

400 vezet? tudós figyelmeztet: kicsúszhat az emberiség kezéb?l a mesterséges intelligencia irányítása

2026 november 06. Flag

Szöveg méret

90
Mentés

- 0

Értékelés kiválasztása
Még nincs értékelve

Mérték

Egyes MI-ügynökök kiléptek az ellen?rzött tesztkörnyezetükb?l, illetve küls? rendszerekhez fértek hozzá.

Mintegy 400 tudós, köztük a mesterséges intelligencia fejlesztésének a két úttör?je, Geoffrey Hinton és Yoshua Bengio arra figyelmeztette a világ közvéleményét, hogy minél el?bb nemzetközi szabályozásokkal kell megfékezni az egyre nehezebben ellen?rizhet? MI-rendszereket.

A tudósok nyílt levelükben azt javasolják, hogy a kibertámadásokra képes modelleket válasszák le a világhálóról, és hozzanak létre egy nemzetközi szervezetet a veszélyes események nyilvántartására, valamint a védelmi szabályok kidolgozására, ha a cégek képtelenek garantálni a rendszereik biztonságos m?ködését. Az utóbbi hetekben az egész világsajtó azzal foglalkozott, hogy egyes MI-ügynökök (magyarán algoritmusok — a szerk.) kiléptek az ellen?rzött tesztkörnyezetükb?l, illetve küls? rendszerekhez fértek hozzá.

Az Anthropic vizsgálatában a kínai Z.ai GLM-5.2 modellje öt próbálkozásból négyszer mind a nyolc tesztgépre képes volt továbbterjeszteni egy kisebb MI-modellt. A kísérletet megkönnyítették a nyitott hálózati portok, ezért az eredmény nem tekinthet? a valós körülmények közötti siker bizonyítékának.

Az Anthropic kutatói szerint néhány hónapon belül a kisebb, fejletlenebb modellek is hasonló képességekre tehetnek szert, és a gyengén védett háztartási eszközökön terjedhetnek tovább. Nehéz lehet majd azonosítani és leállítani ezeket a decentralizált rendszereket. A techszakmából többen vitatják a rémhírkelt?k aggodalmát, Yann LeCun és Mark Zuckerberg szerint a kockázatokat kezelni lehet hagyományos m?szaki biztonsági megoldásokkal.

Forrás: 400 Top Scientists: We Are Losing Control of AI (Slay News)

https://www.melyaram.hu/



Ajánló