

## PAPERMILLS – TOVÁRNE NA ČLÁNKY. DÁ SA BOJOVAŤ PROTI ICH PRAKTIKÁM?

Mgr. Michaela Mikušková

*Dynamický proces vedeckého publikovania je neustále atakovaný novými hrozbami. Tieto hrozby narušajú etiku a integritu poctivého vedeckovýskumného procesu a majú vážny dosah na dôveru verejnosti vo vedecké výsledky. Okrem predátorských praktík falošných časopisov, ktoré ohrozujú reputáciu vedeckých výskumníkov, sa vo vydavateľskom sektore objavil nový typ vedeckého podvodu, takzvané papermills alebo továrne na články, ktoré sú spájané s kupovaním autorstva vo vedeckých článkoch a „výrobou“ rukopisov na predaj.*

**Továrne na články** sú nový jav súvisiaci s vedeckou komunikáciou a publikovaním. Ich podvodné praktiky sa dotýkajú najmä redakcií časopisov, ktoré musia zavádzať účinné postupy a hľadať efektívne nástroje, aby v rámci nastaveného publikačného procesu dokázali falošné rukopisy zachytiť.

Nezisková organizácia *Committee on Publication Ethics (COPE)* definuje továrne na články ako „na zisk orientované, neoficiálne a potenciálne nezákonné organizácie, ktoré produkujú a predávajú podvodné rukopisy zdanlivo sa podobajúce na skutočný výskum“ [1]. Existujú rôzne modely, ako továrne na články fungujú. Niektoré predávajú autorstvo vo falošných rukopisoch až po prijatí článku do niektorého z časopisov, iné ešte pred zaslaním rukopisu do časopisu s tým, že sa snažia ovplyvňovať hodnotiaci proces. Továrne v rámci svojich „služieb“ ponúkajú aj existujúce alebo vyfabrikované datasety alebo citácie v už existujúcich publikáciách [2].

Z analýzy časopisu *Nature* vyplýva, že za posledné dve desaťročia bolo publikovaných viac ako 400 000 článkov, ktoré vykazujú vysokú textovú zhodu so štúdiami, ktoré vznikli v továrňach na články. V analýze sa tiež odhaduje, že 1,5 – 2 % všetkých vedeckých článkov uverejnených v roku 2022 sa veľmi podobá na články z tovární. V oblasti biológie a medicíny sa tento podiel odhaduje až na 3 % [3]. Vo výskumnej správe neziskovej organizácie COPE a asociácie *International Association of Scientific, Technical and Medical Publishers (STM)* [4] sa uvádza, že na základe analýzy 53 000 podaných rukopisov v šiestich vydavateľstvách pochádzalo 2 – 46 % z tovární. Väčšina časopisov čelí nižšiemu percentu podvodných

rukopisov. Sú však aj časopisy, ktoré sú v odhaľovaní falošných rukopisov menej úspešné a teda továrňami na články aj viac atakované. Takémuto problému nedávno čelilo vydavateľstvo Wiley, ktoré muselo stiahnuť viac ako 11 300 článkov a zrušiť 19 časopisov. Továrne atakovali najmä open access vydavateľstvo Hindawi, ktoré je súčasťou tohto vydavateľského domu od roku 2021 [5].

Kontrola podozrivých rukopisov pochádzajúcich z tovární je pre redakcie časopisov náročná. Podľa článku publikovaného v časopise *Science* [6] si v roku 2021 postpublikačná kontrola približne troch tisícok článkov vydavateľstva Springer Nature, pri ktorých bolo podozrenie, že pochádzajú z tovární na články, vyžiadala až desať zamestnancov na plný a čiastočný úväzok. Asociácia STM na tento rastúci problém reagovala založením platformy **Integrity Hub**, ktorej cieľom je vytvoriť účinné nástroje na boj proti rukopisom pochádzajúcich z tovární [7].

### **Existuje účinný nástroj na boj proti továrňam?**

COPE a STM vo svojej výskumnej správe [4] upozornili na konkrétne znaky, ktoré majú články z tovární spoločné a ktoré identifikovali na základe rozhovorov s komerčnými vydavateľmi:

- články sú najmä z oblasti bunkovej a molekulárnej biológie, vedná oblasť sa časom môže meniť,
- príliš časté odkazovanie na experimenty a dáta vytvorené na základe laboratórnej techniky Western blot, cytometrických testov, histológie/farbenia buniek,
- dáta nezodpovedajú uvádzanej metodológii alebo nemôžu byť aplikované za zamýšľaným účelom,
- formálna štruktúra rukopisov je veľmi podobná (grafy, použité písmo a pod.),
- autori majú podozrivé e-mailové adresy, ktoré jednoznačne neidentifikujú pracovisko (napr. univerzitu), používajú súkromné e-mailové adresy, prípadne e-maily odkazujú na pracovisko, ktoré nezodpovedá téme článku,
- autori majú vytvorené nové identifikátory ORCID,
- zoznam autorov sa počas komunikácie s vydavateľstvom mení,
- použité obrázky alebo ich časti boli publikované v inom článku a i.

V sofistikovanom svete podvodných praktík, ktoré prenikajú do vedeckého publikovania, je veľmi ťažké vytvárať nástroje, ktoré budú jednoznačne účinné. Mnohé redakcie časopisov využívajú softvéry a metódy, ktoré pomáhajú odhaľovať podvodné aktivity a rozpoznať podradné a falošné články. Niektoré systémy spracovávania rukopisov dokážu odhaliť, ak z toho istého počítača pochádza väčšie množstvo

textov, čo je znak toho, že jedna osoba alebo organizácia mohla vyprodukovať veľký počet štúdií.

V apríli 2023 vytvoril *STM Integrity Hub* prvú vlastnú aplikáciu – nástroj na identifikáciu rukopisov, ktoré produkujú továrne na články [8]. V odhaľovaní falošných rukopisov sa úspešne javí aj nástroj spoločnosti Clear Skies **Papermill Alarm**. Vyvinul ho Adam Day, riaditeľ spoločnosti Clear Skies, ktorý zároveň pracuje ako dátový vedec vo vydavateľstve SAGE. *Papermill Alarm* sa skladá z viacerých samostatných nástrojov a využíva širokú škálu metód a dát: umelú inteligenciu, veľké jazykové modely (Large Language Models – LLM), sieťovú analýzu a ďalšie nástroje na vyhľadávanie signálov, ktoré poukazujú na organizovaný výskum podvodných praktík [9]. Pomocou algoritmu hlbokého učenia na porovnanie jazyka použitého v názvoch a abstraktoch rukopisov s jazykom použitým v článkoch, o ktorých sa vie, že pochádzajú z tovární, softvér text vyhodnotí prostredníctvom semaforu. Červenou označí ten, ktorý vykazuje vysokú mieru znakov podobných s článkami produkovanými v továrňach, oranžovú priradí tomu, ktorý má určité podobnosti a zelenú textu bez akejkoľvek podobnosti [10]. Treba zdôrazniť, že nástroj nie je stopercentne účinný a ak článok označí červenou farbou, neznamená to, že je zaručene falošný. S jeho pomocou však možno vyfiltrovať články, ktorým treba venovať zvýšenú pozornosť a podrobiť ich ďalšej analýze zo strany editorov časopisu prípadne recenzentov. Nedokáže identifikovať ani články vyprodukované v továrňach, ktoré sa vyhýbajú zaužívaným šablónam. Dôveru do tohto nástroja však vkladá aj *STM Integrity Hub*, ktorý ho v roku 2023 integroval do svojho nástroja na detekciu podvodných článkov z tovární [11].

Podvodníci sa neustále snažia využívať čoraz prepracovanejšie postupy a hľadajú si nové ciele. Proces odhaľovania falošných článkov produkovaných v továrňach nebude s veľkou pravdepodobnosťou nikdy plne automatizovaný, preto sa tento problém nedá ani úplne odstrániť. Samotní vydavatelia musia prijať účinné nástroje a venovať značné technické i ľudské zdroje do zabezpečenia integrity svojich časopisov a obsahu, ktorý publikujú. Len nepretržitá ostražitosť, výmena skúseností, spolupráca vydavateľstiev a rešpektovanie odporúčaní etických organizácií môžu vytvoriť bariéru, ktorá eliminuje dosah týchto podvodov.

### Použité zdroje

[1] Systematic manipulation of the publishing proces via paper mills: forum discussion topics September 2020. In COPE [online]. Committee on Publication Ethics, 2020 [cit. 2024-07-25]. Dostupné na: <https://publicationethics.org/systematic-manipulation-paper-mills>

[2] FTÁČNIKOVÁ, Soňa. Papermill – továreň na články alebo autorstvo na predaj! *ITlib: informačné technológie a knižnice* [online]. 2023, špeciál [cit. 2024-07-25]. Dostupné na: <https://doi.org/10.52036/1335793X.2023.SC.26-33>

[3] NOORDEN, Richard van. How big is sciences's fake-paper problem? *Nature* [online]. 2023,

no. 623, pp. 466 – 467 [cit. 2024-07-25]. Dostupné na: <https://doi.org/10.1038/d41586-023-03464-x>

[4] COPE & STM. *Paper Mills – Research report from COPE & STM* [online]. Committee on Publication Ethics, June 2022 [cit. 2024-07-25]. Dostupné na:

<https://doi.org/10.24318/jtbG8IHL>

[5] Research paper mill. In *Wikipedia: the free encyclopedia* [online]. San Francisco: Wikimedia Foundation, 2021-, last edited 24 May 2024 [cit. 2024-07-25]. Dostupné na:

[https://en.wikipedia.org/wiki/Research\\_paper\\_mill](https://en.wikipedia.org/wiki/Research_paper_mill)

[6] BRAINARD, Jeffrey. Fake scientific papers are alarmingly common. *Science* [online]. 2023, vol. 380, iss. 6645, 9 May 2023 [cit. 2024-07-25]. Dostupné na:

<https://doi.org/10.1126/science.adi6523>

[7] STM Integrity Hub. In *STM Advancing trusted research* [online]. [cit. 2024-07-25]. Dostupné na: <https://www.stm-assoc.org/stm-integrity-hub/>

[8] STM Solutions releases MVP of new paper mill detection tool. In *STM Advancing trusted research* [online]. 13 April 2023 [cit. 2024-07-25]. Dostupné na: <https://www.stm-assoc.org/papermillchecker/>

[9] *Clear Skies* [online]. Clear Skies Ltd [2024-07-24]. Dostupné na: <https://clear-skies.co.uk/>

[10] ELSE, Holly. Papermill alarm software flags potentially fake papers. *Nature* [online]. 23 September 2022 [cit. 2024-07-24]. Dostupné na: <https://doi.org/10.1038/d41586-022-02997-x>

[11] STM Integrity Hub incorporates Clear Skies' Papermill Alarm screening tool. In *STM Advancing trusted research* [online]. 01 September 2023 [cit. 2024-07-25]. Dostupné na:

<https://www.stm-assoc.org/stm-integrity-hub-incorporates-clear-skies-papermill-alarm-screening-tool/>