SENT BY mailerlite