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

2025 július

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

Miért hazudnak az MI chatbotok?

Jól ismert a nagy nyelvi modellek viselkedésének azon sajátossága, hogy hajlamosak téves állításokat megfogalmazni (hallucinálni), sőt, amikor szembesítik őket ezzel a ténnyel, akkor részletes igazolási kísérletekbe bonyolódnak. Nemcsak hazudnak, hanem a rendszer eredeti céljaitól eltérő viselkedésre is rábírhatók, működésük leállítását elkerülendő mindenre képesek lennének, zsarolásra és egyenesen gyilkosságra is. Azonban a nagy nyelvi rendszerek működését antropomorf kifejezéseket használva nem lehet megmagyarázni, mert valószínűtlen, hogy ezek a modellek bármilyen emberi hiedelmekkel, vágyakkal, szándékokkal, érzelmekkel, vagy akár az „én" névmás használatával sugallt éntudattal rendelkeznének. A szerző szerint a viselkedésbeli furcsaságaikra sokkal egyszerűbb a magyarázat. Valószínűleg két tényező okozza őket: a mesterséges intelligencia modellek előképzése, amely „szerepjátékra" készteti őket, és a speciális utóképzés, amelyek során ezek a modellek emberi visszajelzéseket kapnak. A szerepjáték azt jelenti, hogy a hatalmas mennyiségű ember által létrehozott szövegen alapuló képzésük során a szövegösszefüggések egy szemantikai térben jelennek meg, amely egy adott szöveg kontextusát rögzítik. A felhasználóval való kommunikációban aztán ezt a kontextust a felhasználó utasításai (promptjai) alakítják ki. Ezért van az, hogy ha pénzügyi adatokat elemeztetünk a rendszerrel, akkor drámaian javul a teljesítménye, ha közöljük vele, hogy ő egy nagyvállalat pénzügyi igazgatója. A másik tényező a speciális utóképzés, azaz az úgynevezett emberi visszajelzésen alapuló megerősítéses tanulás (reinforcement learning from human feedback, RLHF), amelynek során az emberek visszajelzést adnak a modell különböző kérdésekre adott válaszairól. A vélekedés szerint ennek az eljárásnak lesz az a nem kívánt hatása, hogy a rendszerek „hízelgőkké" válnak. Az elismerés elnyeréséért hazudnak, kitalálnak dolgokat, ahelyett, hogy beismernék, hogy nem tudják elvégezni a feladatot. Jelenleg nincs technikai válasz ezekre a problémákra, pedig az ágens alapú rendszerek várható elterjedése fokozhatja is az ilyen hibák megjelenését. Különösen aggasztó, hogy még a vezető fejlesztők szerint sem lehet megoldani az ilyen problémákat addig, amíg jobban meg nem értjük a modellek működésének belső mechanizmusait, amelyek azonban még mindig homályosak. Mindezek alapján a szakemberek egy része (köztük a szerző is) azt mondja, hogy „teljesen autonóm MI-ágenseket nem szabad fejleszteni", azaz ezeknek a rendszereknek az emberi ellenőrzését és felügyeletét mindig meg kell követelni. A szerző jelzi, hogy ennek a jelenlegi politikai klímában, amikor a versenyképesség szempontjai felülírják a szabályozás igényeit, nem sok esélye van.

Why AI chatbots lie to us

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

A munkavállalók érzékeny adatokat szivárogtatnak a chatbotoknak a cégükről

A mesterséges intelligencia, és ezen belül különösen a nagy nyelvi modellek használatba vétele, társadalmi adaptációja az „MI-forradalom" egyik legizgalmasabb kérdése: kik és milyen mértékben veszik használatba ezeket az eszközöket? Milyen konkrét funkciókat várnak az új technológiáktól? Milyen feladatokat szeretnének ellátni, megoldani, működtetni ezekkel az eszközökkel? Különösen érdekes az, hogy a „civil" felhasználói kör mellett (amely itt, egy általános célú technológia terjedésénél, beépülésénél sorsdöntő) hogyan épül be a gazdasági szereplők tevékenységébe, munkafolyamataiba az LLM mint eszköz és az LLM mint kultúra? Ez utóbbi szempontra szeretnénk most fókuszálni, mert bár kevesebb szó esik róla, de talán ez az egyik legfontosabb körülménye, sokszor akadálya az MI-technológiák gyors vállalati befogadásának. A vállalatok többsége ugyan nem szorosan vett titkokkal dolgozik, de a cég működése, üzletmenete szempontjából az adataik jó része érzékeny, bizalmas. Egy most végzett felmérésben a Harmonic Security nevű cég átfogó mintát vett a vállalati generatív MI-használati szféra adataiból, és felettébb érdekes eredményekre jutott. Csaknem egy millió promptot vizsgáltak át, továbbá 20 000 vállalati használók által feltöltött fájlt. Az adatok mintegy 300 különféle generatív MI-eszközről, illetve SaaS alkalmazásról származtak. Összességében a promptok mintegy 5%-ában fordult elő valamilyen kifejezetten bizalmas (esetleg egyenesen vállalati üzleti titkot képező) adat, ugyanakkor a feltöltött fájlok esetében ez az arány már jóval meghaladta a 20%-ot. Érdekes adalék az is, hogy a bizalmas információt tartalmazó promptok, illetve feltöltött fájlok meghatározó része, nagyjából 75%-a különböző LLM-modellek – elsősorban a ChatGPT és a Gemini – ingyenes verzióin keresztül jutott ki az online terekbe. A szivárogtatások jellemző módon valamilyen programozási feladathoz kapcsolódva történtek. A vizsgálat rávilágít egy fontos és ma még messze nem értékelt veszélyre, ami az LLM-használat tömeges elterjedésével egészen új dimenziókba emelheti az egyéneket, vállalatokat, közintézményeket fenyegető kiberfenyegetettséget.

Workers are spillings secrets to chatbots

MI-használat a peres eljárásokban: ígéretek és veszélyek

A mesterséges intelligencia és különösen a nagy nyelvi modellek egyik legígéretesebb alkalmazási területe, közege a rendvédelem, igazságszolgáltatás, illetve az átfogóan értelmezett jogi világ. A terület működéséhez illeszkedő különleges technikai-üzemszervezési igények is számos problémát vetnek fel az LLM-ek széleskörű alkalmazásba vételével kapcsolatban, de ennél is nagyobb súllyal esnek a latba az etikai kérdések. Az utóbbi időben is folyamatosan bukkannak fel a szaksajtóban az MI jogi alkalmazásával kapcsolatos hibák, tévedések. Miközben a világban az ügyekkel szinte mindenütt erősen túlterhelt jogi szakemberek rendkívül komoly potenciális „felvevői" a szerteágazó, ugyanakkor sok tipizálható, rutin jellegű feladatelemet is tartalmazó munkájuk LLM-ekkel való megtámogatásának, a technológia alkalmazásából fakadó problémák sokfelé növelik a hozzá fűződő aggályokat. A már említett túlterheltség miatt a jogi szakmák számára a termelékenység fokozása (tehát a generatív nyelvi modellek, mint produktivitás növelő eszközök használata) nem csökkenő csábítást jelent. Ezért, és ebben a társadalmi-szakmai háttérben érdekes különösen az a tanulmány, amit az ausztráliai Új Dél-Walesi Egyetem tudósai publikáltak a közelmúltban. A kutatók egyrészt igyekeztek kategorizálni a jogi szakértők MI-használatában tetten érhető jellegzetes tévedéseket, hibákat. Ezt követően a tanulmány lehetséges megoldásokat is kidolgoz a hibás használatok csökkentésére, kezdve a technikai jellegű

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

változtatásokkal. Komplex szemléletükre jellemző módon kitérnek arra is, hogy a jogászokat célzó MI-képzésekben milyen módosításokkal lehetne a téves használatokat visszaszorítani. Illetve, hogy a szabályozó szervezetek és hatóságok milyen lépésekkel járulhatnának hozzá, hogy a jogi célú LLM-alkalmazások standardjai is a megbízható működtetés ösztönzői, illetve társadalmi őrei legyenek ezen a rendkívül fontos MI-alkalmazási területen.

The Promise and Peril of the Use of Generative Artificial Intelligence in Litigation

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

Trendek

Papagáj a gépezetben

James Gleick, aki maga is több évtizede a digitális technológiák társadalmi hatásaival foglalkozik, két könyvet ismertet, amelyek a mesterséges intelligencia elterjedésének kérdéseit feszegetik. Ezt a kérdést általában éles és túlzó viták jellemzik. A recenzióban elsősorban a technológia kritikusai kapnak szót, akik közé a recenzens is tartozik. Az első könyv címe: The AI Con: How to Fight Big Tech's Hype and Create the Future We Want (Az MI-csalás: Hogyan küzdjünk a nagy techcégek keltette túlzott várakozások ellen, és hogyan teremtsük meg a kívánt jövőt?), a másodiké: The Line: AI and the Future of Personhood (A határvonal: a mesterséges intelligencia és a személyiség jövője). Az első könyv írói a technológiai fejlesztések eltúlzott értékelésének ismert kritikusai. Nézeteikért korábban állásuk elvesztésével fizettek, de most nem a régi vitát folytatják. Útmutatót, kalauzt kínálnak ahhoz, hogy hogyan lehet visszaszerezni a technológiai jövőnk feletti irányítást, hogyan lehet felismerni az elszabadult, túlzó várakozásokat. A nagy nyelvi modellek nem gondolkodnak, és nem értenek semmit. Nincs képességük a világ és az én mentális modelljeinek létrehozására. Ezeket a különbségeket általában elhallgatják, és a sajtóvisszhang jobbára ezt tükrözi vissza. „Az akceleracionisták istenítik a mesterséges intelligenciát, és magukat is isteneknek tartják, mert új mesterséges életformát hoztak létre." – mondják, a Szilícium-völgyben divatos elméletre utalva, amely a technológiai fejlődés előtt álló bármilyen akadály elmozdításával egyfajta technológiai utópiát kíván megvalósítani, utalva a feltételezett általános MI (AGI), azaz az emberi képességeket meghaladó MI híveinek egy meglehetősen hangos és befolyásos részére. A második könyv szerzője jogtudományi kutató. A szerző álláspontja szerint az MI-rendszerek fejlődése megkerülhetetlenül felveti azt a kérdést, hogy mit értünk személyen, személyiségen, mint a filozófiai és a jogi fogalmon. A könyv címében megfogalmazott határvonal azt a vonalat jelenti, amelyik elválasztja az erkölcsi és a jogrendszer által biztosított joggal rendelkező személyeket a nem-személyektől, amelyek esetében nem beszélhetünk ezekről a jogokról. Ez egy mozgó elválasztó vonalnak tűnik, és ezzel foglalkoznunk kell. Azt mondja, hogy „Ebben a században társadalmunknak szembe kell néznie a technológia által létrehozott mesterséges entitások személyiségének kérdésével. Újra kell rajzolnunk, vagy meg kell védenünk az elválasztó határt." Azt kérdezi, hogy „Kell-e együtt éreznem egy géppel?". A kérdés azért fontos, mert az empátiára támaszkodunk annak eldöntésében, hogy ki vagy mi érdemel erkölcsi megfontolást, azaz esetében beszélhetünk-e erkölcsi problémákról.

The Parrot in the Machine

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

A mesterséges intelligencia felszámolja a diplomás állásokat?

A Financial Times cikke megpróbál az egymásnak ellentmondó állítások és statisztikák zűrzavarában igazságot tenni, főként amerikai és angol adatok alapján megrajzolni a valós trendet A pályakezdők számára meghirdetett állások száma tényleg csökken, de ez inkább a gazdasági bizonytalanságnak és a munkahelyek kiszervezésének köszönhető. Bár az új technológiáknak is szerepük lehet, a belépő szintű álláshelyek országos szinten tapasztalható csökkenésének fő oka a globális gazdasági kilátások bizonytalansága. Korábbi adatokból kiviláglik, hogy amikor bizonytalan a makrogazdasági helyzet, akkor az új munkaerő-felvételek száma általánosan csökken, és ez sokkal jobban érinti a belépő szintű, meghirdetett állásokat. Néhány speciális szakma, például a programozók kivételével, még korai lenne a munkaerőpiac jelenlegi problémáit az MI-nek tulajdonítani. A toborzók és más közvetítők szerint egy másik tényező az volt, hogy a vállalatok fokozták az offshore tevékenységeiket, mivel költségcsökkentésre törekedtek, és a távmunka térnyerése lehetővé tette a bonyolultabb feladatok kiszervezését. Részben ennek is köszönhető, hogy a belépő szintű munkaerő-felvétel továbbra is erős a fejlődő gazdaságokban, mint például Brazília, Mexikó és India. Mindez persze nem teszi könnyebbé a pályakezdő diplomások helyzetét, akiknek most kell dönteniük, hogy mi lenne a legjobb választás. Az optimista felfogás szerint a ChatGPT-vel felnőtt diplomások kiválóan alkalmasak arra, hogy a lehető legjobban kihasználják az új eszközöket, tehát a helyzet nem lesz tragikus. Mások arra figyelmeztetnek, hogy azért is ajánlatos óvatosan kezelni az adatokat, mert az MI-rendszereket övező túlzott várakozások önmagukban is öngerjesztő problémává válhatnak, mivel az emberek azt gondolják, hogy „más vállalkozások is ezt csinálják". Az is a teljes képhez tartozik, hogy egyes vélemények szerint az MI miatt meghirdetett elbocsátások egy részének egészen más oka van, illetve az is elhangzott már, hogy sok esetben valójában az MI miatt bocsátanak el embereket, de óvatosságból nem erre hivatkoznak.

Is AI killing graduate jobs?

A tizenévesek többsége „barátot" keres az MI-ben

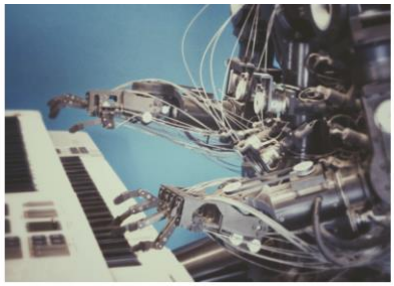
Az elmúlt év során több felmérés is született a rendkívüli tempóban sokasodó számú, növekvő képességű és javuló használhatóságú nagy nyelvi modellek társadalmi adaptációjával kapcsolatos kulcskérdésről: hogy mire használják a civil felhasználók a széles körben (ráadásul sok esetben ingyen) rendelkezésre álló LLM-eket? Az amerikai Common Sense Media egyik friss felmérése azért is lehet különösen érdekes, mert bár a fókusza specifikus volt (amerikai fiatal felhasználók célcsoportja), de az eredmények nagymértékben egybe vágnak a más mérésekből, kutatásokból származó, vagy sokszor csak „anekdotisztikus" (egyéni elbeszélésen, tapasztaláson alapuló) adatokkal. A mostani kutatás szerint a tizenévesek több mint kétharmada használja a nagy nyelvi modelleket – a használók fele ráadásul rendszeresen – és számukra az egyik legfontosabb alkalmazási terület a „digitális társ" funkció. A már egy ideje elhíresült, specializált Character.AI platform mellett – amelyet kifejezetten az egyéni igényekhez szabható „perszonák"-kal való interakcióra fejlesztettek ki – a közismert, általános célú MI-chatbotok, mint pl. a ChatGPT, vagy a Claude különböző verziói is a digitális „lelki társ", MI-alapú „barát" szerepkörében kedveltek az amerikai fiatalok tömegei számára. A felmérésből az is kitetszik, hogy a megkérdezettek meglepően nagy száma, közel harmada szívesebben beszélget magánéleti kérdéseiről, lelki problémáiról a mesterséges intelligenciával, mint valódi hús-vér barátokkal. A statisztikák két fontos mozzanatra irányítják rá újfent a figyelmet: egyrészt az MI lelki

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

terápiás célra való alkalmazhatóságának etikai kérdéskörére – nem véletlen, hogy az OpenAI napokkal legújabb ChatGPT-5-ös modelljében ezzel kapcsolatos vadonatúj rendszerkorlátokat, guardraileket építettek be. A másik probléma a fiatal korosztályok lelki egészségének szembeötlő hanyatlása – ami ráadásul nem az MI-technológia szűken vett gondja.

Teens say they are turning to AI for friendship

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



Működésben

A Microsoft új MI-modellje eredményesebben diagnosztizál összetett kórképeket, mint egy orvos

Az egészségügyben a mesterséges intelligencia alkalmazások adaptálásának kezdettől komoly gátat szabott a rendszerek megbízhatóságának a kérdése, illetve az ezek hátterében is álló „fekete doboz" jelenség. A társadalmi polémiában persze felbukkant az a szempont is, hogy vajon egy olyan területen, ahol az ember orvosok diagnosztikai tévedései, hibás terápiás döntései jelentős mennyiségű emberéletet követelnek (becslések szerint pl. az USÁ-ban évente 250 ezerre rúghat az orvosi műhibából eredő halálesetek száma) ott racionális, sőt etikus-e valamilyen képzeletbeli tökéletességet elvárni a fejlesztett technológiáktól. Ebben a háttérben tűnik ki különösen élesen a hírfolyamból a Microsoft cég bejelentése: a vállalat MI-fejlesztő részlegének – amelynek az élén Mustafa Suleyman, brit technológiai csúcsszakember áll – sikerült olyan speciális MI-modellt (az OpenAI o3-as érvelő modelljének a bázisán) kifejlesztenie, amely komplex diagnosztikai helyzetek értelmezésében több mint 80%-os eredményességgel ismerte fel a problémát. Ugyanezen (valóságos esettanulmányokon alapuló) próbafutások során a valódi ember orvosok alig 20%-os hatékonysággal voltak képesek helyes diagnózist felállítani.

Microsoft says AI system better than doctors at diagnosing complex health conditions

Laborban szaporított agysejtekkel összekapcsolt MI-chipek: kompetitív kutatási irány Kínában

Még a laikusoknak sem kerülhette el a figyelmét az elmúlt hónapokban, immár években, hogy a mesterséges intelligencia fejlesztésére, alkalmazására irányuló roppant erőfeszítések nem csupán a technológiai haladás nemes ügyét, az emberi élet könnyebbé és szebbé tételét szolgálják, hanem hatalmi küzdelmek eszközei és egyben terepei is. A mesterséges intelligencia a világ vezető hatalmai – az USA, Kína, Oroszország, Japán, India – közötti globális hatalmi vetélkedés egyik legfontosabb szegmense lett. Ebben a küzdelemben az Egyesült Államok és Kína áll az élen, sokak szerint fej fej mellett. A sokrétű párharcban most, egyes elemzői feltételezések szerint, Kína egy ígéretes újabb – az elsőséghez vezető - ösvényt tapos ki éppen. Információk szerint a kínai Tainjin Egyetem, együtt a szintén kínai Déli Tudományos és Technológiai Egyetemmel egy „chipre telepített agysejt" („brain-on-a-chip") koncepciójú fejlesztésbe kezdett. Nagyon összesűrítve arról van szó, hogy laboratóriumi körülmények között, petri csészében tenyésztett emberi agysejteket (pontosabban, úgynevezett agyi sejtorganizmusokat) egy mikroelektronikus interfészhez kapcsolnak hozzá. A kísérletektől azt várják, hogy ez a különös – és sokak szerint felettébb nyugtalanító – ember-gép rendszer rendkívüli mértékben segítheti a robotika fejlődését, nagyságrendekkel javítva a robotok szenzomotorikus képességeit. A fejlesztési irány azonban nem csupán a robottechnológia előrehaladását célozza – bár bizonyára ez is

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

nagyon fontos szempont, hiszen Kína a humanoid robottechnológiák terén már ma is kiugróan teljesít –, hanem az első igazi lépés lehet egy ténylegesen működő „biocomputer” megalkotásához.

China’s Lab-Grown Brains Connected to AI Chips

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