A Nemzeti Közszolgálati Egyetem Információs Társadalom Kutatóintézete havi hírlevele a mesterséges intelligencia alkalmazásáról, társadalmi hatásairól és kérdéseiről

2025 április

Az NKE ITKI honlapja: itki.uni-nke.hu

NEMZETI KÖZSZOLGÁLATI EGYETEM
LUDOVIKA

TARTALOMJEGYZÉK

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

Etika és jog

Az MI 2025-ben: a Stanfordi Emberközpontú Mesterséges Intelligencia Intézet (HAI) jelentése

A Stanford Egyetem Emberközpontú Mesterséges Intelligencia Intézete (HAI) minden évben nagyszabású jelentésben foglalja össze az MI-fejlesztések előző évi eredményeit. Az idei jelentés legfontosabb megállapításai a következők: A mérési rendszerek szerint az MI-rendszerek teljesítménye folyamatosan javul. Jelentős előrelépések történtek a kiváló minőségű videók előállításában, és bizonyos programozási feladatokban az MI-ügynökök még az embereket is felülmúlják. Az Egyesült Államok továbbra is vezető szerepet tölt be a legfontosabb MI-modellek területén, de Kína kezdi csökkenteni a szakadékot. 2024-ben az amerikai intézmények 40 figyelemre méltó mesterséges intelligencia modellt készítettek, Kínában ez a szám 15 volt, Európában pedig 3. Míg az Egyesült Államok megtartotta vezető szerepét a mennyiségben, a kínai modellek gyorsan csökkentették a minőségi szakadékot. Az MI-rendszerek egyre hatékonyabbá, megfizethetőbbé és elérhetőbbé válnak. Az egyre nagyobb képességű kis nyelvi modelleknek köszönhetően egy GPT-3.5 szinten teljesítő rendszer következtetési költsége több mint kétszáznyolcvanad részére csökkent 2022 novembere és 2024 októbere között. Hardver szinten a költségek évente 30%-kal csökkentek, míg az energiahatékonyság évente 40%-kal javult. 2024-ben az MI-ipartól származott a legfontosabb MI-modellek közel 90%-a, szemben a 2023-as 60%-kal, miközben az akadémiai/kutatói szféra továbbra is a leggyakrabban idézett kutatások legfőbb forrása. A modellek mérete továbbra is gyorsan növekszik – a betanításhoz szükséges számítási kapacitás öthavonta, az adathalmazok nyolchavonta, az energiafogyasztás pedig évente megduplázódik. A teljesítménybeli különbségek azonban csökkennek. A jelentés alapjául szolgáló széleskörű adatbázis szabadon elérhető.

The 2025 AI Index Report

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

Az Egyesült Arab Emírségek integrált, MI-alapú törvényhozási rendszert hoz létre

Az olajban gazdag Öböl-menti arab országok a felhalmozott gazdagságot, az olaj utáni korszakra készülve részben a high-tech iparok kiépítésére fordítják. Ebbe a sorba illeszkedik az Egyesült Arab Emírségekből származó bejelentés arról, hogy MI-alapú törvényhozási, jogalkotási ökoszisztémát hoznak létre. Az új mesterséges intelligencia rendszer le fogja fedni a teljes jogalkotási környezetet, összekapcsolva a törvényeket a bírósági határozatokkal, a közigazgatás intézkedéseivel és a közszolgáltatásokkal, valós idejű adatokkal látva el a rendszert arról, hogy a törvények végrehajtása hogyan folyik a valóságban. Egy megfogalmazás szerint úgy kell ezt elképzelni, mint a kormányzásban alkalmazott élő GPS-et, amely a terepen történtek alapján folyamatosan újraszámolja az útmutatásait. Az ökoszisztéma proaktív megközelítést vezet be a jogalkotási tervezésben, lehetővé téve a kormány számára, hogy megelőzze a globális geopolitikai, gazdasági, társadalmi és technológiai környezet gyors változásait. Az új modell magában foglal egy nagy országos jogalkotási adatbázist, amely nemcsak a szövetségi és tagállami jogszabályokat integrálja, hanem a bírósági határozatokat, a kormányzati folyamatokat, a szolgáltatásokat és a digitális közigazgatási rendszereket is. Úgy tervezték, hogy valós időben figyelemmel kísérje a globális jogalkotási fejleményeket, és felmérje azok relevanciáját az Egyesült Arab Emírségek szempontjából. Emellett képes lesz azonosítani a nemzeti jogi keretrendszer hiányosságait vagy ellentmondásait. A tervezők szerint ez az összetett rendszer lehetővé teszi az Egyesült Arab Emírségek számára, hogy a hagyományos statikus törvények helyett élő, reagáló szabályozást alakítsanak ki. Képes lesz arra is, hogy úgy biztosítsa mesterséges intelligencia által vezérelt szabályozás rugalmasságát és sebességét, hogy közben megőrzi az állam alapítóinak értékeit. A várakozások szerint a rendszer akár 70 százalékkal is felgyorsíthatja a jogalkotási folyamatot, mesterséges intelligencia eszközök használatával csökkentve a kutatáshoz, a szerkesztéshez, az értékeléshez és a törvényhozáshoz szükséges időt és energiát. A kezdeményezés részeként a kabinet alárendeltségében egy speciális szervezet (Regulatory Intelligence Office) jön létre. Ez az új iroda fogja vezetni a rendszer tervezését, megvalósítását és koordinálását a szövetségi és tagállami, helyi hatóságokkal, valamint a magánszektor érdekelt feleivel partnerségben. A hivatalos vélemény szerint a lépés tükrözi az Egyesült Arab Emírségek elkötelezettségét egy újgenerációs kormányzás kiépítése iránt, amely az emberi meglátásokat ötvözi a mesterséges intelligencia hatékonyságával, hogy gyorsabb, intelligensebb és alkalmazkodó képesebb jogszabályokat hozzon létre.

UAE Cabinet chaired by Mohammed bin Rashid, approves launch of first integrated regulatory intelligence ecosystem in UAE Government

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

Az Unió jogalkotói a mesterséges intelligencia jogalkotás „felhígításának” veszélyeire figyelmeztetnek

Az Európai Unió világelsőként ünnepelt, bár koránt sem kiforrott, sok szempontból még tökéletlen, de mégis átfogó mesterséges intelligencia szabályozó eszköze, a nehezen megszületett AI Act jelentős módosítása már egy ideje napirenden van. Az ügy hátterében az áll, hogy az idén februárban hivatalba lépett amerikai kormányzat az MI-fejlesztések és alkalmazások terén markánsan más filozófiát képvisel, mint elődje: Trump kormányának a felfogásában a mesterséges intelligencia az USA globális hegemón szerepe megőrzésének egyik, ha nem a legfontosabb kulcsa. Egyértelműen nemzetbiztonsági kérdés, és ennek megfelelőn sajátos fejlesztési megközelítés igényel. Míg az európai szabályozás – az Unió társadalom felfogásához, politikáihoz illeszkedő módon – a tématerület szabályozásánál (is) az egyéni, közösségi és társadalmi biztonság szempontjait igyekezett előmozdítani és őrizni, addig ezt az új amerikai politikai vezetés a fejlesztéseket erősen korlátozó tényezőnek tekinti. Amely ráadásul – a technológiai fejlődés jellegéhez igazodva – döntően amerikai tulajdonú óriásvállalatok működésének, továbbfejlődésének gátjaként jelenik meg. Ez az aggály – hogy a személyiségi jogok, kisközösségi és csoport érdekek, az esély- és nem-egyenlőség biztosítása indokolatlanul korlátozza az MI-fejlesztés tempóját és hatókörét – nem csupán a tengerentúli erő- és érdekcsoportok megközelítésében játszik központi szerepet. Európán belül is (szinte az AI Act első változatainak kialakulásától kezdve) kitapintható az erős vélekedés, hogy ez a szabályozási koncepció visszahúzza az innovációt. A nagypolitika különös alakulása, a hatalmi konstrukciók, keretrendszerek érezhető drámai átformálódása most olyan helyzetet teremt, ahol az egymástól gyors ütemben „elhidegülő” Amerika és Európa befolyásos csoportjainak koncepciói, törekvései találkoznak. Az USA katonai, diplomáciai, politikai „kivonulása” nyomán felerősödő európai autonómia gondolat képviselői körében ugyanis a geopolitikai érdekek primer képviselete éppen olyan fontossá kezd válni – akár más szempontok rovására is –, mint az új amerikai kormánykörök előtt. Ennek megfelelően, az Európa stratégiai autonómiáját előtérbe helyező EU Bizottság egy olyan vitafolyamatot kezdeményezett, amelynek eredményeként az Unió rendkívül szigorú AI Act alkalmazása számos fontos területen (a mostani kötelező erő helyett) önkéntes vállalásként „terhelheti” a technológiai vállalati szektor szereplőit. A törvény eredeti kidolgozói közé tartozó uniós törvényhozók egy csoportja azonban a szabályozás ilyen „felhígításának” súlyos veszélyeire figyelmeztet most.

EU lawmakers warn against ’dangerous’ moves to water down AI rules

„Veszélytelenség” és „biztonságosság”: az MI-alkalmazás összemosódó kulcsfogalmai

A magyar nyelvet beszélők egy kicsit gondban vannak, ha az angol „safety”, illetve „security” kifejezéseket kell pontosan lefordítaniuk. A helyzet érdekessége azonban az, hogy – meglepő módon – az angol anyanyelvűeknek is gondot okoz a két fogalom pontos értelmezése, és éppen a mesterséges intelligencia technológiák területén okoz rendkívül komoly félreértéseket, ellentmondásokat ez a szemantikai képlékenység. Érdemes rögtön leszögezni: itt nem valamiféle nyelvészkedésről, bölcsész okoskodásról van szó, hanem véresen komoly szabályozási, fejlesztéspolitikai, sőt geopolitikai tétje van mindennek. Ezért is okozhatott némi értetlenséget, amikor az év elején, a párizsi MI-csúcson az újonnan hivatalba lépő amerikai kormányzat, és hozzá csatlakozva a brit politikai vezetés a mesterséges intelligencia-fejlesztések (és -alkalmazások) tekintetében a „safety”-ről egyértelműen a

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

„security”-re tette át a politikai hangsúlyt. Vagy, ha úgy tetszik, az MI „safety” kérdése immár alapvetően MI „security” felfogásban kapott jelentőséget a vezető angol-szász hatalmak megközelítésében. A mesterséges intelligencia fejlesztési irányainak meghatározásában komoly szerepet kapó amerikai szakmai körök (döntően a korábbi demokrata párti kormányzat liberális felfogását tükröző módon) attól tartanak, hogy a fejlesztés-politikai hangsúlyváltás azzal jár majd, hogy az állambiztonsági, nemzetbiztonsági szempontok kerülnek előtérbe az új technológiák fejlesztése és alkalmazása során, és így óhatatlanul háttérbe szorulnak az eddig hangsúlyosan megjelenő hagyományos „safety” elvárások. Feloldva a rébuszt: arról van itt szó, hogy a nemzet belső és külső biztonsága és nagysága (a „security”) mellett csupán másodlagosan jelentkeznek (ha egyáltalán jelentkeznek) a személyiségi jogokkal, az egyéni szabadságjogokkal (vagy a liberális koncepcióban szintén „egyéni emberi szabadságjog” kategóriájába sorolt olyan szempontokkal, mint például a nemi identitások kérdése) kapcsolatos szempontok és elvárások. A szakmai és politikai körök egy része attól tart, hogy ez a fejlesztés-koncepcióban bekövetkezett irányváltás háttérbe szorítja az MI-rendszerek és modellek „elfogultsága” („bias”) korlátozásának mindent felülíró eddigi prioritását. A megváltozott koncepció hátterében vélhetőleg nem is elsősorban az új amerikai kormányzat és köreinek markánsan más ideológiai beállítottsága áll. Hanem sokkal inkább a világrendet gyorsan és gyökeresen átformáló mozgások, amelyeken belül az új MI-technológiáknak kulcsfontosságú szerep juthat. Ennek megfelelően a globális „MI-versenyfutás” megnyerése fontosabb szempontként jelentkezik, mint hogy a kifejlesztett technológiák a társadalmi biztonság szélesebben értelmezett szempontjainak érvényesítése esetleg lassítsa, korlátozza az MI-fejlesztéseket.

Untangling AI safety and security

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

Trendek

A generatív MI az egészségügyben: az átalakító innováció lehetőségei és felelősségei

A mesterséges intelligencia, és ezen belül a generatív MI és a nagy nyelvi modellek, technológiai forradalmat jelentenek. Ez a nagy változás olyan pillanatban éri az egészségügyet, amikor az ellátórendszernek egyébként is jelentős problémákkal kell(ene) megbirkóznia. A GenMI-ben megtestesülő innovációknak az az ígéretük, hogy jobb, személyre szabott ellátást tesznek lehetővé a nehezülő feltételek ellenére is. Az Amerikai Egyesült Államok Nemzeti Akadémiáinak részét képező National Academy of Medicine kötete a generatív MI-ben rejlő átalakító potenciált vizsgálja, különös tekintettel arra, hogy hogyan lehet alkalmazni ezeket a rendszereket a klinikai döntéshozatalban, a szervezeti hatékonyság javításában és a betegek gyógyítási folyamatba történő bevonására. Bár a korai tapasztalatok ígéretesek, a GenMI eszközök valós hatására vonatkozó bizonyítékok továbbra is korlátozottak. Továbbá a GenMI egészségügyben és biomedicinában történő alkalmazásai egyedi kockázatokat vetnek fel. Az említett alkalmazási lehetőségek eléréséhez ezért számos fél és szakterület együttes erőfeszítésére lesz szükség. Az Akadémia kidolgozott egy sor közös kötelezettségvállalást, amelyek megbízható keretet biztosítanak az egészségügy és az egészségügyi szolgáltatások számára, hangsúlyozva a hozzáférhető, megfizethető, átlátható, elszámoltatható és adaptív egészségügyi rendszereket. Ennek a törekvésnek részeként az összes érintett terület (beleértve a kormányzatot is) képviselői számára műhelybeszélgetéseket szerveztek, hogy biztosítsák a közös gondolkodást és megértést. A főbb témák a nagy nyelvi modellek és a GenMI természete és az egészségügyi vonatkozásai voltak. Alapvetően a lehetséges előnyökre, a kockázatokra, valamint az egészségügyben szükséges irányelvekre és védőkorlátokra (guardrails) összpontosítottak. A kiadvány ezekre a megbeszélésekre és a legfontosabb tanulságokra épít, felvázolva a GenMI lehetséges kockázatait, ígéreteit és a jövő feladatait.

Generative Artificial Intelligence in Health and Medicine: Opportunities and Responsibilities for Transformative Innovation

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

Sok évtizedre lesz még szükség a transzformatív MI eléréséhez

Az Epoch AI kutatóintézet a mesterséges intelligencia fejlődési pályáját és irányítását befolyásoló legfontosabb trendeket és problémákat kutatja, folyamatosan mérik és értékelik az egyre újabb rendszerek teljesítőképességét. Mindennek hátterében ott található az a fejlesztők köreiben nagyon gyakori elképzelés, hogy az emberi képességeket elérő és meghaladó MI-rendszerek eljövetele nagyon közel van már. Ez a tanulmány – bár osztozik az előfeltevésekben – néhány józan szempontot fogalmaz meg. Nincsenek olyan trendek, amelyeket extrapolálva bekövetkezne, hogy nagyon rövid időn belül (néhány éven belül) eljutnánk a transzformatív MI-hez. A teljes távmunka automatizációját leghamarabb évtizedes távlatban várhatjuk (extrapolálva), és szerző várakozása szerint ezek a trendek lassulni fognak. A szerző nem fogadja el, hogy pusztán a szoftveres rendszerek teljesítményének javulása elég lenne hihető mechanizmusnak. Pedig ezen alapszik az MI-rendszerek növekedési ütemének hirtelen és drámai felgyorsulása, és ebből következtetnek arra, hogy 10 évnél jóval rövidebb idő is elég lesz a távmunka automatizálásához. A kísérleti számítások és a valós világ viszonyai olyan szűk keresztmetszeteket jeleznek, amelyek lehetetlenné teszik a nagyon gyors haladást. Az intézet úgy gondolja, hogy a Moravec-paradoxon miatt nagyon tévesek a szűkkörű írási, kódolási vagy érvelési feladatokból származó intuíciók arról, hogy milyen gyorsan tudnának gondolkodni a mesterséges intelligencia rendszerek. A gyakorlatban arra számíthatunk, hogy ahogy az MI-rendszereket arra fogjuk használni, hogy ügynöki, multimodális (pl. kép, hang) és hosszabb összefüggéseket figyelembe vevő feladatokat hajtsanak végre, úgy lesznek egy lassabbak és drágábbak. Mivel az MI-ügynökök térnyerése valószínűsíthető, ezért ez a tendencia is folytatódni fog. A Moravec-paradoxon szerint a hagyományos feltételezésektől eltérően a magas szintű gondolkodás nagyon kicsi számítási teljesítményt vesz igénybe, az alacsony szintű szenzomotoros készségek ellenben óriási számítási erőforrásokat igényelnek. Népszerűen megfogalmazva: a nehéz feladatok könnyűek, a könnyű feladatok pedig nehezek. Mindezek alapján a szerző úgy gondolja, hogy a távmunka teljes automatizálása 10 éven belül valószínűsíthető, amennyiben egyenesen extrapoláljuk az MI-cégek bevételeinek jelenlegi növekedési ütemét, és feltételezzük, hogy ez nem fog lassulni. Ennek azonban csak 30% körüli esélyt adna.

The case for multi-decade AI timelines

„Varázsszavak", avagy a természetes nyelvű kódolás jelene és jövője

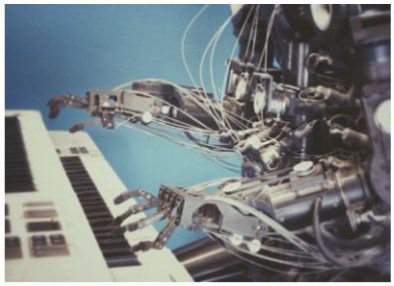
MI-szakmai körökben rohamosan terjed – igaz, még csak néhány hete – egy újabb bűvszó, vagy, ahogy mondani szokták: hype. Bár az ilyen ügyekkel kapcsolatban mindig valós a kockázat, hogy tényleg nem csupán valamiféle kérészéletű fellángolásról van-e szó, ennél a témánál azonban feleslegesnek látszik ez az óvatosság. Ugyanis a „vibe coding", annak ellenére, hogy február utolsó napjaiban robbant be a szakmai közbeszédbe, valójában egy jó ideje velünk levő, nyilvánvaló jelenséget takar. Andrej Karpathy, a Tesla AI volt vezetője, a piacon meghatározó OpenAI cég egyik alapítója fogalmazta meg két évvel ezelőtt: „A legújabb, csúcsszereplő programozási nyelvet úgy hívják: angol." Nevezetesen arról van szó, hogy a számítógépes programozáshoz, ahhoz, amihez évtizedeken át nagyon specializált szaktudásra volt szükség, azt ma már – elsősorban a nagy nyelvi modelleknek hála – bármely angolul (vagy magyarul, vagy urdu nyelven, esetleg szingalézül) tudó ember, minden különös félkészülés és előkészület nélkül megteheti, saját emberi nyelvén. Ez a „játszi könnyedségű, szórakozás számba menő programozás", azaz a „vibe coding" kifejezés lényege. És ez valójában nagyon nagy dolog. Jól illeszkedik

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

ahhoz a történethez, amely 2022 novemberében, egy ChatGPT nevű nagy nyelvi modell nyilvánossá tételével indult, egyszerre valóságos technológiai (sőt technológiai és társadalmi) forradalmat elindítva. A vibe coding témája hamar divatossá vált, néhány napon belül már eligazodni sem lehet majd az ebben a tárgykörben született írások kavalkádjában, de most még hasznos lehet beleolvasni az MI-terület egy másik fenegyereke, Ethan Mollick most megjelent írásába. Mollick ugyanis fogta magát, és egy „ellenőrzött kísérletsorozatban" mérte fel, hogy vajon mennyire valós, működő gyakorlat immár ez a divatossá váló vibe coding? Az Anthropic közkedvelt, nagy teljesítményű nyelvi modellje a Claude Sonnet 3.7 változata lett a „segítőtársa", hogy kipróbálja a saját nyelven való programozás jelenlegi állását: lehetőségeit, és még fennálló korlátait.

Speaking things into existence

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu



Működésben

Energia és MI: rohamosan növekvő villamos-energiaigény és lehetőség az energetikai szektor működésének átalakítására

A mesterséges intelligencia és az energiatermelés és -felhasználás trendjei önmagukban is számos kérdést vetnek fel. Fokozottan így van ez akkor, amikor együttes, egymásra gyakorolt hatásukat próbálják felmérni. Rengeteg becslés, vélemény látott ezekről napvilágot, és a Nemzetközi Energiaügynökség (International Energy Agency, IEA) elérkezettnek látta az időt, hogy az előrejelzéseket szilárdabb, egzaktabb alapokra helyezze. A frissen megjelent Energia és mesterséges intelligencia (Energy and AI) című különjelentés a jelenleg elérhető legátfogóbb, adatvezérelt globális elemzés az energia és a mesterséges intelligencia közötti egyre növekvő összefüggésekről. A jelentés új adatkészletekre, valamint a politikai döntéshozókkal, a technológiai szektorral, az energiaiparral és nemzetközi szakértőkkel folytatott széleskörű konzultációkra támaszkodik. Az előrejelzések szerint a világ adatközpontjainak villamosenergia-igénye 2030-ra több mint kétszeresére, körülbelül 945 terawattórára (TWh) fog nőni, ami valamivel több, mint Japán teljes villamosenergia-fogyasztása napjainkban. A mesterséges intelligencia lesz ennek a növekedésnek a legjelentősebb mozgatórugója, mivel a mesterséges intelligenciára optimalizált adatközpontok villamosenergia-igénye várhatóan több mint négyszeresére fog nőni 2030-ra. Az adatközpontok növekvő villamosenergia-igényének kielégítésére az energiaforrások széles skáláját fogják kiaknázni – bár a megújuló energiaforrások és a földgáz várhatóan vezető szerepet töltenek majd be, mert költségek tekintetében versenyképesek és a kulcsfontosságú piacok könnyen elérhetőek. A jelentés szerint a mesterséges intelligencia felerősíthet bizonyos energiabiztonsági problémákat, miközben segíthet másokat kezelni. Az energiaszolgáltatók elleni kibertámadások száma az elmúlt négy évben megháromszorozódott, és a mesterséges intelligencia miatt egyre kifinomultabbá váltak. Ugyanakkor a mesterséges intelligencia egyre fontosabb eszközzé válik az energiaszolgáltató vállalatok számára az ilyen támadások elleni védekezésben. Bár az adatközpontok villamosenergia-igényének növekedése várhatóan növeli a kibocsátásokat, ez a növekedés az energiaszektor egészét tekintve csekély lesz, és a jelentés szerint potenciálisan ellensúlyozhatja az MI által lehetővé tett kibocsátáscsökkentés, ha a technológia széleskörűen elterjed. Érdekesség, hogy a jelentéssel párhuzamosan az IEA egy új MI-ügynököt is közzétett, amely lehetővé teszi, hogy az olvasók interakcióba lépjenek a jelentéssel. Az ügynök a jelentés weboldalán érhető el, és a jelentés tartalmával kapcsolatos kérdésekre egyszerű és párbeszédes módon válaszol.

Energy and AI

Az Anthropic tudósai kiderítették, hogyan „gondolkodik” az MI – az eredmény nyugtalanító

A mesterséges intelligencia modellekkel kapcsolatban az egyik legnagyobb probléma a rendszerek „fekete doboz” jellege. Lényegében senki (beleértve a rendszereket megalkotó szakembereket) sem tudja egészen pontosan, hogy mi történik a szisztéma belsejében. Könnyű belátni, hogy ez nem elhanyagolható akadályt, nehézséget jelent a technológia széleskörű társadalmi elterjedésében. Az olyan szakterületeken, mint például az egészségügy (azon belül is a gyógyítás), illetve az igazságszolgáltatás, a szélesen vett jogszolgáltatás (egyébként éppen ez a kettő az, ahol az MI-technológiák talán a legrégebb óta kapnak kisebb-nagyobb szerepet), ennek az átláthatatlanságnak komoly jelentősége van. Miközben a transzparencia hiánya érezhetően visszafogja az MI-elterjedését ezeken a területeken, az ellentábor érvelésében is van némi igazság: hogy mindezeknél a hagyományos emberi döntéshozatal sok esetben legalább annyira áttekinthetetlen. Azzal kapcsolatban, hogy a tudásunk jelen pillanatban mennyire korlátozott az általunk létrehozott MI-rendszerek pontos működésével kapcsolatban, az utóbbi bő egy évben számtalan anekdotisztikus adalék került elénk. A rendszerutasításaiknak csupán látszólag megfelelő, és éppen ellentétes módon működő modellek esete közismert. Ahogy az is, amikor egy-egy modell egy adott munkamenetben a „lustálkodást választja”. A mindenképpen több ismeretet, pontosabb rálátást igénylő terület kérdéseinek megoldásához most érdekes kutatási adatokkal járulhat hozzá az Anthropic fejlesztő cég egyik szakmai csapata. A nemrég két publikációban is közre adott eredmények azzal kecsegtetnek, hogy – talán első ízben – sikerült bepillantást nyerni egy rendszer belsejébe, és pontosan feltérképezni, hogyan dolgozza fel az információt, illetve hogyan hoz döntéseket egy nyelvi modell. Az idegtudományi kutatásokból kölcsönzött új módszerekkel a kutatók pontosan követni tudják a rendszerekben mozgó, neuron-szerű jelenségek pályáit. A kísérleti munka során nyert egész sor felfedezés között olyan meghökkentő jelenségek is bizonyítást nyertek, mint például az előre gondolkodás képessége. Meghökkentő látni a „visszafelé érvelés” folyamatát is, amikor a modell a kérdésből visszafelé épít fel egy „logikus” gondolatsort. Nem kevésbé izgalmas az a felfedezés, ami arra utal, hogy adott esetben a modell miért „lódít”. Az új vizsgálati módszer (angol elnevezésével „circuit tracking”) első ízben ad tárgyszerű leírást a sok problémát okozó hallucináció előfordulásának okairól és konkrét módozatáról.

Anthropic scientists expose how AI actually ’thinks’ – and discover it secretly plans ahead and sometimes lies

Könyvkalózkodás: a nagy MI-fejlesztők is előszeretettel használják a LibGent

A szerényen fizetett kutatók, oktatók szerte a világban előszeretettel gyűjtenek a munkájukhoz nélkülözhetetlen szakirodalmat olyan repozitóriumokból, ahol a feltöltött anyag jogtisztaságáról nem adnak az egyes művek mellé nagypecsétes igazolást – igaz, a használók közül nem is kér ilyet senki. Ezek a jelentős méretű gyűjtemények – amelyek közül a legnagyobb, de mindenképpen a legismertebb a legendás LibGen – különös megítélési kettősségben élik életüket. A nagy könyv- és tudományos folyóirat kiadók szemében ezek a könyvkalózok súlyosan veszélyeztetik, károsítják a világ tudományos publikációs iparát. Mások azonban – köztük a sokszor végletes mértékben kihasznált szerzők – egyfajta intellektuális Robin Hoodnak tekintik az ilyen könyves helyek üzemeltetőit, a tudomány szabadságharcosainak, és az emberi progresszió igazi támogatóinak. Anélkül, hogy egyik, vagy másik oldal pártjára, igazsága mellé állnánk, azt nyugodtan kijelenthetjük: a Globális Délnek nevezett fejlődő világ tudományos élete, köz- és felsőoktatása ma nem lenne ugyanaz ezek nélkül a tudományos kincsestárak nélkül. Azonban nem csupán az intellektus szegénylegényei (és –leányai) fordulnak előszeretettel ezekhez a repozitóriumokhoz a munkájukhoz, fejlődésükhöz nélkülözhetetlen szakmai nyersanyagért. Ahogy az utóbbi időkben egyre-másra pattant ki, a dollár százmilliókkal kistafírozott gazdag és hatalmas technológiai óriások is előszeretettel dézsmálgatják a tudás és információ eme

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu

ingyenes éléskamráit. E mögött nyilván erős profitoptimalizálási törekvések is állhatnak, de ami egyértelműen erre a sajátos beszerzésre ösztönzi a nagy MI-fejlesztőket, az a még elérhető „szűz" adatkészletek rohamos fogyatkozása. Márpedig, ahogy az közismert még a laikusok számára is, az adat (méghozzá emberi ésszel szinte felfoghatatlan mennyiségű adat) nélkül lényegében aligha létezhetne a szédítő karriert befutó LLM-technika. A problémán egyelőre nem változtat érdemben számos ígéretes törekvés és fejlesztés, a radikálisan kisebb adatkészlettel is elboldoguló kis nyelvi modellektől a szintetikus tréningadat előállításon át a már meglevő nagymodelleken történő LLM-betanításig. Az illegális könyvrepozitóriumokból kinyert adatokon olyan csúcscégek nevelgették új modelljeiket, mint például a Facebook mögött álló Meta. Ahogy az azóta megindult szerzői jogi perek tanúsítják, a cég nagyon sikeres nyílt forrású nyelvi modellje, a Llama kiképzésében jelentős mértékben használt fel például a LibGen-től megszerezhető szakmai adatkészleteket. Jogi szakértők azt gyanítják, hogy az ilyen tréningadat-szerzési gyakorlat sokkal elterjedtebb, mint ahogy azt gondolnánk. A Zuckerberg-vállalkozás ilyen irányú gyakorlata csupán a jéghegy csúcsa.

The Unbelievable Scale of AI's Pirated-Books Problem

Az MI jó eséllyel „jósolja meg" a következő egészségügyi vészhelyzet helyszínét és időpontját, legyen szó balesetről, vagy rosszullétről

A jogi terület, valamint a pénzügyi- és bankszektor mellett (sőt, azokat többnyire jócskán megelőzve) az egészségügy az a társadalmi szolgáltatási terület, ahol a mesterséges intelligencia technológiák a legrégebb óta és a legszélesebb körben terjednek. A területhez ennek megfelelően szinte naponta kapcsolódnak olyan MI-újítások, fejlesztések, amelyek figyelemre tarthatnak számot. Egy a közelmúltban bemutatott izraeli rendszer még ebből a bőséges kínálatból is kiemelkedik. Izrael MI-csúcstechnológiái már egy ideje ismertek a szakmai közvélemény előtt, ám ezek többsége sajnálatos módon a háborúhoz kapcsolódó megfigyelő, célmegjelölő, döntéstámogató, harc-automatizáló rendszer. A Földközi-tengeri ország mindenesetre az innovációk, start-upok, csúcstechnológiai fejlesztések egyik világközpontja, és a háborús célú fejlesztések többségére amúgy is jellemző „kettős felhasználhatóság" eredményeként a biztonságpolitikai, hadi célú kutatások nyomán jó néhány figyelemre méltó polgári technológiai is „kipörög". Az izraeli hadsereg jelhírszerző egysége, a katonai MI-fejlesztésekért is felelős 8200 számú egység most egy olyan figyelemre méltó újítással lépett ki a civil közvélemény elé, ami forradalmasíthatja a polgári egészségügyi ellátást. Méghozzá a krízishelyzeti ellátások nagyon kritikus területén. Krízishelyzeten ebben az esetben nem csupán a kiterjedt vészhelyzeteket (katasztrófákat) kell értenünk, hanem minden olyan eseményt – balesettől a rosszullétig –, amely a mentőszolgálat mozgósítását követeli meg. A közel-keleti ország Egyesült Hatzalah nevű polgári mentőszervezetének 18 évnyi eset-adattömegére építve egy olyan (térkép alapon kijelző) mesterséges intelligencia alkalmazást fejlesztettek most ki, amely néhány perc, esetleg negyed óra előre tartással jósolja meg a bekövetkező váratlan egészségügyi vészhelyzeteket, beleértve az esemény helyszínét is. Az előzetes működési tapasztalatok szerint a rendszer nagyjából 85%-os valószínűséggel (pontossággal) jelzi előre a mentőszolgálat központi irányítóinak például egy váratlan szívroham-eset bekövetkeztét és helyszínét. A néhány perces „predikciós előretartás" jelentéktelennek tűnhet, de egészségügyi vészhelyzetekben szó szerint életek múlhatnak ezen a néhány értékes percen.

Breatkthrough AI predicts where and when the next medical emergency will happen

1083 Budapest, Ludovika tér 2. | +36 1 432 9000
POSTACÍM 1441 Budapest, Pf. 60. | nke@uni-nke.hu, uni-nke.hu