

ÚJ KORSZAKOT NYIT-E A MESTERSÉGES INTELLIGENCIA A DEZINFORMÁCIÓBAN?

*Az MI szerepének értelmezése a dezinformációs lánc
mint új elméleti keret segítségével*

Pintér Róbert*

1. Bevezetés

1.1. Hihetünk-e a fülünknek és a szemünknek?

A World Economic Forum 2024-es globális kockázati jelentése szerint a mesterséges intelligencia (MI) által generált dezinformációt és félretájékoztatást (*misinformation*) tartották az emberek a második legfontosabb olyan tényezőnek, ami világméretű válsághelyzetet idézhet elő a következő két évben. Első helyre a szélsőséges időjárást, harmadikra pedig a társadalmi és/vagy politikai polarizációt tették a kutatásban részt vevők.¹ Arról, hogy mit jelenthet egy ilyen dezinformáció okozta válsághelyzet, Martin Ford már 2022-ben érzékletesen írt az MI-ről szóló könyvében: képzeljük el, hogy két nappal az amerikai elnökválasztás előtt kiszivárog egy valódinak tűnő generált hangfelvétel az egyik jelöltről, amiben nevetve magyarázza, hogy mindig is fajgyűlölő volt:

„[m]iközben a felvételtől felrobban az internet, és a televíziócsatornák is újra meg újra leadják, a közösségi média világában eluralkodik a káosz és a felháborodás. A politikus ádáz harcok árán jutott el a jelölésig, és feldühödött ellenfelei most a visszalépését követelik.”²

* Egyetemi docens, Budapesti Corvinus Egyetem, Adatelemzés és Informatika Intézet.
ORCID: <https://orcid.org/0000-0001-7807-7997>

¹ The Global Risks Report. World Economic Forum (WEF), 2024.
https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

² Martin Ford: *Jövönk a robotok korában – Hogyan alakítja a mesterséges intelligencia az életünket?* Budapest, HVG, 2022. 207.



Sokan attól tartottak, hogy a 2024-es, világszerte nagyjából 2 milliárd embert és 60 országot érintő választásokon a mesterséges intelligencia segítségével generált dezinformáció a fentihez hasonló módon érdemben befolyásolhatja majd a szavazói döntéseket és ezzel aláássa a demokratikus működést.³ Az amerikai kongresszus 2023-ban törvényjavaslatot is előterjesztett, amely előírja az MI által generált tartalmak egyértelmű jelölését a politikai hirdetésekben.⁴ Ennek hatására a Meta (a Facebook anyavállalata) és más platformok is kötelezővé tették az MI-vel készített vagy módosított tartalmak megjelölését. Az intézkedések a dezinformáció veszélyeire reagáltak, különösen a *deepfake*⁵ és a személyre szabott politikai hirdetések kapcsán.⁶ A Pew Research Centre 2024 augusztusi felmérése szerint az amerikaiak 39%-a aggódott az MI választásokra gyakorolt potenciálisan káros hatásai miatt, ami mind a két politikai oldal szavazóira egyaránt jellemző volt.⁷

Egy-két, nagyobb hírverést keltő esetet leszámítva⁸ azonban az MI által generált tartalmak felhasználása szerencsére jóval alatta maradt a várakozásoknak és nem igazolta az előzetes félelmeket. A romániai elnökválasztás megsemmisített első fordulójának kivételével mindenhol rendben lezajlottak a 2024-es választások. Egyébként a romániai esetben sem az MI-vel generált tartalmak okozták a problémát, hanem a 2008-as Obama kampány óta⁹ politikai kommunikációs szempontból immár hagyományosnak tekinthető közösségimédia-platformok használata, elsősorban a TikTok-on zajló összehangolt dezinformációs kampány.¹⁰

Ennek ellenére a generatív MI-vel kapcsolatos félelmek továbbra is meghatározóak és az aggodalmak eljutnak a nagyközönséghez is, mivel olyan, széles körben ismert szakemberek is fenntartásuknak adnak hangot, akikre világszerte odafigyelnek. Sir Tim Berners-Lee, a WWW atyja egy interjújában kijelentette, hogy „sokkoló, mérgező

³ Ld. például Marina Adams: How AI-generated disinformation might impact this year's elections and how journalists should report on it. *Reuters Institute*, 2024. március 15. <https://tinyurl.com/yc3rrxtz>; vagy Vittoria Elliott: 2024 Is the Year of the Generative AI Election. *Wired*, 2024. május 30. <https://www.wired.com/story/2024-is-the-year-of-generative-ai-elections/>

⁴ 118th US Congress. S. 1596, 2023. május 13. <https://tinyurl.com/pxnvs4zm>

⁵ MI-vel szerkesztett vagy generált képek, videók és hangfelvételek, amelyek valós vagy nem létező embereket ábrázolnak.

⁶ Social Media, Disinformation, and AI: Transforming the Landscape of the 2024 U.S. Presidential Political Campaigns. *The SAIS Review of International Affairs*, 2025. január 14. <https://tinyurl.com/muuv722u>

⁷ Shanay Gracia: Americans in both parties are concerned over the impact of AI on the 2024 presidential campaign. *Pew Research Center*, 2024. szeptember 19. <https://tinyurl.com/wjemsv3r>

⁸ Ilyen jól dokumentált esetek voltak, amikor Biden elnök állítólag nem-szavazásra buzdított telefonon az előválasztáson, vagy egy már halott, de MI-vel feltámasztott indiai politikus támogatásáról biztosította az egyik jelöltet, illetve hírességek, például Eminem vagy Taylor Swift is látszólag politikai állást foglaltak és nem csak az amerikai választásokon. További példák elérhetőek a *Wired* magazin gyűjteményében arról, hogy hogyan használták a generatív MI-t a 2024-es választásokon: Vittoria Elliott: The Wired AI Elections Project. *Wired*, May 30, 2024. <https://www.wired.com/story/generative-ai-global-elections/>

⁹ Bruce Bimber: Digital Media in the Obama Campaigns of 2008 and 2012: Adaptation to the Personalized Political Communication Environment. *Journal of Information Technology & Politics*, Vol. 11., No. 2. (2014) 130–150. <https://doi.org/10.1080/19331681.2014.895691>

¹⁰ Théophane Hartmann: Commission opens TikTok investigation over Romanian presidential elections disinformation. *Euractiv*, 2024. december 17. <https://tinyurl.com/y8shvk3y>

algoritmusok” működnek a közösségi médiában és a megoldás a programozók kezében van. Szerinte a világháló „a dezinformáció, a polarizáció és a vállalati visszaélés játszótérévé változik”.¹¹ Az izraeli sztártörténész, Yuval Noah Harari szerint az MI a demokráciához szükséges társadalmi szintű párbeszédet lehetetleníti el. Szerinte a demokratikus berendezkedés mindaddig kisebb közösségekre (törzsekre, városállamokra) szorítkozott, amíg nem volt meg a technológiai lehetősége a társadalmi szintű párbeszédnek (vagyis a sajtó megjelenése előtt). A modern demokrácia alapja a párbeszéd, ha ez az MI miatt ellehetetlenül, az a demokrácia végét jelentheti.¹²

A dezinformáció és az MI kapcsolata nem csak külföldön, hanem Magyarországon is széles tudományos érdeklődésre tart számot. Nemrégiben például konferenciát is szenteltek a témának.¹³ A Mérték Médiaelemző Műhely 2023-ban dezinformációról szóló tanulmánygyűjteményt adott ki.¹⁴ Aczél Petra és Veszelszki Ágnes szerkesztésében a *deepfake*-ről jelent meg könyv számos szerző közreműködésével.¹⁵ Krekó Péter összeesküvés elméletekről és dezinformációról szóló könyve pedig több kiadást is megért, ami a téma iránti szélesebb, laikus közönség érdeklődését is jól mutatja.¹⁶ Krekó és munkatársai *Dezinformáció és Mesterséges Intelligencia Kutatócsoport* néven önálló kutatócsoportot is felállítottak a téma pszichológiai vetületeinek vizsgálatára az ELTE-n.¹⁷

Az MI-vel kapcsolatos aggodalmak a generatív MI, elsősorban a ChatGPT 2022 végi berobbanásával és használatának széles körben való elterjedésével erősödtek fel. A *deepfake* tartalmak kapcsán egyre gyakrabban hangoztatott állítás, hogy ezentúl ‘nem hihetünk a szemünknek és a fülünknek’,¹⁸ ami egyúttal azt is sugallja, hogy új korszakot hoz az MI a dezinformációban. Jelen tanulmányomban azt vizsgálom meg, hogy az MI milyen szerepet játszik a dezinformációban és vajon tényleg új korszakot jelenthet-e.¹⁹

¹¹ Sokkoló, mérgező algoritmusokra figyelmeztet a világháló atyja. *HVG*, 2024. december 8.

https://hvg.hu/tudomany/20241208_tim-berners-lee-interju-kozosségi-media-veszelyei

¹² Yuval Noah Harari: *Nexus – Az információs hálózatok rövid története a kőkorszaktól az MI-ig*. Budapest, Animus, 2024.

¹³ Nóra Falyuna – Katalin Feher – Márton Demeter – Gábor Szüdi – Joseph E. Uscinski: Against Disinformation: Bridging Science and Public Discourse. *Információs Társadalom* Vol. 24., No 2. (2024) 68–84. <https://dx.doi.org/10.22503/inftars.XXIV.2024.2.4>

¹⁴ Çakır-Somlyai Özge (szerk.): *A dezinformáció jelensége – tanulmánygyűjtemény*. (Mérték Füzetek 33.) Budapest, Mérték Médiaelemző Műhely, 2023. <https://tinyurl.com/dfpurnw3>

¹⁵ Aczél Petra – Veszelszki Ágnes (szerk.): *Deepfake: a valótlán valóság*. Budapest, Gondolat, 2023.

¹⁶ Krekó Péter: *Tömegparanoia 2.0 – Összeesküvés-elméletek, álhírek és dezinformáció*. Budapest, Athanum, 2023.

¹⁷ A kutatócsoportról bővebben ld. <https://tinyurl.com/3xv38xjz>

¹⁸ Ld. pl. Nagy Krisztina: Hihetünk-e még a szemünknek? Mit és hogyan tanítsunk a dezinformáció és a mesterséges intelligencia kapcsolatáról az Alfa és a Z generációnak? Konferencia előadás. *Mesterséges intelligencia – filozófia, jog és dezinformáció konferencia*, Budapest, 2024. november 20. <https://tinyurl.com/4mzcm9nw>

¹⁹ A tanulmányban nem foglalkozom érdemben a dezinformáció és az MI szabályozási vonatkozásaival, bár érintőlegesen több ponton is előkerül a kérdés. Akit ez a téma mélyebben érdekel, annak a következő cikket javaslom kiindulópontnak: Bilge Azgin – Sevki Kiralp: Surveillance, Disinformation, and Legislative Measures in the 21st Century: AI, Social Media, and the Future of Democracies. *Social Sciences*, Vol. 13., No. 10. (2024) 510. <https://doi.org/10.3390/socsci13100510>

Az egészen biztos, hogy sok mindenre képes a generatív MI, ami korábban nem volt lehetséges, legfeljebb csak nagy költségek és erőforrások mozgósítása révén. A generatív MI alkalmas szövegek, képek, mozgóképek és hangfelvételek létrehozására és módosítására. Ez a képesség az utóbbi években bárki számára hozzáférhetővé vált az interneten elérhető generatív MI eszközök (elsősorban a ChatGPT, illetve az olyan képgeneráló alkalmazások, mint a DALL-E vagy a Midjourney és például az ElevenLabs fémjelezte hanggeneráló szolgáltatások) elterjedésével. Csakhogy a technológia képességének megváltozása miatt önmagában még nem feltétlenül köszönt be új éra a dezinformációban. Hiszen elsősorban nem azért *nem csapjuk be egymást*, mert ne lennénk képesek rá (vagyis korábban ne lettek volna meg az eszközeink hozzá), hanem mert a viselkedési normák kölcsönös betartása és a bizalom az alapja az emberi társadalmak működésének. A kérdés – új korszakot hoz-e az MI a dezinformációban – vizsgálata tehát elsősorban nem az MI képességeiről szól, hanem azok emberi felhasználásáról, illetve ennek következtében a minket körülvevő világ működésének esetleges megváltozásáról.

1.2. A dezinformáció fogalma és a dezinformációs lánc

Írásomban a dezinformáció médiatudományi definíciójánál tágabban értelmezem a fogalmat. „Az Európai Bizottság Álhír- és Online Dezinformációs Csoportjának 2018-as definíciója szerint a dezinformáció olyan »hamis, pontatlan vagy félrevezető információ, amelyet a nyilvánosságnak való károkozás vagy a profitszerzés szándékával terveznek meg, mutatnak be és terjesztenek.«”²⁰ Meglátásom szerint ugyanis, mint látni fogjuk, a motivációk ennél jóval tágabbak lehetnek, miközben nem minden dezinformációt szánunk a széles nyilvánosságnak – gondoljunk csak az unokázós csalásokra. Az én értelmezésemben a dezinformáció olyan információ, amit függetlenül valóságtartalmának mértékétől, a terjesztés formájától, közlésének módjától, címzettjétől, befogadásától és ténylegesen elért hatásától, azzal a céllal hoz létre valaki, hogy azt szándékosan félrevezetésre használja. Így nem csak a hírek esnek bele ebbe a fogalomba, ez alapján kerülöm is az 'álhír' mint kifejezés általános használatát, miután az túlságosan leszűkítő. De a Barts Livia által meghatározott dezinformáció meghatározást – „[d]ezinformációról akkor beszélünk, amikor hamis vagy megtévesztő információt osztanak meg tudatosan, hogy kárt okozzanak”²¹ – sem tartom teljesen elfogadhatónak. Ugyanis, mint látni fogjuk, a dezinformáció mögött sokkal színesebb motivációkat találni, mint pusztán a *károkozás*, így a dezinformáció jelentős részét kizárnánk a vizsgálódásból, ha elfogadnánk ezt a meghatározást. A legközelebb Z. Karvalics László megközelítése áll hozzám: „~szándékosan felépített megtévesztő, félrevezető tartalom; a valóságnak a kiadó számára előnyös értelmezése vagy meghamisítása, amelyet a befogadó számára megalapozott és ellenőrzött információként tárnak, és a hatás elérése érdekében hihetővé tesznek.”²²

²⁰ Bajomi-Lázár Péter: Álhírek és igaz hírek – Definíciós kérdések. In: Çakır-Somlyai (szerk.) i. m. 8.

²¹ Barts Livia: Vizuális információs zavarok – típusok és módszerek. In: Çakır-Somlyai (szerk.) i. m. 35.

²² Z. Karvalics László: *Informatorium – Word Guide to Contemporary Information Culture*. Budapest, Tinta Publishing House, 2017. 142. (saját fordítás P.R.)

A dezinformáció fogalompárja a miszinformáció (félretájékoztatás vagy téves információ). Az én értelmezésemben ez olyan dezinformáció, ahol a terjesztésben részt vevő nem tud arról, hogy dezinformációt terjeszt. Vagyis ugyanaz az információ a dezinformáción túl miszinformáció is lehet a detektálástól, a befogadástól és a terjesztés módjától függően.

A cikkben a kutatási kérdés vizsgálatának részeként javaslatot teszek a dezinformáció egyfajta új értelmezési keretére, amit *dezinformációs láncnak* neveztem el. Azt gondolom ugyanis, hogy a dezinformáció szisztematikus vizsgálata segíthet annak eldöntésében, vajon újat hozhat-e a dezinformációban a mesterséges intelligencia. A dezinformációs lánc gondolata gyakorlatilag annak a nyolc hónapos kutatómunkának az eredménye, amit a cikk elkészítése igénybe vett. A fogalmat a szakirodalmak olvasása és a dezinformációs jelenségek értelmezése közben alakítottam ki, annak felismerése nyomán, hogy ezek a jelenségek felfűzhetők egyetlen láncra, ami támogatja azok kapcsolatának jobb megértését.

A következő fejezetekben tehát a dezinformációs lánc hat kulcs lépésén megyek végig és vizsgálom meg azokat egyenként, hogy jelenleg mit tudunk róluk és milyen változások érhetők tetten bennük az MI-vel való kapcsolatuk alapján. A dezinformációs lánc főbb lépései (a lépések kezdőbetűit összeolvasva MoLTe(De)BeHa):

1. Motiváció: a dezinformációt folytató motivációjának azonosítása
2. Létrehozás: a dezinformációs üzenet létrehozásának aktusa
3. Terjesztés: a dezinformáció terjesztésének folyamata
4. (Detektálás: a dezinformáció tényének detektálása – amely nem minden esetben történik meg, ezért itt zárójelben szerepel)
5. Befogadás: a célszemély(ek) részéről a dezinformáció percepciója, feldolgozása és befogadása
6. Hatás: végül a dezinformációval elért hatás, illetve a szélesebb (nem szándékolt) társadalmi következmények.

Mint később látni fogjuk ez a fajta megközelítés jelentősen kibővíti Wardle és Derakhshan információs zavarokról szóló leírását (aminek része a dezinformáció is), ahol a szerzők három fázist különítettek el az információs zavarok vonatkozásában, ezek a létrehozás (az üzenet létrehozása) – (újra)termelés (üzenet, mint média termék) – terjesztés (üzenet nyilvános megosztása).²³

2. A dezinformációs lánc hat kulcs lépése

A jelenség súlyának ismertetése, a központi kutatási kérdés bemutatása és a fogalmi tisztázás után vegyük végig a dezinformációs lánc egyes lépéseit és vizsgáljuk meg abban az MI lehetséges szerepét, különös tekintettel a generatív MI-re.

²³ Claire Wardle – Hossein Derakhshan: *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Strasbourg, Council of Europe Report, 2017. <https://tinyurl.com/4z7se79j>

2.1. Első lépés: motiváció

Bár, mint fentebb láttuk, a szakirodalom hajlamos elsősorban politikai motivációkra visszavezetni a dezinformációt, valójában még a politikai hatással bíró dezinformáció mögött is lehet egészen más indíték. Például a 2016-os amerikai elnökválasztás alatt macedón fiatalok kiterjedt dezinformációs hálózatot üzemeltettek, ami mögött pusztán a meggazdagodás vágya állt. A jól dokumentált esetben az autodidakta, angolul is rosszul tudó ‘vállalkozók’ azt használták ki, hogy az online hirdetési platformok az internetes forgalom mérete alapján fizetnek a hirdetést futtató oldalaknak.²⁴ Mivel az amerikai választáshoz kötődő álhírekkel nagy forgalmat lehetett generálni, ezért dezinformációs oldalakat hoztak létre, amelyek óriási látogatottságot produkáltak. A történetben a veszezi fiatalok szemfülességén és etikátlanságán túl legalább ennyire fontos az interneten bevett reklám alapú üzleti modell szerepe és azon cégek felelőssége, amelyek nem tiltják a dezinformációra alkalmas tartalom mellett hirdetések futtatását. Nagy valószínűséggel persze ez az eset (gazdasági okokból felépülő politikai dezinformációs gépezet) inkább kivételnek, semmint általánosnak tekinthető. De a célzott, akár személyre szabott dezinformációs aktivitások mögött is sokszor gazdasági haszonszerzés áll (ilyenek például a *deepfake* hirdetések, a pénzhamisítás, a festmények hamisítása vagy az unokázós csalás).

Természetesen a legtöbb, nagy hatással bíró dezinformáció mögött politikai motivációt, zavarkeltést, vagy egyenesen a demokratikus intézmények működésének aláásását találni. Ugyancsak ide sorolható a zaj és információs túltelítettség, tényrelativista helyzet létrehozása. Ezeknél a cél, hogy „olyan »fehér zajt« – információs túlterheltséget – teremtsenek, amelyben nehéz megkülönböztetni az igaz és a hamis állításokat”.²⁵ Miután a cikkben később a politikai indíttatású dezinformáció áll majd az elemzés középpontjában, így erről itt nem írnék bővebben.

Egy további fontos motiváció lehet a dezinformációra a szórakoztatás és a humor. Ezek egyik legismertebb formája a *mockumentary* (fiktív eseményeket ábrázoló, de dokumentumfilmként bemutatott műsor, például reality sorozatok) és az áldokumentumfilmek. Utóbbi „kritikára, szatírára és önreflexióra alkalmas formátum”, ami arra készítt, hogy reflektáljunk a „normáinkra, értékeinkre, ideológiáinkra és a létünkre”.²⁶ Ennek alesete a társadalomkritika megfogalmazása vagy a dezinformációnak való kitettség ábrázolása dokumentumfilmes eszközökkel. Magyarországon ennek legismertebb példája az Igazi Mao²⁷ című áldokumentumfilm volt, de más országokban is ismert, lásd például az Olajfalók²⁸ című cseh vízgátfilmet. De a dokumentumfilmben

²⁴ Heather C. Hughes – Israel Waisman-Manor: The Macedonian Fake News Industry and the 2016 US Election. *PS: Political Science & Politics*, Vol. 54., No. 1. (2021) 19–23.
<https://doi.org/10.1017/S1049096520000992>

²⁵ Bajomi-Lázár i. m. 7.

²⁶ Jamriskó Tamás: Mockumentary – Műfaj a valóság határán – Audiomanipulációs gyakorlatok. In: Čakır-Somlyai (szerk.) i. m. 56.

²⁷ Az Igazi Mao (1995): <https://www.imdb.com/title/tt0109184/>

²⁸ Olajfalók (Ropáci – 1988): <https://www.imdb.com/title/tt0096010/>

látható dezinformáció ölthet további formákat is, mint amilyen például a fogyasztói társadalom kritikáját bemutató Cseh álom²⁹ című alkotás.

Társadalmi vonatkozásait illetően fontos motiváció lehet még a dezinformációra a szexuális vágykeltés vagy az erre alkalmas tartalmakkal való visszaélés. E mögött gyakorta szakítás miatti bosszút, gyerekes csínytevést vagy üzleti érdeket találni. Az interneten megosztott *deepfake* pornó komoly problémát jelent, mivel traumatizálhatja azokat, akiket az ilyen videókra vagy képekre montíroznak és tönkretelheti az életüket.³⁰ De komoly törést okozhat azok életében is, akiket ilyen tartalmak készítése miatt perbe fognak – a közelmúltban idehaza és világszerte is hallani ilyen esetekről iskolások körében.³¹

Ha megvizsgáljuk a hamisítások történetét, akkor számos további motivációt is láthatunk:³²

- a hatalom megszerzése vagy megtartása (például Constantinus állítólagos adománylevele a katolikus egyháznak)
- presztízs és tudományos elismertség (tudományos hamisítások, például a formosai – tajvani – útleírás George Psalmanazar tollából, vagy a piltdowni ember, amivel Charles Dawson írta be magát a tudománytörténetbe)
- életben maradás (pl. okmányok hamisítása a II. világháborúban)
- a titkosszolgálatok, katonai hírszerzés megtévesztő műveletei (háborús veszteségek eltitkolása és az ellenség manipulálása például a második világháború partraszállásának helyszínét illetően)

Tehát a motivációt illetően elég sokszínű képet látni. De vajon tud-e az MI maga dezinformációt létrehozni, tehát lehet-e saját motivációja mások megtévesztése? Az MI-kutatás elmúlt közel hetven évét domináló standard modell ezt a lehetőséget kizárja. E szerint ugyanis az MI „a helyes dolgot művelő ágensek tanulmányozására és építésére összpontosított. A helyes dolgot az ágensnek általunk kiosztott célkitűzése adja meg.”³³ Vagyis az MI-nek nem lehet saját célkitűzése és így motivációja sem, azt tőlünk, emberektől kapja. Ugyanakkor időről-időre röppennek fel szennzációs hírek, amelyek ennek az ellenkezőjéről szólnak,³⁴ de az MI saját motivációja inkább kivételnek és bizonyításra váró tulajdonságnak tűnik, mintsem valóságnak.

²⁹ Cseh álom (Český sen – 2004): <https://www.imdb.com/title/tt0402906/>

³⁰ Karen Hao: Deepfake Porn is Ruining Women's Lives. Now the Law May Finally Ban it. *MIT Technology Review*, 2021. február 12. <https://tinyurl.com/5n746f2h>

³¹ Ld. pl. RTL: Gyermekepornográfiának is minősülhet az általános iskolai osztálytársról készült deepfake videó. *RTL*, 2024. május 16. <https://tinyurl.com/2xdrh9k2>; vagy Caroline Haskins: Florida Middle Schoolers Arrested for Allegedly Creating Deepfake Nudes of Classmates. *Wired*, 2024. március 8. <https://tinyurl.com/y8pcz95f>

³² A példák a brüsszeli Európai Történelem Háza utazó kiállításából származnak: *Valódi hamisítvány: A hamisítvány és a hamisítás története*. <https://tinyurl.com/4aheym2p>

³³ Stuart Russell – Peter Norvig: *Mesterséges intelligencia – Modern megközelítésben*. Budapest, Panem Könyvek, 2023. 5.

³⁴ Például Billy Perrigo: Exclusive: New Research Shows AI Strategically Lying. *Time*, 2024. december 18. <https://time.com/7202784/ai-research-strategic-lying>

Annak ellenére, hogy a motivációk sokrétűek a dezinformáció mögött, a cikk fókuszaltságának megőrzése érdekében és a terjedelmi korlátok miatt a továbbiakban elsősorban a politikai motivációval jellemezhető dezinformációs jelenségeket vizsgálom meg mélyebben. Ezek ugyanis azok, amelyek egyúttal nagyobb közönséghez is eljutnak és átfogó társadalmi hatással is bírnak, valamint a szakirodalom is jelentős részben ezekre szűkíti le a dezinformáció vizsgálatát.

2.2. Második lépés: létrehozás

A dezinformációs lánc második lépését, a dezinformáció létrehozását érdemes tágan értelmezni. Wardle tanulmányában az információs zavarok hétlépcsős skáláját különíti el, ami alapján kijelenthető, hogy számos módon lehetséges dezinformációt létrehozni.³⁵ Az első a szatíra vagy paródia, amely nem ártó szándékú, de a kontextusából kiemelve félreérthető lehet.³⁶ A második a hamis kapcsolat, amikor a cím és a tartalom nincs összhangban, csökkentve az ezt a gyakorlatot folytató médiába vetett bizalmat. A harmadik a félrevezető tartalom, ahol manipulált képek vagy adatok torzítják el a valóságot. A negyedik a hamis kontextus, amikor egy valódi tartalmat félrevezető módon használnak fel, például egy videójátékból vett képpel illusztrálnak egy valódi háborús hírt. Az ötödik az impozitor tartalom, amikor egy hiteles forrást hamisan tüntetnek fel, például más nevében kommunikálnak. A hatodik a manipulált tartalom, például amikor utólag eltüntetnek valakit egy képről. A hetedik pedig a hamis tartalom, ahol mindent a dezinformáció céljából hoznak létre. Mint látható Wardle lépcsője egyre aktívabb bevonódást feltételez, míg az elején egy félreértésből is lehet dezinformáció, addig az utolsó lépésnél már minden a cél érdekében kreált.

A létrehozás, mint dezinformációs lépés a generatív mesterséges intelligencia képességének rohamos javulásával kezdett egyre nagyobb figyelmet kapni. A generatív MI segítségével ugyanis létrehozhatók szövegek, képek, hangok, mozgóképek is, amiket akár élő telefon- vagy videóhívásban is felhasználhatnak. Létező megoldások közé tartozik a videón látható arcok kicserélése és a *lip syncing* (ajkak szinkronizálása) a kimondott tartalommal összhangban. A The Verge az álhírek lehetséges jövőjét illusztrálандó már 2018-ban olyan videót tett közzé, amelyben egy amerikai komikus, Jordan Peele beszél, de a néző Obamát látja és hallja helyette.³⁷ A technológia képessége azóta csak javult, 2024 februárjában egy hongkongi bank vezető alkalmazottja 25 millió dollárt utalt át csalóknak egy olyan online meeting után, ahol rajta kívül mindenkit – a cég pénzügyi igazgatóját és minden további kollégát – MI személyesített meg.³⁸ Az MI-vel

³⁵ Claire Wardle: Fake News. It's Complicated. *First Draft*, 2017. február 16. <https://firstdraftnews.org/articles/fake-news-complicated/>

³⁶ Fontos, hogy a szatíra vagy paródia nem tekinthető önmagában dezinformációnak, mivel nem félrevezetés céljából hozták létre. Félrevezető akkor lehet, ha a kontextusából kiemelik, például amikor azt valóságosnak próbálják feltüntetni.

³⁷ James Vincent: Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news. *The Verge*, 2018. április 17. <https://tinyurl.com/2f7mb9b9>

³⁸ Heather Chen – Kathleen Magramo: Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'. *CNN*, 2024. február 4. <https://tinyurl.com/yeyvxyvc>

készített hamis videók annyira hitelesek lehetnek, hogy akár magukat a videóra hamisított szereplőket is elbizonytalaníthatják egy pillanatra, ha találkoznak a felvétellel, hogy vajon tényleg mondtak-e olyat, amit a szájukba adtak.³⁹

Mivel ma már akár pár másodpercenyi hang és mozgókép is elég lehet valakiről a hamis megszemélyesítéshez, fontos, hogy az MI-modellek korlátozva legyenek, szabályozott módon és csak vízjelezve lehessen tartalmakat létrehozni velük. Ugyanakkor feltehető, hogy az általában korlátozottan felhasználható zárt MI-modellekkel szemben az *open source* MI modellek sokkal szabadabban használhatóak. Sőt, a kellő erőforrással bíró, dezinformációt rendszeresen végző szervezetek feltehetően saját céljaiknak megfelelően feltanított vagy módosított MI eszközökkel is rendelkezhetnek, amelyeket nem korlátoznak etikai elvek vagy szabályozásoknak való megfelelési kényszer.⁴⁰

Természetesen egy MI által generált, vagy bármilyen eszközzel létrehozott dezinformáció csak akkor éri el a hatását, ha nem, vagy nehezen detektálható. Ahogy az az áldokumentumfilmekből is látható, a műfaji sajátosságokhoz, kommunikációs stílushoz kell alkalmazkodni ahhoz, hogy valami hihető legyen. Bajomi-Lázár szerint a dezinformációban is kell lennie hihető elemnek, tehát „az álhírek és az igaz hírek szembeállítására hamis dichotómiát teremt, mert egy hír sosem teljesen igaz vagy teljesen hamis”.⁴¹ Vagyis a létrehozásának fontos eleme a hihetőség kontextusának megteremtése.

Az MI által létrehozható dezinformáció paradoxona viszont az, hogy bár a generatív MI megjelenésével drasztikusan megnőtt a dezinformációhoz használható eszközök száma és szinte bárki számára elérhetővé váltak, mégsem tapasztaltuk 2022 óta az ilyen jellegű dezinformáció megsokszorozódását:

„Eddig azonban nem láttuk a mesterséges intelligencia alapú dezinformáció várt áradatát. A 2024-es európai parlamenti választások is viszonylag nyugodtan zajlottak, nagyobb mesterséges intelligencia által generált dezinformációs kampányok nélkül, és úgy tűnik, hogy a korábbi választások is viszonylag immunisak voltak a mesterséges intelligencia által generált dezinformáció hatására.”⁴²

Vagyis a várakozások ellenére a generatív MI okozta korszakváltás elmaradni látszik – legalábbis, ami a létrehozást illeti. Ahhoz, hogy ennek okait megértsük, a dezinformációs lánc további lépéseit kell megvizsgálnunk – mint látni fogjuk ugyanis, nem

³⁹ Ezen a rövid videón például Joe Biden elnök arról beszél 2023-ban, hogy saját magáról látott egy *deepfake* videót és az egy pillanatra őt is megtévesztette, amíg rá nem jött, hogy ő sosem mondott ilyesmit: Biden Says He's Seen 'Mind-Blowing' AI Deep Fakes of Himself. *Bloomberg Quicktake*, 2023. október 31. <https://www.youtube.com/watch?v=CIDXoUoL9Qo>

⁴⁰ Noémi Bontridder – Yves Pouillet: The role of artificial intelligence in disinformation. *Data & Policy*, Vol. 3., e32 (2021) <https://doi.org/10.1017/dap.2021.20>

⁴¹ Bajomi-Lázár i. m. 6.

⁴² Judit Bayer: Disinformation and AI in the AV Sector. In: *AI and the audiovisual sector: navigating the current legal landscape*. European Audiovisual Observatory, Strasbourg, 2024. 81. <https://tinyurl.com/3stnzt88> (saját fordítás, P.R.)

a dezinformáció létrehozása jelenti a legnagyobb kihívást a dezinformációt folytatni kívánó szereplők számára.

2.3. Harmadik lépés: terjesztés

A dezinformáció terjesztése az egyik leginkább kutatott téma a dezinformációs lánc egyes lépései kapcsán. Nem véletlenül, ugyanis ez a kulcsa annak, hogy egyrészt detektálásra kerül-e a dezinformáció (erről részletesen lásd a következő részt), másrészt milyen lesz a befogadása és az hogyan hat vissza magára a terjesztésre, végül, hogy milyen hatást ér el a dezinformáció (ezekről is hosszabban írok később). Tehát a terjesztés, detektálás, befogadás és hatás szorosan egymásba fonódnak és csak az elemezhetőség miatt választottam külön őket. Összefonódásuk tetten érhető például a miszinformáció (félretájékoztatás) és a dezinformáció fentebb már taglalt különbségében is (a miszinformáció a dezinformáció akaratlan terjesztése).

Krekó szerint a dezinformációt terjeszthetik meggyőződéses összeesküvés hívók, akik hatalom és eszközök nélkül frusztráltan marginalizálódtak. De terjeszthetik anyagi érdekből is, mint láttuk a macedón fiatalok példáján. Végül a hatalomban levő populista vezetők hatalmuk megszilárdítása végett. Viszont, ami kiemelten fontos, hogy az elitek és a tömegek együttműködnek a dezinformáció terjesztésében:

„Éppen ezért téves az a megközelítés, amely az álhírek »fogyasztóit« a meggyőzés folyamatának ártatlan áldozataiként láttatja. Lehet, hogy a dezinformációt az elit gerjeszti, de az elitek és a »nép« együttműködnek egymással a törzsi mítoszok előállításában és fenntartásában.”⁴³

Mint később látni fogjuk a dezinformáció kapcsán a közösségi média és az ahhoz köthető MI alapú algoritmusok jelentik a korszakalkotó lényegi újdonságot. Ami ugyan nem köthető az új generatív MI-hez, de a közösségi média algoritmusai miatt mégis MI-központú jelenség. Ináncsi szerint a közösségimédia-tartalmak terjedésében a kulcs a moderáció (erről a detektálásról szóló következő részben írok bővebben is), ugyanakkor:

„[a] használt moderációk (egyéni szinten történő felülvizsgálatok, algoritmus alapú moderáció, vagy teljes moderálásmentesség, illetve az alternatív moderációk) nem alkalmasak arra, hogy megfelelően szűrjék a felhasználók által megosztott tartalmakat.”⁴⁴

Ennek oka több, egymással rokon jelenség, amelyeket *astroturfingnek*, szimulált tömegnek, összehangolt nem hiteles viselkedésnek és részvételi propagandának nevezünk. „Az astroturfing a szakirodalomban az online visszaélések egy altípusaként

⁴³ Krekó i. m. 296–297.

⁴⁴ Ináncsi Mátyás: Közösségi média és dezinformáció. *Military and Intelligence Cybersecurity Research Paper*, 2023/7. 16. <https://tinyurl.com/bdfepxc3>

jelenik meg, amikor politikai kampányok grassroots, spontán viselkedésnek tüntetik fel azt, amit a valóságban egy ember vagy egy szervezet hajt végre.⁴⁵ Tehát egyetlen, dezinformációt folytató személy vagy szervezet több hamis közösségimédia-profil is kezelhet, amivel egyrészt elfedi a dezinformáció koordinációját, másrészt miközben spontánnak tűnő tömeget hamisít a tevékenység mögé, a meglévőnél jóval nagyobb társadalmi relevanciát is sugall. Bár a jelenség a közösségi média előtt is létezett (például fizetett tüntetők), de a közösségi média részvételi jellege miatt sokkal könnyebbé tette a dezinformációs kampányok menedzselését. Ezt a tevékenységet a közösségimédia-platformokat üzemeltető cégek összehangolt, nem hiteles viselkedésnek tekintik – például, amikor fiókok százai vagy ezrei osztják meg közel egyidőben ugyanazt az üzenetet.

Az igazi problémát azonban az jelenti, hogy a közösségi médiában a dezinformációt nem csak a létrehozók terjesztik koordináltan, hanem ebbe bárki – akár véletlenül is – bekapcsolódhat, aminek következménye, hogy „a közösségi média révén a »propagandista« és a »célközönség« közötti határ elmosódik, a korábban passzív közönség pedig egyre jelentősebb szerepet játszik a propagandisztikus tartalmak terjesztésében.”⁴⁶ Ezt nevezi a szakirodalom részvételi propagandának.⁴⁷ Ugyanakkor az a jelenség évtizedek óta ismert, hogy a politikai vagy marketing kommunikációban a létrehozó a célcsoport tagjainak erőforrásait használja az üzenetek terjesztésére. Csakhogy ez a közösségi média előtt legfeljebb e-mailt és SMS-t jelentett és csak tudatos mozgósítás árán tudott igazán tömeges méreteket ölteni.⁴⁸ Ami új, hogy „az üzenetek továbbítását programozott botok és valós emberek együttesen végzik el. Ez utóbbi pedig azzal jár, hogy lehetetlenné válik szétválasztani, hol végződik a hamisítás (koordináció) és hol kezdődik a hitelesség (organikus viselkedés).”⁴⁹ Ezzel a hagyományos tömegmédián túl létrejött egy olyan új tér, ahol a politikai (és gazdasági) üzenetek hatékonyan eljuttathatók és terjeszthetők a közönség aktív bevonásával. Mindez persze kétélű:

„[a] részvételi kultúra így egy lehetőség, de egy kitettség is a politika számára: minél többen felkapják, annál hitelesebbé válik az információ, a megfelelő üzenetek kiválasztása és továbbítása felett azonban az intézményes szereplők könnyen elveszíthetik a kontrollt.”⁵⁰

Ennek a hatékonyságát két további dolog is jelzi. Egyrészt az így terjesztett dezinformációs tartalmak akár a Google keresési lista élére is tudnak kerülni (kikerülve

⁴⁵ Német Szilvi: Szűrkezőzés médiumok, állampolgári kampányok és részvételi propaganda a közösségi média platformokon. In: Çakır-Somlyai (szerk.) i. m. 68–69. [A továbbiakban: Német (2023a).]

⁴⁶ Uo. 73.

⁴⁷ Alicia Wanless – Michael Berk: Participatory Propaganda – The Engagement of Audiences in the Spread of Persuasive Communications. *Social Media & Social Order, Culture Conflict 2.0 Conference*, Oslo. 2017 november. <https://tinyurl.com/ycxbyyh8>

⁴⁸ Ld. pl. Sükösd Miklós – Dányi Endre: E-mail és SMS a 2002-es választási kampányban. *eDemokrácia Műhely*, év nélkül. <https://www.edemokracia.hu/program/smskampany.php>

⁴⁹ Német (2023a) i. m. 73.

⁵⁰ Uo. 74.

a pusztán közösségi médiából) és onnantól kezdve hitelesíti őket a Google, mint az internetes keresés legfontosabb nyugati szereplője.⁵¹ Másrészt ezek a dezinformációs tartalmak „gyorsabban, messzebbre és szélesebb körben” terjednek, mint a cáfolatuk vagy az igazi hírek.⁵² Ezeket ugyanis nagyobb eséllyel és hamarabb osztják meg az emberek, mivel „a hamis hírek gyakran érdekesebbek, fontosabbnak tűnnek, rájátszanak az előítéleteinkre, azaz sokkal inkább megosztásra sarkallnak, mint az igazság.”⁵³

A szűrőbuborék (*filter bubble*) és a visszhangkamra (*echo chamber*) fogalmak is segíthetnek megérteni a közösségi médiában terjedő részvételi propaganda sikerességének okait, bár évtizedes vita folyik mindkét jelenségről.⁵⁴ Röviden, a közösségi média személyre szabott módon, az érdeklődésünk és korábbi aktivitásaink mentén szűri a nekünk szánt információkat (szűrőbuborék). Ezen felül pedig olyan személyekkel vesz körbe minket, akik a saját magunk által is hangoztatott, hozzánk hasonló véleményeket osztják (visszhangkamra). Ugyan a közösségi média használata a tájékozódásban nem feltétlenül kizárólagos, de a két jelenség abban mégis meghatározó, hogy mit láthatunk magában a közösségi médiában, így közvetve arra is hatással bír, hogy mivel keveredünk interakcióba, például mit lájkolunk vagy osztunk meg és így terjesztünk tovább. Mint később látni fogjuk, emiatt a befogadásra és a társadalmi következményekre is érdemi hatásuk lehet.

A botrányaiba belebukott Cambridge Analytica gyakorlata arra mutatott rá, hogy a dezinformációk terjesztésének hatékonyságát növeli a személyre szabás, avagy a mikrocélzás (*micro targeting*) is.⁵⁵ A felhasználók aktivitásainak követésével a közösségimédia-platformok lehetővé teszik hirdetőik és tartalomkészítőik számára, hogy meghatározott demográfiai vagy érdeklődési csoportoknak lehessen üzeneteket eljuttatni, így a dezinformáció könnyebben találja meg az arra fogékony közönségét. Az ilyen magas fokú célzás nagyobb manipulációt képes elérni és megsokszorozza a megtevesztő tartalmak hatását.⁵⁶ Nem véletlen, hogy a Cambridge Analytica által használt dezinformációs technikákat az információs hadviselés eszközei közé sorolhatjuk.⁵⁷

Természetesen a közösségi médiának a dezinformáció terjesztésében való felértékelődésével nemcsak a kutatók vannak tisztában, hanem azok is, akik dezinformációs kampányokat folytatnak. Például Oroszország több közösségimédia-platformon is aktív és szisztematikus módon terjeszti a dezinformációt álhírgyárak segítségével. Ennek

⁵¹ David Gilbert: How Disinformation From a Russian AI Spam Farm Ended up on Top of Google Search Results. *Wired*, 2024. július 9. <https://tinyurl.com/5n86ptb7>

⁵² Pavel Nikolic: A visegrádi négyek és a dezinformáció. *CEU Democracy Institute*, év nélkül. <https://cmds.ceu.edu/visegrad-negyek-es-dezinformacio>

⁵³ Uo.

⁵⁴ Erről kiváló cikk olvasható Gálik Mihály tollából: Gálik Mihály: A hálózati hírmédia sajátosságai, különös tekintettel a visszhangkamra és a szűrőbuborék-jelenségre. In *Medias Res*, 2019/2. 330–342. <http://real.mtak.hu/108287/1/imr-2019-02-08-jav-1.pdf>

⁵⁵ Jessica Dawson: Microtargeting as information warfare. *The Cyber Defense Review*, Vol. 6., No. 1. (2021) 63–80. <https://www.jstor.org/stable/26994113>

⁵⁶ Bontridder–Poullet i. m.

⁵⁷ Dawson i. m.

egyik központja a szentpétervári Internet Research Agency,⁵⁸ ami 2013 óta működik és kiterjedt hálózatot tart fenn fizetett propagandisták (mai kifejezéssel trollok) és hamis közösségi fiókok alatt rejtőző bothadseregek segítségével.⁵⁹ De az amerikai titkosszolgálat szerint a döntően meghatározó orosz tevékenységen túl Kína és Irán is folytatott dezinformációs tevékenységet például a legutóbbi elnökválasztás idején, hogy aláássa az amerikai demokratikus intézmények működését és társadalmi megosztottságot szítson.⁶⁰

2.4. Negyedik lépés: (detektálás)

A dezinformációs lánc negyedik lépése a detektálás, ami annyiban tér el a többi lépéstől, hogy nem mindig történik meg (emiat szerepel a fejezet címében is zárójelben). Mindez érthető is, hiszen a létrehozó célja a megtévesztés, amit könnyebben ér el, ha nem azonosítja dezinformációként az üzenetét. Fontos, hogy a detektálás csak a terjesztéssel és a befogadással szoros összefüggésben értelmezhető, ezt a lépést csak az elemezhetőség miatt választottam külön ezekről.

A detektálás több szinten zajlik, mikroszinten maguk az egyes befogadók is érzékelhetnek valamit dezinformációnak az olyan gyanús jelek alapján, mint a szerző, a dátum vagy a hivatkozások hiánya, esetleg a korábban már bizonyítottan dezinformációt terjesztő források szerepeltetése.⁶¹ Ugyanakkor egyéni szinten rendkívül nehéz az összehangolt nem hiteles viselkedés detektálása, mivel jellemzően egyetlen profillal látjuk ezeket a tartalmakat és nem látjuk a terjedési mintázatokat, a részt vevő hálózat működését – erre később még visszatérek. Az egyén így a dezinformációt elsősorban annak tartalma felől hiteltelenítheti. Szerencsére ebben nincs egyedül, 2023-ban közel 400 független tényellenőrző szolgáltatás volt elérhető, de a fogyasztók megfelelő médiatudatossága és eszközhasználatával ezek hatékonysága alacsony maradhat.⁶² Mivel a részvételi propaganda megerősödése miatt a dezinformáció terjesztése rendszerint a hagyományos tömegmédia megkerülésével a közösségi médiában zajlik, felértékelődik a külső tényellenőrzés szerepe (azaz, ami nem szerkesztőségekhez köthető, amelyek a publikálás előtt jó esetben leellenőrzik egy információ hitelességét).

A tényellenőrzés kihívása három rétegű, egyrészt kevés a tényellenőrzéshez szükséges szervezet és emberi kapacitás, szemben az interneten áramló hatalmas mennyiségű (dez)információval szemben. Másrészt a tényellenőrzés a közösségi médiában nem előzi meg a terjesztést, mint a hagyományos média esetén, hanem versenyt fut

⁵⁸ Az IRA tevékenységéről ld. még: Philip N. Howard – Bharath Ganesh – Dimitra Liotsiou – John Kelly – Camille François: *The IRA, Social Media and Political Polarization in the United States, 2012-2018*. Computational Propaganda Research Project. University of Oxford, 2018. <https://tinyurl.com/3xmfx2nw>

⁵⁹ Çakır-Somlyai Özge: Dezinformáció és demokrácia – A jelenlegi helyzet és a jövőbeli lehetőségek vizsgálata. In: Çakır-Somlyai (szerk.) 109.

⁶⁰ *100 Days Until Election 2024*. Office of the Director of National Intelligence, 2024. július 29. <https://tinyurl.com/2rakp347>

⁶¹ Veszelszki Ágnes: Az álhírek extra- és intralingvális jellemzői. *Századvég*, 2017/2. 51–82. <https://tinyurl.com/3b3ka943>

⁶² Horváth Kata: Álhírek leleplezése – Tényellenőrzés a világban és Magyarországon. In: Çakır-Somlyai (szerk.) i. m. 77.

azzal és mint fentebb láthattuk, a dezinformáció érdekessége folytán gyorsabban és messzebbre terjed. Harmadrészt a tényellenőrzés maga is a terjesztési folyamat részévé válik, a tényellenőrzés eredményének megítélése a dezinformációt akaratlanul vagy tudatosan terjesztő felhasználóktól függ, vagy elfogadják azt vagy sem. Ezért aztán nem elég az objektív tényellenőrzés eredménye ahhoz, hogy egy dezinformáció 'lebukjon' és onnantól kezdve mintegy varázsütésre ne terjedjen tovább. Német kutatása alapján például a tényellenőrzés műfajának megítélésében a politikai szempontok domináltak a Lakmusz tényellenőrző Facebook oldalának esetében: „a fősodró média napirendjébe illeszkedve a tényellenőrző cikkek kevesebb bírálatot váltottak ki az olvasók körében, míg a politikai táborok közötti polarizációt növelhették.”⁶³

Ugyanakkor nem mindenki szerint jó ötlet tényellenőrzéssel, vagy a hagyományos médiában bevált emberi moderálással küzdeni a dezinformáció ellen. Az X (korábban Twitter) és a Facebook 2024-2025-ben szakított a korábbi gyakorlatával és már nem alkalmaz tényellenőröket, ehelyett egy úgynevezett *community notes* rendszerre álltak át. Ennek lényege a folkszónómia, vagyis a felhasználók bevonása a tartalmak megítélésébe és az általuk jelentett tartalmak vizsgálata. Ennek hatékonysága azonban megkérdőjelezhető.⁶⁴ A Facebook esetében egyelőre nem lehet tudni, hogy vajon a 2025 januárjában hivatalba lépő Trump elnökkel való jó kapcsolat kialakítása áll az áttérés hátterében, vagy a Facebook tényleg tarthatatlannak tekintette a dezinformáció terjedésének tényellenőrzés alapú kontrollálását.

Ez már csak azért sem egyértelmű, mert van, aki már évek óta a tényellenőrzés hiányosságaira figyelmeztet. Tim O'Reilly, a webkettő fogalom atyja szerint a közösségi médiában terjedő csalásokat emberi eszközökkel nem lehet kivédeni a csalások kiterjedtsége és gyorsasága miatt. Ennek oka, hogy csak egy gép képes belépni egy másik gép OODA (*Observe – Orient – Decide – Act* / Megfigyel – Elemez – Dönt – Cselekszik) ciklusába és megzavarni azt, egy ember nem elég gyors hozzá. O'Reilly szerint azonban az algoritmusok mégis több mindent megtehetnének a hírek kapcsán, például:

- ahogy a részvényeknél számolnak fundamentális értéket, ami a tényleges érték mögött van, úgy lehetne a híreknél is számolni ilyesmit,
- megkereshetnék a hírek eredeti forrásait, hogy nem mondanak-e ellent a megosztott hírtartalmának,
- megnézhetnék, hogy mennyien hivatkoznak rá, hitelesnek tekinthető-e a közlő oldal,
- ellenőrizhetnék, hogy statisztikailag releváns módon találják-e az információkat (például a bűnügyi statisztikákat minden esetben ugyanannyi lakosra vetítve közlik-e),

⁶³ Német Szilvi: „Értelmezni se tudom, mi az a liberális tényellenőr” – Facebook-hírkomentárok a tényellenőrző újságírásról. *Médiakutató*, 2023/3. 79. <https://tinyurl.com/y7bnpwzy> [A továbbiakban: Német (2023b).]

⁶⁴ Yuwei Chuai – Haoye Tian – Nicolas Pröllochs – Gabriele Lenzini: Did the Roll-Out of Community Notes Reduce Engagement With Misinformation on X/Twitter? *Proceedings of the ACM on Human-Computer Interaction*, Vol. 8., No. CSCW2. (2024) 1–52. <https://doi.org/10.1145/3686967>

- megíthetnék, hogy kattintásvadász-e a címük, azaz olyan, ami nem tükrözi rendesen a tartalmat,
- ellenőrizhetnék, hogy több, független forrás is lehozta-e már a hírt,
- megkérdezhetnék a felhasználókat, hogy biztos tovább akarnak osztani egy posztot anélkül, hogy rákattintottak volna a benne szereplő linkre,
- végül összevethetnék a cikk tartalmát az eredeti forrással, például, ha törvényre hivatkozik, akkor a törvény szövegével.⁶⁵

Persze ezeket az emberek is megtehetnék. De O'Reilly szerint az emberek ezeket a gyakorlatokat eddig sem követték és nem is fogják megváltoztatni a hírolvasási szokásaikat csak úgy. Ráadásul, ha valami olyat olvasnak, amit el akarnak hinni, akkor pláne nem fognak energiát fektetni abba, hogy megkeressék annak esetleges cáfolatát.

És ezzel el is érkeztünk a detektálás kapcsán az egyik legfontosabb, ha nem a legfontosabb kérdéshez: tényleg lehet MI-vel küzdeni az MI által támogatott dezinformáció ellen? Az utóbbi évtizedben szinte MI 'fegyverkezési verseny' bontakozott ki a dezinformáció terjesztésében és detektálásában, illetve korlátozásában.⁶⁶ Sedova és munkatársai a RICHDATA piramis leírásával visszafejtették a közösségi média alapú politikai dezinformációs kampányok működését, miközben akaratlanul egyfajta 'receptet' is adtak ahhoz, eszközöket is nyújtottak az ellene való küzdelemhez. A piramis leírja a dezinformációs kampányok létrehozását, felépítését, terjesztését és a közönség bevonását:

„A dezinformációs kampányokat több szakaszra bontjuk. A felderítés [Reconnaissance] során az operátorok felméri a környezetet, és megértik a közönséget, amelyet manipulálni próbálnak. Infrastruktúrára [Infrastructure] van szükségük – hírvivőkre, hihető személyiségekre, közösségi média fiókokra és csoportokra –, hogy a narratíváikat továbbítsák. Az üzeneteik magva, gyökerezése és növekedése érdekében elengedhetetlen a tartalom [Content Creation and Hijacking] szüntelen áramlása, a posztoktól és hosszú olvasmányoktól kezdve a fotókig, mémekig és videókig. Az internet áramlatában való megjelenés [Deployment] után a dezinformáció ezen egységeit botok, platform-algoritmusok és social-engineering technikák erősítik fel [Amplification], hogy a kampány narratíváit terjesszék. A dezinformáció szétrobbantása azonban nem mindig elegendő: a széles körű hatás a tudatlan felhasználókkal való tartós kapcsolattartáson keresztül, trollkodással [Troll Patrol] érhető el – ez a dezinformációs közelharc megfélemlője. A végső szakaszban a dezinformációs művelet azáltal válik aktuálissá [Actualization], hogy megváltoztatja a tudat-

⁶⁵ Tim O'Reilly: *WTF – Miért rajtunk múlik, hogy mit hoz a jövő?* Budapest, Typotex, 2018.

⁶⁶ Fátima C. Carrilho Santos: Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis. *Journalism and Media*, Vol. 4., No. 2. (2023) 679–687.
<https://doi.org/10.3390/journalmedia4020043>

lan célpontok gondolkodását, vagy akár cselekvésre mozgósítja őket, hogy káoszt szítsanak.”⁶⁷

A RICHDATA annak belátásában is segít, hogy több ponton lehetne beavatkozni MI segítségével, ezek közül az első pont kétségkívül a dezinformáció detektálása lenne. Bayer szerint több fajta módon lehet az MI-t a dezinformáció detektálására használni: a terjesztők azonosításában, a terjedési mintázat detektálásában és a tartalom felismerésében.⁶⁸ Ezeket szerinte érdemes vegyíteni, például lehetséges detektálni a botokat, amelyek közel egyszerre osztják meg ugyanazt az üzenetet, gyanús az is, ha egy közösségimédia-tartalomnak több megosztása, mint kedvelése van, mert rendszerint fordítva szokott lenni. Viszont azzal, hogy a botok keverednek a valós emberekkel, sokszor nagyobb követői táborokkal rendelkező influenszerekkel és celebritásokkal (akik akaratlanul vagy meggyőződésből terjeszthetik a dezinformációt), meglehetősen nehézkessé válik az MI használata ezeknek a hibrid dezinformációs hálózatoknak az automatizált kezelésében (például, hogy mely profilokat *shadow banning*olják⁶⁹, függesszék fel vagy töröljék véglegesen). Ezen túlmenően Bayer szerint egyelőre az MI még nem elég fejlett ahhoz, hogy a tartalom elemzése alapján azonosítsa be a dezinformációt. Ehhez a számtalan beszélt emberi nyelv finomságait (például ironia), kulturális kontextusait és politikai vonatkozásait is ismernie, és ezt a tudást magabiztosan alkalmaznia is kellene. A legnagyobb kihívás azonban az, hogy ezeknek a mesterséges intelligencia által támogatott módszereknek a kombinálása sem elég hatékony, amikor azonnali beavatkozásra lenne szükség a futótűzszerűen terjedő dezinformációval szemben, például „amikor a dezinformáció a tőzsdei folyamatokat kívánja manipulálni, háború idején vagy közvetlenül a választások előtt.”⁷⁰ Egyelőre a gyors és megbízható MI-alapú dezinformációdetektáló rendszerek kifejlesztése még várat magára, bár a nagy nyelvi modellekkel már elkezdődtek a tesztek.⁷¹

De felmerül a kérdés, hogy egyáltalán szükségünk van-e a detektálásra, változik-e tőle valami, érdekli-e az embereket, ha kiderül egy információról, hogy hamis? O'Reilly szerint a történelem, a politika vagy az üzlet nem fekete-fehér, nem igaz vagy hamis, hanem olyan mértékben igazak vagy hamisak, amennyire annak hisszük azokat – ezt nevezi ő meggyőződésnek.⁷² Az emberek pedig nem feltétlenül azért hisznek el valamit, mert az igaz, hanem mert el akarják hinni. Ezzel pedig elérkezünk a befogadás kérdéséhez.

⁶⁷ Katerina Sedova – Christine McNeill – Aurora Johnson – Aditi Joshi – Ido Wulkan: *AI and the Future of Disinformation Campaigns Part 1: The RICHDATA Framework*. Center for Security and Emerging Technology, December 2021. 1. <https://tinyurl.com/5n78ucm2> (saját fordítás P.R.)

⁶⁸ Judit Bayer: AI and disinformation – The technological and legal limits of content regulation in the context of disinformation. *Legal Tech*, 2022/2. 1–8. <https://tinyurl.com/34pphen>

⁶⁹ A *shadow banning* (magyarul talán *árnyék némitásnak* lehetne fordítani) lényege, hogy továbbra is lehet használni a fiókot, de a közzétett tartalmakat csak az azokat közzétevő felhasználója látja, más nem, és erről az érintett nem is feltétlenül szerez tudomást.

⁷⁰ Bayer (2022) i. m. 3.

⁷¹ Bayer (2024) i. m. 89.

⁷² O'Reilly i. m. 261.

2.5. Ötödik lépés: befogadás

A dezinformációs lánc ötödik lépésénél, a befogadás esetében a kiindulópontot az emberi észlelés és információfeldolgozás megértése adja. Ahogy Krekó kiemeli az evolúció során az szolgálta a túlélésünket, ha a saját érzékszerveinkre hagyatkoztunk és alapvetően gyanakvóak voltunk. Ez azonban elvezetett ahhoz, hogy mára viszonylag könnyen manipulálhatóak lettünk, főleg mivel „a legtöbb információt közvetetten szerezzük, és a szövegeket, a képeket, illetve a hangot egyaránt könnyebb manipulálni, mint korábban bármikor.”⁷³

Persze ettől még kell, hogy a dezinformációnak legyen hihető része, mint azt korábban a létrehozás lépésénél már említettük:

„az álhírekben is van némi igazság: egy teljesen légből kapott hírt a befogadók nem hinnének el. A Central Park állatkertjéből sosem szöktek meg a vadállatok, de valóban van ott állatkert, Cion bölcséinek sosem volt titkos jegyzőkönyvük, de cionista világkongresszust valóban tartottak... és így tovább. A koholmányokat az ismert elem teszi hihetővé, mert a befogadó így az új információt be tudja illeszteni a világról alkotott képébe.”⁷⁴

De a dezinformáció esetében a hihetőségen túl legalább ennyire kulcsfontosságú, hogy mit akarunk és kinek elhinni. Miért hisszük el a tévhiteket és ragaszkodunk hozzájuk makacs módon? Hammer szerint⁷⁵ Haidt moráleméleti kategorizációja mentén lehet megérteni a jelenséget:⁷⁶

1. az emberek úgy érzik joguk van eldönteni, hogy kinek és mit higgyenek el;
2. az emberek számba veszik a kockázatokat és veszélyeket, amik az általuk fontosnak ítélt személyeket (pl. család) és dolgokat fenyegetik;
3. vannak dolgok, amikről az emberek nem hajlandók lemondani vagy veszélynek kitenni őket (pl. a vallásuk vagy természeti környezetük);
4. fontos, hogy mit gondolnak az emberek arról, mások mennyire bánnak méltányosan velük;
5. miről vagy kiről érzik, hogy hűséggel tartoznak neki (pl. lakóhelyük közössége, football csapatuk stb.);
6. végül meghatározó az is, hogy milyen tekintélyt ismernek el.

Vagyis a befogadás kontextusa nélkül nem lehet megérteni magát a dezinformációt. Azok a dezinformációk hullanak termékeny talajra, amelyek feljogosítják az embereket,

⁷³ Krekó i. m. 267–268.

⁷⁴ Bajomi-Lázár i. m. 8.

⁷⁵ Hammer Ferenc: Osztálytermi fórum a dezinformációról. In. Çakır-Somlyai Özge (szerk.): *A DEZINFORMÁCIÓ JELENSÉGE – tanulmánygyűjtemény*. Budapest, Mérték Füzetek 33. szám, 2023. 94–95. https://mertek.eu/wp-content/uploads/2023/12/Mertek-fuzetek_33_1.pdf

⁷⁶ Jonathan Haidt: *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.

hogy higgyenek bennük, az emberek számára fontos személyekkel és dolgokkal kapcsolatos veszélyekről szólnak, (látszólagos) méltánytalanságra hívják fel a figyelmet, olyanoktól erednek, akik iránt hűségesek a dezinformációval megcélzott személyek és általuk elfogadott tekintély is hitelesíti a tartalmukat. Krekó hasonló következtetésre jut összeesküvésekről és dezinformációról szóló könyvében, amikor a következőt írja: az „hogy mit tekint valaki összeesküvés-elméletnek vagy álhírnek, nagyban függ már előzetesen meglévő véleményétől, illetve attól, hogy milyen tekintélyeket fogad el hiteles forrásnak.”⁷⁷

Mindez azt is megmagyarázza, hogy a közösségi média algoritmusok működésén túl miért lehetnek annyira sikeresek a visszhangkamrák és a szűrőbuborékok. Eltérő módon reagálunk ugyanis ugyanazokra az információkra attól függően, hogy mi fontos nekünk, emiatt mindenki (legalább egy kicsit) a saját valóságában él, ami már a közösségi média megjelenése előtt is jellemző volt ránk, emberekre.

Emiatt felmerül az a kérdés is, hogy szükséges-e egyáltalán a médiának (vagy bármelyik külső tényellenőrző szervezetnek) *deepfake* ellenőrző kapacitást kiépítenie. És ez nem pusztán azért merül fel, mert ezek egyelőre nem működnek megbízható módon: az egyik szoftver azt mondja valamiről, hogy *deepfake* hamisítvány, a másik pedig azt, hogy nem az.⁷⁸ Hanem mert emiatt a bizonytalanság miatt a végén az olvasónak vagy a nézőnek magának kell úgyis eldöntenie, hogy kinek higgyen. Az emberek pedig egyre inkább abban hisznek, amiben akarnak, kiválogatják azokat az információkat, amik megerősítik a véleményüket, hitüket, ideológiai beállítódásukat. Ahogy Krekó írja: az összeesküvés elméletekben hívók befogadása megváltozik. Nem azért hiszik el valamiről, hogy igaz, mert az annyira tökéletesen manipulált *deepfake* lenne, hanem mert el akarják hinni.⁷⁹

De a dezinformáció személyes befogadásában legalább ennyire fontos a közösségi média működési mechanizmusa és üzleti modellje, illetve annak következményei a befogadás körülményeire. Ináncsi szerint a közösségi média célja az erős érzelmi reakciók kiváltása, mivel ez növeli a felhasználói interakciókat. A viták elősegítik a tartalomtermelést, ami a platformok működésének alapja és a dezinformáció terjedését fokozza, hogy az álhíreket megosztó felhasználók erősebb érzelmi reakciókat mutatnak. Mindez kihívást jelent a moderáció számára is, mivel, ha a közösségimédia-platformok az interakciót vitákra építik, az kedvezhet a félrevezető tartalmak elterjedésének.⁸⁰ Tehát a közösségimédia-oldalak üzleti modelljük miatt (ami az ingyenességen és a reklámbevételen alapul) abban érdekeltek, hogy minél hosszabb ideig fenntartsák a figyelmet, mert azt értékesítik a reklámozók felé. Ezért algoritmusai annak adnak

⁷⁷ Krekó i. m. 261.

⁷⁸ Laura Stroebe – Mark Llewellyn – Tricia Hartley – Tsui Shan Ip – Mohiuddin Ahmed: A systematic literature review on the effectiveness of deepfake detection techniques. *Journal of Cyber Security Technology*, Vol. 7., No. 2. (2023) 83–113.
<https://www.tandfonline.com/doi/full/10.1080/23742917.2023.2192888>

⁷⁹ Krekó i. m. 26.

⁸⁰ Ináncsi i. m. 14–15.

teret, ami ott tartja a felhasználókat, ezek pedig a megosztó, érzelmeket, bevonódást és vitát kiváltó tartalmak, köztük maga a dezinformáció is.

A korábbiak során rendre külön megvizsgáltam az MI szerepét a dezinformációs lánc adott lépésében, illetve az MI által támogatott lehetséges beavatkozási pontokat (bár a dezinformáció elleni küzdelemről később részletesen is írok). A dezinformáció percepciója, feldolgozása és befogadása kapcsán azonban a már fentebb említetteken túl nincs kiemelt szerepe az MI-nek, mivel a befogadás teljes egészében emberi folyamat. Az MI sokkal inkább meghatározó tud lenni a befogadással szorosan összefonódó terjesztés és detektálás lépéseiben, mint magában a befogadásban.⁸¹

2.6. Hatodik lépés: hatás és (nem szándékolt) következmények

„Az álhírek és összeesküvés-elméletek nem elsősorban önmaguk miatt érdekesek, hanem az általuk kiváltott hatás miatt.”⁸² A dezinformációs lánc működésének vizsgálata tehát fontos ugyan, de pusztán azt a célt szolgálja, hogy közelebb férhessünk a társadalmi hatások mélyebb megértéséhez.

A dezinformációnak több fajta módon is kategorizálhatjuk a hatását, például:

- van tétje (pl. háborús uszítás) vagy nincs igazi tétje (pl. a pufi dzseki pápa képe⁸³);
- mikroszintű: személyes – mezoszintű: egyes intézményekre korlátozódó – makroszintű: társadalmi – metaszintű: például annak vizsgálata, hogy okoz-e korszakváltást;
- lokális, regionális, országos vagy nemzetközi hatású;
- illegális vagy nem illegális;
- milyen mértékben okoz kárt: a teljesen ártalmatlantól a globális szintű károkozásig;
- hol, milyen területen fejt ki a hatását (pl. csoportközi erőszak, terrorizmus, demokratikus folyamatok, egészségügy, környezetvédelem, politikai szélsőségek).⁸⁴

A cikk terjedelmi korlátai miatt az alábbiakban a nagy téttel bíró, társadalmi vagy metaszintű változásokat okozó, országos vagy nemzetközi hatású, illegális és globális károkozásra képes dezinformációs tevékenységekre koncentrálok.

Ezek alapján kijelenthető, hogy a dezinformáció „drasztikus következményekkel is járhat, melyre példa amikor az Egyesült Államokban a Capitol(sic!) ostromához és

⁸¹ Az egy érdekes kérdés, hogy ez a jövőben vajon megváltozhat-e. Ray Kurzweil 2024-ben megjelent könyvében (*Singularity is nearer – When we merge with AI*. Vintage, 2024.) a 2040-es évek közepére teszi az ember(i agy) és a gép(i gondolkodás) egyesülését. Ha ez megvalósulna, akkor ez kétségtelenül megváltoztatná az információk befogadásának folyamatát is.

⁸² Krekó i. m. 105.

⁸³ Gergely Máté: Pufi dzseki Ferenc pápa és társai: így leplezzük le az MI-képeket. *Világgazdaság*, 2024. április 13. <https://tinyurl.com/3r4hry7a>

⁸⁴ Ezekről bővebben ld. Krekó i. m.

megszállásához vezetett az álhír terjedés.”⁸⁵ Bajomi-Lázár szerint „[a] társasmédia-platformokon terjesztett hamis információk szerepet játszhattak a 2016-os amerikai elnökválasztás és a brexit végeredményében is.”⁸⁶ A dezinformáció képes lehet aláásni a demokratikus intézményeket és magát a demokráciát is, „[a] 2016-os események, köztük a Brexit-népszavazás és az amerikai elnökválasztás aláhúzták a dezinformáció veszélyét és a demokratikus társadalmak működését megzavaró potenciálját.”⁸⁷

Bayer szerint a dezinformációs kampányok egyre szélsőségesebbé váltak az utóbbi években.⁸⁸ 2015–2016-ban az amerikai választások és a Brexit-népszavazás befolyásolásával kezdődött, majd a COVID-19-cel kapcsolatos álhírek sokak egészségügyi döntéseit eltorzították. A 2021. január 6-i Capitolium elleni támadás során öt ember meghalt, míg az Oroszország, Ukrajna és a Nyugat közötti feszültségeket tápláló dezinformáció háborúhoz vezetett.

Barts szerint az internet és a közösségi média alapjaiban változtatta meg az információ terjesztését, lehetővé téve a dezinformáció gyors és szervezett terjedését.⁸⁹ Információs túlterheltség esetén az emberek inkább a saját csoportjuk tagjaira támaszkodnak, nem pedig szakértőkre.⁹⁰ Fontos, hogy a dezinformáció ismételése ismeretlenségérzetet kelt, ami idővel növeli az elfogadás esélyét (ezt nevezzük ‘illuzórikus igazsághatásnak’). Emellett az információ és forrása gyakran különválnak az emlékezetünkben, így a hamis adatok akkor is befolyásolhatják az attitűdjeinket, ha később kiderül róluk, hogy dezinformáción alapultak.

2016-ban Trump első megválasztásától nem függetlenül az Oxford Dictionaries az év szavának választotta a *post-truth* kifejezést. A fogalom ugyan nem 2016-ban született, hanem feltehetően 1992-ben, de az elmúlt évtized második felétől datálható széles körben való elterjedése, miután egyik évről a másikra hűsszorosára nőtt a használata. Jelentése szerint „[o]lyan körülményekre vonatkozó melléknév, amelyekben az objektív tények kevésbé befolyásolják a közvéleményt, mint az érzelmi vonzalmak.”⁹¹ A *post-truth*ra jellemző a dezinformáció és az összeesküvés elméletek széleskörű térnyerése. Ezek együtt elvezettek ahhoz, hogy még megosztottabbak legyenek a társadalmak és ugyanarról a jelenségről teljesen mást gondoljanak a választók bal és jobb oldalon.

A politika nem csak Magyarországon polarizálódik egyre inkább, Trump 2016-os megválasztása óta beléptünk az új tribalizmus korszakába,⁹² ahol a saját táborot kell erősíteni, még akkor is, ha nem értünk vele maradéktalanul egyet, mert a másik tábor sokkal rosszabb, ha ők hatalomra kerülnek (hatalmon maradnak) az felér egy ‘világvé-

⁸⁵ Ináncsi i. m. 16.

⁸⁶ Bajomi-Lázár i. m. 8.

⁸⁷ Çakır-Somlyai i. m. 114.

⁸⁸ Bayer (2022) i. m. 1.

⁸⁹ Barts i. m. 36.

⁹⁰ Wardle–Derekhsan i. m.

⁹¹ ‘Post-truth’ declared word of the year by Oxford Dictionaries. *BBC*, 2016. november 16.
<https://www.bbc.com/news/uk-37995600> (saját fordítás P.R.)

⁹² Thomas J. Froehlich: A disinformation-misinformation ecology: The case of Trump. In: Ján Višňovský – Jana Radosinská (szerk.): *Fake News is Bad News-Hoaxes, Half-Truths and the Nature of Today's Journalism*. Intechopen, 2021.

gével'. Emiatt a politikai vezetők egy részén képződik egy 'teflon bevonat' – mindegy mit mondanak, mindegy mit csinálnak, a támogatók szemében a kritika csupán a másik tábor ármánykodása lesz. A demokráciában bízók számára ugyanakkor jó hír lehet, hogy szoros eredmények esetén elég a két táboron kívüli bizonytalan szavazókat sikeresen megszólítani, akik feltehetően sokkal inkább hallgatnak az észérvekre. Például a látszólag áttörő győzelem ellenére a Biden–Trump vagy a Trump–Harris választás is csupán néhány csatátér államban és pár százezer választón dőlt el.⁹³

A *post-truth*, a politikai polarizáció és a tribalizmus mellett látszólag kevésbé meghatározó a *deepfake* elterjedésének egy nem szándékolt következménye, a bizalom csökkenése, aminek valójában sokkal átfogóbb hatása lehet. Vaccari és Chadwick⁹⁴ Egyesült Királyságban végzett kvantitatív kutatása alapján a *deepfake* technológia és a dezinformáció komoly kihívást jelent a hírek hitelessége és a közvélemény bizalma szempontjából is. A mesterséges intelligencia által létrehozott szintetikus videók még akkor is növelhetik a bizonytalanságot és csökkenthetik a hírekbe vetett bizalmat a közösségi médiában, ha nem tévesztik meg a felhasználókat. Mindez hosszú távon káros hatással lehet a demokratikus társadalmakra, mivel megváltoztatja az online kultúrát, kevésbé együttműködő és felelősségteljes viselkedési formákat, illetve normákat hoz létre.

Ha szűken nézzük a generatív MI-t, akkor elmondható, hogy az MI manipulációnak pusztán az elvi lehetősége bármit képes hiteltelenné tenni, mert könnyen rámondják bármire, hogy *deepfake*. Ehhez nem is kell *deepfake*-nek lennie, elég pusztán az a tudat, hogy az lehet. Ugyanígy „[a]z álhírek kifejezés gyakori használata a vezetők – tekintélyelvű és demokratikus vezetők egyaránt – által aláásta a média hitelességét, keletet teremtve a vezetők számára a velük szemben kritikus tudósítások lejáratására.”⁹⁵ Tehát egyelőre nem is annyira magától a *deepfake*-től kell tartani, hanem az általa közvetetten okozott bizalmatlansági és hitelességi válság miatt.

3. Befejezés

3.1. Korszakváltás?

A dezinformációs lánc egyes lépéseinek vizsgálata után összegezzük az eredményeket, majd térjünk vissza a cikk címében is megfogalmazott dilemmához, azaz, hogy új korszakot hoz-e a mesterséges intelligencia a dezinformációban. Először röviden tekintsük át a dezinformációs lánc lépéseit és az azokkal kapcsolatos főbb jelenségeket, fogalmakat, amelyeket fentebb részletesen is leírtam:

⁹³ Ernesto F. L. Amaral: Popular Vote and Voter Turnout in the 2016, 2020, and 2024 US Presidential Elections. *Center for Open Science*, 2024. <https://doi.org/10.31219/osf.io/gbsdy>

⁹⁴ Cristian Vaccari – Andrew Chadwick: Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, Vol. 6., No. 1. (2020) <https://doi.org/10.1177/2056305120903408>

⁹⁵ Çakır-Somlyai i. m. 114.

1. táblázat: A dezinformációs lánc lépései, kapcsolódó fogalmai és jelenségei

A dezinformációs lánc lépése	Kapcsolódó fogalmak és jelenségek
motiváció	sokrétű motivációk, tényrelativizmus, információs zaj, <i>mockumentary</i> és áldokumentumfilm, manipuláció, tudományos hamisítás
létrehozás	generatív MI, <i>deepfake</i> , Wardle-féle hét lépcsős skála, arc kicserélése, <i>lip syncing</i> , eszközök elérhetőségének demokratizálódása
terjesztés	<i>misinformation</i> (félretájékoztatás), közösségimédia-algoritmusok, moderáció, <i>astroturfing</i> , összehangolt nem hiteles viselkedés, részvételi propaganda, álhírek gyorsabb terjedése, szűrőbuborék, visszhangkamra, álhírgyárak, <i>micro-targeting</i> , információs hadviselés
(detektálás)	gyanús jelek, tényellenőrzés, <i>community notes</i> , folkszonómia, RICHDATA piramis, MI megoldások (terjesztők azonosítása, terjedési mintázat detektálása, tartalom felismerése), <i>shadow banning</i>
befogadás	dezinformáció igazságtartalma, <i>deepfake</i> és dezinformáció ellenőrző kapacitás, Haidt moráleméleti kategorizációja, a befogadás kontextusa, összeesküvés elméletek, közösségi média üzleti modellje, ingyenesség, figyelem fenntartása, érzelem kiváltása, közösségimédia-algoritmusok, ember és gép egyesülése
hatás és következmény	álhírek terjedése, amerikai elnökválasztás, Brexit, demokratikus intézmények aláásása, demokratikus működés megzavarása, hírekbe vetett bizalom csökkenése, online kultúra és viselkedési normák megváltozása, hitelesség megítélése, illuzórikus igazsághatás, <i>post-truth</i> korszak, politikai polarizáció, tribalizmus, média hitelességének aláásása

Mindezek alapján az egyes lépések mentén az MI (korszakalkotó) szerepét illetően az alábbiak látszanak:

- motiváció: a motivációk ugyan nem változnak, de azzal, hogy demokratizálódik a dezinformációra, *deepfake* készítésére alkalmas eszközök elérése, elvileg egyre többen készíthetnek ilyen tartalmakat;
- létrehozás: a hihető dezinformáció létrehozása sokkal egyszerűbbé válik az MI alkalmazásával, de a generatív MI megjelenése óta a várakozások ellenére egyelőre elmaradt a dezinformációs tartalmak robbanásszerű megnövekedése;
- terjesztés: színlelt tömeg (botok és fizetett trollok) segítségével széles körben terjeszthető a dezinformáció, a *microtargeting* segítségével célzottan juttathatók el a dezinformációt tartalmazó üzenetek a befogadókhoz, akik meggyőződésből vagy tévedésből tovább terjeszthetik azokat (részvételi propaganda);
- detektálás: az MI eszközök segítségével több fajta módon is lehet a dezinformáció ellen küzdeni (terjesztők azonosítása, terjedési mintázat detektálása, tartalom felismerése), de ezek egyelőre nem elég megbízhatók és gyorsak;
- befogadás: mivel a befogadás a közönség oldalán történik meg, szűk értelemben véve ebben játszanak egyelőre a legkevésbé szerepet az MI eszközök (azok inkább a létrehozásban, terjesztésben és detektálásban meghatározóak);
- hatás és következmény: ahogy a befogadás esetén, a hatásban sincs közvetlen szerepe az MI-nek, közvetve azonban óriási változásokat indukálnak, ami akár új korszakot is jelenthet a dezinformációban (erről bővebben lásd lentebb).

A fentebbiek alapján tehát a létrehozás, a terjesztés és részben a detektálás terén van egyelőre közvetlen szerepe az MI-nek a dezinformációs láncban. De mindezek révén közvetve mégis a befogadás és a széleskörű társadalmi következmények terén hoz érdemi változást a mesterséges intelligencia megjelenése a dezinformációban. Ahogy Bayer Judit találóan összefoglalja:

„A mesterséges intelligencia szerepe különösen jelentős a dezinformációval való összefüggésben. E technológiai megoldások létezése előtt a dezinformáció terjesztése költséges, munkaigényes és lassú volt. Az algoritmusok, különösen a gépi tanulás alkalmazása felerősítheti a dezinformáció létrehozását, megjelenését és hatókörét, ezáltal hatását is.”⁹⁶

Állításom szerint a mesterséges intelligencia kapcsán a terjesztés és a társadalmi hatás terén tapasztalható érdemi (negatív) változás az MI dezinformációs láncban betöltött szerepében, nem pedig a dezinformáció létrehozásában, ami a generatív MI megjelenése óta az érdeklődés középpontjába került. Hiába lehet ugyanis sokkal hihetőbb dezinformációs tartalmakat generálni, ha a létrehozók nem lennének képesek érdemben terjeszteni azokat, akkor sokkal kisebb hatást lennének képesek kifejteni. Ráadásul, mint láttuk, az emberek nem azért hisznek el valamit, mert annyira hihető, hanem mert el akarják hinni, s ebben a terjesztés-(detektálás)-befogadás lépéseknek sokkal fontosabb szerepe van, mint magának a létrehozásnak. Ha MI fémjelezte korszakváltásról beszélhetünk, a dezinformációban az nem a generatív MI-nek, hanem egyértelműen a közösségi média MI-algoritmusainak köszönhető. De ez a változás sem ment volna végbe az emberek aktív közreműködése nélkül. Például a detektálásban is kulcsfontosságú a *human-in-the-loop* megközelítés, nem automatizálható teljesen a dezinformáció észlelése és kezelése, moderátoroknak kell felügyelni (verifikálni) az MI eszközök munkáját.⁹⁷ Valójában egy olyan *hibrid dezinformációs láncról* van szó, amelyben maguk az emberek játszanak kulcsszerepet a terjesztésben, detektálásban, befogadásban és a társadalmi változások indukálásában is. A közönség, az influenszerek, celebek és politikusok szinte elválaszthatatlanok a botoktól, fizetett trollaktól és a közösségimédia-algoritmusok működésétől.

Ha korszakhatárt kellene meghatározni a cikkben taglalt jelenségeknek, akkor az 2016 lehetne, amikor a Brexit sokkolta a világot, elnökké választották Trumpot és a *post-truth* kifejezés az év szava lett. Az, hogy ez tartós változás, jól jelzik a COVID-járvány idején terjedő álhírek és az oltásellenesség, illetve a Capitolium 2021 januári ostroma, majd Trump újraválasztása 2024-ben.

De vajon hozhat-e újabb korszakot az MI további fejlődése a dezinformációban? Krekó szerint a generatív MI megjelenéséig alapvetően a terjesztésben játszott fontos

⁹⁶ Bayer (2022) i. m. 1.

⁹⁷ Gianluca Demartini – Stefano Mizzaro – Damiano Spina: Human-in-the-loop Artificial Intelligence for Fighting Online Misinformation: Challenges and Opportunities. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, Vol. 43., No. 3. (2020) 65–74. <https://tinyurl.com/24tzb3s4>

szerepet az MI a közösségimédia-algoritmusok révén.⁹⁸ Idővel azonban a tartalomgyártás is automatizálhatóvá válhat és akkor már nem is kell feltétlenül ember hozzá. De erre az esetre is igaz, hogy nem a dezinformáció az érdekes, hanem az általa kifejtett társadalmi hatás, tehát az embereket nem lehet kihagyni a dezinformációból.

Az várható, hogy a jövőben az MI-t még intenzívebben használják majd a dezinformációnak szánt tartalom létrehozásában is. Az MI fejlesztése 2024 óta az önállóbb ágensek felé mutat, amelyek el tudják végezni a rájuk bízott feladatokat egyfajta MI asszisztensként,⁹⁹ így bevezethetők akár dezinformáció tömeges kreálására is. Ez megerősítheti a 2016-ra datálható, közösségi médiához köthető korszakváltást, de feltételezően nem hoz létre új korszakot mindaddig, amíg a közösségi médiára fókuszáló terjesztés nem változik érdemben. Az viszont már a tudományon túlmutató, sci-fi felé hajló kérdés, hogy vajon az MI képes lesz-e valaha önállóan, saját szándék alapján tömegesen manipulálni az embereket dezinformációval – ez lehetne ugyanis a következő korszak, de reméljük, hogy ilyesmire soha nem kerül majd sor.

3.2. Mi lehet az MI szerepe a dezinformáció elleni küzdelemben?

Az általam javasolt dezinformációs lánc alkalmazása nem csak abban segíthet minket, hogy könnyebben legyünk képesek azonosítani és megérteni a mesterséges intelligencia szerepét a dezinformációban, hanem hogy azonosíthassuk azokat a specifikus beavatkozási pontokat és lehetséges ellenlépéseket, amivel a dezinformáció ellen *szisztematikusan* fel lehet lépni. Ezek alapján véleményem szerint több szintű a probléma kezelésének a lehetősége:

1. dezinformáció készítéséhez kapcsolódó motiváció gyengítése (pl. a *deepfake porn* készítésének kriminalizálása vagy a kattintás alapú reklámbevételekből élő álhírgyárak monetizációs képességének korlátozása);
2. dezinformáció létrehozásának megnehezítése (pl. generatív MI-platformok felhasználási körének korlátozása), az MI által létrehozott tartalmak vízjelzése – bár sajnos ezeket egyelőre könnyű hamisítani;¹⁰⁰
3. a dezinformáció detektálásának és címkézésének megerősítése, elterjesztése és tudatos használatának elősegítése, a közösségimédia-platformok dezinformáció detektálási gyakorlatának szabályozása;¹⁰¹
4. a dezinformáció terjedésének megakadályozása vagy legalább lassítása, hatókörének korlátozása (pl. moderálás, hátrébb sorolás, *shadow banning* stb.);
5. a dezinformáció befogadásának alakítása, a társadalom felkészítése a dezinformációra egyéni és társadalmi szinten: pl. közös biztonsági szó az unokázó családok ellen, a tudatos médiahasználat elősegítése, a médiaműveltség oktatása, információs írástudás szintjének javítása;

⁹⁸ Krekó i. m. 96.

⁹⁹ Az OpenAI MI első ágensét 2025 elején jelentette be: <https://openai.com/index/introducing-operator/>

¹⁰⁰ Bayer (2024) i. m. 90.

¹⁰¹ Azgin–Kiralp i. m.

6. ha mégis elterjedt, akkor a dezinformáció negatív hatásának csökkentése.¹⁰²

Mindebben segítségünkre lehet az MI is, többek között a dezinformáció felderítésében és a terjedés korlátozásában.¹⁰³ De a dezinformáció elleni küzdelem feladata elsősorban nem az MI-é, hanem közös ügy: „[a] dezinformáció elleni küzdelem nemcsak a kormányok és a médiumok felelőssége, hanem az egyéneké, a családoké és a közösségeké is.”¹⁰⁴ A felhasználókon túl össze kell(ene) fognia a közösségimédia-platformok tulajdonosainak és fejlesztőinek, a szabályozásért felelős döntéshozóknak, kutatóknak és újságíróknak, hogy megfelelő kereteket dolgozzanak ki a dezinformáció kezelésére a létrehozás, terjesztés és detektálás terén.

Bármilyen lépéseket teszünk is azonban, még kell tanulnunk együttélni azzal, hogy az MI felerősíti a *deepfake* generálását és terjesztését. De fontos, hogy ettől még ne fatalista állítás legyen az, hogy ‘nem hihetünk a szemünknek és fülünknek’, tehát nem minden eleve hazugság és nem is vagyunk teljesen tehetetlenek ez ellen. Ez ugyanis a *post-truth* korba való belépés kritikátlan elfogadása lenne, ahol bárkinek lehet saját ténye, mivel azt is olyannak tekintik, mint a véleményt és annak van igaza, aki a leghangosabban és a legkitartóbban kiabál. Ha nem hihetünk a *deepfake* miatt a saját érzékszerveinknek, az még inkább megnöveli a szkepszis szintjét, a mindenben való szisztematikus kétkedés viszont megágyaz az összeesküvés elméleteknek:

„ez egyébként a post-truth korszak egyik nagy paradoxona: a manipuláció ellen éppen az egészséges szkepszis és gyanakvás vértézheti fel az újságírókat és az állampolgárokat. Azonban a mindent eluráló, cinizmusba forduló szkepszis ad lehetőséget a közvélemény összeesküvés-elméleteken keresztüli manipulációjára”¹⁰⁵

Tehát a ‘nem hihetünk az érzékszerveinknek’ állítás akkor jótékony hatású, ha mindenben csak egy *egészséges* kétkedést alakít ki. Vagyis ne higgyünk el bármit, csupán azért, mert hihetőnek látszik, ellenőrizzük le az információ kontextusát is. De egyúttal tudjuk azt is, hogy bizalom nélkül nem működik egyetlen emberi társadalom sem.

¹⁰² Krekó Péter a dezinformáció és az összeesküvés elméletek elleni küzdelem négy legfontosabb eszközét az alábbiakban látja: megelőző csapás (előzetes kínálatcsökkentés), válaszcspás (utólagos kínálatcsökkentés), gyógyítás (utólagos keresletcsökkentés) és immunizálás (előzetes keresletcsökkentés). Krekó i. m. 275. Ezek összhangban vannak az általam a 3–6. pontokban jelzett lehetséges beavatkozásokkal.

¹⁰³ Bayer (2024) i. m. 91.

¹⁰⁴ Çakır-Somlyai i. m. 106.

¹⁰⁵ Krekó i. m. 268.