ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
https://nv.nltu.edu.ua
https://doi.org/10.36930/40320612
Article received 19.12.2022 p.
Article accepted 29.12.2022 p.
UDC 004.[02+04+58+62+77+8+91]
ISSN 1994-7836 (print)
ISSN 2519-2477 (online)
@ Correspondence author
T. P. Diak
tetiana.p.diak@lpnu.ua

***Т. П. Дяк*[1], *Ю. І. Грицюк*[1], *П. П. Горват*[2]**
[1] Національний університет "Львівська політехніка", м. Львів, Україна
[2] Ужгородський національний університет, м. Ужгород, Україна

ПРОБЛЕМА ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИН НА ВЕБ-САЙТАХ МЕРЕЖІ ІНТЕРНЕТ

Проаналізовано наявні підходи до вирішення проблеми виявлення фейкових новин у мережі Інтернет, розглянуто екосистему новин як бізнес-модель їхньої появи, ознайомлення та поширення, що передбачає комплекс взаємопов'язаних сутностей – виробників новинної інформації її користувачів і розповсюджувачів, які сукупно можуть вирішити різноманітні завдання потенційних учасників на єдиному інтегрованому полі. З'ясовано, що мережа Інтернет має очевидні переваги над традиційними ЗМІ у розповсюдженні новин, такі як миттєвий доступ читачів до потрібної інформації, безкоштовне її розміщення, відсутність обмежень у стилі подання та різноманітність формату – текстова, графічна та мультимедійна. Однак, їхня неврегульованість будь-яким редакційним наглядом, а також державними органами з інформаційної безпеки призвели до того, що пересічному читачу часто важко визначити достовірність інформації в деяких опублікованих новинах. Встановлено, що серед вітчизняних фахівців заслуговують уваги ґрунтовні публікації в основному професійних журналістів, у яких вони висвітлюють як різну хибну інформацію, так і повну дезінформацію. Не відстають від них і молоді дарування, які у своїх критичних дописах розвінчують міфи про силу і міць північного сусіда, а також різні фейки про ті чи інші резонансні події. Зазначену проблему за останнє десятиліття з успіхом почали досліджувати закордонні вчені, які домоглися чималих результатів як у практичному, так і теоретичному планах. Досліджено, що головним завданням виявлення фейкових новин є автоматизована їх ідентифікація на ранніх стадіях появи, а також відсутність або мала кількість так званої позначеної (маркованої) інформації для машинного навчання відповідних моделей, призначених для ідентифікації фейкових новин, а також подальшого їх аналізу. Тому багато закордонних дослідників пропонують все нові та нові методи і засоби для виявлення фейкових новин, які з плином часу прогресують у вирішенні цієї проблеми з різним ступенем точності отриманих результатів. З'ясовано, що за терміном екосистемне мислення знаходиться деякий світогляд, цілеспрямоване мислення та відповідні дії людей, залучені в цій системі. Екосистема новин як бізнес-модель їхньої появи, ознайомлення та поширення, передбачає комплекс взаємопов'язаних сутностей – виробників новинної інформації її користувачів і розповсюджувачів, які сукупно можуть вирішити різноманітні завдання потенційних учасників на єдиному інтегрованому полі. Визначено, що існує певний набір методів і засобів, реалізованих у відповідних програмних системах, які найкраще підходять для вирішення проблеми виявлення фейкових новин у мережі Інтернет. Проте, більшість сучасних дослідників намагаються виробити свої підходи із застосуванням комбінацій унікальних і вже апробованих методик, щоб успішно вирішити зазначену проблему. Проаналізовано можливості сучасних програмних систем на підставі інноваційного фреймворку Transformer, який використовує зміст новин, їх контент і соціальний контекст для аналізу їхніх корисних характеристик, а також для прогнозування ймовірності появи серед них фейків. Розроблена модель, маючи в своїй основі архітектуру Transformer, легко піддається машинному навчанню за наборами позначених новин, що допомагає швидко виявляти фейки в новинній інформації.

Ключові слова: інтелектуальний аналіз текстів; комп'ютерна лінгвістика; штучний інтелект; нейронна мережа; генетичний алгоритм; оптимальне рішення; рейтинговий трекер.

Вступ / Introduction

Удосконалення технологій мережі Інтернет і мобільного зв'язку привело до того, що соціальні веб-сервіси та медіа-ресурси стають все більш масштабними та глибоко інтегрованими в наше повсякденне життя [8, 51, 68]. Завдяки легкому до них доступу потенційні користувачі мають звичку отримувати від них інформацію та обмінюватися нею у своїх групах, а також висловлювати й поширювати різні дописи [12, 19, 46, 55, 67]. Шкода, але через відкритість соціальних мереж, велику кількість її користувачів і, як наслідок, різноманітних

Інформація про авторів:
Дяк Тетяна Петрівна, канд. пед. наук, доцент, кафедра прикладної лінгвістики. **Email:** tetiana.p.diak@lpnu.ua; https://orcid.org/0000-0003-0919-9792
Грицюк Юрій Іванович, д-р техн. наук, професор, кафедра програмного забезпечення. **Email:** yurii.i.hrytsiuk@lpnu.ua; https://orcid.org/0000-0001-8183-3466
Горват Петро Петрович, канд. фіз.-мат. наук, доцент, завідувач кафедри комп'ютерних систем та мереж. **Email:** petro.horvat@uzhnu.edu.ua; https://orcid.org/0000-0002-3972-0115
Цитування за ДСТУ: Дяк, Т. П., Грицюк, Ю. І., Горват, П. П. Проблема виявлення фейкових новин на веб-сайтах мережі Інтернет. Науковий вісник НЛТУ України. 2022, т. 32, № 6. С. 78–94.
Citation APA: Diak, T. P., Hrytsiuk, Yu. I., & Horvat, P. P. (2022). The problem of fake news detection on Internet websites. *Scientific Bulletin of UNFU*, 32(6), 78–94. https://doi.org/10.36930/40320612

джерел інформації часто трапляються так звані фейкові новини [7, 21, 26, 72, 86, 91]. Їх можуть поширювати спеціально навчені фахівці, щоб ввести пересічних читачів якщо не в оману, то посіяти серед них хибні думки [28, 33, 35, 44, 49]. Така інформація може завдати серйозної моральної шкоди як безпосередньо цим читачам, так і суспільству загалом, а також спричинити значні політичні та економічні втрати [51, 57, 60, 63]. Причина в тому, що пересічні користувачі не мають часу та навиків перевіряти достовірність прочитаної інформації. Тому перед відповідними державними службами постає нагальне завдання виявляти такі фейкові новини в соціальних мережах, якось їх позначати чи навіть видаляти, тобто гарантувати слабко орієнтованим читачам отримання повсякчас достовірної інформації.

Одним із шляхів вирішення цієї проблеми є мережа спеціалізованих веб-сервісів, призначених розпізнавати такі фейкові новини [13, 67, 70, 37, 81, 88, 90]. Основна концепція цих застосунків полягає в ідентифікації інформації із блогів, статей чи різних новинних ресурсів і відповідній її класифікації на певні категорії, наприклад – фейкова чи достовірна. Дане завдання належить до інтелектуального аналізу текстів (ІАТ, англ. *Text Mining*) – одного з напряму інтелектуального аналізу даних (англ. *Data Mining*) та штучного інтелекту (англ. *Artificial Intelligence*) – розділу комп'ютерної лінгвістики (англ. *Computer Linguistics*) та інформатики, що опікується формалізацією дієвих проблем і завдань, які має виконувати людина [6, 33, 34, 60, 63, 66, 87]. Метою ІАТ є отримання інформації з набору текстових документів, ґрунтуючись на застосуванні ефективних методів машинного навчання та оброблення природної мови [10, 14, 15, 20, 25, 28, 57, 61, 80, 81, 77]. Text Mining використовує всі ті ж підходи до перероблення інформації [6, 18, 40, 61], що й інтелектуальний аналіз даних, однак різниця між ними проявляється тільки в остаточних методах. Окрім цього, Data Mining має справу зі сховищами та базами даних, а Text Mining – з електронними бібліотеками, репозиторіями текстових документів і потоковими текстами мережі Інтернет.

Тому розроблення наукового підходу до автоматизації процесу ідентифікації новинної інформації в мережі Інтернет і його імплементація у відповідний веб-сервіс для можливості використання цільовою аудиторією є актуальною проблемою сьогодення. Робота спрямована на удосконалення уже наявних підходів до вирішення даної проблеми шляхом застосування сучасних методів ІАТ і штучного інтелекту, що забезпечить потенційних користувачів зручним сервісом для перевірки підозрілої інформації.

Об'єкт дослідження – виявлення фейкових новин у мережі Інтернет.

Предмет дослідження – методи і засоби аналізу текстової інформації на веб-сайтах мережі Інтернет, що дасть змогу її ідентифікувати, класифікувати, позначати чи навіть видаляти, а також покращить як стабільність роботи інформаційної системи, так і зменшить її ресурсовитратність.

Мета роботи – проаналізувати наявні підходи до вирішення проблеми виявлення фейкових новин у мережі Інтернет, розглянути екосистему новин як бізнес-модель їхньої появи, ознайомлення та поширення, що передбачає комплекс взаємопов'язаних сутностей – видавців, інформації та користувачів, які сукупно можуть вирішити різноманітні завдання потенційних учасників на єдиному інтегрованому полі.

Для досягнення зазначеної мети необхідно вирішити такі *основні завдання дослідження*:

- проаналізувати стан дослідженості проблеми виявлення фейкових новин вручну та автоматизованої ідентифікації новинної інформації в мережі Інтернет і різні програмні додатки, що реалізують відповідні методи;
- розглянути особливості екосистемного мислення як бізнес-модель діяльності будь-якої компанії та екосистему новин як модель діяльності виробників (ЗМІ – редакційні компанії, журналісти, блогери), поширення новинної інформації (зміст і контент новин) та користувачі (соціальний контекст);
- проаналізувати відомі програмні системи для виявлення фейкових новин у мережі Інтернет, навести як їх позитивні відмінності, так і деякі недоліки, зазвичай, присутні за умови використання найдосконаліших методів;
- провести обговорення результатів дослідження, головним завданням якого є автоматизована ідентифікація фейкових новин на ранніх стадіях їх появи, а також мала кількість так званої позначеної (маркованої) інформації для машинного навчання відповідних моделей.
- зробити висновки за результатами виконаного дослідження та надати рекомендації щодо їх практичного використання.

Аналіз останніх досліджень та публікацій. Ще не так давно людство традиційно взнавало новини з надійних інформаційних джерел, так званих засобів масової інформації (ЗМІ) – щоденних газет, радіо і телебачення, які, зазвичай, дотримувались строгого кодексу публіцистики і доброчесності журналістики [51, 68]. Однак, у кінці 1990-их років мережа Інтернет запропонувала новий спосіб отримання, розміщення інформації та обміну нею спочатку з дотриманням невеликих редакційних стандартів, а згодом – й без них. Різні вебсайти і соціальні мережі стали важливим джерелом новин для багатьох людей [7, 26, 72, 91]. Згідно з даними Ranktracker [68], станом на червень 2022 року мережею Інтернет користуються 4,66 мільярда людей у всьому світі, це означає, що близько 90% інтернет-користувачів використовують повсякчас соціальні мережі, водночас Facebook, будучи провідною платформою соціальних мереж, налічує близько 2,7 мільярда активних користувачів. Отже, існують очевидні переваги мережі Інтернет у розповсюдженні новин, такі як миттєвий доступ читачів до потрібної інформації, зазвичай безкоштовне її розміщення, відсутність обмежень у стилі подання та різноманітність формату – текстова, графічна та мультимедійна. Однак, ці платформи здебільшого не врегульовані будь-яким редакційним наглядом, а ні державними органами [67, 37, 81, 90]. Тому часто важко визначити, правдива чи фейкова інформація в деяких опублікованих новинах.

Серед вітчизняних фахівців заслуговують уваги ґрунтовні публікації в основному професійних журналістів, у яких вони висвітлюють як різну хибну інформацію, в т. ч. й фейкові новини, так і повну дезінформацію. Особливо актуальними є публікації, які стосуються воєнного стану в Україні, починаючи ще з 2014 року, а також військових дій як на лінії розмежування зокрема, так і на території країни загалом. Такими критичними і, водночас, повчальними публікаціями радують нас час від часу журналістські розслідування та дописи таких знаменитостей, як Георгія Почепцова, Марини Довженко, Ярослава Зубченка, Марини Дорош, Отара Довженка, Гали Скляревської, Володимира Малинки, Вадима Міського та ін. Їхній здобуток становить від 150 до 250 публікацій на різні тематики, актуальність яких бу-

ла як під час висвітлення різних подій у той чи інший період, а також вони представляють інтерес для різних служб, в т.ч. й тих, які відповідають за забезпечення інформаційної безпеки суспільства та національну безпеку держави.

Не відстають від них і молоді дарування, які у своїх критичних дописах розвінчують міфи про силу і міць північного сусіда, а також різні фейки про ті чи інші події як у країні загалом, так і на фронті зокрема. Хоча їхні публікації не такі й численні (від 20 до 150), проте варті уваги таких журналістів, блогерів і просто дописувачів як Гончарової Катерини, Красовської Зої, Гордієнко Тетяни, Ломакіної Ангеліни, Білоусенко Ольги, Дуцик Діани, Толокольнікової Катерини, Андрейців Ірини, Бахтєєва Бориса, Грабовського Сергія, Долженкової Інни, Дачковської Марії, Бурковського Петра, Остап Світлани, Рябоштан Ірини, Куляса Ігоря, Єременко Світлани, Рихліцького Володимира, Куликова Андрія, Неймаш Тетяни, Горчинської Олександри, Вишницької Альони, Будівської Галини, Виговської Наталії та ін.

Однак, нас цікавить все ж таки проблема автоматизації процесу ідентифікації новинної інформації в мережі Інтернет. Шкода, але серед вітчизняних науковців важко назвати таких, праці яких би заслуговували уваги. Тому зазначену проблему почали досліджувати закордонні вчені. Оскільки головним завданням виявлення фейкових новин є автоматизована їх ідентифікація на ранніх стадіях появи, то дану проблему досліджували такі науковці: Abu-Salih B. та Wongthongtham P. [1], Baly R. та Karadzhov G. [3], Cao J. та Qi P. [5], Devlin J. та Chang M. W. [7], Helmstetter S. та Paulheim H. [14], Horne B. та Adali S. [16, 17, 18, 19], Jacob Devlin, Ming-Wei Chang. [8], Jin Z. та Cao J. [22, 23], Jwa H. та Oh D. [24], Kai Shu, Amy Sliva. [71], Karimi H. та Roy P. [27], Lewis M. та Liu Y. [32], Liu Y. та Wu Y. F. B. [34, 35], Nakamura K. та Levy S. [38], Nguyen V. H. та Sugiyama K. [44], Papadopoulos S. та Kompatsiaris Y. [48], Patil D. R. та Patil J. B. [50, 51], Pizarro J. [54], Popat K. та Mukherjee S. [56, 57], Przybyla P. [60], Qi P. та Cao J. [62], Radford A. та Wu J. [64], Shu K. та Wang S. [70, 72], Sudhakar Murugesan та Kaliyamurthie K. P. [37], Vosoughi S. та Roy D. [78], Wang Y. та Yang W. [80, 81], Wu X. та Lode M. [84], Yang K.-C. та Niven T. [86], Ying L. та Yu H. [88], Zhou X. та Zafarani R. [90, 91] та ін.

Водночас, іншою проблемою виявлення фейкових новин є відсутність або мала кількість так званих позначених (маркованих) даних для машинного навчання відповідних моделей, призначених для їх ідентифікації та подальшого аналізу. Вирішенню цієї немаловажної проблеми присвятили свої праці такі науковці: Anderson C. W. [2], Boididou C. та Papadopoulos S. [4], De Maio C. та Fenza G. [6], Gruppi M. та Horne B. D. [12], Hoens T. R. та Polikar R. [15], Huang Q. та Zhou C та ін. [20], Jiang S. та Chen X. [21], John Dougrez-Lewis та Maria Liakata [10], Kai Nakamura, Sharon Levy [39], Kaliyar R. K. та Goswami A. [25, 26], Khanam Z. та Alwasel B. N. [28], Liu C. та Wu X. [33], Mohammadrezaei M. та Shiri M. E. [36], Naseem U. та Razzak I. [40, 41, 42], Nishant Rai та Deepika Kumar [45], Paschalides D. та Kornilakis A. [49], Perez-Rosas V. та Mihalcea R [53], Potthast M. та Kiesel J. [59], Pushp P. K. та Srivastava M. M. [61], Qian F. та Gong C. [63], Raza S. та Ding C. [65, 66, 67], Silva R. M. та Santos R. L. S. [73], Vijjali R. та Potluri P. [76], Wanda P. та Jie H. J. [79], Williams A. та Nangia N. [83], Xian Y. та Akata Z. [85], Yang S. та Shu K. [87], Zellers R. та Holtzman A. [89] та ін.

Проаналізуємо деякі розробки зазначених вище науковців дещо детальніше.

Згідно з визначенням [67], фейкові новини (від англ. *Fake News*) – підробка чи імітація новин (маніпулятивне спотворення фактів, дезінформація), які створено з ігноруванням редакційних норм, правил і процесів, прийнятих у ЗМІ для забезпечення їх достовірності. Зазвичай, вони не витримують жодних, навіть поверхневих перевірок на відповідність реальним подіям, але, незважаючи на це, мають потужний вплив на свідомість великої кількості людей.

Особливість фейкових новин у тому, що їхнє поширення у соціальних мережах має характеристики, які експерти [31, 70, 78] визначають як феноменальні, а саме:

- швидкість їхнього розповсюдження майже в шестеро вища, ніж реальних новин [70];
- ймовірність їхнього репосту на 70 % вища від репосту реальних новин [71];
- умовна "глибина" поширення [31] практично у десятеро більша за реальні новини й сягала довжини ланцюжка у 19 перепостів, водночас як для реальних новин це значення не перевищує 10 репостів [78].

Фейкові новини мають такі базові ознаки [67]:

1) Примітивізм – масовий продукт завжди достатньо спрощено створюють, позаяк без цього він не зможе отримати значного поширення [55].
2) Низький статус читачів – призначені для аудиторії, яка за різних причин не перевірятиме достовірність наданої інформації.
3) Надемоційне подання – розраховане на негативний психологічний стан потенційних читачів, які не піддаватимуть критичному аналізу надану інформацію.
4) Практично не мають продовження, позаяк розраховані виключно на оперативну та миттєву маніпуляцію суспільною думкою.

Отже, фейкові новини – це інформація, яка є хибною або спотвореною, зазвичай вводить в оману її споживачів, однак її подають як висвітлення справжніх подій [70]. Термін "фейкові новини" став популярним під час президентських виборів у США 2016 року. Після цього такі відомі компанії, як Google, Twitter і Facebook вжили конкретних заходів для боротьби з ними [4, 14, 41]. Однак, через експоненційне зростання інформації на новинних онлайн-порталах і на сайтах соціальних мереж, розрізняти справжні та фейкові новини навіть досвідченим експертам з часом стає надзвичайно важко [71]. Тому в сучасних умовах інформатизації суспільства методи виявлення фейкових новин поділяють на ручну перевірку фактів і автоматичну їх ідентифікацію з подальшим інтелектуальним аналізом.

Веб-сайти для перевірки фактів, такі як Reporterslab, Politifact та ін. [90], зазвичай покладають свої сподівання на людське судження, щоб визначити правдивість деяких новин. Навіть такий гігант, як Amazon Mechanical Turk, часто використовує так званий краудсорсинг для виявлення фейкових новин у онлайнових соціальних мережах. Оскільки краудсорсинг (англ. *Crowdsourcing*, crowd – "гурт" і sourcing – "походження, виробництво") – це передача своїх виробничих функцій (деяких процесів і завдань) певній розподіленій робочій силі, яка може виконувати ці завдання віртуально, то їхні методи перевірки фактів використовують тільки основні засоби (так звані ярлики "правда/неправда" чи маркери – справжня чи фейкова) для визначення якості

тої чи іншої інформації, що малоефективно для встановлення її як достовірності, так і релевантності [71]. Тому ручні методи перевірки фактів мають деякі обмеження [88]:

1) виявлення та повідомлення про кожну фейкову новину, створену в мережі Інтернет, займає багато часу;
2) важко масштабувати масив новостворених новин, особливо в соціальних мережах;
3) часто упередження фактчекерів (з англ. *Fact Checking* – перевірка фактів), наприклад, статі, раси чи національності можуть вплинути на основний етикет інтерпретації правди.

Методи автоматичної ідентифікації фактів є альтернативою ручним методам їх виявлення, проте на сьогодні такі методи широко використовують для визначення достовірності інформації. Згідно з цим підходом, характерні ознаки фейкових новин зазвичай визначають з характеристик, пов'язаних з їхнім змістом [12] або із соціального контексту через взаємодію користувачів [38, 54, 72]. Наприклад, у роботі [60] автор досліджував методи, які можуть ідентифікувати онлайн-документи різного змісту, особливо це стосується новинних публікацій на підставі стилю їх подання. Він показав, що сучасні класифікатори тексту загального призначення, незважаючи на їхню хорошу продуктивність, при спрощеному аналізі його змісту насправді давали низькі прогнози у навчальних вибірках. Для того, щоб досягти справді стилістичного прогнозу, автор зібрав вибірку із 103 219 документів зосереджених у 223 онлайн-джерелах попередньо, позначених медіа-експертами відповідними ярликами. Також він розробив реалістичні сценарії оцінювання новин та два нових класифікатори тексту – нейронну мережу (англ. *Neural Network*) та модель на підставі стилометричних ознак. Отримані результати дослідження довели, що запропоновані ним класифікатори тексту зберігають високу точність визначення його змісту під час аналізу документів на раніше невідомі теми (наприклад, нові події) та з раніше невідомих джерел (наприклад, нових веб-сайтів).

Методи аналізу тексту на підставі його контенту [4, 16, 59, 60, 73] використовують різні типи інформації, наприклад, дописи, повідомлення чи статті, їхні заголовки, а також джерела їх надходження, призначені для створення відповідних класифікаторів фейкових новин. Більшість таких методів використовують стиліметрію (наприклад, сегментацію речень, лексемизацію та тегування POS (англ. *Part-Of-Speech Tagging* – розмічування частин мови) і лінгвістичні особливості (наприклад, лексику, набір слів і частоту їх появи, схеми відмінків) публікацій для вловлювання оманливих сигналів або підготовки суспільної думки. Наприклад, Horne і Adah у роботі [18] виділяють стиліметрію та психологічні особливості новин з їхніх заголовків, щоб відрізнити фейкові новини від реальних. Водночас, Przybyla та ін. у роботі [60] розробили класифікатор тексту на підставі його стилю, у якому використали двонаправлену довготривалу короткочасну пам'ять (англ. *Long Short-Term Memory,* LSTM) для аналізу новинних публікацій, що базується на особливостях їх появи. Zellers та ін. у роботі [89] розробили модель нейронної мережі для визначення достовірності новин, використовуючи текст публікації та стиль її написання. Деякі інші роботи [73, 91] в текстах новин розглядають структуру лексем, набори використаних слів, синтаксис, частини мови, TF-IDF метрики, латентні теми для визначення ознак достовірності новин на підставі їхнього контенту [4, 5, 18, 67, 72, 73].

Часто TF-IDF (від англ. TF – *Term Frequency* – частота слова, IDF – *Inverse Document Frequency* – зворотна частота документа) як статистичний показник використовують для оцінювання важливості слів у соціальному контексті документа [41, 44, 67, 70, 72], що є частиною їхньої колекції чи репозиторію [73]. Деякі науковці вважають [91], що вагомість (значущість) слова пропорційна кількості його вживань у документі та обернено пропорційна частоті появи цього слова у інших документах відповідної колекції. Також показник TF-IDF широко використовують в задачах аналізу текстів та інформаційного пошуку в мережі Інтернет. Його часто застосовують як один з критеріїв релевантності документа до пошукового запиту, а також під час розрахунку міри спорідненості документів внаслідок їх кластеризації.

Отже, загальна проблема методів ідентифікації фактів, заснованих на їх змісті, контенті чи соціальному контексті [41, 67, 72], полягає в тому, що стиль, платформа та теми фейкових новин постійно змінюються [16, 60]. Моделі, під час їх машинного навчання на одному наборі даних, можуть погано працювати на новій інформації з іншим змістом, стилем подання чи мовою написання [73]. Окрім цього, цільові аудиторії споживачів фейкових новин з часом оновлюються, деякі ярлики стають застарілими, а інші потребують повторного маркування [60]. Більшість методів, заснованих на змісті тексту, не можливо адаптувати до цих змін, що вимагає повторного вилучення функцій аналізу новин і повторного маркування даних на підставі нових функцій [72]. Ці методи також вимагають великої кількості навчальних даних для виявлення фейкових новин [54]. Адже відомо, що на той час, коли для навчання моделі вже зібрано достатньо даних, автори фейкових новин вже поміняли стиль їх подання чи їх поширено вже занадто далеко [38]. Оскільки лінгвістичні особливості тексту, які використовують в методах на підставі його змісту, переважно специфічні для певної мови, тому їх універсальність також обмежена [16].

Щоб усунути недоліки методів ідентифікації фактів, засновану на підставі його контенту [5, 18, 67], багато науковців значну частину своїх досліджень [35, 72, 79] почали зосереджувати на соціальних контекстах для виявлення фейкових новин [44, 67, 70]. Методи їх аналізу досліджують соціальну взаємодію користувачів [54] і виділяють релевантні функції, що відображають їхні публікації (огляд/допис, коментарі, відповіді) і деякі особливості роботи соціальних мереж – відносини читачів до них і навпаки. Наприклад, Liu і Wu у роботі [35] пропонують застосовувати класифікатор тексту на підставі нейронних мереж, який використовує твіти в соціальних мережах, послідовності ретвітів і профілі користувачів Twitter для визначення достовірності новин. Інші дослідження [40, 41, 54] вивчають почуття користувачів щодо фейкових новин у соціальних мережах і виявляють зв'язок між їхнім настроєм і виявленням фейків. У цих роботах також розглядають облікові записи та поведінку підставних користувачів, щоб побачити, чи можуть вони виявити опосередковані позиції.

Деякі підходи до автоматизованого процесу ідентифікації новин, засновані на соціальних контекстах [41, 67, 72], використовують методи ідентифікації фактів на

підставі позицій читачів і на поширенні новин [35, 72, 87]. Методи, засновані на позиціях читачів, використовують точки зору користувачів на публікації у соціальних мережах, щоб визначити їх істину. Тут користувачі висловлюють свою позицію щодо наданої інформації явно або опосередковано [38]. Відверті позиції читачів є прямим вираженням їхніх думок, які, зазвичай, доступні в коментарях до публікацій у соціальних мережах. Деякі дослідження [35, 72] здебільшого використовували голоси за/проти, великий палець вгору/вниз для виділення явних позицій [6, 35]. З іншого боку, опосередковані позиції користувачів щодо наданої інформації, як правило, засновані на мовних особливостях публікацій у соціальних мережах [61]. Деякі дослідження [26, 60] використовують тематичне моделювання для вивчення прихованих позицій читачів. Інші дослідження [79, 87] розглядають облікові записи та поведінку підставних користувачів, щоб з'ясувати, чи можуть вони виявляти свою опосередковану позицію. Щоб дізнатися приховані позиції читачів з тем публікацій, деякі дослідники [30] використовують так зване тематичне моделювання (англ. *Thematic Modeling*) – спосіб побудови моделі колекції текстових документів, яке визначає, до яких тем належить кожен з документів.

Перехід з простору термінів у простір знайдених тем чи тематик публікацій допомагає лінгвістам вирішувати синонімію та полісемію термінів, а також дещо ефективніше вирішувати такі завдання як тематичний пошук, класифікацію, сумаризацію та анотацію колекцій документів і новинних потоків інформації. Для цього використовують тематичні моделі (англ. *Topic Model*), які з наданої колекції текстових документів визначають, до яких тем належить кожен із них, і які слова (терміни) утворюють кожну тему [77].

Тематичне моделювання як вид статистичних моделей для знаходження прихованих тем документів, що трапляються в їхній колекції, знайшло своє застосування в таких областях, як машинне навчання та оброблення природної мови [30, 77]. Дослідники використовують різні тематичні моделі для аналізу документів, що знаходяться в архівах, для аналізу зміни тем у їхніх наборах. Інтуїтивно розуміючи, що документ належить до певної теми, в документах, присвячених одній темі, можна натрапити на деякі слова частіше за інші. Наприклад, слова "транспортний засіб" і "постраждалі" трапляються частіше в документах про надзвичайні події, водночас як "командування" і "відділення" – в документах про військові події. Зазвичай, документ стосується кількох тем в різних пропорціях. Наприклад, для документу, в якому 10 % теми становить командування, а 90 % теми – транспортний засіб, можна припустити, що слів про цей засіб в 9 разів більше, однак, це зовсім не означає, що зміст документа стосується військових дій. Тематичне моделювання відображає таку інтуїцію в математичній структурі моделі аналізу тексту, яка дає можливість дослідникам на підставі вивчення колекції документів і дослідження частотних характеристик слів у кожному документі зробити висновок, що кожен документ – це деякий баланс тем [77].

Методи ідентифікації фактів на підставі їхнього поширення [20, 21, 34, 63] використовують інформацію, пов'язану з фейками, наприклад, як користувачі їх розповсюджують між собою у мережі. Загалом, вхідними даними для цих методів може бути або їхній каскад точок (пряме подання поширення новин), або самовизначений граф (опосередковане подання, що фіксує інформацію про їх розповсюдження) [90]. Не дивлячись на те, що ці методи використовують графи та багатовимірний каскад точок для виявлення фейкових новин [20, 34], проте вони все ще перебувають на початковій стадії свого розвитку.

Раніше, для виявлення фейкових новин багато дослідників [47] пробували використовувати нейролінгвістичне програмування НЛП (англ. *Neuro-linguistic programming,* NLP), так званий науковий напрям у психотерапії та практичній психології, що вивчає закономірності суб'єктивного досвіду людей через розкриття механізмів і способів моделювання їх поведінки і передачі виявлених моделей іншим людям. Хоча його ефективність з часом була спростована [62], проте використовуючи класичні методики NLP та Deep Learning (глибинне навчання) нейронних мереж, деякі науковці [5, 24, 33] змогли застосувати неконтрольовані методи для групування новин за темами, виконувати пошук недостовірної інформації та здійснювати її класифікацію за певними ознаками.

Окрім аналізу текстових документів, часто для виявлення фейкових новин використовують візуальну інформацію як доповнення для визначення їх правдивості. Значна кількість робіт [4, 5, 22, 23, 62] присвячена дослідженню зв'язків між зображеннями та достовірністю твітів/лайків під ними. Однак, у роботі [5] візуальна інформація була створена вручну, що значно обмежило можливості відповідних моделей видобувати складний контент з наявних даних [4, 73]. Збираючи автоматично візуальну інформацію з різних публікацій, Jin et al. у роботі [22] запропонував підхід, де використав глибинне навчання нейронної мережі. У своєму дослідженні автор спробував поєднати візуальні функції високої якості зображень з текстовим і соціальним контекстом дописів під ними [41, 44, 67, 70, 72].

З розвитком мультимедійних технологій багато зловмисників [5, 23, 62] намагаються використовувати зміст зображень чи відео, щоб ввести в оману читачів шляхом швидкого їх розповсюдження, що робить візуальний контент важливою частиною фейкових новин [4, 5, 18, 73]. Часто зображення, додані до новинних публікацій, є не тільки фейковими, які майстерно підроблено, але й реальними, які успішно використовують для подання нерелевантних подій. Отже, цілеспрямоване використання деяких характеристик зображень у фейкових новинах є важливою та складною проблемою для їх виявлення. У реальному світі зображення фейкових новин можуть істотно відрізнятися від зображень реальних подій як на фізичному, так і на семантичному рівнях, що можна чітко відобразити в частотній та піксельній області відповідно. В роботі [62] автори пропонують нову структуру багатодоменної візуальної нейронної мережі MVNN (англ. *Multi-domain Visual Neural Network*), в якій об'єднано візуальну інформацію з частотних і піксельних доменів для виявлення фейкових новин. Зокрема, вони розробили нейронну мережу на базі CNN для автоматичного захоплення складних шаблонів зображень потенційно фейкових новин у частотній області. Також використали багатогалузеву модель CNN-RNN для вилучення візуальних характеристик з різних семантичних рівнів у піксельній області. Механізм уваги читачів вони використовували для динаміч-

ного злиття ознак частотних і піксельних доменів. Масштабні експерименти, проведені на реальному наборі зображень, демонструють, що MVNN перевершує наявні методи з точністю щонайменше на 9,2% і допомогає підвищити продуктивність мультимедійного виявлення фейкових новин понад 5,2%.

Нещодавно для автоматичної ідентифікації фейкових новин було застосовано так зване трансферне навчання (англ. *Transfer Learning*) – дослідницька задача в машинному навчанні, зосереджена на зберіганні знань, отриманих під час розв'язання однієї задачі, та застосуванні цих знань до іншої, але пов'язаної задачі [24, 33]. Хоча таке навчання показало обнадійливі результати під час оброблення зображень і в завданнях НЛП, його застосування для виявлення фейкових новин все ще недостатньо досліджене. Проблема в тому, що навіть ідентифікація новинної інформації є делікатним завданням, під час якого трансферне навчання моделі має мати справу із семантикою, прихованими значеннями змісту тексту, його контентом і соціальними контекстами дописів [41, 44, 67, 70, 72], які супроводжують зображення і, в основному, є недостовірними.

Підсумовуючи зазначене вище, вважаємо, що зміст новин і їхній контент, а також соціальний контекст медіа-ресурсів, профілі зловмисних користувачів і їхні дії можна використовувати для ідентифікації новинної інформації. Однак ці підходи створюють значні проблеми під час їх безпосередньої реалізації. Наприклад, збір соціальних контекстів з дописами користувачів під зображенням є обширною тематикою дослідження, проте ці дані не тільки великі, але й неповні, часто зашумлені та неструктуровані, що робить відповідні методи виявлення фейкових новин, зазвичай, неефективними.

Результати дослідження та їх обговорення / Research results and their discussion

Значна частка потенційних читачів новин прагнуть взнавати їх зміст і контент онлайн з мережі Інтернет, де вони мають доступ до різних новинних ресурсів [66, 67]. Щоб допомогти цим читачам знайти потрібну й релевантну новинну інформацію широко використовують сучасні системи рекомендацій новин NRS (англ. *News Recommendation System*). Розроблені на підставі неї програмні системи призначені усувати наявність підозрілої чи навіть хибної інформації та пропонувати споживачам тільки достовірні новини, які можуть їх зацікавити. Однак, розробникам таких систем доводиться вирішувати проблему появи так званих фейкових новин, з якими NRS стикається під час ідентифікації наявної інформації, що вимагає запровадження різних дієвих для її аналізу рішень шляхом використання найсучаснішого програмно-апаратного забезпечення.

1. Екосистемне мислення та екосистема новин. На сьогодні більше не існує технологій, які б розробляли ізольовано для вирішення якогось одного завдання. Зазвичай, кожен продукт і технологічне рішення, що знаходиться за ним, вбудовують в складну систему взаємодій з іншими продуктами, учасниками ринку, регуляторними нормами, логістикою, платформами і т.д. Хоча кожному з нас відоме слово екосистема з різних соціальних контекстів [67, 72], але далеко не всі розуміють, що за цим терміном знаходиться деякий світогляд, цілеспрямоване мислення та відповідні дії людей [69].

Багато фахівців вважають [2, 52, 69], що екосистема – новий формат мислення. Однак, термін "екосистемне мислення" – не нова ментальна модель. Наприклад, для керівників бізнесу ця модель стала особливо популярною завдяки розвитку інформаційних технологій та їх повсякденне використання – будь-то на виробництві, транспорті чи в побуті. Екосистема як бізнес-модель їхньої діяльності передбачає комплекс взаємопов'язаних товарів чи наданих послуг, які можуть вирішити різноманітні завдання потенційних клієнтів на єдиному інтегрованому полі. Великі IT- та фінансові компанії, такі як Microsoft, Apple, Google чи Amazon, діяльність яких влаштована саме за таким принципом, широко використовують екосистемне мислення.

Деякі фахівці [52, 69] поділяють екосистемні та платформні компанії, але останні вже давно вийшли за межі одного виду діяльності. Наприклад, Facebook за останнє десятиліття виросла із соціальної мережі настільки, що вже потребує нового бренду, бо старий не відображає зміст її складної структури. Адже в екосистему соціальної мережі Facebook входять багато інших соцмереж і месенджерів, це також ігрова та e-commerce платформа, платіжна система та багато іншого. Наприклад, електронна комерція (від англ. *Electronic Commerce*, скорочено e-commerce), будучи сферою цифрової економіки, містить всі фінансові та торгові транзакції, які проводять за допомогою комп'ютерних мереж, та бізнес-процеси, пов'язані з проведенням цих транзакцій. Основними складовими e-commerce є мобільна комерція, електронний переказ коштів, управління ланцюгами поставок, Інтернет-маркетинг, оброблення онлайн-транзакцій, електронний обмін даними (англ, *Electronic Data Interchange*, EDI), системи управління запасами та автоматизовані системи збору даних. Тому екосистема бізнес-діяльності, як і все інше, не буває постійною, вона з часом розвивається за рахунок збільшення взаємодії між її учасниками, залучаючи все нових партнерів своїм "гравітаційним полем".

Екосистема новин містить три основні сутності [2, 66, 72]: видавці (новинні ЗМІ чи редакційні компанії, які публікують новини), інформація (зміст і контент новин) та користувачі (соціальний контекст їхніх профілів і дій). Як показано на рис. 1, спочатку новини надходять від видавців. Потім вони потрапляють на різні веб-сайти чи онлайн-платформи новин. Користувачі отримують новини з цих джерел, діляться ними на різних платформах (блоги, соціальні мережі). Зв'язки між акаунтами користувачів, посилання один на одного, хештеги та боти становлять соціальний контекст медіа-ресурсів [4, 5, 18, 67, 72, 73].

Також на рис. 1 показано схему типової процедури аналізу новин [43]. Спочатку новинна інформація надходять із екосистеми новин [2], яку часто називають набором даних. Зміст новин, їх контент і соціальний контекст твітів/лайків, у т.ч. й медіа-ресурсів, профілі користувачів і їхні дії надходять до відповідних модулів, де дані попередньо обробляють [41, 44, 67]. Вхідними даними для рівня ідентифікації та аналізу є такі компоненти, як зміст новин, їх контент і соціальні контексти. Результатом роботи цих модулів є векторне подання окремих компонент, які потім об'єднують для отримання єдиного узагальнення, яке передають як вхідні дані до блоку Transformer. В цьому блоці відбувається їх класифікація, а потім так звана крос-ентропія, де вони отримують мітку (фальшиву чи справжню) для кожної новини як остаточний результат.

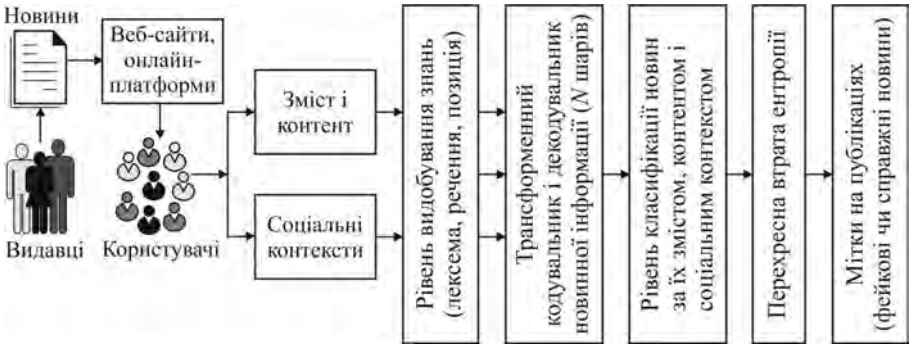

Рис. 1. Схема типової структури розповсюдження новин та їхнього аналізу (розроблено на підставі [43]) / Scheme of a typical structure of news distribution and its analysis (developed on the basis of [43])

Модуль, що аналізує зміст новин (публікацій), видобуває їхній контент з екосистеми новин. Основна частина новини (її зміст) і відповідна побічна інформація представляють новинну інформацію. Основний текст новини – це публікація, яка описує, наприклад, подію, що відбулася. Як правило, спосіб написання новини відображає основний перебіг події та точку зору автора на неї. Стосовно новин, то в цьому модулі розглядають таку додаткову інформацію [67]:

- джерело новин, наприклад, такі компанії як CNN чи BBC, вільні журналісти, блогери тощо;
- заголовок, який описує основну тему публікації, зазвичай створений так, щоб привернути увагу читачів;
- автор новини, буває як пересічний дописувач чи блогер-початківець, так і професійний журналіст чи спеціально навчений фахівець з різних служб – державних чи приватних;
- час публікації новини – показник їхньої "свіжості" або запізнення;
- довідкова інформація, зазвичай, стосується прихильності джерела появи новин до певної категорії її споживачів. Наприклад, джерело новин із різними публікаціями на користь правих відображає партійну упередженість як джерела та авторів, так і їхніх читачів.

До модуля, що аналізує соціальний контекст медіа-ресурсів, надходять публікації, лайки, поширення, відповіді, фоловери та їхні дії [41, 44, 67, 70, 72]. Коли функцій, пов'язаних з аналізом змісту новин, недостатньо або недоступно, соціальний контекст може надати уважному читачу корисну інформацію про достовірність новини. Йдеться про те, що кожен такий контекст подають постом (коментар, огляд, відповідь) і відповідною побічною інформацією (метаданими). Отже, публікація – це об'єкт соціальної мережі, розміщений відповідним джерелом новин. Вона містить для читачів корисну інформацію, щоб зрозуміти їхній погляд на новину. Можна вносити таку додаткову інформацію, пов'язану з соціальним контекстом медіа-ресурсів, профілів користувачів і їхніх дій [67]:

- користувач – людина або бот, яка(ий) зареєстрований в соціальних мережах;
- заголовок чи короткий допис, що зазвичай відповідає змісту новини;
- оцінка, часто числова, надана читачами публікації, які її схвалюють, опротестовують чи ганьблять;
- джерело новин – різні новинні компанії, вільні журналісти, блогери, дописувачі тощо;
- кількість коментарів до публікації – функція, яка забезпечує рівень її популярності;
- співвідношення "за" та "проти" – оцінка схвалення/відхилення публікації безпосередніми читачми;
- загальна оцінка аудиторії – сукупна відповідь всіх читачів на кожну публікацію. Припускають, що новина або її заголовок з оцінкою менше одиниці не є надійною, і навпаки;
- довіра користувачів до соціальних мереж (довіра соціальних мереж до користувачів) – додаткова функція соціального контексту. Ця функція допомагає аналітикам визначити, чи схильний читач поширювати дані новини, чи ні. Наприклад, деякий допис під новиною читача, який не заслуговує довіри, вказує на те, що новина є справжньою чи фальшивою [67].

Встановлення довіри користувачів до соціальних мереж і навпаки не є новою в літературі [1, 42]. У деяких роботах застосовувався підхід, згідно з яким здійснювалось виявлення певної спільноти [48], наприклад, з расистськими поглядами, проводився аналіз настроїв її учасників і застосовувалися методи ранжування їхніх профілів за інтенсивністю висловлювань чи проявом думок. У такий спосіб встановлювали лідерів спільноти, визначали їхню схильність закликати до насильства. Однак, для виявлення фейкових новин, які враховують довіру користувачів до соціального контексту медіа-ресурсів, використовують просту кластеризацію зображень чи відеофайлів [72]. Тут часто застосовують так зване нульове навчання ZSL, щоб визначити довіру користувачів до соціальних мереж.

Наприклад, в роботі [1] автори досліджували настрій читачів (також відомий як аналіз їхніх думок), яке згодом стало основним напрямом їхньої роботи, що мало на меті зрозуміти причину збільшення змісту новинної інформації, створеного дописувачами, для використання соціальних мереж. Наприклад, в онлайн-соціальних мережах (англ. *Online Social Networks*, OSNs) аналіз настроїв читачів використовують в декількох особливих дослідження [58]. У контексті довіри читачів до соціальних медіа-ресурсів були розроблені рамки для аналізу надійності контенту користувачів [18, 72], беручи до уваги їхні загальні почуття до написаного. Однак, щодо дослідження надійності цієї довіри, то їхні намагання провести аналіз настроїв у відповідях на дописи не увінчались значним успіхом. Отже, семантичний аналіз тексту варто об'єднати, щоб покращити результуюче відчуття від проведеної роботи.

Зазвичай, ZSL (англ. *Zero-Shot Learning*) – це механізм, за допомогою якого комп'ютерну програму вчать розпізнавати об'єкти на зображенні або отримувати інформацію з супровідного тексту без позначення навчальних даних [85]. Наприклад, поширеним підходом до кластеризації зображень є навчання моделі з нуля на даних, що стосуються конкретного завдання. Механізм ZSL дає можливість виконати це завдання без будь-якого попереднього спеціального навчання. При цьому модель ZSL може виявляти невідомі кластери, на які вона ніколи не потрапляла під час навчання, на підставі попередніх знань з домену джерела зображень [84] або допоміжної інформації [61].

Щоб визначити рівень довіри соціальних мереж до користувачів, спочатку групують дії кожного з них (ко-

ментарі, публікації, відповіді), а потім вводять цю інформацію в класифікатор ZSL, який часто створюють на підставі архітектури Transformer [7, 24, 45, 67, 76]. Тобто, прикріплюють заздалегідь підготовлену контрольну точку (вагомість коефіцієнтів моделі на стадії навчання) для величезного набору даних, внаслідок чого отримують багатожанровий висновок природної мови MNLI (англ. *Multi-genre Natural Language Inference*) [83] з цим класифікатором.

За допомогою моделі ZSL попередньо навчену контрольну точку можна налаштувати для виконання спеціального завдання, наприклад? перевірки довіри користувачів до результатів певного дослідження, виконаного деякою компанією. Тут користувачів, зазвичай, класифікують у завчасно визначені категорії, які вказують на різні рівні довіри до них. Загалом визначають п'ять рівнів довіри (категорій): новий користувач, дуже ненадійний, ненадійний, надійний, дуже надійний. Також використовують попередні знання про точно налаштовану модель ZSL та її контрольні точки, а також семантику допоміжної інформації для уточнення характеристик відомих категорій користувачів. Запропонована у роботі [85] модель ZSL може виокремлювати нові кластери користувачів, які згодом переводять в класи (категорії). Пізніше цю інформацію вносять як слабкі мітки в модель виявлення фейкових новин.

Збір даних з багатьох новинних джерел часто використовують для аналізу інформації від великих спільнот чи думок їхніх учасників, які розміщують їх у соціальних мережах. Для цього застосовують методи і засоби MNLI, позаяк це великомасштабний проект, який охоплює низку жанрів усного та письмового тексту. Довіра соціальних мереж до користувачів і краудсорсинг було детально розглянуто в різних дослідженнях [3, 46]. Згідно з їхніми даними, велика кількість краудсорсингових даних дає можливість моделі MNLI виявляти зв'язки між довірою користувачів і тим, як вони висловлюють свою думку про соціальну мережу. Зрозуміло, надзвичайно важко збирати думки так званого натовпу людей (англ. *Crowd of People*), а також прямі відгуки безпосередніх користувачів. Тому тут отримують переваги попередньо навченої моделі MNLI з точки зору розміру та тривалості їх навчання, а також переваги точності аналізу новинного тексту.

Отже, провівши аналіз таких термінів, як екосистемне мислення та екосистему новин, можна зробити такі проміжні висновки.

З'ясовано, що за терміном екосистемне мислення знаходиться деякий світогляд, цілеспрямоване мислення та відповідні дії людей, залучені в цій системі. Зазначений термін стосується й екосистеми новин як бізнес-модель їхньої появи, ознайомлення та поширення, що передбачає комплекс взаємопов'язаних сутностей – виробників (новинних ЗМІ чи редакційних компаній), новинної інформації (зміст і контент новин) та її користувачів (соціальний контекст їхніх профілів і дій), які сукупно можуть вирішити різноманітні завдання потенційних учасників на єдиному інтегрованому полі.

Наведено схему типової процедури аналізу новин, де спочатку новинна інформація надходить із відповідної екосистеми, яку часто називають джерелом даних. Зміст новин, їх контент і соціальний контекст твітів/лайків, у т.ч. й медіа-ресурсів, профілі користувачів і їхні дії як окремі компоненти надходять до відповідних модулів, де їх попередньо обробляють. Результатом роботи цих модулів є векторне подання окремих компонент, які потім об'єднують для отримання єдиного узагальнення, яке передають як вхідні дані до блоку Transformer. В цьому блоці відбувається їх класифікація, а потім так звана крос-ентропія, де вони отримують мітку (фальшиву чи справжню) для кожної новини як остаточний результат.

Для встановлення довіри користувачів до соціальних мереж і навпаки часто застосовують підхід, згідно з яким виявляють певну спільноту, що мають деякі негативні поглядами. Для цього аналізують настрої її учасників і застосовують методи ранжування їхніх профілів за інтенсивністю висловлювань чи проявом думок. У такий спосіб встановлють лідерів спільноти, визначають їхню схильність закликати до насильства. Однак, для виявлення фейкових новин, які враховують довіру користувачів до соціального контексту медіа-ресурсів, використовують просту кластеризацію зображень чи відеофайлів. Тут часто застосовують так зване нульове навчання ZSL, щоб визначити довіру користувачів до соціальних мереж.

Визначено, що ZSL – механізм, за допомогою якого відповідну модель вчать розпізнавати об'єкти на зображенні або отримувати інформацію з супровідного тексту без позначення навчальних даних. Поширеним підходом до кластеризації зображень є навчання моделі з нуля на даних, що стосуються конкретного завдання. При цьому модель ZSL може виявляти невідомі кластери, які не траплялися під час навчання, використавши для цього попередні знання з домену джерел зображень або допоміжної інформації.

2. Програмні системи для виявлення фейкових новин у мережі Інтернет. Проаналізувавши останні дослідження та публікації щодо методів ідентифікації новинної інформації, можна однозначно стверджувати, що існує певний набір методів і засобів, які найкраще підходять для вирішення проблеми виявлення фейкових новин у мережі Інтернет. Встановлено, що більшість сучасних дослідників намагаються виробити свої підходи із застосуванням комбінацій унікальних і вже апробованих методик, щоб успішно вирішити зазначену проблему. Тому далі детально проаналізуємо найбільш відомі та ефективні програмні системи, призначені виявляти фейкові новини, наведемо як їх позитивні відмінності, так і деякі недоліки, зазвичай, присутні навіть в найдосконаліших системах.

Система на підставі архітектури Transformer. Shaina Raza і Chen Ding [65, 66, 67] розробили інноваційний фреймворк, у якому використали найсучасніші можливості його машинного навчання. Розроблена авторами ПС на підставі фреймворку використовує зміст новин, їх контент і соціальні контексти для аналізу їхніх корисних характеристик [41, 44], а також для прогнозування ймовірності появи фейкових новин. Розроблена модель, маючи в своїй основі архітектуру Transformer [7, 24, 45, 67, 76], легко піддається машинному навчанню за наборами позначених новин, що допомагає швидко виявляти фейки в новинній інформації. У роботі [67] також використано додаткову інформацію (метадані) із змісту новин, їх контенту і соціальих контекстів під час використання моделі, що дало можливість краще класифікувати новинну інформацію. У роботах [7, 24, 45] було застосовано системний підхід до дослі-

дження зв'язку між профілем автора новин та їхньою достовірністю. Запропоновано нову модель, засновану на архітектурі Transformer [76], з використанням так званого нульового навчання для визначення рівня довіри соціальних мереж до її користувачів. Перевага цього підходу в тому, що він може визначити довіру як до постійних, так і нових користувачів, а також може виявити зловмисників, які часто змінюють свою тактику, щоб привернути увагу до себе потенційних прихильників, або вразливих користувачів, які поширюють дезінформацію. Застосовано модель слабкого нагляду (англ. *Weak Supervision*) для маркування публікацій, яка значно пришвидшує роботу над її машинним навчанням, використовуючи для цього великі та складні набори даних. Завдяки такому підходу позначені новини можна миттєво отримувати з відомих джерел їхнього поширення та оновлювати моделі в режимі реального часу [67].

Система Check-It. Demetris Paschalides та ін. [49] запропонували плагін Check-It для браузерів, що допомагає розпізнавати правдиві новини (рис. 2). Розроблена ними ПС на підставі цього плагіну містить модульне ПЗ, яке підтримує процедуру ідентифікації фейкових новин на підставі різноманітних сигналів. Також ця система використовує двостадійний процес вибору функцій, які здійснюють L2-регулярізацію для зменшення ступеня перенавчання моделі (англ. *Model Overfitting*), та генетичний алгоритм (англ. *Genetic Algorithm*, GA), щоб визначити оптимальну кількість цих функцій, які можуть успішно навчати логістичну регресію (англ. *Logistic Regression*, LR) з низьким ресурсом і малою пам'яттю. Проведені масштабні експерименти показують, що запропонований метод вибору функцій перевершує найсучасніші його альтернативи. Ретельний аналіз та порівняння з іншими найсучаснішими роботами автори провели на реальних даних, результати якого показали, що система Check-it забезпечує точність прогнозу ідентифікації фейкових новин понад 90 %.

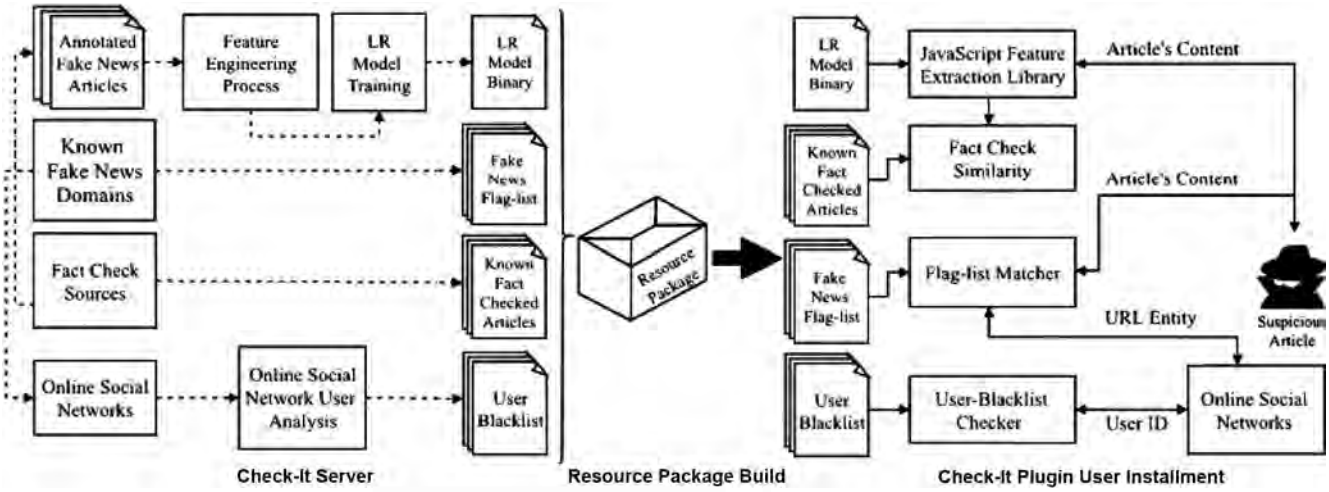

Рис. 2. Архітектура системи Check-It / Architecture of the Check-It system [49]

Систему Check-It можна встановлювати у декількох браузерах, у т.ч. й з Google Chrome, Mozilla Firefox тощо. Загалом архітектура системи містить чотири основні модулі:

a) модуль Flag-list Matcher зіставляє джерела появи новинних публікацій із відомими фейковими доменами новин і джерелами перевірки фактів;
b) модуль Fact Check Similarity порівнює новинну публікацію із перевіреними відомими опублікованими фактами, позначеними як фейкові, від організацій, що їх перевіряють;
c) модуль Online Social Network User Analysis відповідає за аналіз поведінки користувачів у соціальних мережах і створює "чорний список" розповсюджувачів фейкових новин;
d) модель LR (англ. *Logistic Regression*), будучи класифікатором новинної інформації, навчають на лінгвістичних ознаках, які видобувають з наборів даних фейкових новин за допомогою двостадійного процесу вибору функцій.

Щоб оцінити достовірність новинної публікації, система Check-it виконує певну послідовність кроків, кожен з яких використовує інший сигнал для її оцінювання. Розглянемо ці кроки дещо детальніше.

Крок 1. Сигнал переліку позначок домену (англ. *Domain Flag-list Signal*). Переліки позначок стосуються добре відомих доменів для поширення дезінформації (наприклад, Kaggle, OpenSources і Greek-Haxes), де їх коментують та підтримують журналісти-експерти, редактори, політологи та соціологи. Перелік позначок є одним із найпростіших способів початкового й, водночас, швидкого оцінювання достовірності новинної публікації. Хоча на цому кроці не перевіряють правдивість самих публікацій, що надходять з різних веб-сайтів, однак тут ідентифікують їх авторів, які постійно беруть участь у кампаніях з дезінформації чи поширенні пропаганди. З цією метою система Check-It підтримує колекцію відповідних доменів відомих авторів фейкових новин.

Крок 2. Сигнал перевірки подібності (англ. *Fact Check Similarity Signal*). Ряд приватних ініціатив і організацій, як-от Politifact, Snopes і MediaBiasCheck, спрямовані на боротьбу з пропагандою та містифікаціями, що поширюється мережею Інтернет. На цих сайтах, зазвичай, наймають професійних журналістів, політичних експертів або навіть відомих діячів з їх політичного спектру, щоб проводити дослідження та коментувати правдивість публікацій. Після встановлення достовірності публікації ці веб-сайти оприлюднюють свої висновки та пов'язану з ними інформацію (URL тощо). Система Check-It використовує веб-сайти для перевірки джерел появи/перевірки фактів, перехресно перевіряючи кожну публікацію, оброблену її модулями, а також відомі перевірені факти в публікаціях і генерує інформативне попередження, коли публікація потрапляє в перелік фейкових новин на цих веб-сайтах.

Крок 3. Онлайн-сигнал соціальної мережі OSN (англ. *Online Social Network Signal*). Хоча зловмисники створюють хибний зміст новин з наміром завдати шкоди їх читачам, соціальні мережі (OSN) в мережі Інтернет забезпечують засоби для його поширення. Недавні дослідження [10, 20, 40, 42] показали, що платформи OSN (наприклад, Twitter) стали механізмом для організації масових кампаній дезінформації учасників мережі. Оскільки OSN відіграють важливу роль у розповсюдженні фейкових новин, він є ще одним сигналом у наборі інструментів плагіну Check-It. Ідея використання цього сигналу полягає у застосуванні онлайн-аналізу користувачів соціальної мережі та створенні їхнього динамічного "чорного списку", який зіставляє ідентифікатори користувачів із показником хибності, що вказує на ймовірність публікації як фейкові новини. Використовуючи такий перелік, система Check-It може попереджати порядних користувачів про публікації підозрілих дописувачів. Зазвичай, плагін Check-It підтримує тільки Twitter через його величезну популярність і легкість доступу до потоку даних за допомогою Twitter Streaming API. Зокрема, розроблена авторами система використовує твіти з потоку Twitter, ідентифікує URL-адреси з доменів відомих фейкових новин і застосовує ймовірнісну модель користувача на підставі DeGroot для розрахунку оцінки хибності його дій, створюючи на виході їх "чорний список" [4, 14, 41, 54].

Крок 4. Сигнал текстового аналізу (англ. *Textual Analysis Signal*) зосереджує свої дії на мета-інформації, отриманій з опублікованих новин, оброблених системою Check-It, або пов'язаних з ними. Цей сигнал покладається на фактичний зміст публікації (заголовок і основний текст), використовуючи методи оброблення природної мови (NLP) для отримання лінгвістичних особливостей, які зазвичай використовують у фейкових новинах. Отримані результати використовують для машинного навчання моделі логістичної регресії (LR) на наборі анотованих новинних публікацій, щоб передбачити їх правдивість. У системі Check-It з введених публікацій видобувають фейкові новини за допомогою реалізованої бібліотеки з двостадійним процесом вибору функцій JavaScript.

Багаторівнева мультимодальна мережа перехресної уваги (англ. *Multi-modal Cross-attention Network,* MMCN). Ефективне виявлення фейкових новин останнім часом привернуло значну увагу багатьох дослідників. Хоча їхні роботи [88, 90, 91] зробили значний внесок у прогнозування ймовірності появи фейкових публікацій, однак при цьому вони приділяють менше уваги використанню зв'язку (подібності) між текстовою та візуальною інформацією в новинних публікаціях. Надання важливості такій подібності значно допомагає ідентифікувати фейкові новини, які, наприклад, намагаються використовувати нерелевантні зображення, щоб привернути увагу читачів.

У роботі [91] автори запропонували метод виявлення фейкових новин з урахуванням схожості SAFE (англ. *Similarity-Aware Fake News Detection Method*), який досліджує мультимодальну (текстову та візуальну) новинну інформацію. Вони вважають, що нейронні мережі потрібно застосовувати для окремого виділення текстових і візуальних функцій, що сукупно дасть змогу ефективно виявляти фейкові новини. Тому потрібно досліджувати зв'язок між вилученими функціями в різних модальностях. Такі подання текстової та візуальної інформації разом із їхніми зв'язками потрібно вивчати спільно та використовувати для прогнозування ймовірності появи фейкових новин. Запропонований авторами метод дає змогу розпізнавати хибність новинних публікацій за їх текстом, зображеннями або їх невідповідностями. Також вони провели масштабні експерименти на великих реальних даних, які демонструють ефективність запропонованого методу.

L. Ying та ін. [88] розробили багаторівневу мультимодальну мережу перехресної уваги MMCN (англ. *Multi-modal Cross-attention Network*), де намагалися проаналізувати якомога більше ознак, які можуть впливати на визначення правдивості публікації (рис. 3). Ця мережа спільно моделює багатомодальну інформацію та багаторівневу семантику дописів у єдину наскрізну структуру для виявлення фейкових новин. Вона розроблена для аналізу включення мультимодальної інформації, наявної практично в кожній публікації, яка може використовувати зв'язки між словами речення та областями зображення для доповнення та просування одне до одного, а також для високоякісного мультимодального подання. Окрім цього, багаторівневу семантику текстової інформації, інтегровану з візуальним вмістом, автори використали для генерування багаторівневих семантичних характеристик за допомогою багаторівневої мережі кодування, об'єднаних для формування всебічного подання новинної інформації.

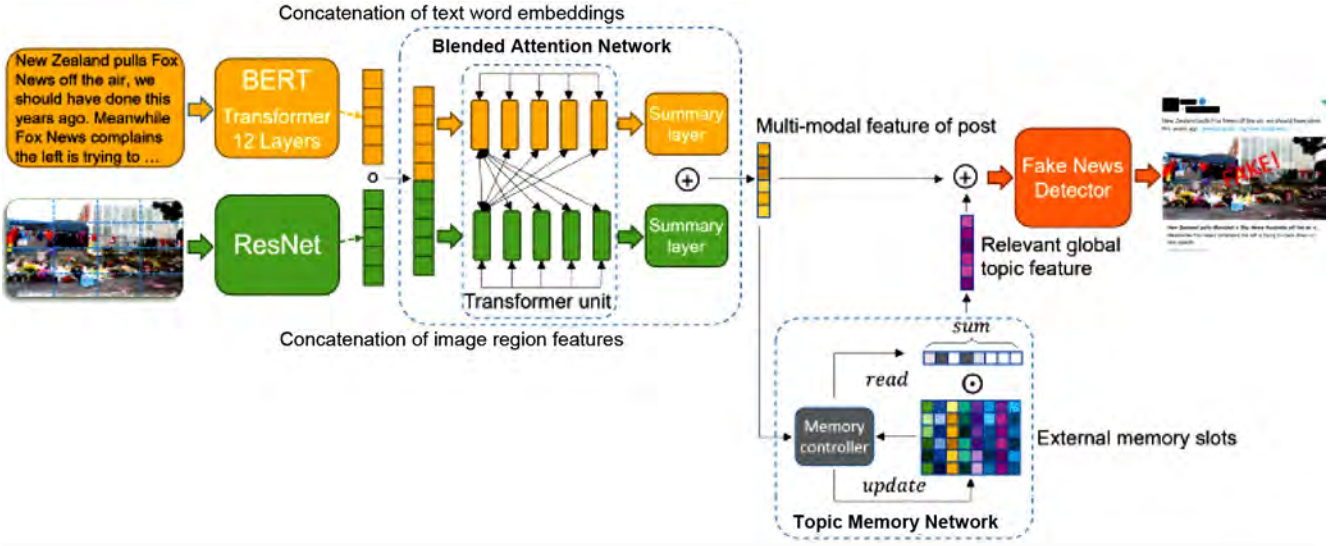

Рис. 3. Загальна структура Multi-modal Cross-attention Network / The general structure of the Multi-modal Cross-attention Network [88]

Розроблена архітектура програмної системи містить такі модулі:

- Мережа кодування тексту та зображень (англ. *Text and Image Encoding Network*). Маючи допис із мультимодальним вмістом, який містить текст і зображення, використовують лексеми фрагментів тексту як фрагменти в текстовій модальності. Попередньо навчену модель BERT використовують для отримання вбудованих токенів фрагментів слів. Водночас, для кожного зображення в публікації використовують попередньо навчену модель ResNet50, щоб вибрати функції, що відповідають розділеним областям. При цьому на етапі машинного навчання попередньо навчені моделі фіксуються.
- Мультимодальне представлення функцій (англ. *Multi-modal Feature Representation*). На підставі виділених фрагментів представлень для текстових слів і областей зображення розроблено мультимодальну мережу перехресної уваги, щоб спільно моделювати взаємозв'язки між модальністю та інтрамодальністю новинної інформації. Беручи до уваги ці зв'язки, можна посилити та доповнити особливості різних модальностей виділених фрагментів. Модуль text CNN і операцію об'єднання потім використовують для агрегування отриманих подань фрагментів у багатомодальні функції, що нагадує модель набору візуальних слів.
- Тематична мережа пам'яті (англ. *Topic Memory Network*). Щоб використати багаторівневу семантику в наведених публікаціях, розроблено нову багаторівневу мережу кодування для моделювання та спільного вивчення великої кількості багаторівневої семантики мультимодального вмісту, яка об'єднує текстову інформацію з візуальним її змістом. Зафіксовані багаторівневі семантичні характеристики поєднують разом, щоб сформувати повне подання виділених фрагментів.
- Мережа виявлення фейкових новин (англ. *Fake News Detection Network*). Основна мета – класифікувати кожну публікацію в соціальних мережах як фейкову або правдиву. Класифікатор приймає вивчені мультимодальні функції як вхідні дані, а потім передає їх у повністю під'єднану мережу з відповідною функцією активації для класифікації публікацій у попередньо визначені категорії.

Загальну структуру MTMN показано на рис. 3. Вхідні дані складаються з текстового вмісту дописів і прикріпленого зображення. Текстові слова та області зображень потім відповідно кодують попередньо навченою моделлю BERT і моделлю ResNet50. Мережу Blended Attention Network використовують для моделювання міжмодальних і внутрішньомодальних зв'язків, агрегування текстових і візуальних фрагментів і включення різних модальностей, щоб нарешті отримати їх мультимодальне подання. Ґрунтуючись на цих поданнях, мережа пам'яті тем новин спільно вивчає глобальні особливості латентних тем із функціями новинних публікацій, які спільно використовують між темами, і зберігає вивчені їх глобальні особливості. Функції публікації поєднуються з відповідними релевантними глобальними функціями тем, створеними шляхом зчитування їх пам'яті, для формування остаточних представлень виділених фрагментів.

Гібридна модель для виявлення фейкових новин (англ. *A Hybrid Model for Detect Fake News*) була досліджена Indhumathi Gurunathan [13], який використав методи машинного навчання для її розроблення. З усіх відомих складових новинної інформації, що можуть слугувати для аналізу її вмісту, в цій моделі було використано всі доступні – заголовки, контент, пов'язані зображення, соціальні контексти з лайками читачів, які поширюють фейкові новини, і упередженість джерела/автора, щоб виявити оманливі публікації. Цікавим результатом дослідження гібридної моделі є те, що проведений аналіз підкреслює взаємозв'язок між політичною упередженістю видавця новин та довірою до нього. Фактично, проаналізувавши інформацію, зібрану з mediabiasfactcheck.com – найповнішого медіа-ресурсу в мережі Інтернет, було показано, що гіперпартійні джерела появи новин частіше поширюють оманливі історії, ніж інші джерела. Окрім цього, було з'ясовано, що можна уникати зчитування новин, щоб визначити їх правдивість, оскільки за допомогою аналізу упередженості видавців, взаємодії користувачів і пов'язаних із новинами зображень можна досягти аналогічних результатів (достовірність 0,90 проти 0,88 і середня точність 0,79 проти 0,78).

Отже, провівши аналіз програмних систем для виявлення фейкових новин у мережі Інтернет, можна зробити такі проміжні висновки.

З'ясовано, що існує певний набір методів і засобів, реалізованих у відповідних програмних системах, які найкраще підходять для вирішення проблеми виявлення фейкових новин у мережі Інтернет. Проте, більшість сучасних дослідників намагаються виробити свої підходи із застосуванням комбінацій унікальних і вже апробованих методик, щоб успішно вирішити зазначену проблему.

Проаналізовано можливості програмної системи на підставі інноваційного фреймворку Transformer, який використовує зміст новин, їх контент і соціальний контекст для аналізу їхніх корисних характеристик, а також для прогнозування ймовірності появи серед них фейків. Розроблена модель, маючи в своїй основі архітектуру Transformer, легко піддається машинному навчанню за наборами позначених новин, що допомагає швидко виявляти фейки в новинній інформації.

Проаналізовано можливості програмної системи Check-It, а також відповідний плагін, який містить модульне ПЗ, що підтримує ідентифікацію фейкових новин на підставі різноманітних сигналів. Також ця система використовує двостадійний процес вибору функцій, які здійснюють L2-регулярізацію для зменшення ступеня перенавчання моделі, та генетичний алгоритм, щоб визначити оптимальну кількість цих функцій, які можуть успішно навчати логістичну регресію з низьким ресурсом і малою пам'яттю.

Проаналізовано багаторівневу мультимодальну мережу перехресної уваги, в якій використано зв'язки (подібності) між текстовою та візуальною інформацією в новинних публікаціях, що значно допомагає ідентифікувати фейкові новини, які, наприклад, намагаються використовувати нерелевантні зображення, щоб привернути увагу читачів.

Проаналізовано гібридну модель для виявлення фейкових новин, у якій використано методи машинного навчання для її розроблення. З усіх відомих складових новинної інформації, що можуть слугувати для аналізу її вмісту, в цій моделі було використано всі доступні – заголовки, контент, пов'язані зображення, соціальні контексти з лайками читачів, які поширюють фейкові новини, і упередженість джерела/автора, щоб виявити оманливі публікації.

Обговорення результатів дослідження. Фейкові новини є справжньою проблемою сучасного світу, з кожним днем вона стає більш масштабною, їх все важче і важче виявляти навіть кваліфікованими фахівцями [67]. Головним завданням виявлення фейкових новин є

автоматизована їх ідентифікація на ранніх стадіях появи. Іншою проблемою є відсутність або мала кількість так званої позначеної (маркованої) інформації для машинного навчання відповідних моделей, призначених для ідентифікації фейкових новин і подальшого їх аналізу. Тому багато дослідників [16, 22, 42, 65, 80] пропонують все нові та нові програмні системи для виявлення фейкових новин, які пробують вирішувати ці проблеми з різним ступенем точності отриманих результатів. Запропоновані ними моделі таких систем використовують інформацію зі змісту публікацій, їхнього контенту чи соціальних контекстів для виявлення фейків [41, 70, 72]. Робота таких моделей зазвичай базується на архітектурі Transformer, яка містить дві частини: кодувальну для ідентифікації та аналізу новинної інформації; декодувальну, призначену передбачати майбутню появу фейкових новин на підставі минулих спостережень. Також у моделях присутні багато функцій для аналізу змісту новин, їхнього контенту чи соціального контексту [4, 18, 73], щоб допомогти їм краще класифікувати отриману інформацію. Окрім цього, багато авторів [22, 67, 72] пропонують ефективну техніку маркування новин для вирішення проблеми машинного навчання відповідних моделей. Отримані ними експериментальні результати на реальних даних показують, що розроблені моделі можуть виявляти фейкові новини з достатньою точністю протягом декількох хвилин після їх поширення (раннє виявлення), ніж базові розробки. Обговоримо їх дещо конкретніше.

Виявлення фейкових новин є підзавданням класифікації тексту [33], тому їх часто визначають як завдання класифікації справжньої чи хибної інформації [65, 67]. Тому й термін "фейкові новини" стосується хибної або оманливої інформації, яка виглядає як справжні новини. Такі новини мають за мету обдурити або ввести в оману довірливих людей. Фейкові новини подають у різних формах, таких як наживка – заголовки, що вводять в оману, дезорієнтація – зі злим наміром ввести громадськість до хибних дій, дезінформація – хибні дані незалежно від мотивів її подання, містифікація, пародія, сатира, чутки, слухи та інше [90].

Нещодавні дослідження [17, 72, 90] показують, що швидкість поширення фейкових новин є безпрецедентною, а результатом їх дій є принаймні широкий розголос, якщо не заклик до певних дій. Яскравий тому приклад – поширення антивакцинальної дезінформації й чутки, які некоректно порівнювали кількість зареєстрованих виборців у 2018 році з кількістю голосів, поданих на виборах президента у США 2020 року. Наслідки таких новин помітні під час рухів громадськості проти вакцинації, які перешкоджали глобальній боротьбі з COVID-19, або завдана шкода майну та загибель людей (чотирьох охоронців) під час заворушень біля Капітолію (Конгрес США, 7 січня 2021 р.), організовані прихильниками Трампа після програшу ним президентських виборів. Тому надзвичайно важливо зупинити поширення фейкових новин на ранніх стадіях їх появи та, за можливості, знешкодити їхніх авторів.

Значна прогалина в багатьох дослідженнях полягає в тому, що вони зосереджені насамперед на автоматизованій ідентифікації фейкових новин, а не на прогнозуванні їх появи в той чи інший період часу. Фундаментальні роботи [35, 72] щодо раннього виявлення фейкових новин зазвичай їх ідентифікують щонайменше через 12 годин після розповсюдження, що є надто пізно, позаяк 3/4 потенційних їх читачів встигли вже ознайомитися з ними, а серед них принаймні 2/3 поширили її як мінімум у своїй групі [78]. Ефективна модель має мати можливість виявляти фейкові новини на ранніх стадіях, що є мотивацією до відповідних дій багатьох дослідників за останні 2-3 роки.

Інша проблема, яку було згадано в цій роботі, – це дефіцит позначеної новинної інформації (новин, маркованих як справжні або фейкові) у сценаріях реалізації реального світу. Наявні сучасні автоматизовані системи [24, 57, 72] зазвичай використовують повністю марковані дані для класифікації фейкових новин. Однак, дані реального світу, ймовірно, будуть здебільшого мішаними, серед яких максимум 25-30 % марковані [35]. Враховуючи практичні обмеження, такі як відсутність значної кількості фахових експертів у галузі інформаційної безпеки для маркування недостовірних новин, вартість їх виявлення вручну та складність вибору належної мітки для кожної з них, виникає нагальна потреба розроблення ефективних методів машинного навчання відповідних моделей для автоматизованої ідентифікації новинної інформації на ранніх стадіях їх появи. Одним з альтернативних підходів є використання галасливих, часто обмежених можливостями або неточних новинних джерел для контролю за маркуванням великих обсягів навчальних даних. Ідея полягає в тому, що виставлені навчальні мітки можуть бути не зовсім точними або частково правдивими, але такі дані можна використовувати для машинного навчання моделі. Згодом, після її додаткового навчання на реальних даних, така модель буде практично придатною для прогнозування часу появи фейкових новин. Таку схему використання тренувальних міток на тій чи іншій інформації ще називають технікою слабкого нагляду (англ. *Weak Supervision Technique*) [81].

Зазвичай моделі для виявлення фейкових новин навчають на поточних даних (доступних протягом певного часу), які можуть навіть не стосуватися майбутніх подій. Багато мічених даних із перевірених фейкових новин незабаром старіють, позаяк втрачають свою цінність через появу нових подій. Наприклад, модель, яку навчали на фейкових новинах до виникнення коронавірусу COVID-19, може не класифікувати їх належно під час його безпосередньої появи. Однак, цю модель можна легко донавчати вже на реальних даних, які стосуються різних фактів про перебіг коронавірусу. Таку проблему моделі з цільовою концепцією (наприклад, новини як "справжні" чи "фальшиві"), коли основний зв'язок між вхідними даними та цільовою функцією їхнього аналізу змінюється з плином часу, ще називають дрейфом концепції (англ. *Drift Concept*) [15]. У своїх публікаціях багато науковців [6, 19, 32, 35, 55] досліджували, чи впливає такий дрейф на точність роботи їхньої моделі під час ідентифікації новинної інформації, і якщо так, то як можна пом'якшити його для підвищення достовірності отримання прогнозів.

Багато науковців [24, 33, 57, 72, 89] розглядають проблему, яка стосується раннього виявлення фейкових новин і дефіцит позначених даних для автоматизованої ідентифікації новинної інформації. Вони пропонували нову структуру моделі, засновану на архітектурі глибокої нейронної мережі. Наявні роботи розглядають такі функції моделі, як аналіз змісту новин та їхнього кон-

тенту [24, 26, 89], аналіз їх соціального контексту [33, 35, 36, 57, 72, 87], або обидві разом [44, 57, 72]. Автори цих робіт намагалися виявляти фейкові новини на ранніх стадіях їх появи, тобто через кілька хвилин після їх публікації на сайті мережі. Також вони вирішували проблему нестачі маркерів на новинній інформації, яка часто виникає в сценаріях реального світу через її велику кількість. Окрім цього, їхні моделі можуть боротися з дрейфом концепції, що є актуальним для їх машинного навчання на штучно згенерованих даних, яку згодом можна додатково навчати вже на реальній інформації.

Деякі науковці [7, 32, 64, 72], натхненні моделлю двонаправленого та авторегресійного трансформера (BART) [32] від Facebook, яку успішно використовують в задачах моделювання мови, пропонують застосувати глибокий двонаправлений кодер і декодер зліва направо під управлінням однієї уніфікованої моделі для вирішення завдань виявлення фейкових новин. У своїх дослідженнях вони вирішили працювати з моделлю BART [7] замість найсучаснішої моделі BERT [32], яка продемонструвала свої можливості в завданнях NLP (оброблення природної мови), наприклад, відповіді на запитання та отримані мовні висновки. Також була розроблена GPT-2 модель [64], яка має вражаючі властивості авторегресії (часового ряду). Основна ефективність її використання полягає в тому, що модель BART [7] поєднує унікальні функції (двонаправлену та авторегресійну) як для генерування новинного тексту, так і часового моделювання, які авторам знадобилася для досягнення їхніх цілей.

Нагадаємо, що в 2019 році багато хто став свідком блискучого використання машинного навчання. Модель GPT-2 від американської компанії OpenAI [64, 84], що займається розробленням і ліцензуванням технологій на підставі машинного навчання, продемонструвала вражаючу здатність писати зв'язні та емоційні тексти, що перевершують наші уявлення про те, що можуть генерувати сучасні мовні моделі. GPT-2 не є особливо новою архітектурою – вона дуже нагадує Трансформер-Декодер (англ. *Decoder-only Transformer*). Відмінність GPT-2 в тому, що це справді величезна мовна модель на підставі архітектури Transformer, навчена на значному наборі даних.

Розроблена свого часу модель BART [7] відрізнялася від оригінальної [32] такими особливостями: на відміну від оригінальної, яка приймає одне речення/документ як вхідні дані, їхня модель у частині кодера містить значний набір функцій для аналізу змісту новин, їх контенту [4, 72] та соціального контексту [41, 70]; розроблена модель використовує декодер, щоб отримати передбачення не тільки з попередніх текстових послідовностей (новинних публікацій), як у оригінальній моделі BART [32], але й із попередньої поведінки користувачів (як вони реагують на ці публікації). Запропонована модель [7] дає змогу виявляти фейкові новини завчасно шляхом тимчасового моделювання поведінки користувачів. Проте, поверх оригінальної моделі BART [32] автори додали один лінійний шар, щоб краще класифікувати новини як справжні чи фейкові.

Отже, за результатами виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – розроблено новий підхід до вирішення проблеми автоматизованої ідентифікації, аналізу та маркування новинної інформації, в основу якого закладено новітні методи машинного навчання, що дасть змогу отримати більшу точність виявлення хибної інформації, зробить роботу програмної системи дещо стабільнішою і менш ресурсовитратною.

Практична значущість результатів дослідження – розроблений підхід з машинним навчанням моделі для вирішення задач із можливою зміною набору даних протягом певного часу можна використати для побудови автоматизованої ідентифікації, аналізу та маркування новинної інформації.

Висновки / Conclusions

Проаналізовано наявні підходи до вирішення проблеми виявлення фейкових новин у мережі Інтернет, розглянуто екосистему новин як бізнес-модель їхньої появи, ознайомлення та поширення, що передбачає комплекс взаємопов'язаних сутностей – виробників новинної інформації її користувачів і розповсюджувачів, які сукупно можуть вирішити різноманітні завдання потенційних учасників на єдиному інтегрованому полі. За результатами виконаного дослідження можна зробити такі основні висновки.

1. З'ясовано, що мережа Інтернет має очевидні переваги над традиційними ЗМІ у розповсюдженні новин, такі як миттєвий доступ читачів до потрібної інформації, зазвичай безкоштовне її розміщення, відсутність обмежень у стилі подання та різноманітність формату – текстова, графічна та мультимедійна. Однак, веб-сайти, де розміщують новини, здебільшого не врегульовані будь-яким редакційним наглядом, а також державними органами з інформаційної безпеки. Тому пересічному читачу часто важко визначити, правдива чи фейкова інформація в деяких опублікованих новинах.

2. Встановлено, що серед вітчизняних фахівців заслуговують уваги ґрунтовні публікації в основному професійних журналістів, у яких вони висвітлюють як різну хибну інформацію, в т. ч. й фейкові новини, так і повну дезінформацію. Не відстають від них і молоді дарування, які у своїх критичних дописах розвінчують міфи про силу і міць північного сусіда, а також різні фейки про ті чи інші резонансні події. Шкода, але серед вітчизняних науковців важко назвати таких, праці яких би заслуговували уваги. Зазначену проблему за останнє десятиліття з успіхом почали досліджувати закордонні вчені, які домоглися чималих результатів як у практичному, так і теоретичному планах.

3. Досліджено, що головним завданням виявлення фейкових новин є автоматизована їх ідентифікація на ранніх стадіях появи. Водночас, іншою проблемою є відсутність або мала кількість так званої позначеної (маркованої) інформації для машинного навчання відповідних моделей, призначених насамперед для ідентифікації фейкових новин, а також подальшого їх аналізу. Тому багато закордонних дослідників пропонують все нові та нові методи і засоби для виявлення фейкових новин, які з плином часу прогресують у вирішенні цієї проблеми з різним ступенем точності отриманих результатів.

4. З'ясовано, що за терміном екосистемне мислення знаходиться деякий світогляд, цілеспрямоване мислення та відповідні дії людей, залучені в цій системі. Зазначений термін стосується й екосистеми новин як бізнес-

модель їхньої появи, ознайомлення та поширення, що передбачає комплекс взаємопов'язаних сутностей – виробників (новинних ЗМІ чи редакційних компаній), новинної інформації (зміст і контент новин) та її користувачів (соціальний контекст їхніх профілів і дій), які сукупно можуть вирішити різноманітні завдання потенційних учасників на єдиному інтегрованому полі.

5. Для встановлення довіри користувачів до соціальних мереж і навпаки часто застосовують підхід, згідно з яким виявляють певну спільноту, що мають деякі негативні поглядами. Для цього аналізують настрої її учасників і застосовують методи ранжування їхніх профілів за інтенсивністю висловлювань чи проявом думок. У такий спосіб встановлють лідерів спільноти, визначають їхню схильність закликати до насильства. Однак, для виявлення фейкових новин, які враховують довіру користувачів до соціального контексту медіа-ресурсів, використовують просту кластеризацію зображень чи відеофайлів. Тут часто застосовують так зване нульове навчання ZSL, щоб визначити довіру користувачів до соціальних мереж.

6. З'ясовано, що існує певний набір методів і засобів, реалізованих у відповідних програмних системах, які найкраще підходять для вирішення проблеми виявлення фейкових новин у мережі Інтернет. Проте, більшість сучасних дослідників намагаються виробити свої підходи із застосуванням комбінацій унікальних і вже апробованих методик, щоб успішно вирішити зазначену проблему.

7. Проаналізовано можливості сучасних програмних систем на підставі інноваційного фреймворку Transformer, який використовує зміст новин, їх контент і соціальний контекст для аналізу їхніх корисних характеристик, а також для прогнозування ймовірності появи серед них фейків. Розроблена модель, маючи в своїй основі архітектуру Transformer, легко піддається машинному навчанню за наборами позначених новин, що допомагає швидко виявляти фейки в новинній інформації.

References

1. Abu-Salih, B., Wongthongtham, P., Chan, K. Y., & Zhu, D. (2019). CredSaT: credibility ranking of users in big social data incorporating semantic analysis and temporal factor. *Journal of Information Science*. Vol. 45, Issue 2, pp. 259–280. https://doi.org/10.1177/0165551518790424
2. Anderson, C. W. (2016). News ecosystems . In: Witschge, T, Anderson, CW, Domingo, D and Hermida, A. (Eds.) *The Sage Handbook of Digital Journalism. SAGE, London*, UK. 410–423. URL: https://uk.sagepub.com/en-gb/eur/the-sage-handbook-of-digital-journalism/book244110
3. Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2020). Predicting factuality of reporting and bias of news media sources. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2018, pp. 3528–3539.
4. Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O., & Kompatsiaris, Y. (2018). Detection and Visualization of Misleading Content on Twitter. *International Journal of Multimedia Information Retrieval*, 7(1), 71–86. https://doi.org/10.1007/s13735-017-0143-x
5. Cao, J., Qi, P., Sheng, Q., Yang, T., Guo, J., & Li, J. (2020). Exploring the Role of Visual Content in Fake News Detection. *Disinformation, Misinformation, Fake News Social Media*, pp. 141–161. https://doi.org/10.1007/978-3-030-42699-6
6. De Maio, C., Fenza, G., Gallo, M., Loia, V., & Volpe, A. (2020). Cross-relating heterogeneous Text Streams for Credibility Assessment. In: *IEEE Conference on Evolving and Adaptive Intelligent Systems*, 2020-May. https://doi.org/10.1109/EAIS48028.2020.9122701
7. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint. *Computer Science. Computation and Language*. http://arxiv.org/abs/1810.04805.
8. Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, & Toutanova, Kristina. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1 (Long and Short Papers), pp. 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics. https://doi.org/10.18653/v1/n19-1423
9. Django. (2022, 9th September). Django Introduction. MDN Web Docs. *Server-side website programming*. URL: https://developer.mozilla.org/en-US/docs/Learn/Server-side/Django/Introduction
10. Dougrez-Lewis, J., Liakata, M., Kochkina, E., & He, Yu. (2021). Learning Disentangled Latent Topics for Twitter Rumour Veracity Classification. In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP* 2021, pp. 3902–3908, Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-acl.341
11. GitHub. (2022). TensorFlow. URL: https://github.com/tensorflow/tensorflow
12. Gruppi, M., Horne, B. D., & Adalı, S. (2020). NELA-GT-2019: A large multi-labelled news dataset for the study of misinformation in news articles. *arXiv Preprint. Computer Science. Computation and Language*. http://arxiv.org/abs/2003.08444v2.
13. Gurunathan, I. (2021, 2th May). A Hybrid Model to Detect Fake News. *Computer Science Graduate Projects and Theses*. 17. URL: https://scholarworks.boisestate.edu/cs_gradproj/17
14. Helmstetter, S., & Paulheim, H. (2018). Weakly supervised learning for Fake News detection on Twitter. In: 2018 IEEE/ACM *International Conference on Advances in Social Networks Analysis and Mining* (ASONAM), pp. 274–277.
15. Hoens, T. R., Polikar, R., & Chawla, N. (2012). V: Learning from streaming data with concept drift and imbalance: an overview. *Progress in Artificial Intelligence*, 1, 89–101.
16. Horne, B. D., Dron, W., Khedr, S., & Adali, S. (2018). Assessing the news landscape: a Multi-module toolkit for evaluating the credibility of news. In: *The Web Conference 2018–Companion of the World Wide Web Conference*, WWW 2018, pp. 235–238.
17. Horne, B. D., Nørregaard, J., & Adali, S. (2019). Robust Fake News detection over time and attack. *ACM Transactions on Intelligent Systems and Technology*. https://doi.org/10.1145/3363818
18. Horne, B., & Adali, S. (2017). This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire Than Real News. *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 11, no. 1, pp. 759–766. https://doi.org/10.1609/icwsm.v11i1.14976
19. Horne, Benjamin; Gruppi, M. (2021). NELA-GT-2020: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. *arXiv Preprint. Computer Science. Computation and Language*. http://arxiv.org/abs/2102.04567. (2021). https://doi.org/10.7910/DVN/CHMUYZ
20. Huang, Q., Zhou, C., Wu, J., Liu, L., & Wang, B. (2020). Deep spatial–temporal structure learning for rumor detection on Twitter. Neural Computing and Applications https://doi.org/10.1007/s00521-020-05236-4
21. Jiang, S., Chen, X., Zhang, L., Chen, S., & Liu, H. (2019). User-characteristic enhanced model for Fake News detection in social media. In: *CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 634–646.
22. Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: *Proceedings of the 25th ACM International Conference on Multimedia*, pp. 795–816.
23. Jin, Z., Cao, J., Zhang, Y., Zhou, J., & Tian, Q. (2017). Novel Visual and Statistical Image Features for Microblogs News Veri-

fication. In: *IEEE Transactions on Multimedia*. Vol. 19(3), No. 3, pp. 598–608. https://doi.org/10.1109/TMM.2016.2617078
24. Jwa, H., Oh, D., Park, K., Kang, J. M., & Lim, H. (2019). exBAKE: automatic Fake News detection model based on Bidirectional Encoder Representations from Transformers (BERT). *Applied Sciences,* 9, 4062. https://doi.org/10.3390/app9194062
25. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake News detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*. Vol. 80, pp. 11765–11788. https://doi.org/10.1007/s11042-020-10183-2
26. Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2020). FNDNet–a deep convolutional neural network for Fake News detection. *Cognitive Systems Research*, 61, 32–44. https://doi.org/10.1016/j.cogsys.2019.12.005
27. Karimi, H., Roy, P., Saba-Sadiya, S., & Tang, J. (2018). Multi-source Multi-class Fake News detection. In: *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1546–1557. Santa Fe, New Mexico, USA. Association for Computational Linguistics. URL: https://aclanthology.org/C18-1131
28. Khanam, Z., Alwasel, B. N., Sirafi, H., & Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. *IOP Conference Series: Materials Science and Engineering*. Vol. 1099, No. 1, pp. 012040. https://doi.org/10.1088/1757-899x/1099/1/012040
29. Komolafe, A. (2022, 14th November). Retraining Model During Deployment: Continuous Training and Continuous Testing. MLOps Blog. URL: https://neptune.ai/blog/retraining-model-during-deployment-continuous-training-continuous-testing
30. Korshunov, A., & Gomzin, A. (2012). Thematic modeling of texts in natural language (journal). Proceedings of the Institute for System Programming of the Russian Academy of Sciences.
31. Kyrychenko, A. (2019). Elon Musks OpenAI created an algorithm that writes Fake News. hromadske. URL: https://hromadske.ua/posts/openal-ilona-maska-stvoriv-algoritm-yakij-pishe-fejkovi-novini. [In Ukrainian].
32. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv Preprint. Computer Science. Computation and Language.* https://doi.org/10.18653/v1/2020.acl-main.703
33. Liu, C., Wu, X., Yu, M., Li, G., Jiang, J., Huang, W., & Lu, X. (2019). A two-stage model based on BERT for short Fake News detection. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 11776 LNAI, pp. 172–183. https://doi.org/10.1007/978-3-030-29563-9_17
34. Liu, Y., & Wu, Y. F. B. (2018). Early detection of Fake News on social media through propagation path classification with recurrent and convolutional networks. In: *32nd AAAI Conference on Artificial Intelligence*, AAAI 2018, pp. 354–361.
35. Liu, Y., & Wu, Y. F. B. (2020). FNED: A Deep Network for Fake News Early Detection on Social Media. *ACM Transactions on Information Systems*. Vol. 38, Issue 3, pp. 1–33. https://doi.org/10.1145/3386253
36. Mohammadrezaei, M., Shiri, M. E., & Rahmani, A. M. (2018). Identifying fake accounts on social networks based on graph analysis and classification algorithms. *Security and Communication Networks*. https://doi.org/10.1155/2018/5923156
37. Murugesan, S., & Kaliyamurthie K. P. (2022). Estimation of precision in Fake News detection using novel bert algorithm and comparison with random forest. *Authorea*. https://doi.org/10.22541/au.165237518.82791368/v1
38. Nakamura, K., Levy, S., & Wang, W. Y. (2019). r/fakeddit: a new multimodal benchmark dataset for fine-grained Fake News detection. *arXiv Preprint. Computer Science. Computation and Language*. http://arxiv.org/abs/1911.03854
39. Nakamura, K., Levy, S., & Wang, W. Ya. (2019). r/fakeddit: A new multimodal benchmark dataset for fine-grained Fake News detection. *arXiv preprint arXiv:1911.03854*.
40. Naseem, U., Razzak, I., & Eklund, P. W. (2020). A survey of preprocessing techniques to improve short-text quality: a case study on hate speech detection on twitter. *Multimedia Tools and Applications*, 80, 35239–35266. https://doi.org/10.1007/s11042-020-10082-6
41. Naseem, U., Razzak, I., & Hameed, I. A. (2019). Deep context-aware embedding for abusive and hate speech detection on Twitter. *Australian Journal of Intelligent Information Processing Systems*. Vol. 15, No. 4, pp. 69–76. URL: https://www.researchgate.net/publication/340756139_Deep_Context-Aware_Embedding_for_Abusive_and_Hate_Speech_detection_on_Twitter
42. Naseem, U., Razzak, I., Khushi, M., Eklund, P. W., & Kim, J. (2021). COVIDsenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis". In: *IEEE Transactions on Computational Social Systems*, Vol. 8, no. 4, pp. 1003–1015. https://doi.org/10.1109/TCSS.2021.3051189
43. Neptune.ai documentation. About neptune.ai. URL: https://docs.neptune.ai/about/intro/
44. Nguyen, V. H., Sugiyama, K., Nakov, P., & Kan, M. Y. (2020). FANG: leveraging social context for Fake News detection using graph representation. *International Conference Information Knowledge Management Proceedings*. https://doi.org/10.1145/3340531.3412046
45. Nishant, R., Deepika, K., Naman, K., Chandan, R., & Ahad, A. (2022). Fake News Classification using transformer based enhanced LSTM and BERT. *International Journal of Cognitive Computing in Engineering*. Vol. 3, pp. 98–105. https://doi.org/10.1016/j.ijcce.2022.03.003
46. Nørregaard, J., Horne, B. D., & Adalı, S. (2019). NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In: *Proceedings of 13th International Conference on Web and Social Media*, ICWSM 2019, pp. 630–638. https://doi.org/10.7910/DVN/ULHLCB
47. O'Connor, Joseph. (2013). NLP Workbook: A Practical Guide to Achieving the Results You Want Paperback. Red Wheel; Workbook, Reprint edition, 304 p. URL: https://www.amazon.com/NLP-Workbook-Practical-Achieving-Results/dp/1573246158
48. Papadopoulos, S., Kompatsiaris, Y., Vakali, A., & Spyridonos, P. (2012). Community detection in social media. *Data Mining and Knowledge Discovery*, 24, 515–554 (2012). https://doi.org/10.1007/s10618-011-0224-z
49. Paschalides, D., Kornilakis, A., Christodoulou, C., Andreou, R., Pallis, G., Dikaiakos, M., & Markatos, E. (2019). Check-It: A plugin for Detecting and Reducing the Spread of Fake News and Misinformation on the Web. WI 19: IEEE/WIC/ACM International Conference on Web Intelligence. *Thessaloniki Greece*. New York, NY, USA. https://doi.org/10.1145/3350546.3352534
50. Patil, D. R. (2022). Fake News Detection Using Majority Voting Technique. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.2203.09936
51. Patil, D. R., & Patil, J. B. (2015). Survey on Malicious Web Pages Detection Techniques. *International Journal of u – Service, Science and Technology*. Vol. 8, No 5, pp. 195–206. https://doi.org/10.14257/ijunesst.2015.8.5.18
52. Pavelko, Volodymyr. (2022). Entry to the ecosystem mission. TransformWISE. *Transformation Portal*. URL: https://transformwise.com/portal/article-introduction-to-ecosystem-thinking/
53. Perez-Rosas, V., & Mihalcea, R. (2013). Sentiment analysis of online spoken reviews. ISCA. Archive. Interspeech 2013. https://doi.org/10.21437/interspeech.2013-243
54. Pizarro, J. (2020). Profiling Bots and Fake News Spreaders at PAN19 and PAN20 : Bots and Gender Profiling 2019, Profiling Fake News Spreaders on Twitter 2020," 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 626–630. https://doi.org/10.1109/DSAA49011.2020.00088
55. Pocheptsov, G. (2018). There are many fakes, the theory of fakes is even bigger. *DM media sapiens. Media analytics. Articles*. URL: https://ms.detector.media/mediaanalitika/post/21769/2018-09-09-feykov-mnogo-teoryy-feykov-eshche-bolshe/. [In Ukrainian].
56. Popat, K., Mukherjee, S., Strötgen, J., & Weikum, G. (2016). Credibility assessment of textual claims on the web. In: *International*

Conference on Information and Knowledge Managenent Proceedings, 24–28-October-2016, pp. 2173–2178. https://doi.org/10.1145/2983323.2983661
57. Popat, K., Mukherjee, S., Yates, A., & Weikum, G. (2018). DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.1809.06416
58. Poriaa, S., Cambriab, E., & Gelbukhc, A. (2016, 15th September). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*. Vol. 108, pp. 42–49. https://doi.org/10.1016/j.knosys.2016.06.009
59. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A Stylometric Inquiry into Hyperpartisan and Fake News. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.1702.05638
60. Przybyla, P. (2020). Capturing the Style of Fake News. *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34, no. 01, pp. 490–497. https://doi.org/10.1609/aaai.v34i01.5386
61. Pushp, P. K., & Srivastava, M. M. (2017). Train Once, Test Anywhere: Zero-Shot Learning for Text Classification. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.1712.05972
62. Qi, P., Cao, J., Yang, T., Guo, J., & Li, J. (2019). Exploiting multi-domain visual information for Fake News detection. In: *19th IEEE International Conference on Data Mining*. https://doi.org/10.48550/arXiv.1908.04472
63. Qian, F., Gong, C., Sharma, K., & Liu, Y. (2018). Neural User Response Generator: Fake News Detection with Collective User Intelligence. In: *IJCAI, International Joint Conference on Artificial Intelligence*, pp. 3834–3840. https://doi.org/10.24963/ijcai.2018/533
64. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. *OpenAI Blog*. 1, 9. URL: https://paperswithcode.com/paper/language-models-are-unsupervised-multitask
65. Raza, S., & Ding, C. (2019). News Recommender System Considering Temporal Dynamics and News Taxonomy. 2019 *IEEE International Conference on Big Data*, pp. 920–929. https://doi.org/10.1109/BigData47090.2019.9005459
66. Raza, S., & Ding, C. (2021). News Recommender System: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, 55, 749–800. https://doi.org/10.1007/s10462-021-10043-x
67. Raza, S., & Ding, C. (2022). Fake News detection based on news content and social contexts: a transformer-based approach. *International Journal of Data Science and Analytics*. Vol. 13, pp. 335–362. https://doi.org/10.1007/s41060-021-00302-z
68. Rose-Collins, Felix. (2022, Jul 20). Statistics of social media: Social media remake – and the axis of why. *Rank tracker. Social services*. URL: https://www.ranktracker.com/uk/blog/social-media-stats-social-media-is-taking-over-heres-why/
69. Savruk, Helena. (2022). Thinking about ecosystems: what is needed for large-scale changes. *School of strategic architect*. URL: https://www.ssa.kmbs.ua/misliti-ekosistemami-sho-potribno-dlya-masshtabnih-zmin
70. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: a data repository with news content, social context, and spatiotemporal information for studying Fake News on social media. *Big Data*. Vol. 8, No. 3, 171–188. https://doi.org/10.1089/big.2020.0062
71. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD *Explorations Newsletter*. Vol. 19, Issue. 1, pp. 22–36. https://doi.org/10.1145/3137597.3137600
72. Shu, K., Wang, S., & Liu, H. (2019). Beyond news contents: The role of social context for Fake News detection. In: WSDM 2019–*Proceedings of 12th ACM International Conference on Web Search Data Mining*, pp. 312–320. https://doi.org/10.1145/3289600.3290994
73. Silva, R. M., Santos, R. L. S., Almeida, T. A., & Pardo, T. A. S. (2020). Towards automatically filtering Fake News in Portuguese. *Expert Systems with Applications*. Vol. 146, 113–199. https://doi.org/10.1016/j.eswa.2020.113199
74. Starkova, Anna. (2022). History of e-commerce. Part 1. *Turumburum*. URL: https://turumburum.ua/blog/e-commerce-nachalo-chast-1/
75. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In: NIPS17: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010.
76. Vijjali, R., Potluri, P., Kumar, S., & Teki, S. (2020). Two stage transformer model for Covid-19 Fake News detection and fact checking. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.2011.13253
77. Vorontsov, K. V. (2013). Probabilistic topic modeling. URL: www.machinelearning.ru
78. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*. Vol. 359. Issue 6380, pp. 1146–1151. https://doi.org/10.1126/science.aap955
79. Wanda, P., & Jie, H. J. (2020). DeepProfile: finding fake profile in online social network using dynamic CNN. *Journal of Information Security and Applications*. Vol. 52, 102465. https://doi.org/10.1016/j.jisa.2020.102465
80. Wang, Y., Sohn, S., Liu, S., Shen, F., Wang, L., Atkinson, E. J., Amin, S., & Liu, H. (2019). A clinical text classification paradigm using weak supervision and deep representation. *BMC Medical Informatics and Decision Making*. Vol. 19, 1–13. https://doi.org/10.1186/s12911-018-0723-6
81. Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., & Gao, J. (2020). Weak supervision for Fake News detection via reinforcement learning. In: AAAI 2020–34th AAAI Conference on Artificial Intelligence, pp. 516–523. https://doi.org/10.48550/arXiv.1912.12520
82. What is TensorFlow.js? (2022). Learn to build anything with Google. URL: https://developers.google.com/
83. Williams, A., Nangia, N., & Bowman, S. R. (2017). A broad-coverage challenge corpus for sentence understanding through inference. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.1704.05426
84. Wu, X., & Lode, M. (2020). Language Models are Unsupervised Multitask Learner (summarization). *OpenAI Blog*. Vol. 1, 1–7. https://bit.ly/3vgaVJc
85. Xian, Y., Akata, Z., Sharma, G., Nguyen, Q., Hein, M., & Schiele, B. (2016). Latent embeddings for zero-shot classification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 69–77. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.1603.08895
86. Yang, K.-C., Niven, T., & Kao, H.-Y. (2019). Fake News Detection as Natural Language Inference. *arXiv Preprint. Computer Science. Computation and Language*. https://doi.org/10.48550/arXiv.1907.07347
87. Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019). Unsupervised Fake News detection on social media: a generative approach. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33, No. 01, pp. 5644–5651. https://doi.org/10.1609/aaai.v33i01.33015644
88. Ying, L., Yu, H., Wang, J., Ji, Y., & Qian, S. (2021). Fake News Detection via Multi-modal Topic Memory Network. In: *IEEE Access*. Vol. 9, pp. 132818–132829. https://doi.org/10.1109/ACCESS.2021.3113981
89. Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2020). Defending against neural Fake News. *NeurIPS Proceedings*. URL: https://proceedings.neurips.cc/paper/2019/hash/3e9f0fc9b2f89e043bc6233994dfcf76-Abstract.html
90. Zhou, X., & Zafarani, R. (2021). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM*

Computing Surveys. Vol. 53, Issue 5, pp. 1–40. https://doi.org/10.1145/3395046
91. Zhou, X., Wu, J., & Zafarani, R. (2020). SAFE: Similarity-Aware Multi-modal Fake News Detection. *arXiv Preprint. Computer Science. Computation and Language.* https://doi.org/10.48550/arXiv.2003.04981

***T. P. Dyak*[1], *Yu. I. Hrytsiuk*[1], *P. P. Horvat*[2]**
[1] *Lviv Polytechnic National University, Lviv, Ukraine*
[2] *Uzhhorod National University, Uzhhorod, Ukraine*

THE PROBLEM OF FAKE NEWS DETECTION ON INTERNET WEBSITES

The available approaches to solving the problem of detecting fake news on the Internet are analyzed in the paper. The news ecosystem is considered as a business model for its emergence, delivering and broadcasting, which involves a complex of interconnected entities such as news producers, users and distributors, which can collectively solve various tasks of potential system participants on a single integrated field. In the course of research the Internet is found to have obvious advantages over traditional mass media in news distribution such as readers' instant access to the necessary information, free data placement, no restrictions on presentation style, and also a variety of formats, in particular text, graphics, and multimedia. However, the lack of any editorial oversight, as well as government information security authorities, has meant that it is often difficult to determine the reliability of some published news for the average reader. Thorough publications which are delivered mainly by professional journalists covering both reliable information and complete misinformation merits attention among Ukrainian experts. Young gifted journalists are not lagging behind debunking myths about the strength and power of the northern neighbour, as well as various fakes about certain high-profile events in their critical posts. Some foreign scientists have been successfully studying the problem over the last decade, and they have already achieved significant results both in practical and theoretical terms. Hence, the main task of detecting fake news is considered to be its automated identification in the early stages of emergence, as well as the absence or small amount of indicated information for machine learning of the corresponding models. Therefore, many foreign researchers offer more brand new methods and tools for detecting fake news, which eventually progress in solving this problem. Their software systems contain combinations of unique and proven techniques to successfully solve the stated problem of detecting fake news on the Internet. The possibilities of modern software systems are analyzed based on the innovative transformer framework, which uses the news content and its social context to analyse useful characteristics, as well as to predict the probability of the emergence of fakes among them. The developed model being based on the transformer architecture is easily amenable to machine learning on sets of indicated news, which helps quickly identify fake news information. The capabilities of the Check-It software system were analyzed, as well as the corresponding plug-in, which contains modular software that supports the identification of fake news based on various signals. Also, this system uses a two-stage feature selection process that performs L2-regularization to reduce the degree of model overtraining, and a genetic algorithm to determine the optimal number of these features that can successfully train low-resource, low-memory logistic regression.

Keywords: Text Mining; Computer Linguistics; Artificial Intelligence; Neural Network; Genetic Algorithm; optimal solution; Rank tracker.