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

2023 augusztus

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 ENSZ is vizsgálja a biztonságos MI-kormányzás lehetséges kereteit

A mesterséges intelligencia gyorsuló fejlődésével egyre nagyobb számban jelennek meg a jogi, megbízhatósági, átláthatósági és biztonsági problémák, melyek ugyanakkor ösztönzik az ezekről folytatott társadalmi viták létrejöttét is. Az MI előretörése ma már túlmutat az országhatárokon, az általa jelentett veszélyek nemzetközi szinten jelentkeznek és az egyes országok külön-külön nem tudnak megbirkózni velük. Az MI mostanra közös szabályozást igényel, amely alapján az államok és a nem állami szereplők azonos normákhoz tartják magukat és számíthatnak arra, hogy mindenki biztonságosan használja a technológiát. Ez a megközelítés az MI területén egymással versengő országok közötti geopolitikai kapcsolatokat is javíthatná. Az MI szabályozásáról való közgondolkodás globális szintre emelkedését mi sem bizonyítja jobban, mint hogy július 18-án az ENSZ Biztonsági Tanácsa is napirendjére tűzte az MI biztonsági kérdéseinek széleskörű vitáját. A Tanács az ülésen megvitatta azokat az irányítási struktúrákat, amelyek szükségesek a technológia nemzetközi békére és biztonságra vonatkozó kockázatainak mérsékléséhez. A Biztonsági Tanács soros elnökségét adó Nagy-Britannia által kezdeményezett ülésen konkrét határozat nem született ugyan, de a tagállamok – köztük a világ három vezető politikai és katonai nagyhatalma, az Egyesült Államok, Oroszország és Kína is – megállapodtak abban, hogy az MI-technológiák az egész emberiség békéjét és biztonságát fenyegető veszélyt jelenthetnek, ezért szabályozásuk a legmagasabb multilaterális szinten volna kívánatos. Az ügy horderejét jól mutatja az is, hogy a Tanács az ülésen a mesterséges intelligencia által jelentett egzisztenciális fenyegetést az emberiségre a nukleáris fegyverek veszélyeihez hasonlította és a technológia hasonló módon történő szabályozására szólított fel.

Overcoming geopolitical tension to govern AI for a peaceful future

A titkos hadműveletek törvényi szabályozása jelenthetné a mintát az MI nemzetbiztonsági célú alkalmazásaihoz

Az új MI-technológiák két módon is megjelenhetnek a modern államok nemzetbiztonsági szempontrendszerében: egyrészt az MI-alkalmazások új nemzetbiztonsági kockázatok forrásai lehetnek például az ellenséges MI-rendszerek kibertámadásokra, információs hadműveletekre való felhasználása révén, illetve a saját rendszerek kibernetikai vagy információs megtámadása következtében. A világ legjelentősebb agytrösztjeként számon tartott Brookings Institution szakértője most arra dolgozott ki javaslatot, hogy az Egyesült Államok hivatalos szervei és szervezetei által történő egyes offenzív jellegű,

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

nem nyilvános jellegű mesterséges intelligencia használatok szabályozásánál az USA jelenleg hatályban levő, és a különféle erőszakszervezetek által végrehajtott fedett (titkos) hadműveletek lefolytatásánál irányadó jogszabályokat vegyék követendő mintának. Az Egyesült Államok Kongresszusában mindeddig elhangzott törvényjavaslatok közül csak néhány érinti a nemzetbiztonsággal kapcsolatos MI rendszereket és egyik sem hoz létre semmiféle keretszabályozást az ilyen eszközökre vonatkozóan. Márpedig az amerikai hírszerző és katonai ügynökségek által kifejlesztett és használt MI rendszerek alighanem ugyanolyan jelentős kockázatot jelentenek, mint a nyilvánosan elérhető mesterséges intelligenciák. A szerző szerint a rejtett műveletekről szóló jogszabály számos eleme jól átültethető lenne egy új jogszabályba, amely a nemzetbiztonsági MI olyan magas kockázatú felhasználásával foglalkozik, amelyre egyébként a meglévő jogszabályok nem terjednek ki.

Regulating National Security AI Like Covert Action?

A jogalap nélküli gazdagodás és a tartalomszűrés esetleges összefüggései

A közösségi médiatereket elárasztó káros vagy egyenesen veszélyes tartalmak visszaszorítása modern társadalmaink egyik legsúlyosabb gondja, egyben legnehezebb feladata. Két szakértő, Ayelet Gordon-Tapiero és Yotam Kaplan legújabb tanulmányukban most annak a lehetőségét igyekeztek körüljárni, hogy miképp lehetne a „jogalap nélküli gazdagodás” (unjust enrichment) jogelvét az online tartalomszűrés problémaköréhez kapcsolni és ezzel áttörést elérni az internetes platformüzemeltetők aktív szerepvállalásában. A jogalap nélküli gazdagodás fogalma azon az alapgondolaton nyugszik, miszerint a jogalap nélkül (unjust) vagy károkozó módon (wrongful) végzett tevékenységeknek nem szabad nyereségessé válniuk. Az elv mögött meghúzódó logika egyszerű: mindaddig, amíg egy helytelen cselekvés jövedelmező, addig fenn is marad. Márpedig – vélik a tanulmány szerzői – éppen ezt a problémát lehet azonosítani napjaink platform krízisével kapcsolatban. Az egyetlen hatékony módja tehát annak, hogy a platformok felhagyjanak a káros tartalmak népszerűsítésével vagy tűrésével, ha gondoskodunk arról, hogy az ne legyen számukra profitábilis. A jogalap nélküli gazdagodás joga elősegítheti ezt a célt a nyereség elvonásának doktrínája révén, amely lehetővé teszi a bíróságok számára, hogy megfosszák a jogsértőket - ez esetben a platformok működtetőit - a jogtalanul szerzett nyereségtől.

Unjust Enrichment by Algorithm

A techvilág óriáscégei összefogtak a felelősségteljes MI érdekében

A világ négy - a mesterséges intelligencia kutatása és fejlesztése terén is - vezető vállalata, az OpenAI, a Google, a Microsoft és az Anthropic nemrégiben bejelentette a Frontier Model Forum elnevezésű iparági testület megalakulását, amely céljául a

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

határterületi MI modellek biztonságos és felelős fejlesztésének biztosítását tűzte ki. A fórum tagsága kizárólag azokra a vállalatokra korlátozódik, amelyek határterületi modelleket, azaz olyan nagyméretű gépi tanulási modelleket fejlesztenek és alkalmaznak, amelyek meghaladják a legfejlettebb létező modellek jelenlegi képességeit és sokféle feladatot képesek elvégezni. A szervezet tehát önszabályozó módon működik és csak a jelentősen nagyobb teljesítményű MI modellek potenciális kockázataival szándékozik foglalkozni. A fórum az elkövetkező évben három kulcsfontosságú területre kíván összpontosítani a határterületi MI modellek biztonságos és felelős fejlesztésének érdekében. Az első a tudásmegosztás és a legjobb gyakorlatok előmozdítása az ipar, a kormányok, a civil társadalom és a tudományos élet körében, különös tekintettel a biztonsági előírásokra és gyakorlatokra. A második terület az MI biztonsággal kapcsolatos kutatásainak támogatása, ezen belül is egy olyan nyilvános könyvtár kialakítása, amely tartalmazza az iparág legjobb gyakorlatait és követelményeit, valamint a műszaki értékeléseket és a referenciaértékeket. A harmadik terület a vállalatok, a kormányok és más érintett felek közötti információmegosztás és együttműködés elősegítése. A fórum további célkitűzései között szerepel az olyan alkalmazások kifejlesztésére irányuló erőfeszítések támogatása, amelyek segíthetnek a társadalom legnagyobb kihívásainak kezelésében, mint például az éghajlatváltozás mérséklése, a rák korai felismerése és megelőzése, valamint a kiberfenyegetések elleni küzdelem.

A new partnership to promote responsible AI

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

Trendek

Mérföldkövet jelenthet az MI folyamatos tanulásában a „katasztrofális felejtés” problémájának megoldása

Amerikai tudósok vizsgálják a gépi tanulás egy különös és erősen korlátozó velejáróját, a „katasztrofális felejtés” jelenségét, ami annyit jelent, hogy a mesterséges intelligencia rendszerek, miközben új feladatokat tanulnak meg, elveszítik a korábbi feladatok során megszerzett információkat. A kutatók arra az érdekes felismerésre jutottak, hogy a mesterséges neurális hálózatok - akárcsak az ember - jobban megjegyzik azokat az információkat, amelyeket egymástól jelentősen eltérő feladatok során sajátítottak el, mint a hasonló jellegű feladatokból származókat. A szakemberek most azt remélik, hogy a kutatások eredményeként, az emberi tanulási folyamatok másolásával, az MI-rendszerek folyamatos tanulási képességei jelentős mértékben javulhatnak, valamint csökkennek a gépi- és az emberi tanulási folyamatok jellege közötti különbségek, ami potenciálisan kifinomultabb teljesítményű MI-rendszerek kifejlesztéséhez vezethet. Annak megértéséhez, hogy a mesterséges ágensek miért alakítanak ki lyukakat a saját kognitív folyamataikban, az Ohiói Állami Egyetem csapata azt elemezte, hogy a "folyamatos tanulásnak" nevezett folyamat mennyire befolyásolja az általános teljesítményüket. A „folyamatos tanulás” (continual learning) a gépi tanulási folyamatok egyik fontos mozzanata: az a jelenség, amikor a rendszert arra programozzák, hogy folyamatosan tanuljon egy egymást követő elemekből álló feladatsoron úgy, hogy az egyes feladatok során megszerzett tudást tovább viszi és felhasználja a következő feladat jobb megértésére. Bár az autonóm rendszereket kihívást jelent megtanítani az olyan dinamikus, élethosszig tartó tanulásra, amire az ember alkalmas, ezeknek az új ismereteknek a birtokában a tudósok gyorsabb ütemben tudják bővíteni a gépi tanulási algoritmusokat, amelyek aztán könnyebben alkalmazkodhatnak a változó környezethez és a váratlan helyzetekhez. Lényegében az lenne a cél, hogy ezek a rendszerek egy nap utánozni tudják az emberek tanulási képességeit.

Overcoming ’Catastrophic Forgetting’: A Leap in AI Continuous Learning

A kínai nagy nyelvi modellek rövid története

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

Wudaokou-nak hívják azt a negyedet Pekingben, amely számos egyetemnek és tudományos intézménynek ad otthont. Többek között a Beijing Academy of Artificial Intelligence (BAAI) is itt található, amely egy, a mesterséges intelligencia K+F területén aktív magán alapítású, nonprofit tudományos központ. Ehhez az intézményhez kötődik a WUDAO-nak nevezett nagy nyelvi modell megalkotása, amelyet sokan egyfajta kínai válaszként értékeltek a nagy nyelvi modellek kibontakozó forradalmi fellendülésére. A történet 2018-ban kezdődött, ugyanis ekkor jelent meg a Google BERT (Bidirectional Encoder Representations from Transformers) nyelvi modelljét bemutató tanulmány, melynek hatására számos kínai vállalat újult erővel kezdte fejleszteni saját modelljeit. Ezek teljesítménye sokszor elmaradt a média által sugallt paraméterektől, érdekes volt ugyanakkor látni, hogy számos kínai egyetem, cég, intézmény, tudományos intézet dolgozik a nagy nyelvi modellek területén szoros együttműködésben. Mindazonáltal alapvető probléma volt az elégtelen hardver-kapacitások visszatartó hatása, illetve a kockázati tőkebefektetők tartózkodó viselkedése. Az OpenAI ChatGPT 3.5-ös rendszerének megjelenése gyökeres változást eredményezett a kínai fejlesztők körében is, mivel rájöttek, hogy az általuk létrehozott modellek nem érik el a legjobb nemzetközi rendszerek szintjét. A cikk a WUDAO modell fejlesztésének történetét tárgyalja, amely egyfajta válasz volt a ChatGPT 3.5-re. Az elnevezés kínai nyelven felvilágosodást jelent, de játékosan a Wudaokou negyed nevére is visszautal. A fejlesztésekhez nagymértékben hozzájárult a pekingi önkormányzat támogatása is, melyet a kínai kockázati tőke is követni kezdett.

The Wudaokou Origins of China’s Large Models

A szintetikus adatok új trendje

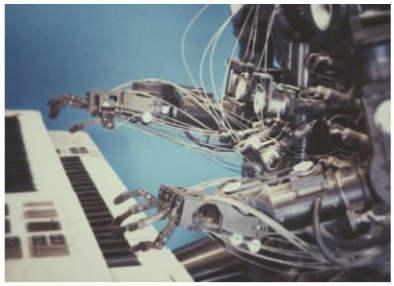
Míg a legtöbb MI-modell ember által készített adatokra épül, néhány vállalat hamarosan olyan adatokat kezd használni – vagy próbálja kitalálni, hogyan használhatná fel ezeket –, amelyeket maga a mesterséges intelligencia hozott létre. Az így előállított adatokat szintetikus adatoknak nevezik és a nagy nyelvi modellek (Large Language Model) betanítása érdekében alapvetően két okból jöhet szóba az alkalmazásuk. Egyrészt a most alkalmazott adatfeldolgozási szisztéma rendkívül drága. Másrészt jelenleg a chatbotokat, például az OpenAI ChatGPT-jét és a Google Bardját működtető LLM-ek betanítása elsősorban az interneten keresztül elérhető adatokkal történik, amelyről ezt követően második lépésként emberek adnak visszajelzést és javítják az információ hiányosságait. Ez nemcsak drága folyamat, hanem azzal a problémával is meg kell birkóznia, hogy a legfejlettebb nagy nyelvi modellek tanításához lényegében már most felhasználják az összes ténylegesen rendelkezésre álló, ember által létrehozott adatot, amit azt jelenti, hogy a még erősebb rendszerek létrehozásához szinte biztosan többre lesz szükség. A szintetikus adatok használata olcsóbb is lenne és az elérhető adatok mennyisége sem lenne korlátozva. A szintetikus adatok használatának van azonban árnyoldala is. Az MI által generált adatok megbízhatósága könnyen megkérdőjelezhető, mivel még az ember által létrehozott anyagokon kiképzett MI-rendszerek is jelentős ténybeli hibákat követnek el. Már most is vannak vállalatok - például a Cohere -, amelyek

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

szintetikus adatokon képezik ki rendszereiket, csak nem verik nagy dobra ezt a tényt és sokasodnak a jelei annak, hogy az OpenAI is hasonlóra készül. Ugyanakkor léteznek már olyan startupok is, amelyek arra szakosodtak, hogy szintetikus adatokat adjanak el más cégeknek.

AI Developers are Quietly Training AI Using AI-Generated Data

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



Működésben

A Harvard Egyetem, népszerű programozási kurzusában egyenesen ösztönzi majd az MI használatát

A mesterséges intelligencia rendszerek hatalmas fejlődése nem hagyja érintetlenül az oktatás világát sem. A közelmúltban a Harvard Egyetem bejelentette, hogy a rendkívül népszerű informatikai kurzusuk, a Computer Science 50 részét képező programozási modulban ezentúl mesterséges intelligenciát is alkalmazni fognak és a diákok egy külön erre a célra betanított chatbotot használhatnak majd a tanuláshoz. Idén ősztől tehát a hallgatóknak a mesterséges intelligencia segíthet megtalálni az általuk írt kódokban lévő hibákat, emellett visszajelzést ad a tervezett diákprogramokról, elmagyarázza számukra az ismeretlen kódsorokat és hibaüzeneteket, valamint egyéni kérdésekre is tud majd válaszolni. A kurzus oktatói kísérleteznek a GPT 3.5 és a GPT 4 modellekkel is. David J Malan, a kurzus egyik oktatója szerint azt remélik, hogy az MI segítségével megközelítőleg minden egyes hallgatóra fog jutni egy tanár, mivel a kurzus által biztosított eszközök segítségével a hallgatók a hét minden napján, 24 órában támogatást kaphatnak, és olyan ütemben és stílusban tanulhatnak, ami egyénileg a legmegfelelőbb számukra. Malan hangsúlyozza, hogy az általuk fejlesztett eszköz szellemiségében hasonló lesz a ChatGPT-hez és a GitHub Copilothoz, de az lesz a célja, hogy a hallgatókat a megoldás felé terelje, ahelyett, hogy egyenesen megadná nekik a választ. Bár az oktatók eddig is használtak MI szoftvert a hallgatók által írt programok osztályozásához, az értékelés folyamata így is időigényes volt. Most abban bíznak, hogy a mesterséges intelligencia teljes oktatási folyamatba való bevezetésével több idejük maradhat a hallgatókkal való tartalmasabb, személyesebb kapcsolatokra.

Harvard University to encourage AI use in popular Computer Science coding module

A Meta különféle személyiséggel rendelkező chatbotokat fejleszt; cél a felhasználók megtartása

A Facebook tulajdonos Meta egy sor különböző személyiséget megformáló, mesterséges intelligenciával működő chatbot bevezetésére készül, amelyekkel a közösségi médiaplatformok iránti elkötelezettséget próbálja növelni a felhasználók körében. A Mark Zuckerberg által vezetett technológiai óriás jelenleg olyan chatbotok prototípusait tervezi, amelyek emberszerű beszélgetéseket tudnak folytatni a Facebook közel 4 milliárd felhasználójával. A személyiségeknek (personas) nevezett chatbotok eltérő karakterek formájában jelennének meg: egy változat például Abraham Lincoln

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

amerikai elnököt személyesítené meg, egy másik pedig egy szörfös stílusában adna utazási tanácsokat. A chatbotok bevezetése része annak a törekvésnek, amelyet a vállalat a meglévő felhasználók megtartásáért és az új felhasználók megszerzéséért folytat. A beszélgető robotok nem csupán a vállalat szolgáltatásaihoz kapcsolják hozzá a felhasználókat, hanem egyúttal hatalmas mennyiségű olyan új adatot is gyűjthetnek az interakciók során, amelyek segíthetnek a cégnek relevánsabb tartalommal és személyre szabott hirdetésekkel megcélozni a potenciális ügyfeleket. A megszemélyesített chatbotok használatával más vállalatok is kísérleteznek. Az egymilliárd dollárra értékelt Character.ai nevű start-up nagy nyelvi modelleket használ, hogy chatbotjai olyan személyek stílusában tudjanak beszélgetni, mint például a Tesla vezérigazgatója, Elon Musk vagy a Nintendo játék Mario nevű figurája. Zuckerberg szerint az MI használata nem egyetlen rendszer használatát fogja jelenteni, hanem számos MI-ágenssel tud majd a felhasználó kapcsolatba lépni.

Meta prepares chatbots with personas to try to retain users

Valós időben fordítja az 5000 éves ékírásos táblákat egy új MI-szoftver

Bár hatalmas mennyiségű akkád nyelvű dokumentum maradt fenn ékírásos agyagtáblákon, számos probléma akadályozza ezek feldolgozását. Alapvető nehézséget okoznak a szövegek nyelvi-kulturális sajátosságai, ezért csak nagyon kevés szakember képes értelmezni a fennmaradt szövegeket. Az ókori akkád szövegek fordítása kétlépéses folyamat. Először is, a tudósoknak át kell írniuk az ékírásos jeleket, vagyis fogják az ékírást és a célnyelv hasonló hangzású fonetikáját használva átírják azt. Az ősi könyvtárak megfelelő lefordítása még ilyen nyelvi gazdagság mellett sem kis teljesítmény. A már említett kihívásokon túl nehézségek forrása az akkád nyelv polivalens jellege is. Ez azt jelenti, hogy az ékírásos jeleknek több, egymástól eltérő olvasata is lehet, attól függően, hogy az egyes írásjegyek egy adott mondatban milyen szerepet töltenek be. Ennek a helyzetnek a megváltoztatása érdekében egy régészekből és informatikusokból álló multidiszciplináris csapat kifejlesztett egy olyan mesterséges intelligencia rendszert, amely szinte azonnal képes lefordítani az akkád nyelvű ékírásos szövegeket és olvashatóvá tenni az 5000 éves táblákban őrzött történelmi feljegyzéseket. Bár a kezdeti eredmények ígéretesek, sok még a tennivaló. A tesztmondatok egy részét rosszul fordították le és más MI-modellekhez hasonlóan ez az MI-rendszer is hajlamos a hallucinációkra: olyan válaszokat ad, amelyeknek nincs kapcsolatuk a forrással. Összességében az MI-modell akkor működik a legjobban, ha rövid és közepes hosszúságú mondatokat fordít le. Jobban teljesít az olyan strukturált, kötöttebb műfajokkal, mint például a királyi rendeletek és közigazgatási dokumentumok és nehezebben boldogul a szépirodalmi műfajokkal: mítoszokkal, himnuszokkal és próféciákkal. A kutatók javítani szeretnék a fordítás pontosságát is, nagyobb adatkészleteket használva a rendszer tanításához. Azt remélik, hogy idővel az MI-modelljük virtuális asszisztensként szolgálhat az emberi tudósok számára. Az MI-rendszer gyorsan el tudja majd készíteni a szövegek nyersfordítását, míg a tudósok a nyelvtörténeti és a történelmi ismereteik segítségével finomíthatják azt.

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

New AI translates 5000-year-old cuneiform tablets instantly

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