A kínai MI igazi veszélye: Miért kell Amerikának vezetnie a versenyt az *open source* MI fejlesztésében?

Jared Dunnmon írása a Foreign Affairs 2025. február 28-i számában

(ismertetés)

A kérdés meglepő. Hát nem az lenne a logikus, hogy a MI legújabb „találmányainak" kódolását hétpecsétes titokként őrizzék? És ráadásul ez e fontos, mérvadó folyóiratban jelenik meg… Miért fontos, hogy a MI legfrissebb eredményei, a mögöttük álló kódolás mindenki számára elérhetők legyenek? A szerző ráadásul a lehető legkompetensebb személy – a legutóbbi időkig a Pentagon Defense Innovation Unit-jának a „technikai" igazgatója volt. A témát Kínával és a DeepSeek-kel kapcsolatban tárgyalja – ami még meglepőbb: éppen a nagy vetélytárssal szemben kell feltárni a legújabb fejlesztések kódolását?

Dunnmon szerint „az az ország, amely megalkotja a holnap app-jait és szolgáltatásait formáló nagy nyelvi modelleket, mindent fölülmúló befolyással lesz nem csak az azokban megtestesülő normák és értékek fölött, de a félvezetők „ökoszisztémája" fölött is, mely a MI alapjait képezi. Mivel Kína is és Amerika is úgy gondolja, hogy ezek a technológiák hadiipari előnyökkel rendelkeznek, fontos a hosszútávú vezető szerep megőrzése a MI fejlesztésében."

Lényeges különbség, hogy míg az amerikai modellek megalkotóik tulajdonát képezik, a DeepSeek és a kínai modellek nyíltak, szabad hozzáférésűek, bárki tanulmányozhatja, letöltheti azokat, módosíthatja, kísérletezhet velük. Már Sam Altman és Eric Smith is felhívta rá a figyelmet, hogy hátrányos, ha a legújabb MI-modellek nem szabad hozzáférésűek. Nem elegendő ugyanis az IT vállalatok hatalmas tőkeerejére és szellemi kapacitására támaszkodni; a felhasználók körében is jelentkezhetnek fontos újítások, ötletek.

„A mai nagy nyelvi modelleket és a hagyományos számítógépeket összehasonlítva egy fontos analógiát látunk, ami a nyílt kódolású szoftvernek, az operációs rendszernek tulajdonítható. Annak mintájára, ahogy az operációs rendszer az emberbarát kompjúter-programokat olyan instrukciókká változtatja, amiket a gépi hardver valósít meg, a nagy nyelvi modellek hidat képeznek az emberi nyelv és a gép által feldolgozott információ között."

Az amerikai IT hatalmas eredményei nem kis részben az OSS-nek (Open Source Softwears) tulajdoníthatók. De ez a nyíltság nyilvánvalóan biztonsági problémákat vet fel. A nagy MI-cégek ezért legújabb modelljeiket chatbox-ok és más web-portálok révén teszik hozzáférhetővé. A leggyakrabban használt amerikai LLM-ek közül[1] csak a Meta Llama-ja nyílt kódolású. És mint ahogy arra számítani lehetett, máris kitudódott, hogy azt a kínaiak katonai célokra használják fel.[2] De a DeepSeek példájából látszik, hogy hiába tiltja Amerika a legfejlettebb chipek exportját, a kínai mérnökök is képesek hasonlók kifejlesztésére. A DeepSeek már nem csak az amatőrök, de a profi felhasználók körében is vezet Amerikában.

[1] Nagy Nyelvi Modellek; más néven generatív MI; ezek azok, amelyek válaszolnak a kérdéseinkre.
[2] Bár – amint a DeepSeek-ből látható – maguk is képesek már e technológiák megalkotására.

Márpedig ennek komoly politikai-ideológiai következményei vannak. [A szerző az „értékek" kifejezést használja.][3] A kínai KP kiadta az utasítást, hogy az LLM-eket „a szocializmus alapértékeinek megfelelően kell fejleszteni és el kell kerülni a problematikus információkat és illegális tartalmakat. Ha pl. megkérdezik a DeepSeek-et, hogy mi történt 1989-ben a Tienanmen téren, azt a választ adja, hogy „sajnos, erre nem válaszolhatok". Ha pedig kényes, problémás kérdéseket adnak fel neki, kitér a válasz elől, mellébeszél… Ez sokkal nagyobb veszélyt jelent a liberális értékekekre, mint pl. a TikTok. További kockázatot jelenthet, hogy az amerikai/külföldi felhasználó adatai eljutnak Kínába.

De a legnagyobb kockázatot az jelenti, hogy a kínaiak hozzáférhetnek a Nyugat e téren fennálló legnagyobb előnyéhez: kiépíthetnek egy kínaiak által irányított MI ellátási láncolatot. Hogyan történhet ez meg:

Ma egy amerikai cég, az Nvidia uralja a MI-hoz szükséges chipek piacát az ún. „graphic processing unit"-jával (GPU), mely a mai MI-modellek túlnyomó többségét működteti. De az Nvidia valójában a MI modellek működtetéséhez szükséges szoftverek piacát is uralja: a CUDA a MI fejlesztők által írt programokat a GPU-k számára szóló parancsokká írja át. Ily módon a vállalat a MI modellek számára írt programok „ökoszisztémáját" is dominálja. A DeepSeek-et is Nvidia-chipek felhasználásával gyakoroltatták be, melyekhez a kínaiak állítólag legális módon jutottak hozzá.

Ebből azt a következtetést lehetne levonni, hogy még jobban meg kell szigorítani a hozzáférést az Nvidia chip-jeihez. Igen ám, de miután a kínaiak a DeepSeek-et Nvidia chipekkel trenírozták, utána már azt a kevésbé fejlett saját chipjeikkel is képesek voltak futtatni. Nevezetesen: az SMIC által gyártott chipekkel, melyeket a Huawei fejlesztett ki (Huawei's Ascend Neural Processing Unit chips). A kínai kormány támogatásával hamarosan lehetségessé válik, hogy a nagy nyelvi modelleket saját chip-jeikkel fejlesszék tovább. És mivel a kínai DeepSeek jóval olcsóbb az amerikai modellekhez képest, ennek az lehet a következménye, hogy a DeepSeek-et használók az egész világon áttérnek az SMIC- és Huawei-chipek használatára. A tőkebefektetések tehát feléjük irányulnak, az Nvidia, Intel, Samsung és TSMC [ez tajvani cég] pedig háttérbe szorulnak.

Tehát az, hogy a kínaiak a DeepSeek-et szabad forráskódúvá teszik, elindít egy folyamatot, melynek eredményeként romba dől Amerika chip-gyártó hegemóniája. Sőt, megszűnik vezető szerepe az MI-szektorban – hasonlóan ahhoz, ami a lítium-ionos akkumulátorok és sok más termék esetén történt.

Ahhoz, hogy ezt elkerülje, Amerikának széles körben alkalmaznia kell a nyílt forráskódú MI-modelleket, és ezt az állami szerveknek is elő kell segíteniük. 2022-23-ban 31 md dollár kockázati tőke áramlott az amerikai MI-szektorba, de ebből csak 2,9 md jutott a nyílt forráskódú MI-vállalatoknak. Az amerikai kormánynak ösztönöznie kell az OSS fejlesztésekbe áramló befektetéseket, sőt, közvetlenül is részt kell venniük azokban. A támogatásnak konkrétan arra is irányulnia kell, hogy a fejlesztők az amerikai chipeket használják. Továbbá, állami támogatásban kell részesíteni a már megjelent legújabb MI-

[3] A nagy nyelvi modelleket egy kiterjedt adatbázison trenírozzák; nyilvánvaló, hogy válaszaik az adatbázis tartalmától függenek.

modellek továbbfejlesztését és terjesztését. [Vegyük észre, itt mind olyan javaslatokról van szó, ami a kínai gazdaságpolitika lényegéből fakad: állami támogatást nyújtani a kívánatos fejlesztési irányoknak.]

A kormánynak nem korlátoznia kell a nyílt forráskódú MI-modellek exportját, hanem inkább olyan ösztönzőket kell alkalmaznia, melyek az új MI modellek megalkotásakor a saját chipgyártó háttér felhasználására késztetnek. Támogatni kell a saját, amerikai nyílt forráskódú MI-vállalatokat, mint amilyen a Google Tensor Flow-ja, a Meta PyTorch-ja, vagy a Llama, hogy versenyképesek legyenek a hasonló profilú, államilag támogatott kínai nagy nyelvi modellekkel. Együtt kell működnie az olyan kezdeményezésekkel, mint a Stargate, melynek keretében az Arm, a Microsoft, az Nvidia, az Oracle, az OpenAI, a Softbank és az MGX az elkövetkező négy évben 500 md dollárt szándékoznak befektetni Amerika MI-infrastruktúrájába. Létre kell hozni – az érdekelt vállalatokkal – egy integrált számítástechnikai „környezetet". Végsősoron egy olyan rendszert kell kialakítani, amely kedvezőbb a felhasználók számára, mint a kínai.

Továbbá: támogatni kell azokat az új modelleket, melyek túlmutatnak a jelenleg a csúcsot képező nagy nyelvi modelleken. Ilyenek a Cartesia Structured State Space modellje vagy az Inception diffúziós LLM-jei. Fel kell karolni mindazokat, akik új hardverek és szoftverek fejlesztésén dolgoznak. A nyílt kódolású szoftverek kivitelét akadályozó vámok helyett olyan politikára van szükség, amely arra ösztönzi a nyugati fejlesztőket, hogy saját chip-készletüket alkalmazzák. A trösztellenes eljárásokkal csínján kell bánni – hiszen most a kínaiakkal folyó versenyről van szó. Ugyanakkor fel kell lépni a kínaiak árdömpingje ellen.

Befejezésül a szerző azt nyomatékosítja, hogy a visszalépés a nagy nyelvi modellek nyílt kódolásúvá tételétől stratégiai hiba lenne. Továbbá: a nagy nyelvi modelleket és a mögöttes „tech ökoszisztémát" együttesen kell fejleszteni. „Vagy élen jársz vagy veszítel – nincs más opció."

Bp, 2025. március

Kiss Károly ismertetése