Боротьба з поширенням фейк-ньюз: цифрові інструменти верифікації контенту

К. О. Назаренко

Київський національний університет ім. Тараса Шевченка, м. Київ, Україна
Corresponding author. E-mail: nazarenkokristina@gmail.com

Paper received 30.05.20; Accepted for publication 16.06.20.

https://doi.org/10.31174/SEND-HS2020-231VIII39-12

Анотація. У статті розглянуто методи та інструменти верифікації контенту. Акцентовано увагу на нові сервіси та системи, які здатні надати правдиві дані або визначити фрагменти спотвореної інформації у медіаконтенті, описані інструменти з залученням штучного інтелекту. Визначено 3 групи інструментів з верифікації контенту, що доступні не тільки журналістам, а й звичайному читачеві.

Ключові слова: *фейкові новини, верифікація контенту, фактчекінг, цифрові ресурси, штучний інтелект, діп-фейки.*

Вступ. Проблема продукування й поширення фейкових новин сьогодні постала як ніколи гостро. Її порушують політики, вчені, представники медіакорпорацій, журналісти, а держави розроблюють політичні та правові механізми задля контролю над її розповсюдженням. Хибна інформація у різних формах стала неодмінною частиною сьогоднішнього медіапростору та чинить вплив на всіх його акторів, змінює контексти та сенси. Надзвичайно важливим інструментом протидії розповсюдженню фейків став процес верифікації та фактчекінгу. Створені інструменти перевірки фактів, новин, контенту надають надважливу опцію перевірки інформації будь-кому в незалежності від місця знаходження.

Короткий огляд публікацій за темою. Chokshisept N. «How to Fight 'Fake News' (Warning: It Isn't Easy)», Boudreau M.&Straub D.&Gefen D. «Validation in Information Systems Research: A State-of-the-Art Assessment», Гороховський О. «Фактчек як тренд розслідувань: можливості та перспективи», Шевченко В. «Фактчекінг і верифікація у журналістській роботі», Рябічев В. Л. «Верифікація контенту в соціальних медіа» та ін.

Мета дослідження: визначити технологічні можливості та інструменти верифікації контенту, які є доступними як для журналістів, так і для читача.

Матеріали та методи. Основними методами, застосованими у статті, є контент-аналіз та синтез, а також наукове моделювання. Для досягнення визначеної мети було проаналізовано ресурси, які дозволяють перевіряти інформацію, структуровано знайдені інструменти і зроблені узагальнення щодо специфіки перевірки даних та фактів.

Результати та їх обговорення.

Сьогодні, для потреб авторів, журналістів та редакторів створено кілька сотень сервісів й інформаційних ресурсів різних типів, які надають можливість верифікувати інформацію та отримати необхідні свідчення не залежно від місця знаходження журналіста чи користувача.

Верифікація – процес перевірки контенту на вірогідність або істинність. Близьке до цього поняття «фактчекінг» використовуються в сенсі перевірки фактичних тверджень з метою виявлення недомовок, маніпуляцій, розповсюдження неправдивої інформації [5]. В контексті цього дослідження обидва поняття вважаються комутативними, оскільки мають одну й ту саму остаточну мету – перевірити матеріал на правдивість.

ІТ-індустрія сьогодні розробила низку діджитал інструментів, що дозволяють перевіряти інформацію. Останні можна поділити на такі категорії:

- цифрові механізми та ресурси із перевірки контенту або його частини;
- цифрові джерела відкритих даних;
- цифрові продукти із залученням штучного інтелекту.

І. Отже, серед цифрових механізмів перевірки інформації можна виокремити такі категорії:

Інструменти перевірки особи:

Facebook Graph Search – розумний пошуковик, який існує в межах соціальної мережі Facebook з 2013 року та надає можливість швидко знайти людей, відеочи фото- контент. Пошук можливо здійснювати за будь-якими параметрами, до прикладу, виокремивши людей за вподобаннями певних сторінок, за комунікаційними зв'язками з іншими користувачами мережі чи за місцем знаходження людини у той чи інший період часу або зараз. Окрім того, пошук можна здійснювати навіть якщо ім'я людини не відомо, наприклад, за посадою чи віком [18].

GeoSocial Footprint – ресурс, який об'єднує фрагменти інформації про місцеперебування «цифровими слідами» або публікаціями, які користувач залишає в мережі. Платформа дозволяє відстежити зміну геолокації користувача. Проте, обов'язковою умовою є активована функція GPS [23].

LinkedIn – соціальна мережа, яка містить інформацію про місце роботи та ділові інтереси понад 675 млн осіб з різних куточків світу. Платформа дозволяє отримати додаткову інформацію щодо місця роботи, професійних або наукових здобутків людини [29].

Сервіси, які допомагають верифікувати місця на світлинах:

GoogleMaps – популярна, доступна карта в режимі онлайн, яка надає супутникові зображення та можливість переглянути ландшафт, інформацію про погоду і панорамний вид на 360 градусів на рівні вулиць. Кейси про те, як через супутникові знімки були винайдені факти, що сприяли розслідуванням, можна часто зустріти у медіа [7]. Проте, у червні 2019 року видання The Wall Street Journal опублікувало матеріал, у якому викрило понад 11 млн неіснуючих компаній та адрес, які містилися у GoogleMaps [13]. Тож інформацію з будь-якого

джерела варто перевірити декілька разів.

NASA Earth Observatory – онлайн видання НАСА, засноване у 1999 році, яке дозволяє отримати доступ до супутникових зображень з обсерваторії Землі. Портал діє як світовий архів зображень, з безкоштовними картами, зображеннями і наборами даних [14].

Wikimapia – міжнародний безкоштовний вебсайт, графічна онлайн-енциклопедія, мета якої відзначити всі географічні об'єкти Землі. Проєкт поєднує в собі інтерактивну карту з принципом вільного редагування. Зараз у Wikimapia зареєстровано понад 2,4 млн користувачів, а на карту додано понад 27 млн об'єктів. Всі дані доступні для загального користування [39].

Інструменти перевірки правдивості зображень:

Assembler – інструмент для перевірки автентичності зображень, який допомагає виявляти сфальсифіковані фотографії, навіть ті, що створені за допомогою штучного інтелекту. Assembler аналізує зображення і відзначає передбачувані сліди редагування. Кожен з вбудованих в програму «детекторів» створений для виявлення певного типу маніпуляцій зі знімком (колірних аномалій, областей зображення, які були скопійовані і вставлені кілька разів, використання більше однієї моделі камери для зйомки) [8].

Jeffrey's Exif Viewer – онлайн інструмент, який дозволяє екстрагувати метадані з цифрової фотографії, навіть якщо такі данні відсутні на опублікованому зображенні. Завдяки інструментам сервісу можна визначити дату і час знімання, параметри налаштування камери і, в деяких випадках, GPS координати місця зйомки [27].

Foto Forensics – сервіс, який використовує аналіз рівня помилок (ELA), щоб віднайти частини зображення, які було змінено. ELA шукає відмінності в рівнях якості зображення, висуваючи на перший план ті, де, можливо, були зроблені зміни. Сервіс активно використовується редакціями, журналістами, фактчекерами для встановлення факту маніпуляцій із зображеннями [21].

Google Search by Image – сервіс, який дозволяє прослідкувати історію зображення в мережі Інтернет. Завантаживши зображення або зазначивши його пряму вебадресу, користувач може знайти пов'язані або подібні зображення, вебсайти та інші сторінки, де було використане це зображення [25].

TinEye – пошукова система, сайт, яка надає можливість ретроспективно перевірити зображення та визначити його джерело й випадки його використання. Також, можливо встановити чи існують будь-які інші копії зображення [38].

Це не вичерпний перелік наявних технологічних можливостей перевірки контенту, проте достатній задля ідентифікації «правдивості» чи «фейковості» певних даних чи новин.

Крім того, кожного року по всьому світу створюється дедалі більше профільних ресурсів для перевірки інформації. Наприклад, нещодавно Колумбійський університет запустив ресурс **Emergent**. Цей сайт дозволяє простежити «оригінальність користувацького контенту». Алгоритм здатний аналізувати скільки разів контент згадували в мережі та як саме. Таким чином, цей портал може визначити звідки походить «вірусне поширення» інформації або матеріалу [17].

Для англомовних користувачів доступний спеціальний сервіс перевірки достовірності відео – **Citizen Evidence Lab**, де можна визначити точний час завантаження відео, геолокацію тощо [12].

II. Наступною категорією ресурсів, які дозволяють отримати доступ до різного роду інформації є цифрові джерела відкритих даних.

Серед міжнародних баз та реєстрів відкритих даних можна визначити:

The Grafiti – система, яка шукає графіки, діаграми й статистичні дані в різноманітних джерелах по всьому світу [37].

GapMinder – інформаційний центр, що збирає датасети за різними темами: економіка, освіта, охорона навколишнього середовища та ін. [22].

Бази даних OCCRP – глобальний архів дослідних матеріалів для журналістських розслідувань. Понад 250 наборів даних з 230 країн. Реєстри компаній, витоку документів, архіви судових документів, ліцензії і концесійні угоди, закупівлі, секретні документи, санкційні списки, розпорядчі документи та ін. [31].

Серед міжнародних баз даних можна визначити: реєстри юридичних актів, компаній, нерухомості, кадастрові мапи тощо. Серде них можна згадати: **BAILII** [9], **Infogreffe** [26], **земельний кадастр Іспанії** [16], **HypoDoc** [24] та ін.

Серед цифрових джерел відкритих даних в Україні, можна визначити такі ресурси:

Портал відкритих даних – урядовий ресурс відкритих даних, який передбачає доступ до інформації органів влади з можливістю її наступного використання. Для журналіста цей сайт – базова платформа для ознайомлення з поняттям відкритих даних, принципами їх функціонування, додаткової перевірки інформації [3].

Пошуково-аналітична система. 007 – ресурс, що містить інформацію про використання організаціями та інституціями коштів державного бюджету. На сайті працює пошук за кодом організації ЄДРПОУ, який дозволяє проаналізувати всі транзакції та візуалізувати фінансові зв'язки між державою і компаніями [4].

E-data – сайт, що надає доступ до сервісів open budget і spending. Open budget показує детальні схеми доходів та витрат державного бюджету України за кожен місяць. Spending показує трансакції казначейства, звітність та договори розпорядників, державних цільових фондів, державних та комунальних підприємств [15].

ProZorro – розміщує дані про публічні закупівлі та тендери в Україні [35].

Opendata – ресурс, що дозволяє моніторити реєстраційні дані українських компаній та інформацію з судового реєстру. Сервіс також працює у різних месенджерах [33].

SaveEcoBot – бот, який функціонує в месенджера, дозволяє проаналізувати дозволи компаній на потенційно небезпечну для навколишнього середовища діяльність (наприклад, викиди певних речовин тощо) [36].

III. Сьогодні сприяти розв'язанню проблеми поширення фейкової інформації активно почали й цифрові продукти із залученням штучного. Адже виникнення та поширення «Deep fake», мабуть, одного з найбільш ефективних та високотехнологічних способів

спотворення інформації, кинуло виклик концепції об'єктивної та всебічної журналістики. «Deep fake» – технологія на основі штучного інтелекту, яка використовується для створення або змінення відеовмісту.

Штучний інтелект, на думку Антоненко В. М., Рогушина Ю.В. – це наукова дисципліна, мета якої автоматизувати інтелектуальну діяльність людини [1].

Cape – продукт британського стартапа Bloomsbury AI, який у 2018 році придбав Facebook. Cape – механізм, який має навички машинного читання та може відповідати на прочитані питання. Тобто, він здатний вивчати джерела новин і визначати хибні повідомлення. Нейромережа вивчає текст новини та перевіряє посилання автора на джерела, використовуючи при цьому технології розробки текстів з використанням обробки природної мови (NLP) [32].

Як визначають Ю.І. Лехан, О.А. Пастух, обробка природної мови (Natural Language Processing) – це міждисциплінарна галузь, яка стоїть на перетині комп'ютерних наук, штучного інтелекту та обчислювальної лінгвістики. Основним проблемним полем є забезпечення взаємодії між комп'ютерами та людськими (природними) мовами [2].

Інтелектуальний аналіз тексту (Text mining) – напрям інтелектуального аналізу даних та штучного інтелекту, метою якого є отримання високоякісної інформації з колекцій текстових документів за допомогою застосування методів машинного навчання та обробки природної мови. Основна задача Text mining полягає в тому, щоб виявити інформацію, яка, можливо, невідома і прихована в контексті іншої інформації. Це досягається за допомогою різних методологій аналізу. Обробка природної мови – одна з них, вона виконує лінгвістичний аналіз, що допомагає машині «читати» текст.

Factmata – стартап, який займається протидією дезінформації в Мережі за допомогою штучного інтелекту. Компанія визначає свою місію у тому, аби «допомогти світові зрозуміти якість, надійність та безпеку онлайн інформації – створення кращого Інтернету для всіх».

Компанія розробляє ряд інструментів, як для звичайних користувачів інтернету, так і для компаній, і ЗМІ (Forbes, Bloomberg, The Guardian, The New York Times. The Times). Ядром проєкту є відкрита новинна платформа, яка міститься у відкритому доступі. Роль штучного інтелекту в проєкті полягає в тому, щоб перевіряти надійність і якість опублікованих повідомлень. Кожний фрагмент тексту з будь-якого посилання проходить семантичну перевірку на пропаганду насильства, політичну ангажованість, сексизм, расизм, екстремізм, мову ненависті, агресивну риторику. Штучний інтелект визначає вебсайти або статті, які поширюють неправдиву інформацію з наміром комерційної вигоди. Окрім штучного інтелекту Factmata працює зі спільнотами та експертами. Завдяки роботі з експертними журналістами та дослідниками соціальних наук, Factmata розробили евристику, яка дозволяє швидко виділити проблематичний зміст у посланні або статті [20].

AdVerif.ai – стартап, який був створений для комерційної мети аби маркетологи мали можливість верифікувати рекламні оголошення та захищати власний бренд. Згодом, із розширенням інтерфейсу, були запропоновані можливості перевірки інформації для видавців та медіа. Інструмент доповнює роботу редакції з можливостями глибокого навчання та обробки природних мов для виявлення шаблонів, які вказують на спам, шкідливе програмне забезпечення або невідповідний вміст. Алгоритми штучного інтелекту використовують онлайн-сховища знань для підтвердження фактів або виділення потенційно підроблених [30]. AdVerif.ai використовує найсучасніші технології обробки природних мов (НЛП) разом з машинним навчанням (Deep Learning) аби виявити закономірності, пов'язані зі спамом і невідповідним вмістом.

Машинне навчання – галузь комп'ютерних наук, яка вивчає методи навчання комп'ютеризованих систем на підставі даних без програмування їх поведінки [11].

Використовуючи технологію Machine Vision, AdVerif.ai також працює із фото та відео матеріалами, аналізуючи чи містить він елементи насильства або вікового контенту.

Висновки. Отже, сьогодні існує ціла низка діджитал інструментів, що дозволяють перевіряти інформацію різного роду журналістам та читачам, а також визначати спотворені елементи у медіаконтенті. Відкриті джерела даних, реєстри та цифрові ресурси здатні надати доступ майже до невичерпного переліку відомостей. У симбіозі з традиційними механізмами перевірки інформації, такі інструменти верифікації можуть зіграти важливу роль у протидії поширенню фейк-ньюьз, а подальша розробка інструментів з залученням штучного інтелекту, механізмів глибокого навчання та обробки природних мов, у майбутньому можуть забезпечити високий рівень перевірки інформації, що генерується та поширюється.

ЛІТЕРАТУРА

1. Антоненко В. М., Рогушина Ю.В. Сучасні інформаційні системи і технології. Навчальний посібник. – К.: КСУ МГІ, 2005. – 131 с.
2. Лехан Ю.І., Пастух О.А. Обробка та аналіз текстової інформації методами машинного навчання. – Матеріали VII Міжнародної науково-технічної конференції молодих учених та студентів. Актуальні задачі сучасних технологій. – Тернопіль 28-29 листопада 2018. – С. 145.
3. Портал відкритих даних [Е. ресурс] — Режим доступу: https://data.gov.ua/
4. Пошуково-аналітична система.007 [Е. ресурс] — Режим доступу: https://www.007.org.ua
5. Семчишин Я. Фейк vs Факт. Як перевіряти інформацію від публічних осіб [Електронний ресурс]// MadiaLab Online – Режим доступу: http://medialab.online/news/fejk-vs-fakt-yak-pereviryaty-informatsiyu-vid-publichny-h-osib/
6. Ушакова І. О. Інформаційні системи та технології на підприємстві: конспект лекцій / І. О. Ушакова, Г. О. Плеханова. – Харків: Вид. ХНЕУ, 2009. – 128 с.
7. Чумаков Н. Google Maps помогли найти тело жителя Флориды — он пропал более 20 лет назад [Е. ресурс] / Николай Чумаков // Bird in flight. – 2019. – Режим доступу: https://birdinflight.com/ru/novosti/20190913-google-maps-22-years.html
8. Assembler [Е. ресурс] — Режим доступу: https://jigsaw.google.com/assembler/

9. BAILII [E. ресурс] — Режим доступу: https://www.bailii.org
10. Better Language Models and Their Implications [E. ресурс] / [A.Radford, J. Wu, D. Amodei та ін.]. – 2019. – Режим доступу: https://openai.com/blog/better-language-models/#sample5
11. Bishop C.M. Pattern Recognition and Machine Learning / C.M. Bishop. — NY: Springer. — 2006. – Режим доступу: https://adverifai.com/technology/
12. Citizen Evidence Lab [E. ресурс] — Режим доступу: https://citizenevidence.org
13. Copeland R. Millions of Business Listings on Google Maps Are Fake—and Google Profits [E. ресурс] / R. Copeland, K. Bindley // The Wall street Journal. – 2019. – Режим доступу: https://www.wsj.com/articles/google-maps-littered-with-fake-business-listings-harming-consumers-and-competitors-11561042283?shareToken=stf36c9ebf181345f6b6af7084e214a6dc
14. Earth Observatory [E. ресурс] — Режим доступу: https://earthobservatory.nasa.gov
15. E-data [E. ресурс] — Режим доступу: https://e-data.gov.ua/
16. El RMC [E. ресурс] — Режим доступу: http://www.rmc.es/InfoIndicesTitularidad.aspx
17. Emergent. A real-time rumor tracker [E. ресурс] — Режим доступу: http://www.emergent.info
18. Facebook Graph Search [E. ресурс] // Facebook — Режим доступу: https://www.facebook.com/graphsearcher/
19. FactCheck.org [E. ресурс] — Режим доступу: https://www.factcheck.org
20. Factmata [E. ресурс] — Режим доступу: https://factmata.com
21. Fotoforensics [E. ресурс] — Режим доступу: http://fotoforensics.com
22. GapMinder [E. ресурс] — Режим доступу: https://www.gapminder.org
23. GeoSocial Footprint [E. ресурс] — Режим доступу: https://geosocialfootprint.com
24. HypoDoc [E. ресурс] — Режим доступу: http://www.hypodoc.fr
25. Images.Google [E. ресурс] — Режим доступу: https://images.google.ru
26. Infogreffe [E. ресурс] — Режим доступу: https://www.infogreffe.fr
27. Jeffrey's Image Metadata Viewer [E. ресурс] — Режим доступу: http://exif.regex.info/exif.cgi
28. Lee D. Researchers create 'malicious' writing AI [E. ресурс] /Dave Lee. – 2019. – Режим доступу: https://www.bbc.com/news/technology-47249163
29. LinkedIn [E. ресурс] — Режим доступу: https://www.linkedin.com
30. Marr B. Fake News And How Artificial Intelligence Tools Can Help [E. ресурс] / B. Marr // Forbes. – 2018. – Режим доступу: https://www.forbes.com/sites/bernard-marr/2018/05/16/fake-news-and-how-artificial-intelligence-tools-can-help/#7d4a859f271d
31. OCCRP [E. ресурс] — Режим доступу: https://aleph.occrp.org
32. O'Hear S. Facebook is buying UK's Bloomsbury AI to ramp up natural language tech in London [E. ресурс] / Steve O'Hear // Techcrunch. – 2018. – Режим доступу: https://techcrunch.com/2018/07/02/thebloomsbury/
33. Opendatabot [E. ресурс] — Режим доступу: https://opendatabot.ua/
34. Politi Fact [E. ресурс] — Режим доступу: http://www.politifact.com
35. ProZorro [E. ресурс] — Режим доступу: https://prozorro.gov.ua/
36. SaveEcoBot [E. ресурс] — Режим доступу: https://www.saveecobot.com/
37. The Grafiti [E. ресурс] — Режим доступу: https://beta.grafiti.io
38. TinEye. Reverse Image Search [E. ресурс] — Режим доступу: https://www.tineye.com
39. Wikimapia [E. ресурс] — Режим доступу: http://wikimapia.org

REFERENCES

1. Antonenko V.M, Rogushina Y.V. Modern information systems and technologies. Tutorial. - K., 2005.– 131 p.
2. Lekhan Y.I, Shepherd O.A. Processing and analysis of textual information by machine learning methods. – Proceedings of the VII International Scientific and Technical Conference of Young Scientists and Students. Actual problems of modern technologies. – Ternopil, November 28-29, 2018.– P. 145.
3. Open data portal, available at: https://data.gov.ua/
4. Search and analytical system.007, available at: https://www.007.org.ua
5. Semchishin J. Fake vs Fact. How to check information from public figures, available at: http://medialab.online/news/fejk-vs-fakt-yak-pereviryaty-informatsiyu-vid-publichny-h-osib/
6. Ushakova I.O Information systems and technologies in the enterprise: lecture notes / I.O Ushakova, G.O. Plekhanova. – Kharkiv: Ed. KhNEU, 2009. – 128 p.
7. Chumakov N. Google Maps helped to find the body of a resident of Florida - he disappeared more than 20 years ago/ N. Chumakov // Bird in flight. – 2019. – available at: https://bird-inflight.com/ru/novosti/20190913-google-maps-22-years.html

Fake news proliferation struggle: content verification digital tools

K. O. Nazarenko

Abstract. The article discusses the methods and tools of content verification. Emphasis is placed on new services and systems that are able to provide true data or identify fragments of distorted information in media content, described tools involving artificial intelligence. There are 3 groups of content verification tools that are available not only to journalists but also to the ordinary reader.

Keywords: *fake news, verification of information, factchecking, digital resources, artificial intelligence, deep fakes.*