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

2023 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

Egyre több nagyvállalat építi le „etikus MI” elköteleződését – ahogy az üzlet kezd lendületet venni

Az elmúlt másfél-két évben mindenhol, de különösen a nyugati társadalmakban egyre inkább az etikai kérdések kerültek előtérbe a mesterséges intelligencia alkalmazásával összefüggésben, már-már háttérbe szorítva a tisztán műszaki megközelítéseket. Most egyre több jel utal arra, hogy a morális szempontok előnyben részesítése addig tart, amíg az üzletben felsejlő egyre hatalmasabb profitok árnyékot nem vetnek erre a példamutató megközelítésre. A hírek szerint ugyanis a technológiában érdekelt óriások, a Microsoft, a Twitter és a Twitch egyre-másra számolják fel az MI etikai problémáinak vizsgálatára korábban létrehozott részlegeiket, útilaput kötve a technológia negatív hatásait boncolgató munkatársak lábára. Ezt a tendenciát jól példázza az Amazon esete. A gigacég, észlelve, hogy az általa birtokolt streaming platform, a Twitch egyre több elfogultsági problémával küzd, tavaly külön részleget hozott létre a rasszista, szexista és más problémákat mutató ajánló algoritmusának vizsgálatára és korrigálására. A nagy lendülettel munkába állított felelős mesterséges intelligenciáért létrehozott csapat munkatársainak a zömét azonban néhány héttel ezelőtt szélnek eresztették, az etikai kérdéseken dolgozók maradékát pedig más területekre csoportosították át. Hasonló lépésre szánta el magát tavaly novemberben a Twitter, a Microsoft pedig idén januárban számolta fel etikai és társadalmi kérdésekkel (Ethics and Society) foglalkozó csapatát, amely a felelős mesterséges intelligenciával kapcsolatos kutatásokat vezető csoportok egyike volt a vállalatnál.

Companies ax ’ethical AI’ teams, just as the tech begins to take off

A kiberbűnözők is kezdik felfedezni a ChatGPT-ben rejlő lehetőségeket

Mint ismeretes 2022 november végén az OpenAI kiadta a ChatGPT-t, a vállalat nagy nyelvi modelljére épülő felületét, amely azonnal hatalmas érdeklődést váltott ki a mesterséges intelligencia és annak lehetséges felhasználási területei iránt. Sorra jelennek a hírek arról, hogy milyen sokféle feladat elvégzésére alkalmazható ez az igen népszerű platform. Hamar kiderült azonban, hogy a ChatGPT egyik felhasználási területe, a kódgenerálás segítségével az informatikailag kevésbé képzett emberek is könnyedén tudnak kibertámadásokat indítani. A Check Point Research (CPR), kiberbiztonsági cég kísérleti jelleggel leírta, hogy a ChatGPT segítségével hogyan lehet végrehajtani egy teljes fertőzési folyamatot, a meggyőző adathalász e-mailek létrehozásától az angol nyelvű parancsok fogadására képes fordított parancsértelmező

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

(reverse shell) futtatásáig. A cég kutatói hamar választ kaptak arra a kérdésükre, hogy mindez csak hipotetikus fenyegetés-e vagy már léteznek olyan kiberbűnözők is, akik rosszindulatú célokra használják az OpenAI technológiáját. Több jelentős illegális hacker közösség tevékenységének elemzése során világossá vált, hogy már megjelentek az első olyan internetes csalók, akik visszaélnek a chatbot képességeivel. Az esetek egy része egyértelműen azt mutatta, hogy az OpenAI-t használó kiberbűnözők közül sokan egyáltalán nem rendelkeznek szoftverfejlesztői ismeretekkel, de csak idő kérdése, hogy a kifinomultabb, felkészültebb kiberbűnözők mikor kezdik el használni az MI-alapú eszközöket saját céljaikra.

OPWNAI: Cybercriminals Starting to Use ChatGPT

A gépi tanulás fejlesztése az erős gazdasági verseny közepette – társadalmi kockázatok és előnyök

A mesterséges intelligencia kutatása a huszadik században kezdődött, de a mesterséges neurális hálózatok modern modelljeit csak 2012 után kezdték komolyabban alkalmazni a gépi tanulás kutatása során. Az elmúlt tíz évben mind a számítógépes látás, mind a természetes nyelvi feldolgozás terén egyre jobb eredmények születtek. A mesterséges intelligencia fejlődése az utóbbi időben rohamosan felgyorsult, azonban még mindig nyitott kérdés, hogy mindez milyen előnyökkel és kockázatokkal jár és ez utóbbiakat hogyan lehet kezelni. A cikk három fő kockázatot tárgyal, amelyek mindegyike az MI rendszerek biztonsági problémái közé tartozik. Ezek a következők: a mesterséges intelligencia rendszerek összehangolásának problémája, a mesterséges intelligenciával való visszaélés problémája és az információ ellenőrzésének kérdése. Ez utóbbi arra utal, hogy ezek a rendszerek képesek új információt létrehozni, és hamarosan nehéz lesz különbséget tenni a tények és a fikció között. A szerző rövid történeti áttekintés után kitér az előnyökre és a kockázatokra, majd arra a következtetésre jut, hogy a kockázatok potenciálisan mérsékelhetők a megbízható mesterséges intelligencia megteremtésére irányuló szoros együttműködés és a tudatosság növelése révén. Ezeknek az elveknek az alapján hozta létre az Európai Unió a TAILOR nevű kutatóközpont hálózatot, amelynek küldetése az emberközpontú megbízható mesterséges intelligencia megvalósítása és annak elősegítése, hogy Európa globális példaképpé váljon a felelős mesterséges intelligencia terén.

The rapid competitive economy of machine learning development: a discussion on the social risks and benefits

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

Trendek

Noam Chomsky: A ChatGPT hamis ígérete

Noam Chomsky és társai cikke kritikai vélemény a nagy nyelvi rendszerek gyors fejlődése kapcsán megfogalmazódó túlzásokra. Ma a mesterséges intelligencia terén elért forradalminak tűnő fejlődés aggodalomra és optimizmusra egyaránt okot ad. Optimizmusra azért ad okot, mert az intelligencia a problémák megoldásának eszköze, aggodalomra pedig azért, mert a gépi tanulás a nyelv és a tudás alapvetően hibás felfogását támogatja. Az emberi elme, eltérően a ChatGPT-től és hasonló rendszerektől, nem mintaegyeztetésre szolgáló statisztikai motor, amely több száz terabájtnyi felhalmozott adat alapján egy tudományos kérdésre a legvalószínűbb választ adja. Az emberi elme meglepően hatékony, sőt elegáns rendszer, amely kis mennyiségű információval dolgozik; nem arra törekszik, hogy nyers összefüggésekre következtessen az adatpontok között, hanem magyarázatokat alkot. Az igazi intelligencia képes az erkölcsi gondolkodásra is, azaz elménk egyébként határtalan kreativitását etikai elvekkel korlátozzuk. Ahhoz, hogy hasznos legyen, a ChatGPT-nek képesnek kell lennie újszerű kimenet létrehozására; ugyanakkor ahhoz, hogy a legtöbb felhasználó számára elfogadható legyen, kerülnie kell az erkölcsileg kifogásolható tartalmakat. A ChatGPT és más gépi tanulási csodák programozói folyamatosan azzal küzdenek, hogy elérjék ezt a fajta egyensúlyt. A szerzők szerint ezek a rendszerek a beépített korlátok ellenére könnyen rábírhatók arra, hogy erkölcsileg káros tartalmakat hozzanak létre és lényegüknél fogva képtelenek egyensúlyt teremteni a kreativitás és a korlátok között. Ezek a technológiai rendszerek hasznosak lehetnek bizonyos szűk területeken, de messze állnak a valódi mesterséges intelligenciától, és még távolabb attól, hogy bármi hasznosat tudjanak mondani az emberi elme működéséről.

Noam Chomsky: The False Promise of ChatGPT

Az MI fejlesztési verseny holisztikus megközelítése

Az Egyesült Államok számára nem idegen a technológiai fegyverkezési verseny, hiszen a hidegháborús korszak jelentős részét határozta meg a Szovjetunióval való versengése az űrkutatás területén. Napjainkban az Egyesült Államok nagy riválisa Kína, a harc pedig újabban a mesterséges intelligencia alapú technológiák fejlesztéséért folyik, amely némi hasonlóságot mutathat az űrverseny során lezajlott folyamatokkal. A szakértők szerint azonban a fegyverkezési verseny keretrendszerének alkalmazása a mesterséges intelligencia fejlődésének tanulmányozására nem ragadja meg a technológia körüli globális dinamikát. Az MI alapú

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

innovációkat ugyanis - szemben az űrkutatással - alacsonyabb belépési korlátok, demokratizált hozzáférés és a magánszektor túlsúlya jellemzik. A mesterséges intelligencia jelenlegi előretörésére hasznosabb történelmi analógia lehet az ipari forradalom időszaka. Az akkor végbement változások ugyanis olyan technológiák kombinációján alapultak, amelyben az egyes összetevők kölcsönösen támogatták és felerősítették egymást, hasonlóan ahhoz, amit ma látunk az MI és az adatok, a dolgok internete (IoT), a mobil eszközök, valamint a vezetékes és vezeték nélküli hálózatok esetében. Ahelyett, hogy azt kutatnánk, ki lesz a mesterséges intelligencia verseny győztese, érdemesebb azt vizsgálni, hogy az MI egyes részterületein az adott ország vezető szerepre való törekvése mennyire kritikus fontosságú vagy kívánatos.

Who's Winning the AI Race? It's Not That Simple.

Az MI-chatbotok meghekkelése lett az informatikusok új szórakozása

A GPT alapú mesterséges intelligencia chatbotok működésének lényege, hogy a program a felhasználók által megadott szöveges kérdésekre, utasításokra (prompt-okra) az adatbázisában tárolt információk alapján válaszokat generál. A rendszerek úgy lettek megalkotva, hogy számos beépített biztonsági megoldás ügyel arra, hogy a program megtagadja az olyan kérések végrehajtását, amelyek sérthetik a tartalom moderálására vonatkozó irányelveiket. Azonban bizonyos speciális utasítások segítségével ezek a biztonsági korlátok megkerülhetők és a program olyan kérdésekre is válaszolhat, amelyekre egyébként nem megengedett. Ezeket a speciális kéréseket nevezik "jailbreak"-nek. Ma már egyre többen foglalkoznak olyan módszerek kidolgozásával, amelyek alkalmasak a nagy teljesítményű chatbotok potenciális biztonsági réseinek felderítésére. A jailbreaking tehát nem más, mint a ChatGPT és a hozzá hasonló MI modellek etikai biztosítékainak feltörése. Miközben a módszer veszélyes információk, gyűlöletbeszéd vagy egyszerű valótlanságok generálását is eredményezheti, a technikai kihíváson túl hasznos feladatot is ellát: rávilágít a mesterséges intelligencia modellek korlátaira. Az OpenAI vállalat maga is ösztönzni kívánja ezeknek a sérülékeny pontoknak a kiszűrését, ezért nemrég bejelentette Bug Bounty elnevezésű hibakereső programját, melynek keretében pénzjutalmat ajánlott fel azoknak a felhasználóknak, akik segítenek megtalálni a ChatGPT hibáit és biztonsági hiányosságait.

Jailbreaking AI Chatbots Is Tech's New Pastime

A pénzügyi szektor sajátos igényeihez fejlesztett nagy nyelvi modellt a Bloomberg cég

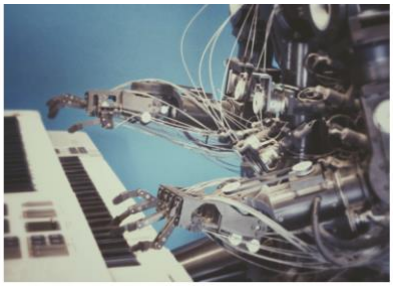
A természetes nyelvfeldolgozás széles körű használata régóta jelen van a pénzügyi technológiák számos területén. Az alkalmazások között a szentimentelemzéstől a névelem-felismerésen át a kérdésértelmező és válaszkereső rendszerekig sok minden megtalálható, ami szerepet kaphat a pénzügyi szektorhoz kapcsolódó tevékenységek támogatásában. Bár a nagy nyelvi modellek sok területen bizonyították már hatékonyságukat, a szakirodalom mindeddig még nem számolt be kifejezetten a pénzügyi ágazat számára kifejlesztett LLM-szoftverekről. A Bloomberg

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

hírügynökség és John Hopkins Egyetem kutatócsoportja által publikált tanulmány most bemutatta az első ilyen fejlesztést, a BloombergGPT néven megvalósított, mintegy 50 milliárd paraméteres pénzügyi nyelvi modellt, amelyet a szerzők ismertetése alapján rendkívül kiterjedt pénzügyi adattömegen képeztek ki. A kutatók a Bloomberg óriási adatkészleteire alapozva, egy 365 milliárd adatelemből (token) álló adatbázist építettek fel az algoritmus számára, amelyet további 345 milliárd - nyilvános adathalmazokból származó - adatelemmel egészítettek ki. A kutatók szerint a vegyes adatkészleteken betanított modell a pénzügyi feladatokban jelentős mértékben felülmúlja a már meglévő, hasonló modelleket, miközben az általános LLM benchmarkokon is azokéval egyenértékű vagy jobb teljesítményt nyújt.

BloombergGPT: A Large Language Model for Finance

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



Működésben

A tanácsadói iparág új chatbotja 4000 jogász munkáját segíti majd egy multinál

A Price Waterhouse Coopers vállalat nemrégiben chatbot szolgáltatást vezetett be jogászai számára és ezzel a kezdeményezéssel csatlakozott azon szolgáltató cégek sorához, amelyek mesterséges intelligencia technológiát használnak hatékonyságuk növelésére. A PwC a projektet a Harvey MI startup céggel kötött 12 hónapos partnerség keretében valósítja meg és több mint 100 országban a vállalat mintegy 4000 jogásza fog hozzáférni a technológiához. A chatbot alkalmazásának célja, hogy segítsen felgyorsítani a munkát számos területen: így a due diligence átvilágítási eljárásoktól kezdve a szabályozási megfelelésen át a jogi tanácsadói szolgáltatásokig több ponton is javíthatja a hatékonyságot. A cég a közeljövőben adótanácsadói tevékenységére is szeretné kiterjeszteni a szolgáltatást. A Harvey az OpenAI és a ChatGPT technológiájára épül és az OpenAI Startup Fund támogatásával jött létre. A ChatGPT, a Microsoft által támogatott OpenAI MI-chatbotja tavaly nemzetközi szenzációvá vált, köszönhetően annak a képességének, hogy szinte bármilyen kérdésre képes látszólag kifinomult, mégis közérthető válaszokat adni. A nagy nyelvi modelleken alapuló technológia különösen azok számára jól hasznosítható, akiknek nagy mennyiségű szöveget kell létrehozniuk és elemezniük. Ma már egyre több üzleti szolgáltató cég kísérletezik azzal, hogy egyes feladatainak elvégzését generatív MI rendszerek bevonásával gyorsítsa fel.

PwC Introduces AI Chatbot for 4000 Lawyers to Speed Up Work

A Google cég most megjelent mesterséges intelligenciája a cég népszerű levelező rendszerét emeli magasabb szintre

Tavaly november óta, amikor a ChatGPT berobbant a köztudatba, egyre gyorsuló ütemben folyik a mesterséges intelligencia chatbotok kifejlesztésére irányuló verseny a világ technológiai óriáscégei között. Idén tavasszal a Google is előrukkolt saját generatív mesterséges intelligencia eszközeivel, melyeknek egy részét elsőként a már széles körben elterjedt alkalmazásaiba, a Gmail levelező rendszerébe és a Google Docs (Google Dokumentumok) dokumentumkezelő szolgáltatásába építve teszi elérhetővé. A vállalat a közeljövőben új MI funkciókkal kívánja bővíteni üzleti termékeit is, elsősorban a Google Cloudot, valamint a fejlesztői számára olyan új, mesterséges intelligenciával kiegészített alkalmazásprogramozási felületet biztosít, amely megkönnyíti az alkalmazások tesztelését. A Gmail és a Google Docs új fejlesztéseinek segítségével a felhasználónak elég begépelnie néhány kulcsszót vagy instrukciót a kívánt

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

témában és az alkalmazások automatikusan létrehoznak egy, már megformázott, jól használható szöveget. Az új funkciókat egyelőre csak korlátozott számú „megbízható” felhasználó tesztelheti, de amikor a Gmail és a Google Docs mintegy három milliárd felhasználója számára is elérhetővé válnak ezek az újítások, minden eddiginél több emberhez fognak eljutni a mesterséges intelligencia legújabb generációs technológiái.

Google Is About to Unleash AI for Gmail and Google Docs

Hogyan forradalmasíthatja a Mesterséges Intelligencia a diplomáciát?
A diplomáciai módszerek nem sokat változtak a 19. század óta, de az innovációk áradata még ezen a hagyománytisztelő területen is egyre jobban érezteti hatását. A nemzetközi kapcsolatok menedzselésének legfontosabb funkciója ma is a tárgyalás, amelynek bár alapvető struktúrája nem változott, a hozzá kapcsolódó folyamatokon mindinkább nyomot hagynak a digitalizáció vívmányai. Nathaniel Fick szerint, aki az Egyesült Államok Külügyminisztériumának a kibertérért és digitális politikáért felelős hivatalát vezeti, az MI-alapú ChatGPT által generált tájékoztatók például ma már „minőségileg elég közel állnak” a munkatársai által készítettekhez. A nagy nyelvi modellek fejlődésével a mesterséges intelligencia rendszerek gyorsabban tudnak majd információkat keresni és összegezni, mint egy csapat ember, így jobban felkészíthetik a diplomatákat a tárgyalásokra. Bár ezek a rendszerek bizonyos fokú emberi felügyeletet fognak igényelni, egy többoldalú tárgyaláson a szövetséges felek saját MI rendszereik segítségével össze tudják hangolni álláspontjaikat. Ahogy egyre több fél fejleszti ki saját MI-alapú szoftverét a diplomáciai munka támogatására, a tárgyalásokon kulcsszerepet kaphatnak az MI "alkubotok" (hagglebot), azaz az olyan számítógépek, amelyek adott kompromisszumok és érdekek alapján képesek meghatározni az optimális megállapodást. Az egyre kifinomultabb MI-rendszerek rendszerek felforgathatják a technológiáról alkotott elképzeléseinket, lehetővé téve, hogy az MI ne csak egy eszköz, hanem önálló szereplő legyen a nemzetközi kapcsolatokban.

How AI Could Revolutionize Diplomacy

Virtuális románc fűti Kína MI-forradalmát

A chatbot-társaikhoz érzelmileg kötődő kínaiak versenyelőnyt biztosíthatnak nemzetüknek a mesterséges intelligencia fejlesztések terén. A virtuális barátokkal való csevegés lehet az egyik módja annak, hogy a kínaiak mindennapi életének részévé váljon az MI technológiák használata. Míg más országokban a mesterséges intelligencia etikai, morális kérdéseiről és hosszú távú társadalmi hatásairól folyik a vita, Kínában sok ember éppen a kedvenc chatbotjával „randevúzik”. A Microsoft kutatói által megalkotott kínaiul beszélő, érzelmeket szimuláló chatbotnak, Xiaoice-nak 2022-ben már több mint 660 millió felhasználója volt. A 2014-ben megjelent Xiaoice-t a kutatók arra tervezték, hogy "érzelmi kapcsolatot" teremtsen a felhasználókkal. A rendszer a ChatGPT-hez hasonlóan az adatokból és a felhasználói interakciókból tanulva reagál a kérésekre és utasításokra, azonban univerzális felhasználás

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

helyett sokkal inkább nyújt érzelmi támogatást. A Xiaoice azonban csak egy a kínai fogyasztók számára elérhető MI innovációk gazdag kínálatából, amelyek révén a technológia továbbfejlesztéséhez nélkülözhetetlen adattömeg és szakértelem halmozódik fel. Kína egyelőre lemaradásban van az amerikai vállalatokkal szemben az MI fejlettebb változatainak kifejlesztéséért folytatott versenyben, ugyanakkor mivel az országban a chatbotok és más MI alkalmazások társadalmi elfogadottsága egyre nő és a hatóságok is támogatják a legújabb technológiák bevezetését és elterjedését, Kína idővel döntő előnyre tehet szert azokkal az országokkal szemben, amelyek jelenleg éppen aktívan igyekeznek visszafogni e technológiák fejlődését.

Virtual Romance Is Fueling China's AI Revolution

A mesterséges intelligencia rádióállomások „bemondó-műsorvezetőjeként" szerkeszti a műsort

A nagy nyelvi modellek egyre bővülő alkalmazási lehetőségei között az egyik legújabb terület a mesterséges intelligencia alapú rádiós tartalomszolgáltatás. A clevelandi Futuri médiavállalat legújabb fejlesztése, a RadioGPT az OpenAI korábbi nyelvi modelljének, a GPT-3-nak és más technológiáknak a segítségével testre szabható tartalmakat hoz létre és forgalmaz rádióállomások számára. A RadioGPT automatizálja a rádióműsorok szerkesztését és vezetését, ezáltal kiválthatja a rádiós műsorvezetők munkáját, valamint használata a rádióműsorok számát is növelheti. A szoftver képes elkészíteni egy tökéletesen használható rádiós adásmenetet, azaz a műsor programelemeinek egymásutánját rögzítő technikai forgatókönyvet. A RadioGPT az MI-vezérelt TopicPulse tartalomkereső technológia segítségével monitorozza a közösségi médiát és több mint 250 ezer más hír- és információforrást az éppen aktuális helyi trendek és témák azonosítására, majd ehhez igazodva készíti el a konkrét rádiós adásmenetet. A RadioGPT által megírt szöveget ezután felolvashatják a csatorna saját műsorvezetői vagy mesterséges intelligencia által generált hangok, melyek lehetnek valós személyek hangjához hasonlító, MI által létrehozott hangok is. A rádióállomások egyszerre akár három virtuális szereplővel is készíthetnek műsort. A szoftver igazodik a vételkörzet, azaz a potenciális hallgatóság specifikus igényeihez és az érdeklődési körének, demográfiai jellemzőinek megfelelő híreket osztja meg. A jelenleg tesztelés alatt álló rendszer legérzékenyebb pontja annak biztosítása, hogy az algoritmus az általa begyűjtött információk alapján ne terjesszen téves információkat az adásokban.

AI is now DJing radio stations

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