

Furcsa kifejezések miatt buknak meg a hamis kutatási dokumentumok¹

Sasvári Péter²

Link: <https://doi.org/10.13140/RG.2.2.29171.71206>

Becsült olvasási idő: 7 perc

Bevezetés

Az alábbi elemzésből kiderül, hogy a furcsa szófordulatok a tudományban elkövetett csalásokra utalhatnak. 2021 áprilisában felkeltette egy informatikusokból álló csoport érdeklődését a folyóiratcikkekben szereplő furcsa kifejezések sorozata. Nem értették, miért használják a kutatók a „hamis tudat”, az „*alapos neurális szervezet*” és a „*hatalmas információ*” kifejezéseket a szélesebb körben ismert „*mesterséges intelligencia*”, „*mély neurális hálózat*” és „*big data*” kifejezések helyett. [1]

Elemzés

A további vizsgálatok kiderítették, hogy ezek a szokatlan megfogalmazások – amelyeket ők „*tortured phrases*”-nek, azaz furcsa kifejezéseknek neveznek – valószínűleg gépi fordítás vagy a plágiumot leplezni próbáló szoftverek eredményei. És úgy tűnik, hogy ezek gyakoriak a számítástechnikai tudományok területén írt tanulmányokban.

A kutatás-integritás kutatók véleménye szerint Cabanac és kollégái a hamis kutatási tanulmányok egy új típusát fedezték fel, és a július 12-én [2] az arXiv-on közzétett munkájuk az érintett szakirodalom tekintetében csak a jéghegy csúcsát jelenti.

A kutatók a Dimensions hivatkozási adatbázisban indexált folyóiratcikkekben számos furcsa megfogalmazásra kerestek rá, hogy felmérjék, hány cikket érint ez a probléma. Több mint 860 olyan publikációt találtak, amelyekben legalább egy ilyen kifejezés szerepelt, és ezek közül 31-et egyetlen folyóiratban, a *Microprocessors and Microsystems*ben (Elsevier folyóirat, <https://www.sciencedirect.com/journal/microprocessors-and-microsystems>) publikáltak.

„*Ez árt a tudománynak. Nem bízhatunk ezekben a folyóiratokban, ezért meg kell találni és vissza kell vonni őket*” – mondja Guillaume Cabanac, a franciaországi Toulouse-i Egyetem informatikusa, aki a tanulmányon dolgozott.

¹ Az alábbi közlemény a Nemzeti Közszolgálati Egyetem Államtudományi és Nemzetközi Tanulmányok Kar gondozásában megjelenő **Államtudományi Hírlevél** Tudományos sarok rovatában jelent meg. A korábbi hírlevelek elérhetőek az alábbi honlapon keresztül: <https://antk.uni-nke.hu/kutatas-tudomanyos-elet/allamtudomanyi-hirlevel/allamtudomanyi-hirlevel-2023>

Az oktatási anyagnak szánt tanulmány **‘Tortured phrases’ give away fabricated research papers**, <https://doi.org/10.1038/d41586-021-02134-0> alapján készült.

² Egyetemi docens, Nemzeti Közszolgálati Egyetem, Államtudományi és Nemzetközi Tanulmányok Kar, Közszervezési és Infotechnológiai Tanszék, 1083 Budapest, Ludovika tér 2. E-mail: Sasvari.Peter@uni-nke.hu

1. táblázat: Furcsa megfogalmazások a számítástechnikai folyóiratokban

Tudományos kifejezés	Furcsa megfogalmazás
Big Data	Colossal information (hatalmas információ)
Artificial intelligence (mesterséges intelligencia)	Counterfeit consciousness (hamis szemlélet)
Deep neural network (mély neurális hálózat)	Profound neural organization (alapos neurális szervezet)
Remaining energy (maradvány energia)	Leftover vitality (maradék életerő)
Cloud computing (felhőalapú számítástechnika)	Haze figuring (pára kalkuláció)
Signal to noise (jel-zaj viszony)	Flag to commotion (zászló a felforduláshoz)
Random value (véletlen érték)	Irregular esteem (szabálytalan megbecsülés)

Cabanac és munkatársai azt gyanították, hogy a furcsa megfogalmazások gépi fordítás vagy a meglévő szöveget átíró szoftverek eredményei, ezért a *Microprocessors and Microsystems* és más folyóiratok kivonatainak egy részét olyan eszközzel futtatták le, amely felismeri, hogy a szövegeket a [GPT mesterséges intelligencia eszközzel](#) [3] generálták-e. Az eszköz által megjelölt *Microprocessors and Microsystems* tanulmányok közül a manuális ellenőrzések során néhányban „hibákat” tártak fel, például értelmetlen szöveget, valamint plagizált szövegeket és képeket.

Az alaposabb vizsgálat érdekében a csoport letöltötte a *Microprocessors and Microsystems* című folyóiratban 2018 és 2021 között megjelent összes cikket; ezt az időkeretet azért választották, mert 2019-ben jelent meg a GPT frissített változata. Különböző tényezők alapján mintegy 500 „megkérdőjelezhető cikket” azonosítottak. Elemzésükből kiderült, hogy a 2021 februárja után megjelent cikkek elfogadási ideje átlagosan ötször rövidebb volt, mint az ezt megelőzően publikáltaké. Ezeknek a cikkeknek nagy hányada kínai szerzőktől származott. A tanulmányok egy részhalmazának pedig azonos volt a beadási, áttekintési és elfogadási időpontja, és ezek többsége a folyóirat különszámaiban jelent meg. A szerzők szerint ez gyanús. A főszerkesztő által felügyelt rendszeres számokkal ellentétben a különszámokat általában egy vendégszerkesztő javasolja és felügyeli, és egy adott kutatási területre összpontosítanak.

Nem a *Microprocessors and Microsystems* volt az egyetlen érintett cím – a kutatók több száz más folyóiratban megjelent cikkekben is találtak bizonyítékot a furcsa kifejezésekre. „Az előzetes vizsgálatok azt mutatják, hogy több ezer furcsa megfogalmazást tartalmazó cikket indexelnek a főbb adatbázisok” – írják, hozzátéve, hogy „más tudományterületek fogalmaival kapcsolatos további furcsa megfogalmazásokat még nem fedtek fel”.

A különszámok vizsgálata

Körülbelül egyidőben azzal, hogy Cabanac és kollégái először vették észre a furcsa kifejezéseket, és tudtukon kívül, a *Microprocessors and Microsystems* szerkesztőjének is aggályai támadtak a folyóirat néhány különszámában megjelent tanulmány szakértői értékelésének integritásával és szigorúságával kapcsolatban.

A folyóirat kiadója, az Elsevier vizsgálatot indított. Ez még folyamatban van, de július közepén a kiadó több mint 400, a folyóirat hat különszámában megjelent cikkel kapcsolatban fejezte ki aggályait.

Eszerint a *Microprocessors and Microsystems* érintett különszámaiban szereplő tanulmányokat egyenként, „*egymástól függetlenül újraértékelik*”, és amint a vizsgálatok lezárultak, a folyóirat további tájékoztatást ad majd a státuszukról.

A kiadó hozzáteszi, hogy a folyóirat „*szerkesztési rendszerének konfigurációs hibája*” miatt sem a főszerkesztő, sem a lapok kezelésére kijelölt szerkesztő nem kapta meg a tanulmányokat jóváhagyásra, ahogyan kellett volna. „*Ez a konfigurációs hiba a rendszer migrációja miatti átmeneti probléma volt, és amint felfedeztük, ki is javítottuk*” – áll a közleményben.

Az Elsevier szóvivője a *Nature*-nek adott nyilatkozatában elmondta, hogy a *Microprocessors and Microsystems* vizsgálata megállapította: a szerzők valószínűleg visszafordító-szoftvert használtak a plágium álcázására, és hogy valószínűleg ez a furcsa megfogalmazások forrása.

A vizsgálat azt is feltárta, hogy 49, Cabanac és kollégái által gyanúsnak ítélt és a folyóirat rendszeres számaiban megjelentetett cikket eredetileg a folyóirat különszámaiba nyújtottak be, és a vendégszerkesztők fogadták el azokat, „*de a szerzők kérésére később a rendszeres számokban jelentek meg*” – áll a közleményben. Ezek a cikkek már az Elsevier vizsgálatának részét képezik – teszik hozzá.

Elisabeth Bik, [egy kaliforniai kutatás-integritás elemző](#), [4] aki arról ismert, hogy felismeri a duplikált képeket a tanulmányokban, azt mondja, hogy Cabanac kutatásának eredményei „*sokkolóak*”. „*Ez a hamis cikkek egy teljesen új és nyugtalanító típusa*” – teszi hozzá.

Összefoglalás

Jennifer Byrne, az ausztráliai Sydney Egyetem molekuláris-onkológiai kutatója, aki szintén a hamis cikkek kiszűrésével foglalkozik, azt mondja, hogy ez valószínűleg a jéghegy csúcsa, mivel a kutatók csak egy kiadó egyetlen folyóiratát vizsgálták meg alaposan. „*Ezeket a cikkeket azért is találták meg, mert nagyon rossz minőségűek voltak, de a szakirodalomban több hihető, mesterséges intelligenciával generált cikk is lehet, amelyeket nehezebb felfedezni*” – teszi hozzá.

Felhasznált irodalom

- [1.] Holly Else (2021): ‘Tortured phrases’ give away fabricated research papers, *Nature* **596**, 328-329 (2021), <https://doi.org/10.1038/d41586-021-02134-0>
- [2.] Guillaume Cabanac, Cyril Labbé, Alexander Magazinov (2021): Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals, Preprint at arXiv <https://arxiv.org/abs/2107.06751>
- [3.] Matthew Hutson (2021): Robo-writers: the rise and risks of language-generating AI, *Nature* **591**, 22-25., <https://doi.org/10.1038/d41586-021-00530-0>
- [4.] Helen Shen (2020): Meet this super-spotter of duplicated images in science papers, *Nature* **581**, 132-136, <https://doi.org/10.1038/d41586-020-01363-z>