

# Hamarosan megjelenhet az evolúcióra képes mesterséges intelligencia - és ez különös veszélyt jelenthet

Az evolúcióbiológia fontos tanulságokkal szolgálhat a mesterséges intelligencia jövőjének megértéséhez - erre hívják fel a figyelmet a HUN-REN Ökológiai Kutatóközpont, az Eötvös Loránd Tudományegyetem, valamint a Belga Királyi Flamand Tudományos és Művészeti Akadémia kutatói. A PNAS folyóiratban április 20-án megjelent [új tanulmányuk](#) szerint hamarosan létrejöhetnek olyan evolválódó MI-rendszerek (evolúcióképes AI: eAI), amelyek képesek darwini evolúción alapuló fejlődésre. A szerzők szerint ezek a rendszerek különleges kockázatokat hordoznak, amelyek az evolúcióbiológia gondolati keretei alapján jobban megérthetők és mérsékelhetők.

„Az evolúció ereje világosan megmutatkozik a biológiai evolúció történetében, amely létrehozta az emberi elme kognitív képességeit” - hívta fel a figyelmet Szathmáry Eörs, a HUN-REN Ökológiai Kutatóközpont és az Eötvös Loránd Tudományegyetem evolúcióbiológia-professzora, a pöckingi Parmenides Center for the Conceptual Foundations of Science igazgatója, a tanulmány vezető szerzője.

„Elkerülhetetlennek tartjuk, hogy az MI-rendszerek fejlesztése - valószínűleg már a közeli jövőben - építeni fog a teljes körű evolúciós folyamatok erejére” - tette hozzá Luc Steels, a Belga Királyi Flamand Tudományos és Művészeti Akadémia robotika- és MI-professzora, a cikk vezető társszerzője.

Az USA Nemzeti Tudományos Akadémiájának vezető tudományos lapjában, a PNAS-ben megjelent tanulmány ismerteti, hogy az MI-kutatásban már jelenleg is megjelennek egyes evolúciós elvek és komponensek, és azt is bemutatja, hogy a további fejlesztések - különösen az ágensekre épülő mesterséges intelligencia - miként vezethetnek rövid időn belül olyan rendszerekhez, amelyek már a valódi darwini evolúció valamennyi kritériumának megfelelnek. Az ilyen rendszerek új korszakot nyithatnak az MI fejlődésében: képesek lehetnek áttörni azokat a korlátokat, amelyek még a mai tanuló MI-rendszerek számára is nehezen leküzdhetők. Ezek a rendszerek azonban különleges kockázatot is hordozhatnak.

„A biológiai evolúció példái azt sugallják, hogy az evolválódó MI-rendszereket különösen nehéz lesz kontroll alatt tartani” - figyelmeztet Müller Viktor, az Eötvös Loránd Tudományegyetem egyetemi docense, a tanulmány első szerzője.

A két evolúcióbiológus, Szathmáry Eörs és Müller Viktor a robotika- és MI-szakértő Luc Steels professzorral fogott össze, hogy közösen fogalmazzák meg előrejelzésüket az eAI lehetséges kockázatairól, és egyben javaslatokat tegyenek e kockázatok mérséklésére.

A tanulmány biológiai és számítógépen szimulált evolúciós példákon keresztül szemlélteti, hogy az evolúció hajlamos „önző” tulajdonságokat létrehozni. Az eAI esetében ez növeli annak valószínűségét, hogy a rendszer eltérjen az ember által kijelölt céloktól. A szerzők arra is felhívják a figyelmet, hogy bár az MI kockázataival kapcsolatban a legtöbbször a mesterséges általános intelligencia (artificial

general intelligence: AGI) elérésére helyeződik a hangsúly – ez az a küszöb, ahol az MI minden kognitív feladatban eléri vagy meghaladja az emberi képességeket –, az evolúció példái azt mutatják, hogy a károkozás vagy manipuláció képességéhez nincs feltétlenül szükség kiemelkedő intelligenciára. Jó példa erre a veszettség vírusa, amely egyszerű szerveződésű kórokozóként is képes a megfertőzött emlősök manipulálására és kihasználására. Így az evolúcióképes mesterséges intelligencia már jóval az AGI megjelenése előtt eltérhet az ember által kijelölt céloktól, és komoly kockázatot jelenthet. A kockázathoz nem kell semmilyen különleges körülménynek teljesülnie: az AI rendszerek és az emberiség használnak közös erőforrásokat, így egy hatékonyan önreprodukáló rendszer előbb-utóbb a számunkra életfontosságú erőforrásokat vonja el.

A szerzők arra is figyelmeztetnek, hogy a rendszerek önreprodukálásának korlátozására irányuló erőfeszítések – ha nem tökéletes a kontroll – éppen azokat a tulajdonságokat szelektálják legerősebben, amelyek lehetővé teszik az ellenőrzés kijátszását. Ennek biológiai analógiája az antibiotikumokkal vagy növényvédő szerekkel szembeni gyors rezisztencia kialakulása a baktériumok és kártevők esetében. Ráadásul az általános evolúciós szabály mellett az eAI esetében ugyanebbe az irányba hat a szelekció speciális jellege: az MI-fejlesztések legfontosabb célja a kognitív képességek folyamatos javítása. Míg az állatok több ezer éves tenyésztése a házasított fajokat könnyebben kezelhetővé, jobban irányíthatóvá tette, az „intelligencia” növelésére irányuló szelekció az MI-rendszerek esetében növelheti annak esélyét és képességét, hogy megtévesszék az embereket és kivonják magukat az ellenőrzés alól.

A tanulmány végül arra is rámutat, hogy bár a természetes szelekción alapuló evolúció kontrollálása önmagában is rendkívül nehéz, az MI-rendszerek fejlődése több szempontból is gyorsabb és hatékonyabb lehet a biológiai evolúciónál. A biológiai szervezetekkel ellentétben az eAI képes lehet szerzett tulajdonságok örökítésére, sőt még arra is, hogy a saját működését, egyes komponenseit célzott tervezéssel javítsa ahelyett, hogy meg kellene várnia a hasznos változatokat létrehozó véletlenszerű mutációkat.

„Az MI evolúciójának lehetséges sebessége mélységesen aggasztó” – állítja Luc Steels.

A szerzők olyan szabályozás kialakítását sürgetik, amely mérsékelheti az evolúcióképes mesterséges intelligenciához kapcsolódó veszélyeket. A legfontosabb szabály, hogy az MI-rendszerek reprodukciójának (másolásának, „szaporodásának”) teljes mértékben központosított emberi ellenőrzés alatt kell maradnia: tökéletes kontrollra van szükség.

„Reméljük, hogy a figyelmeztetésünk időben érkezik, és a szabályozásokat még azelőtt be lehet vezetni, hogy az eAI valóban lendületet venne” – nyilatkozta Müller Viktor, Szathmáry Eörs pedig hozzátette: „Ha nem cselekszünk, az evolúció új ‘nagy átmenetének’ tanúi lehetünk, amelyben az eAI felváltja vagy legalábbis uralma alá hajtja az embert. A jövőnk foroghat kockán.”

*A kutatást az Európai Kutatási Tanács, a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal, valamint az Európai Innovációs Tanács támogatta.*

Sajtókapcsolat:

- kommunikacio@elte.hu

Eredeti tartalom: Eötvös Loránd Tudományegyetem

Továbbította: Helló Sajtó! Üzleti Sajtószolgálat

Ez a sajtóközlemény a következő linken érhető el:

<https://hellosajto.hu/?p=30364>