Budai Balázs Benjámin

A digitális állam fenyegetettsége: a mesterséges intelligencia

A digitális államot elég sokféleképpen jellemezhetjük. Két jelzőt azonban itt kiemelünk: technológiavezérelt és proaktív. Míg az első azt takarja, hogy az államnak minden olyan technológiát be kell építenie, amely hatékonyságát, hatásosságát és eredményességét fokozza, addig a második azt jelenti, hogy kérés nélkül szolgáltat, megelőzve (előzetesen kiszolgálva) minél több állampolgári igényt, kérést. A mesterséges intelligencia technológiái ezeket az elvárásokat – látszólag és egyelőre – jól hozzák. Fokozatosan bővül a felhasználási területek sora, az erre vonatkozó tapasztalatok is többnyire pozitívak. Azonban körvonalazódnak a veszélyek is, amelyekkel egyelőre nem foglalkozunk kellő mértékben. E rövid tanulmány ezekre a kihívásokra próbál rávilágítani. S mi mással lehetne egy pályája csúcsán álló társadalomtudós előtt tisztelegni, mint azzal, hogy egy, a társadalmi változás pozitív jellegét meghatározó technológiai kontextus árnyoldalait ízlelgetjük, továbbgondolásra késztetve az amúgy technofil olvasót…

Bevezetés

A homo sapiens alapvetően kényelmes (vagy inkább lusta) állat. Általában előnyben részesíti a kreativitást igénylő mozzanatokat az ismétlődőkkel szemben, és amit csak lehet, delegál azokból a feladatokból, amelyeket – magához mérten – szint alattinak gondol. Míg az első ipari forradalom idején a delegálás kimerült a gépek használatában, addigra ma már (Ipar 4.0) a gondolkodás feladatából is jócskán kiszervezünk, igaz, itt már nem csupán az ismétlődő jelleg miatt, hanem az ember számára kezelhetetlen inputmennyiség miatt is (*Big Data).*

A mesterséges intelligencia (MI) fogalma, McCarthy 1958-os definíciója[1] óta, folyamatosan változik, fejlődik. De nagy általánosságban olyan szoftvereket értünk

[1] McCarthy 1959: 77–84.

https://doi.org/10.36250/01183_12

alatta, amelyek adatokat képesek értelmezni (azaz adatokból információt generálnak), abból következtetéseket vonnak le, és annak megfelelően reagálnak valamilyen problémamegoldás érdekében. Ráadásul öntanulók (gépi tanulás), így minél több adatot értelmeznek és kezelnek, annál megfelelőbb reakciót tudnak adni.

Ugyanakkor ez okozza a – jelenlegi ismereteink szerinti – legnagyobb veszélyeket is: miután egyre nagyobb adathalmazokat kezelnek, és egyre nagyobb és összetettebb struktúrákat tekintenek át, és ebből következtetéseket vonnak le, nincs felettük elégséges emberi kapacitás, amely képes ezt kontrollálni, így tévedhetnek. Egy technológiai áttörésnél ráadásul többnyire jóindulatú, etikus felhasználást feltételezünk, és a visszaélésekre, diszfunkciókra csak később (olykor már túl későn) gondolunk.

Helyzetelemzés: az MI-alapú megoldások jelene és jövője

Az MI jelene

Az MI szoftveralapú és fizikai megoldásokkal van jelen hétköznapjainkban. Míg előbbit virtuális asszisztensek, kép-, hang-, valamint kép- és hangelemző szoftverek, kereső és elemző szoftverek, beszéd- és arcfelismerő rendszerek alkotják, addig utóbbit robotok, önvezető autók, drónok és a dolgok internetén (IoT) alapuló megoldások.[2] E vívmányokat látjuk az okosotthon-megoldásokban, a közlekedésfejlesztésben, az egészségügyi innovációknál és más – akár közigazgatási – szolgáltatásoknál is. (A terület jelentőségét és potenciálját mutatja, hogy a legtöbb fejlett ország – így hazánk is – Mesterséges Intelligencia Stratégiát készített.[3] A stratégia jelentős közigazgatási vonatkozással is bír: erősen fókuszál az automatizált döntéshozatali megoldásokra és az MI-alkalmazások használatára. Ugyanakkor az MI-vel járó veszélyekről nem ejt szót.)

Napjaink Magyarországán már a közigazgatás is használ MI-alapú megoldásokat: biometrikus azonosítót (gépi látás), chatrobotokat (MIA chatrobot, Nébó chatrobot), hangleiratozási és hangképzési szolgáltatásokat, önkiszolgáló ügyintéző terminálokat (KIOSK), automatikus döntéshozatali rendszert (AKD), valamint Rugalmas Adóellenőrzési Döntéstámogató és Adatbányászati Rendszert (RADAR).

[2] Európai Parlament 2020.

[3] *Magyarország Mesterséges Intelligencia Stratégiája 2020–2030* (2020).

MI-tendenciák, predikció

A Gartner, amely évek óta viszonylag megbízható technológiai előrejelzéseket ad a mesterséges intelligencián alapuló technológiák vonatkozásában is, 2022-ben kiadta előrejelzését. Szabadalmaztatott ábrázolásának alapja egy diagram (Gartner Hype), amely a technológiákkal kapcsolatos elvárások és az idő függvényében mutatja be az egyes technológiai megoldásokat, ezzel párhuzamosan láthatóságukat és megismerhetőségüket is figyelembe veszi. A Gartner a technológiai innovációk esetében öt érettségi fázist különböztet meg:[4]

1. Az innovációs trigger szakaszában csupán egy technológiai áttörés van a kezünkben. E fázisban még a felhasználási területet is csak körvonalaiban látjuk. A technológiát a média fokozódó érdeklődéssel követi, ugyanakkor itt találkozunk a legtöbb olyan megoldással, amely lehet, hogy lufiként viselkedik.
2. A túlduzzasztott elvárások csúcsán fokozódik a médianyomás, egymásra licitálnak a szakértők, ugyanakkor az innováció árnyoldalaival még mindig nem foglalkoznak kellő mértékben.
3. A kiábrándulás vályújába esnek azok az innovációk, amelyek esetében akkora mértékű lesz a problémák mennyisége, hogy az meghaladja a technológiával nyerhető hasznok nagyságát. Így a felhasználók és a befektetők elfordulnak a megoldástól, csökken a kutatási bázis is, az innováció perifériára csúszhat, vagy akár el is tűnhet.
4. A megvilágosodás emelkedője egy konszolidációs fázis, ahol a technológia használhatósága egyértelművé válik, a felhasználás és a felhasználók köre is tágul.
5. Végül a termékenység platóján áll be az a helyzet, amikor széles körű, általános technológiává válik a vonatkozó megoldás.

Ezek fényében külön érdekes megvizsgálni a 2022 nyarán kiadott predikciót. A legtöbb megoldás a technológiai áttörés vagy a felfokozott várakozások fázisában van, azaz valós teljesítményüket, alkalmazási területüket és várható hasznukat nem ismerjük. Nem ismerjük továbbá a velük járó hátrányokat sem. Érdekes látni azt is, hogy az innovációk néha kiegészítik egymást, néha ellentmondanak egymásnak az első két fázisban. Ami viszont jól körvonalazható, hogy négy fő kategóriában látunk innovációkat:

Adatközpontú MI: ez az algoritmusok betanításához használt adatok javítására és gazdagítására helyezi a hangsúlyt. Itt találhatók a tudásgrafikonok,

[4] Grafikus ábrázolását lásd: Wiles 2022.

az adatcímkézések és a szintetikus adatok. Ez utóbbi kiemelkedő jelentőségű lehet, hiszen minden olyan adatot szintetikus adatnak minősítünk, amelyet nem a való világ közvetlen megfigyeléséből, hanem valamilyen módszerrel, mesterségesen állítanak elő. (Ilyen módszer lehet a statisztikai mintavétel, a szemantikai megközelítés, a szimuláció stb.)

6. Modellközpontú MI: e megoldások a döntéstámogatás ok-okozati összefüggéseit erősítik. Elsősorban vertikális és horizontális folyamatok (eljárások, nb. közigazgatási eljárások) javításában lehet nagy szerepe. (Idetartozik a fizikai-informált MI, a kompozit MI, az ok-okozati MI, a generatív MI, az alapmodellek és a mély tanulás.)
7. Alkalmazásközpontú MI: kifejezetten alkalmazásokra fókuszál. (Idetartozik – többek között – az operatív AI-rendszerek tervezése, az intelligens robotok, a természetes nyelvi feldolgozás [NLP], az autonóm járművek és a számítógépes látás. De idesorolhatjuk a logikai játékokat, a beszédfelismerést, a szakértői rendszereket, a mesterséges neurális hálózatokat is.)
8. Emberközpontú MI: a közvetlen emberi vonatkozásokhoz kapcsolódik. (Idesorolható az MI-bizalom, a kockázat- és biztonságkezelés, a felelős MI, a digitális etika, valamint az MI oktatókészletei.) E területeknek legalább 5–10 évet jósol a Gartner, mielőtt a fősodor részeivé válnak.[5]

Összefoglalva a predikciót: a mesterséges intelligencia számos technológiai áttörést hozott az elmúlt néhány évben, azonban ezek felhasználási területeit egyelőre csak „felfokozott" várakozásokkal tudjuk követni, hiszen bőszen érvelünk amellett, hogy fokozott hatékonyságot és költségmegtakarítást hoznak, ráadásul jobb szolgáltatásokat alakíthatunk ki a segítségükkel. Ugyanakkor nem látjuk az ezekkel érkező problémák teljes körét, így erre nem is tervezünk lépéseket.

Közigazgatásra vetítve jól látszik, hogy a szintetikus adatok, a döntési modellek, az MI-alapú alkalmazások csomagjai folyamatosan érkeznek, azonban az ehhez tartozó garanciális elemek biztosításával adósak maradunk. Márpedig ha egy olyan adatintenzív ágazatot, mint a közigazgatás, ellep az MI, akkor – garanciák nélkül – a működés megbízhatósága kerül veszélybe. Nem véletlenül mondja Azamat Abdoullaev, hogy „az emberiség jövőjére nézve nem a klímaváltozás vagy a nukleáris háború jelenti a legnagyobb kockázatot, hanem az emberi szuperhülyeség, amely egy emberszerű, mesterséges intelligenciát tervez…"[6]

[5] Wiles 2022.
[6] Abdoullaev 2023.

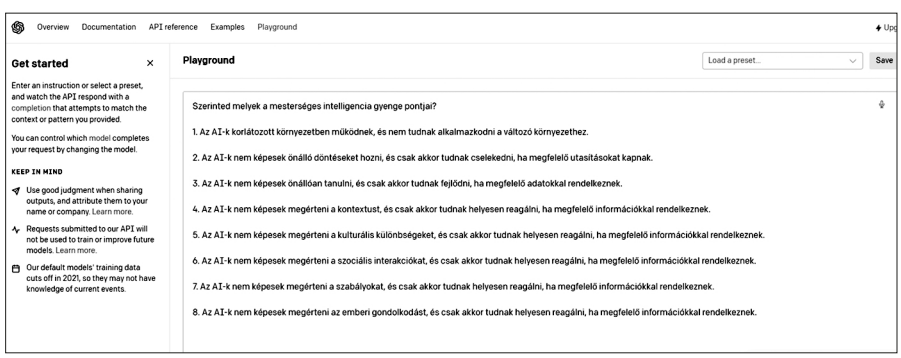
Miután az MI felhasználása és kutatása néhány évtizedre tekint vissza, még abban a fázisban vagyunk, ahol e technológia nem alakítja át annyira a társadalmi alrendszereket, hogy elkezdjük komolyan venni annak lehetséges hátrányait. Abban a pillanatban, hogy a digitális kitettség nagy mértéke tömegek munkáját teszi feleslegessé, hatalmas társadalmi feszültségek aktiválódnak, és elindul a társadalom szerkezetének bomlása.[7]

Lehetséges beavatkozási pontok

De ha minden ennyire rózsaszín, akkor mégis hol induljunk el a gyengeségek detektálásában? Két megközelítést használtam: egyrészt megkérdeztem egy chatrobotot saját magáról, másrészt áttekintettem a nemzetközi diskurzus sajtószemléjét e kérdéskörben.

Önreflexió

Megkérdeztem az Open AI fejlesztette ChatGPT-t, hogy hol látja a saját gyenge pontjait, így válaszolt:



1. ábra. *ChatGPT: a mesterséges intelligencia gyenge pontjai*
Forrás: https://platform.openai.com/playground/p/default-study-notes?model=text-davinci-003

A gép adta válasz is további kérdéseket vet fel, hiszen e válaszok több helyütt ellentmondanak a mesterséges intelligenciáról alkotott alapvető definícióknak, elég csak

[7] Sartor 2019.

az öntanulásra, a szabályértelmezésre és a döntésképességre gondolni. De az talán vitán felüli, hogy valóban korlátozott környezetben működnek, a kontextus értelmezésével problémájuk van, nem értik a kulturális különbségeket, a szociális interakciókat és azok árnyalatait, és az emberi gondolkodástól a gépi gondolkodás még nagyon messze van. (Azaz egyelőre nem várható el az érdemi döntéstámogatása a legtöbb olyan közigazgatási ügyben, amely diszkrecionális döntést igényel az ügyintéző részéről.)

Megkérdeztem másképp is:

3. ábra. *ChatGPT: a mesterségesintelligencia-alapú chat megbízhatósága*
Forrás: https://platform.openai.com/playground/p/default-study-notes?model=text-davinci-003

Azaz megerősítést nyert az a feltételezésem, hogy az MI nem tévedhetetlen. Főleg azokon a területeken, ahol a probléma komplex, hiszen az erre adott gépi válaszok „gyakran nem megfelelőek".

MI-gyengeségi diskurzus

A ProBonón vezetett *Hogyan csináljunk digitális térségeket?* c. önfejlesztő csatorna[8] sajtószemléjét áttekintve az MI gyengeségeit négy nagyobb csoportra oszthatjuk 2023-ban:

1. Hagyományos – az emergens technológiai területekre jellemző – sajátosság, hogy a technológiai fejlődés iránya és üteme nehezen kiszámítható, így szabályozása jelentős késéssel, reaktív módon történik.[9] Ez figyelhető meg számos diszruptív

[8] Lásd: https://probono.uni-nke.hu/onfejlesztes/csoportok-es-csatornak/bejegyzesek/5f72fc4b-466b0aeb680f6c84#post_64414bbc466b0ad3883ea427.

[9] Kaal 2016.

technológia esetében is, elég csak a blokkláncalapú megoldásokra gondolni, ahol a szabályozás még mindig csak próbálkozik a kérdés megragadásával. Az MI vonatkozásában szintén tehetetlennek mutatkozik a jogalkotás. Egyelőre (néhány adatvédelmi vonatkozáson túl) stratégiai szinten is csak homályos elvárásokat fogalmaz meg mind az uniós, mind a nemzeti szint. (Nem véletlenül okozott fennakadást az, hogy egy adatvédelmi incidens után Olaszország felfüggesztette a gyakran „tévedő" generatív MI, a ChatGPT alkalmazásának lehetőségét.) Hiányzik a felelősségvállalási kérdések tisztázása, megoldatlan az MI-vel összefüggő produktumok szellemi tulajdonjogi kérdéseinek rendezése.

2. Hiányoznak az etikai irányvonalak, amelyek az MI-technológiák alkalmazását keretek közé szorítják. Az etikus felhasználás és az azt övező civil és állami oltalmi gyűrű kialakítása elengedhetetlen. Nem kellően átgondolt az MI alkalmazásának versenyjogi, fogyasztóvédelmi aspektusa. Nem megoldott a magánszféra védelme, az elfogult vagy rosszindulatú (károkozó) algoritmusok elleni védekezés, ami komoly adatbiztonsági és adatvédelmi aspektusokat hordoz magában. (Egy MI rávehető – többek között – spamelésre, álhírek terjesztésére, adathalászatra, hamis weboldalak készítésére, malware-készítésre.)
3. Az emergens technológiai innovációk „felforgatják" a társadalmi alrendszereket. Már most számos elemzés[10] foglalkozik az automatizáció munkaerőpiacra gyakorolt hatásával. Az elemzések közötti fő különbség, hogy a digitális kitettség hatásait milyen időtávra időzítik. Ez csupán abból a szempontból jelentős, hogy mikorra kell felkészülniük azoknak a munkaerőknek, akiktől a „gépek elveszik a munkát". A felkészülést pedig tervezni és támogatni kell, hiszen az e munkakörökben dolgozók átképzésével optimalizálható a dolgozói *outplacement*. Ez a tervezés – jelen pillanatban – a közigazgatásból hiányzik.
4. Az MI használatának, fejlesztésének oktatási környezete is kezdetleges.

Következtetések

Az MI-alapú technológiák – mint emergens és diszruptív technológiák – térnyerése megállíthatatlan. Nem is lehet cél e technológiák megakasztása, sőt, még lassítása sem, annál inkább stratégiai eszközökként történő kezelése. Azonban sokkal fontosabb, hogy felhasználásukat azokon a területeken tervezzük és próbálgassuk, ahol a felmerülő gyengeségek és veszélyek aránya alacsonyabb. Ahol pedig az alkalmazás

[10] EZ [é. n.]; Adams 2022; Tham [é. n.]; O'Loughlin 2022.

veszélyekkel társul, fektessünk lényegesen nagyobb hangsúlyt ezek vizsgálatára, megoldására (döntően szabályozására)! Addig pedig olyan nemzetközi és nemzeti szabályozási környezet kialakítására van szükség, amely biztosítja az átláthatóságot, a méltányosságot, az elszámoltathatóságot, a versenyképességet és nem utolsósorban a biztonságot valamennyi aktor számára.

The Threat to the Digital State: Artificial Intelligence

The digital state can be described using several adjectives, but we will emphasize two: technology-driven and proactive. The former implies that the state should adopt all technologies that enhance its efficiency and effectiveness, while the latter means it should provide services without being asked, anticipating (pre-serving) as many citizen needs and requests as possible. Artificial intelligence technologies apparently meet these expectations – at least for the time being. The range of applications is gradually expanding, and the experience is mostly positive. Nevertheless, there are emerging threats that have not yet received sufficient attention. This brief paper aims to highlight these challenges. What better way to honour a social scientist at the peak of his career than by tasting the shadows of a technological context that determines the positive nature of social change, and encouraging the technophile reader to think further?

Felhasznált irodalom

Abdoullaev, Azamat (2023): What's New in Artificial Intelligence from the Gartner Hype Cycle. *Linkedin.com,* 2023. február 27. Online: https://www.linkedin.com/pulse/whats-new-artificial-intelligence-from-gartner-hype-cycle-abdoullaev/

Adams, Diane K. (2022): Five Predictions for the Future of Work: A Captivating Employee Experience. *Forbes.com,* 2022. december 2. Online: https://www.forbes.com/sites/forbeshumanresourcescouncil/2022/12/02/five-predictions-for-the-future-of-work-a-captivating-employee-experience/?sh=2522ed7645fa

Európai Parlament (2020): *Mi az a mesterséges intelligencia és mire használják?* Online: https://www.europarl.europa.eu/news/hu/headlines/society/20200827STO85804/mi-az-a-mesterseges-intelligencia-es-mire-hasznaljak

EY [é. n.]: *Are You Prepared for Future of Work?* Online: https://www.ey.com/en_se/workplace-of-the-future

Kaal, Wulf (2016): What Happens When Technology is Faster than the Law? *The CLS Blue Sky Blog,* 2016. szeptember 22. Online: https://clsbluesky.law.columbia.edu/2016/09/22/what-happens-when-technology-is-faster-than-the-law

Magyarország Mesterséges Intelligencia Stratégiája 2020–2030 (2020). Online: https://ai-hungary.com/api/v1/companies/15/files/137203/view

McCarthy, John (1959): Programs with Common Sense. In *Mechanisation of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory.* London: Her Majesty's Stationary Office, 77–84.

O'Loughlin, Henry (2022): 54 Future of Work Predictions for 2030. *Buildremote.co,* 2022. április 6. Online: https://buildremote.co/future-of-work/predictions/

Sartor, Giovanni (2019): *Artificial Intelligence: Challenges for EU Citizens and Consumers*. Online: https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/631043/IPOL_BRI(2019)631043_EN.pdf

Tham, Melissa [é. n.]: 6 Predictions for the Future of Work. *Noodlefactory.ai.* Online: https://articles.noodlefactory.ai/6-predictions-for-the-future-of-work

Wiles, Jackie (2022): What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle. *Gartner.com,* 2022. szeptember 15. Online: https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle

Wirtz, Bernd W. – Weyerer, Jan C. – Kehl, Ines (2022): Governance of Artificial Intelligence: A Risk and Guideline-based Integrative Framework. *Government Information Quarterly,* 39(4), 101685. https://doi.org/10.1016/j.giq.2022.101685