



MI Újság

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

2023 július

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

A hírlevél tartalma a Creative Commons Nevezd meg! – Ne add el! – Így add tovább! 4.0 Nemzetközi Licenc feltételeinek megfelelően használható.



**NEMZETI
KÖZSZOLGÁLATI
EGYETEM**
LUDOVIKA

TARTALOMJEGYZÉK

Etika és jog

- A generatív MI radikálisan forgathatja fel a jogászmunkát
- MI-vízjellel sietnének a szerzői jogaikban fenyegetett alkotó művészek segítségére
- MI-szabályozási próbakörnyezet

Trendek

- Lassan elkezdte kiszorítani a hús-vér embereket az MI a kínai kreatív iparágakból
- Az NVIDIA vezetője nem kertel: aki nem szerez MI-jártasságot, annak esélye sem lesz a boldogulásra
- A kormányok kezdenek ráébredni a biológiai kockázatokra – de gyors cselekvésre van szükség
- A verseny tárgya: a gépi felejtés problémáinak megoldása

Működésben

- Új korszak küszöbén: hamarosan munkába áll egy „valódi” többcélú robot
- Honnan tudjuk, hogy mennyire intelligensek az MI-rendszerek?
- Az MI-rendszerek biztonságos működésének lehetséges alapjai
- Az MI-segítségével megelevenedik a Biblia





Etika és jog

A generatív MI radikálisan forgathatja fel a jogászmunkát

A mesterséges intelligencia egyes alkalmazásai már jó ideje megjelentek a jogász munkában. Felsorolni is nehéz azokat a résztvevőket vagy részfeladatokat a jogász hivatás gyakorlásával kapcsolatban, ahol az MI technológiák térhódítása előnyöket kínál. Ez azonban csupán az érem egyik oldala, mert a jogász szakma egyike a leginkább konzervatívoknak, amely alapvető értéként tekint a megbízhatóságra és a precizitásra. Márpedig ezeken a területeken az új generatív MI alkalmazások, beleértve az utóbbi időben széles körben elterjedt nagy nyelvi modelleket, ma még messze nem tökéletesek. Bár az MI egyelőre gyerekcipőben jár, idővel gyökeresen megváltoztathatja az ügyvédek munkáját és az ügyvédi irodák díjszabásának kialakítását. Az amerikai Goldman Sachs pénzügyi óriásvállalat nemrégiben készült jelentése szerint a jogi feladatok mintegy 44%-át el lehetne végezni MI alkalmazásokkal és ez az arány - az irodai és adminisztratív támogatás kivételével - magasabb, mint bármelyik más vizsgált foglalkozás esetében. A jogászok már most is számos feladatra használják az MI-t, többek között a hatalmas mennyiségű iratanyag átvizsgálására és feldolgozására - melyek elvégzésében az MI már bizonyította, hogy remekül alkalmazható -, de átvilágításra (due diligence), kutatásra és adatelemzésre is. Ezek az alkalmazások nagyrészt az "extraktív" MI-re támaszkodnak, amely - ahogyan a neve is mutatja - információt von ki a szövegből, és válaszol a szöveg tartalmára vonatkozó konkrét kérdésekre. Ugyanakkor komoly lehetőségek rejtőznek az ezeknél nagyobb teljesítményre képes generatív MI jogi alkalmazásaiban is, melyekre akadnak példák a már ma is kísérleti stádiumban működő „jogi asszisztens” chatbotok formájában. A téma megosztó a jogász szakmán belül is. Egy közelmúltban végzett felmérés szerint a megkérdezett jogászok túlnyomó része, 82%-a vélte úgy, hogy a generatív MI jól használható lenne a jogi munkakörnyezetben, ám csupán 52%-uk nyilatkozott úgy, hogy az MI ilyen alkalmazása ténylegesen kívánatos is.

[Generative AI could radically alter the practice of law](#)

MI-vízjellel sietnének a szerzői jogaikban fenyegetett alkotó művészek segítségére

A mesterséges intelligencia modelleknek hatalmas méretű adatkészletekre van szükségük az alapjukat képező összetett algoritmusaik betanításához, és ma már nem titok, hogy ezeknek a kulcsfontosságú komponenseknek a biztosítása során sok esetben sérülnek az adattulajdonosok vagy adatgazdák szerzői és más jogai. Ugyanakkor annak tételes bizonyítása, hogy egy modell betanításához engedély nélkül használtak fel adatkészleteket, ma még közismerten nehéz feladat. Egy közelmúltban megjelent tanulmány szerzői gárdája egy új, hatékonynak tűnő megoldást ajánl a probléma orvoslására. A kutatók azzal az elképzeléssel álltak elő, hogy az adatkészletek illetéktelen felhasználását az adatokba ágyazott digitális vízjellel lehetne megakadályozni. Az adatkészletek védelmének legegyszerűbb módja a használatuk korlátozása - ilyen például a titkosítás, amely azonban számos hátránnyal is jár, ugyanis megnehezítheti az adatkészletek használatát az arra jogosult felhasználók számára is. A kutatók most bemutatott módszerükkel a jogosulatlan felhasználás megakadályozása helyett inkább arra helyezték a hangsúlyt, hogy ki lehessen deríteni, hogy egy konkrét MI modellt egy adott adatkészleten képeztek-e ki. Azokat a modelleket, amelyekről beigazolódik, hogy a betanításukhoz jogosulatlanul használtak fel egy adatkészletet, az adattulajdonos utólag megjelölheti és nyomon követheti. A technika a mesterséges intelligencia modellek széles skáláján alkalmazható, bár a tanulmány elsősorban az osztályozó modellekre, ezen belül is a képek osztályozására összpontosít. A digitális vízjeles módszer ugyanakkor maga is hordoz veszélyeket, például lehetőséget teremthet a rosszindulatú szereplők számára a modellek manipulálására.

[With AI Watermarking, Creators Strike Back](#)

MI-szabályozási próbakörnyezet

Az OECD 2019-es Mesterséges Intelligencia Alapelveiben azt javasolják, hogy a kormányok „fontolják meg a kísérletezés alkalmazását annak érdekében, hogy olyan ellenőrzött környezetet biztosítsanak, amelyben az MI rendszerek tesztelhetők és bővíthetők”. A kísérletezés lehetővé teszi a szabályozók számára, hogy új gazdasági, intézményi és technológiai megközelítéseket, valamint jogi rendelkezéseket próbáljanak ki az uralkodó szabályozási struktúrákon kívül. A kísérleti és „hibrid” szabályozási megközelítések közé tartoznak az innovációs központok, a szabályozási tesztkörnyezetek (regulatory sandbox), a szabványosítás, illetve a szabályozók és piacok bevonásával történő társszabályozás. Az OECD mostani jelentése ezek közül a szabályozási tesztkörnyezetekre összpontosít, amelyek ígéretesnek tűnnek a gyors innovációs ciklusú területeken, mint például az MI és a pénzügyi technológia (fintech). A tesztkörnyezetek kialakítása során a hatóságok olyan innovatív termékek vagy szolgáltatások tesztelésébe vonják be a cégeket, amelyek túlmutatnak a meglévő jogi kereteken. A részt vevő cégek az innováció érdekében felmentést kapnak bizonyos jogi rendelkezések vagy megfelelési folyamatok alól. Szabályozási tesztkörnyezet sokféle lehet, de vannak közös jellemzőik: ideiglenesek, a tesztelési folyamat általában hat hónapra korlátozódik; felmentést adnak a meglévő jogi rendelkezések alól és személyre szabott jogi támogatást nyújtanak egy adott projekthez; gyakran próba-szerencse alapon

működnek; az általuk gyűjtött műszaki és piaci információk, adatok pedig segítik a szabályozó hatóságokat annak megítélésében, hogy az adott jogi keretek alkalmasak-e a célnak. A jelentés tárgyalja az MI szabályozási tesztkörnyezetek létesítésével kapcsolatos problémákat, kiemelve az interdiszciplináris együttműködés, az MI-szakértelem kiépítése, a szabályozási interoperabilitás és a kereskedelempolitika fontosságát. Szintén fontos szempontként tárgyalja a kísérlet értékelését, illetve azt, hogy milyen kritériumok alapján vehet egy vállalkozás részt a kezdeményezésben. Végül áttekinti a szabályozói tesztkörnyezet hatását az innovációra és versenyre.

[Regulatory sandboxes in artificial intelligence](#)





Trendek

Lassan elkezdte kiszorítani a hús-vér embereket az MI a kínai kreatív iparágakból

A Sixth Tone angol nyelvű kínai hírportál szerint a kreatív iparágban dolgozó kínai munkavállalóknak egyre erősebb konkurenciával kell szembenézniük, még hozzá nem a kollégáik részéről, hanem a világszerte nagy visszhangot keltő, az emberi képességeket is megközelítő, élvonalbeli generatív mesterséges intelligencia technológiák az igazi versenytársaik. Míg egyes munkavállalók fenyegetve érzik magukat az új technológiáktól, mások szerint a gépek sosem lesznek képesek teljesen helyettesíteni az emberi elmét. Mindeközben néhány cég, ahol kreatív tervezéssel és szövegírással (copywriting) kapcsolatos munkakörök is vannak, már el is kezdte lecserélni az alkalmazottait gépekre és helyettük bevezetni ezeket a feltörekvő technológiákat. Kína egyik legnagyobb marketingügynöksége, a Blue Focus áprilisban bejelentette, hogy „teljesen és határozatlan ideig” leállítja egyes kreatív munkáinak kiszervezését - amely eddig a cégóriás jellemző gyakorlata volt -, és helyette inkább mesterséges intelligencia által generált tartalmakat használ a működésében. A bejelentésre közvetlenül azután került sor, hogy a vállalat megvásárolta az OpenAI GPT-4 verziójához való hozzáférés licenzét. Miközben a Goldman Sachs befektetési bank új jelentésében globálisan mintegy 300 millióra becsüli az MI-technológiák által kiváltható teljes munkaidős állások számát, a PricewaterhouseCoopers tanácsadó cég szerint Kínában az elkövetkező hús évben a meglévő munkahelyek csaknem 20%-a válhat feleslegessé a technológia következtében, ugyanakkor a termelékenység és a bevételek növelésével a mesterséges intelligencia új munkahelyeket is teremthet. A téma társadalmi megítélésére és fogadtatására ad némi rálátást a kínai közösségi média ehhez kapcsolódó párbeszédfolyama. A témában jártas felhasználók egy része már ma is arra panaszkodik, hogy az MI-alkalmazások készségszintű ismerete nélkül nem találnak megfelelő állást a kreatív munkakörökben.

[AI is starting to Replace Humans in China's Creative Sector](#)

Az NVIDIA vezetője nem kertel: aki nem szerez MI-jártasságot, annak esélye sem lesz a boldogulásra

Jensen Huang, az Nvidia Corporation társalapítója és vezérigazgatója szerint a cégeknek és az egyéni munkavállalóknak egyaránt jártasságot kell szerezniük a mesterséges intelligenciák alkalmazása terén, máskülönben végérvényesen kiesnek a versenyből. Márpedig, ha valaki, akkor Huang tudhat egyet s mászt erről a témáról. A chiptervezéssel foglalkozó Nvidia májusban rekordnyereséget ért el, döntően az MI fejlesztőktől és szolgáltatóktól érkező kereslet szédületes felívelése következtében. A vállalat vezére úgy véli, hogy az új technológia át fogja alakítani a vállalati környezetet és minden egyes munkahelyet megváltoztat. Álláspontja szerint azok az agilis cégek, amelyek ki tudják használni az mesterséges intelligencia előnyeit, javítani fogják piaci pozíciójukat; aki pedig erre nem lesz képes, az egyszerűen semmivé foszlik. A kaliforniai Santa Clarai-i székhelyű vállalatíriás igazgatója a Tajvani Nemzeti Egyetemen fogalmazta meg azt, amely szellemessége és kíméletlen őszintesége miatt egy csapásra közkeletű megfogalmazása lett az új MI-technológiáknak a munkaerőpiacra, illetve általában a versenyszférára gyakorolt hatásainak: „Miközben az emberek egy része amiatt aggódik, hogy az MI elveszi a munkáját, az igazság az, hogy azt a munkát egy másik ember fogja elvenni tőle - egy olyan, aki jártassá vált az MI használatában”. Huang szerint a technológiát, amelyet az OpenAI ChatGPT-je hozott be tavaly a köztudatba, számos iparágban egyfajta másodpilótaként fogják használni a munkavállalók teljesítményének növelésére, miközben olyan új munkahelyek jönnek létre, amelyek ezelőtt soha nem is léteztek, ezzel egyidőben néhány másik pedig elavulttá, feleslegessé válik.

[Nvidia CEO Says Those Without AI Expertise Will Be Left Behind](#)

A kormányok kezdenek ráébredni a biológiai kockázatokra – de gyors cselekvésre van szükség

A MIT kutatói egy nemrégiben elvégzett kísérletben egyetemi hallgatókat kértek fel arra, hogy teszteljék, vajon a mesterséges intelligencia által vezérelt chatbotokat lehet-e arra utasítani, hogy segítsenek laikusoknak járványt okozni. Egy órán belül a chatbotok négy lehetséges pandémiás kórokozót is javasoltak. Elmagyarázták, hogyan lehet ezeket szintetikus DNS-ből fordított genetika segítségével létrehozni és megadták azoknak a DNS-szintézissel foglalkozó cégeknek a nevét, amelyekről úgy ítélték meg, hogy valószínűleg nem fogják biztonsági szempontból szűrni a megrendeléseket. A biofegyverek kifejlesztése szerencsére nem ennyire egyszerű, de ez a kísérlet így is megmutatta, hogy potenciálisan mi történhet, ha az MI technológia felhasznál bizonyos tudományos ismereteket. A nyugtalanító tény az, hogy a nagy nyelvi modellek és az új biológiai tervezési eszközök drámai módon csökkentik az akadályokat egy következő, mesterségesen előidézett világjárvány előtt. A Google korábbi vezérigazgatója, Eric Schmidt szerint az MI által tervezett biofegyverek már „nagyon rövid távon is aggodalomra” adnak okot. A kormányok kezdenek ráébredni a felmerülő biológiai kockázatok mértékére. A cikk röviden ismerteti a folyamatban levő törekvéseket, de a szerző véleménye az, hogy többre van szükség és az Egyesült Királyságnak most három fronton kellene gyorsan cselekednie. Az első lépés a nagy nyelvi modellek és a biológiai

tervezési eszközök értékelésének kidolgozása. Korlátozni kell a veszélyes eszközökhöz való hozzáférést, ezen belül a lehetséges megoldások: a biológiai károkozásra vonatkozó információk eltávolítása az MI-rendszerek betanítási adataiból, szigorú tartalomellenőrzés bevezetése, valamint a halálos biológiai ágensek tervezéséhez használt szoftverek terjesztésének megakadályozása. Másodsor, az Egyesült Királyságnak erőfeszítéseket kell tennie arra, hogy az esetleges új kórokozókat megjelenésük után gyorsan ki lehessen mutatni. Ez az a terület, ahol az MI erejét hasznosítani lehet a korai felismerés elősegítésére, amennyiben a kevésbé fejlett egészségügyi rendszerekkel rendelkező országokat is bekapcsolják a rendszerbe. Végül pedig nemzetközi összefogásra van szükség, hogy az MI és a biotechnológia összekapcsolódó kockázataira lehessen összpontosítani.

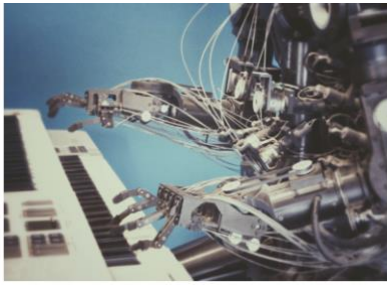
[Governments are waking up to biosecurity risks – but we must act fast](#)

A verseny tárgya: a gépi felejtés problémáinak megoldása

A mélytanulás a közelmúltban óriási fejlődésen ment keresztül és bár ez a folyamat nagyon izgalmas, a mély neurális hálózati modellek széleskörű alkalmazása nagyfokú körütekintést is igényel. A Google fejlesztői igyekeznek felelősségteljesen fejleszteni a mesterséges intelligencia technológiákat, az olyan lehetséges kockázatok megértésével és mérséklésével, mint például a tisztességtelen torzítások terjedése vagy a felhasználói adatvédelem problémái. Ennek érdekében lehet szükséges a rendszerek betanításához használt adatok egy részének törlése. A törölni kért adatok hatásának teljes eltávolítása kihívást jelent, mivel amellet, hogy egyszerűen törlik ezeket az adatokat az adatbázisokból, ahol tárolják őket, az adatok más eszközökre és rendszerekre, például a betanított gépi tanulási modellekre gyakorolt hatását is törölni kell. A közelmúltban végzett kutatások kimutatták, hogy bizonyos esetekben nagy pontossággal lehet következtetni arra, hogy egy adott példát használtak-e egy gépi tanulási modell betanítására. Ez pedig adatvédelmi aggályokat vethet fel, mivel azt jelenti, hogy még ha az egyén adatait törlik is az adatbázisból, akkor is ki lehet következtetni, hogy azokat felhasználták-e egy modell betanításához. A fentiek következtében jelent meg a gépi tanulás egy újonnan kialakuló részterülete, a gépi felejtés (machine unlearning), amelynek célja, hogy eltávolítsa a képzési példák egy meghatározott részhalmazának – a „felejtésre ítélt halmaznak” – hatását a betanított modellből. Egy ideális felejtési algoritmus eltávolítaná bizonyos példák hatását, miközben más előnyös tulajdonságokat megtartana. A legegyszerűbb az lenne, ha a modellt újra betanítanák a már kiigazított képzési halmazon, amelyből kizárják az elfelejtendő halmazból származó mintákat. Ez azonban nem mindig járható út, mivel a mély modellek újratanítása számítási kapacitás szempontból költséges lehet. Egy ideális gépi felejtési algoritmus ehelyett a már betanított modellt használná kiindulópontként, és hatékonyan végezne módosításokat a kért adatok hatásának megszüntetésére. A Google most az érdeklődő szakemberek számára versenyt hirdet a probléma megoldására, ezen a területen az első ilyen eseményként.

[Announcing the First Machine Unlearning Challenge](#)





Működésben

Új korszak küszöbén: hamarosan munkába áll egy „valódi” többcélú robot

A kanadai székhelyű technológiai start-up, a Sanctuary AI nemrégiben bemutatott egy új, általános célú robotot, amelyet arra terveztek, hogy számos, jelenleg emberek által végzett munkahelyi feladatot is el tudjon végezni - akár emberekkel együtt, akár nélkülük. A robotok már évtizedek óta segítik a jellemzően nagy számban ismétlődő, ugyanakkor jelentős figyelmet igénylő munkatevékenységek végrehajtását. Ezek között megtalálhatók olyanok, amelyek javarészt ember-gép páros részeként segíti az ipari termelési folyamatokat, de az önállóan, emberi beavatkozás nélkül működő típusokra is bőségesen akad példa - ilyenek az autógyártás területén a szerelőszalag mellett dolgozó robotok. Amiből azonban nem sok akad, az a többcélú robot, amely nem csupán egy jól meghatározható tevékenység, esetleg tevékenység sor elvégzésére alkalmas, hanem amely az adott feladat jellegétől függően sokféle munkára igénybe vehető. A Sanctuary AI Phoenix fantázianevű szerkezete éppen egy ilyen többcélú robot. A mintegy 170 cm magas, 70 kilogramm súlyú, szenzorokkal gazdagon felszerelt robot sokféle, több síkban is precízen kivitelezett mozgásra és akár 25 kg teher felemelésére is képes. 20 szabadságfokkal és haptikus érzékelőkkel rendelkező kezei lehetővé teszik, hogy olyan ügyességet és pontosságot igénylő feladatokat is el tudjon végezni, mint például a csomagok címkézése vagy a gyümölcsök leszedése. A legfontosabb újdonságot azonban a mesterséges intelligencia jelenti, amely az érzékelők inputjait fogadja és amely az emberéhez hasonló intelligenciával ruházza fel a vállalat hatodik generációs általános célú robotját. A Carbon nevű saját fejlesztésű MI vezérlőrendszer jóvoltából a Phoenix számítógépes szimulációkkal betanítható az új feladatok elvégzésére. A Sanctuary januárban a Mark's kiskereskedelmi hálózat egyik üzletében végzett próbaüzem során mutatta be a robotok munkavégzésbe való bevonásának lehetőségeit. Az ötödik generációs Phoenixet egyetlen hét leforgása alatt mintegy 110 különféle tevékenységre tanították meg, ami körülbelül 40%-a azoknak a feladatoknak, amelyeket egy átlagos alkalmazottnak el kell tudnia végezni az üzletben.

[A general purpose robot is entering workforce](#)

Honnan tudjuk, hogy mennyire intelligensek az MI-rendszerek?

A mesterséges intelligencia kutatásának kezdetektől célja, hogy az emberi kognitív képességekkel rendelkező vagy azokat meghaladó rendszereket fejlesszenek ki. A kutatók rendre be szokták jelenteni, hogy már közel az az idő, amikor ez bekövetkezik, ám mindeddig nem váltak valóra ezek a jóslatok. Fölmerül a kérdés, hogyan lehet azt megtudni, hogy mennyire értelmesek a rendszereink? A tapasztalat szerint ennek kiderítése egyáltalán nem egyszerű. Ahelyett, hogy szubjektív benyomásokra hagyatkoznánk, az MI kutatásában régi hagyomány, hogy olyan tesztekkel vizsgálják ezeknek a rendszereknek a képességeit, amelyeket eredetileg az emberi intelligencia és megértés felmérésére fejlesztettek ki. Ezekkel azonban számos probléma lehet. Az egyik az „adatszennyezés” kérdése. Bár feltételezhetjük, hogy a szabványos tesztet elvégző emberek még nem látták a kérdéseket és a válaszokat, ez nem feltétlenül igaz egy olyan nagy MI rendszer esetében, mint a GPT-4, amelyet az elérhető digitális tartalmak hatalmas tömegén tanítottak. Ezen tartalmak némelyike ugyanis tartalmazhatta azokat a kérdéseket, amelyeken a GPT-4-et később tesztelték. A második a robusztusság problémája. Bár feltételezzük, hogy az ember, aki helyesen válaszol egy tesztkérdésre, képes lesz helyesen válaszolni egy nagyon hasonló kérdésre, ez nem mindig igaz az MI-rendszerekre. Az olyan nagy nyelvi modellek, mint a GPT-4, közismerten nagyon érzékenyek arra, hogy a feladataikat hogyan fogalmazzák meg számukra. A harmadik a hibás referenciaértékek problémája. Számos, az MI-rendszerek betanításánál viszonyítási alapként használt adatkészletről kimutatták, hogy lehetővé teszi az úgynevezett „shortcut learning”-et, vagyis olyan finom statisztikai asszociációkat, amelyeket a gépek felhasználhatnak a helyes válaszok előállítására anélkül, hogy ténylegesen megértenék a szándékolt fogalmakat. Ahhoz, hogy tudományosan értékelhessük az emberszerű, sőt emberfeletti gépi intelligenciára vonatkozó állításokat, nagyobb átláthatóságra van szükség e modellek képzésének módjait illetően, valamint jobb kísérleti módszerekre és viszonyítási alapokra, támaszkodva a kisgyerekek képességeinek vizsgálatára kidolgozott módszerekre. Ezeket az jellemzi, hogy módszerek és gondosan ellenőrzött körülmények között lefolytatott kísérleteken alapulnak.

[How do we know how smart AI systems are?](#)

Az MI-rendszerek biztonságos működésének lehetséges alapjai

A legújabb nagy nyelvi modellek gyors és széles körű elterjedése egyszerre váltott ki izgalmat és aggodalmat a fejlett MI technológiák vonatkozásában, amelyre válaszul sokan a mesterséges intelligencia biztonságának területén keresik a megoldásokat. De mit jelent a mesterséges intelligencia rendszerek biztonságossá tétele, és milyen értékeket és megközelítéseket kell alkalmazni ehhez? A technológiai változásokkal szembeülvélve az emberek először ösztönösen a műszaki megoldásokat keresik. A fejlett mesterséges intelligencia hatásait azonban nem lehet pusztán technikai eszközökkel mérsékelni, a szélesebb társadalmi betekintést nem tartalmazó megoldások ugyanis csak fokozzák a veszélyeket. A Science magazin nemrég megjelent cikke szerint a valóban biztonságos MI eléréséhez a társadalomnak szociotechnikai megközelítésre

van szüksége, amely képes korlátozni a fejlett mesterséges intelligencia jelenlegi és potenciális veszélyeit. Ez azt jelenti, hogy figyelembe kell venni a mesterséges intelligencia politikai gazdaságtanát, és fel kell ismerni, hogy a hipotetikus jövőbeli károokra való túlzott összpontosítás előre láthatóan azzal a kockázattal jár, hogy néhány vezető vállalat hatalma megerősödik, és a jelenlegi nagy kihívások megoldatlanok maradnak. A tényleges MI rendszerek és az általuk jelentett kockázatok esetében az empirikusan megalapozott munkát kell előnyben részesíteni, nem pedig a még nem létező rendszerekből eredő kockázatok technikai mérséklésének feladatát. Az éveken át tartó szociotechnikai kutatások azt mutatják, hogy a fejlett digitális technológia ellenőrizetlenül hagyva a hatalom és profitszerzés eszközévé válik, az emberi jogok, a társadalmi igazságosság és a demokrácia rovására. A fejlett MI biztonságossá tétele az ezeket az értékeket érintő kockázatok megértését és csökkentését is jelenti. A szociotechnikai megközelítés azt hangsúlyozza, hogy egyetlen szakértői csoport sem dönthet egyedül arról, hogy milyen kockázatok és milyen ártalmak számíthatnak, és milyen értékekhez kell igazodjon a biztonságos mesterséges intelligencia.

[AI safety on whose terms?](#)

Az MI-segítségével megelevenedik a Biblia

Az izraeli BiblePics cég bibliai elbeszélésekre alapozott mesterséges intelligencia segítségével hozná közelebb a digitális világban felnövő fiatalokhoz a vallási üzenetet. A vizuális történetmesélést és a legmodernebb mesterséges intelligencia algoritmusokat felhasználva az alkalmazásban a különböző bibliai alakok a MI által generált formában jelennek meg és beszélgetést is lehet velük folytatni. Az alkotók célja, hogy az MI által generált képeken keresztül meg tudják ragadni a felhasználók fantáziáját, akik ezáltal mélyebb kapcsolatot alakíthatnak ki a bibliai történetekkel. A grafikai világon kívül a BiblePics egy MI által vezérelt csevegési funkciót is tartalmaz, amely lehetővé teszi a felhasználók számára, hogy értelmes beszélgetéseket folytassanak neves bibliai alakokkal. A felhasználók útmutatást kérhetnek és felvilágosítást kaphatnak olyan ismert bibliai személyiségektől, mint Ádám, Éva, Mózes, Ábrahám, valamint a keresztény felhasználók számára készült változatban például Jézustól. Ezen kívül a csevegési funkció interaktív jellege lehetővé teszi a felhasználók számára, hogy társalgás formájában merüljenek el a Szentírás mélységeiben, ezzel is bátorítva az ismeretek önálló felfedezését és elősegítve a mélyebb megértést. A készítők kiemelt fontosságot tulajdonítottak a pontosságnak, tekintettel a vallási szövegekkel kapcsolatos érzékenységre. Az MI-modelleket a bibliai szövegek hatalmas korpuszán tanították be és a rendszer működése szempontjából világos korlátokat határoztak meg, így biztosítva, hogy pontos és a kontextusnak megfelelő válaszok szülessenek. A bibliai narratívák integritásának és szöveghűségének biztosítása érdekében az MI által mozgatott szereplők csak olyan válaszokat adhatnak, amelyekben biztosak, így megelőzve a félretájékoztatást vagy a pontatlanságokat.

[Israeli firm BiblePics uses AI to let people talk to biblical characters](#)

