DOI: 10.18372/2225-5036.26.14942

АНАЛІЗ ЗАСТОСУВАННЯ ІСНУЮЧИХ ТЕХНІК РОЗПІЗНАВАННЯ ФЕЙКОВИХ НОВИН ДЛЯ ПРОТИДІЇ ІНФОРМАЦІЙНІЙ ПРОПАГАНДІ

Штефанюк Євгеній, Опірський Іван, Гарасимчук Олег

Національний університет «Львівська політехніка»

ШТЕФАНЮК Євгеній Федорович.

Рік та місце народження: 1994 рік, м. Хуст, Закарпатська область, Україна.
Освіта: Національний університет «Львівська Політехніка», 2018 рік.
Посада: аспірант кафедри захисту інформації Національного університету «Львівська політехніка».
Наукові інтереси: протидія інформаційним впливам, проектування комплексних систем захисту інформації.
E-mail: yevhen.sht@gmail.com.
Orcid ID: 0000-0003-0734-6648.

ОПІРСЬКИЙ Іван Романович, д.т.н., доц.

Рік та місце народження: 1987 рік, м. Сімферополь, АР Крим, Україна.
Освіта: Національний університет «Львівська Політехніка», 2008 рік.
Посада: професор кафедри захисту інформації з 2019 року.
Наукові інтереси: методи і засоби технічного захисту інформації, охорона державної таємниці, проектування комплексних систем захисту інформації, лазерні системи акустичної розвідки, математичні методи та моделі захисту інформації, технічні канали витоку інформації, спецвимірювання.
Публікації: понад 120 наукових публікацій, серед яких наукові статті, монографії, навчальні посібники, тези та матеріали доповідей на конференціях.
E-mail: iopirsky@gmail.com.
Orcid ID:0000-0002-8461-8996.

ГАРАСИМЧУК Олег Ігорович, к.т.н., доц.

Рік та місце народження: 1979 рік, м. Бережани, Тернопільська обл., Україна.
Освіта: Національний університет «Львівська Політехніка», 2001 рік.
Посада: доцент кафедри захисту інформації.
Наукові інтереси: комплексні системи санкціонованого доступу, генерування псевдовипадкових чисел та послідовностей, генерування пуассонівських імпульсних послідовностей, методи і засоби захисту інформації, проектування комплексних систем захисту інформації, сигнальні процесори в системах захисту інформації.
Публікації: більше 80 наукових публікацій, серед яких наукові статті, монографії, навчальний посібник, патенти, тези та матеріали доповідей на конференціях.
E-mail: oleh.harasymchuk@gmail.com.
Orcid ID: 0000-0002-8742-8872.

Анотація. *Проблема виявлення неправдивої (фейкової) інформації, що передається через різні канали в мережі Інтернет стає все більш актуальною. Одним з різновидів такої інформації є цільова пропаганда, яка має конкретну мету та використовує спеціально створені ресурси. Для боротьби з таким інформаційним впливом можна використовувати вже розроблені засоби виявлення фейкових новин. В цій статті розглянуто особливості інформаційної пропаганди та підходів до боротьби з нею; ефективність роботи декількох відомих технік розпізнавання фейкових новин; проведено аналіз ефективності цих технік в контексті можливості їхнього застосування для протидії цілеспрямованим інформаційним впливам. На основі проведеного дослідження обрано найбільш перспективний алгоритм для розпізнавання інформаційної пропаганди в соціальних мережах.*

Ключові слова: *інформаційна безпека, критична інфраструктура, загрози безпеці, когнітивне моделювання, нечітка когнітивна карта.*

Вступ

В останні роки ми спостерігаємо тенденцію до збільшення ролі соціальних медіа в житті користувачів Інтернету. Все більше людей використовують їх та Інтернет-ресурси для отримання останніх новин. Внаслідок цього, однією з головних проблем є визначення рівня довіри до інформації, що поширюється цими ресурсами.

Останнім часом набувають все більшого поширення так звані фейкові новини – статті, які розповсюджують неправдиву інформацію, написані спеціально, щоб ввести користувача в оману. Згідно з деякими джерелами[1,2], в окремих випадках кількість таких фейкових новин може перевищувати число правдивих, що створює так званий ефект information vertigo ("інформаційного запаморочення"), коли користувачі вже не можуть відрізнити правдиву інформацію від вигаданої. Це стає підґрунтям до потужного інформаційного впливу на соціальну думку. Яскравим прикладом можуть слугувати дослідження, що вказують на безпосередній вплив поширення фейкових новин на президентську кампанію США при обранні Дональда Трампа [3]. І якщо окремі фейкові новини покликані просто сформувати хибне уявлення про певний предмет чи подію, то цілеспрямоване поширення неправдивої інформації окремими організаціями чи урядами може вплинути на прийняття політичних рішень чи навіть дестабілізувати ситуацію в соціумі.

Згідно з [4] однією з основних проблем ЄС та США в кіберпросторі є інформаційна пропаганда з деяких пострадянських країн. Вона становить серйозну загрозу як в соціальному, так і в геополітичному контексті. Таким чином, надзвичайно актуальним завданням є розробка ефективних засобів протидії такій інформаційній пропаганді, як з боку окремих організацій, так і з боку урядів країн.

Метою дослідження було дослідити особливості інформаційної пропаганди з країн пострадянського простору; запропонувати підхід для протидії такій пропаганді; проаналізувати існуючі алгоритми розпізнавання фейкових новин та їхню ефективність та обрати найточніший, з точки зору авторів, алгоритм, який дозволив б побудувати ефективну систему для реалізації запропонованого підходу для протидії такій інформаційній пропаганді в соціальних мережах.

Особливості інформаційної пропаганди з пострадянських країн

Інформаційна пропаганда з деяких пострадянських країн має низку особливостей, що вирізняють її від традиційної інформаційної війни. Проведені дослідження дозволили визначити ці особливості і також представити техніки, що ними застосовуються.

До головних особливостей інформаційної пропаганди з пострадянських країн належать[5]:

- надзвичайно великий об'єм;
- неконсистентність;
- велика кількість каналів поширення;
- викривлення реальних фактів, а інколи їх повна фальсифікація.

Розглянемо кожну з них.

Великий об'єм фейкової інформації забезпечує можливість її підтвердження великою кількістю користувачів, що підвищує ступінь довіри до такої інформації. Окрім того, згідно з останніми дослідженнями, при відсутності чи малому інтересі з боку кінцевих користувачів, ступінь переконливості залежить більше від кількості фактів, які підтверджують інформацію, ніж від ступеня правдивості кожного з них [5].

Велика кількість каналів поширення забезпечує можливості донесення пропаганди до користувачів з різних країн, з джерел, які ніяк не асоціюються з країною її походження. До того ж, інформація, отримана з різних джерел виглядає більш правдивою для користувача.

Неконсистентність пропаганди йде всупереч класичним уявленням про ведення інформаційної війни [5], де кожне повідомлення має бути узгоджене з іншими і з загальною ідеологією. Проте, неузгодженість окремих каналів або й окремих повідомлень пропаганди має свою перевагу: отримання інформації начебто з різних точок зору підвищує довіру до джерела інформації.

Викривлення або фальсифікація реальних фактів – є одним з головних чинників, які зумовлюють ефективність такої пропаганди та складність їй протидіяти [5]. Оскільки при такому підході немає потреби, щоб фейкові повідомлення містили хоча б якусь частину правдивої інформації, то пропагандистська машина може миттєво реагувати на будь-які світові події, представляючи їх з вигідної сторони. А оскільки найперше повідомлення складає найбільше враження, то наступні повідомлення можуть бути просто пропущені користувачем. Окрім цього, з'являючись першим, таке повідомлення швидко поширюється користувачами соцмереж, таким чином, збільшуючи об'єм фейкової інформації про подію.

Для організації такого масштабного впливу необхідні великі інформаційні ресурси. У такому випадку це агенції новин в країнах ЄС та США, які працюють на державу-джерело пропаганди, фейкові сторінки в соцмережах, інформаційні веб-сторінки, чат-боти, друковані та телевізійні ЗМІ.

Кожен з цих ресурсів організує донесення пропаганди своїм каналом поширення інформації. І якщо агенції новин, теле-, радіо- та друкована продукція ще може підлягати контролю, то соцмережі та інтернет-ресурси в той час можуть безперешкодно поширювати фейкову інформацію.

Особливістю Інтернет-ресурсів є те, що вони можуть створюватися організаціями, ніяк не пов'язаними з державою-ініціатором пропаганди. Тому єдиним надійним джерелом є аналіз змісту цих ресурсів.

Проведене дослідження фокусувалося на соціальних мережах як каналі поширення неправди-

вих даних, оскільки вони мають велику аудиторію, здатні швидко поширювати фейкову інформацію та реакції користувачів і для них можна організувати ефективне розпізнавання пропаганди за допомогою сучасних алгоритмів. Соціальні мережі можуть використовуватися для декількох цілей – поширення фейкової інформації сторінками-ботами або реальними користувачами, дискредитація існуючих довірених джерел, нагнітання соціальної напруженості.

Для оперативного виявлення таких сторінок і фейкової інформації, яка ними поширюється, доцільно використати алгоритми розпізнавання фейкових новин, про які мова піде в наступному розділі.

Огляд існуючих підходів до розпізнавання фейкових новин.

Аналізуючи літературні джерела щодо проблематики виявлення фейкових новин на основі машинного навчання [6], можна класифікувати підходи до виявлення фейкових новин на дві великі групи: з попереднім навчанням та з самонавчанням. Алгоритми першої групи потребують навчання та перевірки на двох окремих множинах вхідних даних, які дозволяють точно підібрати вагові коефіцієнти і забезпечують високу ефективність кінцевої системи. Алгоритми з самонавчанням не потребують окремого етапу навчання для забезпечення результату. Вони застосовуються тоді, коли ручна класифікація вхідних даних для навчання є дуже трудомістким завданням, а також коли потрібно, щоб система могла сама підлаштовуватися при зміні умов реального середовища застосування.

На основі особливостей даних, які враховуються алгоритмами, їх можна поділити на:

- content-based: враховують текстову інформацію (тобто текст самої новини, поста в соцмережі чи твіту, описи профілів у соціальній мережі);
- social-based: враховують соціальну складову (особливості поширення поста в соцмережі та реакції користувачів на нього);
- combined: використовують обидва підходи.

Оскільки соціальні мережі відіграють одну з провідних ролей у поширенні неправдивої інформації, то дослідження фокусувалося на аналізі алгоритмів виявлення фейкових відомостей саме в соціальних мережах.

Прикладом підходу, який використовує текстову інформацію з поста і профілю автора є фреймворк FAKEDETECTOR [7]. Він базується на результатах дослідження, які показують, що між автором новини, її контентом та темою існує сильна кореляція. Цей фреймворк використовує hybrid feature learning unit (HFLU), що використовується для виявлення зовнішніх та внутрішніх особливостей конкретної новини. Далі FAKEDETECTOR використовує deep diffusive neural network для подальшої обробки векторів цих особливостей. Кожній з сутностей – автору, темі та контенту новини - присвоюється свій рейтинг довіри.

Для виділення зовнішніх та внутрішніх особливостей цей підхід використовує RNN (Recurrent Neural Network).

Отримані вектори особливостей потрапляють до deep diffusive neural network, яка на основі вхідних даних здійснює обчислення рівня довіри для кожної сутності.

Автори цього підходу пропонують навчати нейронну мережу методом зворотного поширення помилки, як найефективнішого методу для даного завдання.

Таким чином, FAKEDETECTOR – фреймворк з попереднім навчанням що здійснює визначення рівня довіри до самого інформаційного контенту, його автора та його теми.

Іншим підходом, який комбінує алгоритми для обробки як текстової інформацію з дописів так і інформації про реакції користувачів на ці дописи, є описаний в [6]. Цей метод враховує особливості поширення фейкових новин в соціальних мережах. Він пропонує застосування двох окремих методів для виявлення фейкового контенту. Перший метод ґрунтується на твердженні, що неправдиві новини можна виявляти за кількістю та особливостями реакцій користувачів на неї. У якості алгоритму, який враховує соціальні особливості поширення фейку автори пропонують використати підхід “harmonic boolean label crowdsourcing (HC) on social signals”, або HC-CB-3. Проте він є ефективним лише тоді, коли цей контент вже деякий час поширювався по соціальній мережі, і на нього відреагувала достатня кількість користувачів. Тільки що створений в соцмережі фейк на початку свого існування може не зібрати достатню кількість коментарів чи реакцій, або вони можуть бути відсутні взагалі, що унеможливлює застосування цього методу. У цьому випадку автори пропонують застосовувати алгоритм аналізу змісту дописів для виявлення особливостей, що характерні для фейкового інформаційного контенту.

Надзвичайно важливим для цього підходу є значення порогу вибору двох алгоритмів, який визначає, при якій кількості реакцій користувачів застосовувати перший, а при якій – другий метод.

Отже, особливістю цього підходу є комбіноване використання двох методів, які враховують як специфіку контенту новини, так і особливості її поширення в соціальній мережі.

Прикладом підходу, який би об’єднував в собі використання як текстової так і соціальної складової в одному алгоритмі, є алгоритм, описаний в [8] – unsupervised framework, або UFD. Він ґрунтується на врахуванні реакції користувачів на певну новину. Вважається, що коментуючи пост, користувач таким чином висловлює свою думку про нього (правдива, на його думку, ця інформація, чи ні), а отже, ці дані можна використовувати в якості фактора при розпізнаванні фейку. Згідно з цим підходом, всі користувачі діляться на дві групи - довірені та недовірені, в залежності від точності їхньої оцінки конкретних дописів, тобто від кількості збігів між їхнім припущенням в реакції і справжньою оцінкою правдивості новини. У залежності від конкретної групи, до якої належить користувач, при аналізі інформаційного контенту його реакція враховується з певним коефіцієнтом.

У якості основного алгоритму автори використовують Collapsed Gibbs Sampling, update правило якого вираховується на основі кількості реакцій ко-

ристувачів з довіреної та недовіреної груп. Це дозволяє алгоритму підлаштовуватися під змінні умови застосування, проте накладає певні обмеження на наявність реакцій на дописи.

Визначення найточнішого підходу для ефективного виявлення неправдивої інформації, розповсюджуваної в соціальних мережах в рамках ведення інформаційної пропаганди.

Головним завданням даного дослідження був аналіз особливостей інформаційної пропаганди, зокрема, поширення неправдивої інформації через соціальні мережі; аналіз сучасних підходів для виявлення фейкових новин в соцмережах та їхньої ефективності; визначення найточнішого підходу для ефективного виявлення неправдивої інформації, розповсюджуваної в соціальних мережах в рамках ведення інформаційної пропаганди.

Можна виокремити ряд особливостей, характерних для цього інформаційного каналу:

- фейкові новини поширюються в соцмережах у формі дописів користувачів;
- користувачі можуть вказувати посилання на новину на певному Інтернет-ресурсі;
- варто зазначити, що користувачі можуть бути як агентами інформаційного впливу, які свідомо створюють та поширюють фейки, так і несвідомими поширювачами неправдивих відомостей.

Цілі, переслідувані пропагандою, можна в загальному поділити на два напрямки:

- формування думки в соціумі стосовно певної події чи процесу – тоді на перший план виходить саме кількість поширень новини;
- загострення напруги в соціумі – тоді важливою стає також полеміка в реакціях користувачів на цю новину.

У загальному цілями для активної пропаганди є відомі і популярні суспільно значимі питання, або останні події в країні, як-от: мови національних меншин в державі або висловлювання лідерів політичних партій.

Отже, для ефективного виявлення таких фейків необхідний комплексний підхід. Сформуємо критерії, яким він повинен відповідати:

- виявляти неправдиві новини на найбільш ранньому етапі, при невеликій кількості реакцій користувачів, чи їх повній відсутності;
- забезпечувати високий коефіцієнт виявлення;
- враховувати як особливості контенту, так і реакції користувачів на дописи;
- оскільки інформаційна пропаганда є цілеспрямованим процесом, підхід повинен враховувати можливі зміни парадигми впливу, а також те, що частина користувачів соцмережі може бути так званими ботами або агентами інформаційного впливу.

У рамках дослідження було проаналізовано ефективність вищезгаданих підходів та алгоритмів для розпізнавання фейкових новин.

Їхнє навчання базувалося на декількох популярний базах даних фейкового контенту: BuzzFeed, PolitiFact та LIAR [9]. З точки зору протидії інформаційній пропаганді найбільш оптимальною є база даних BuzzFeed. Вона містить достатньо різноманітні політичні факти, джерелом яких можуть бути не лише політичні діячі, партії чи організації, а саме в цьому і полягає особливість інформаційного впливу – фейкові новини можуть приписуватися як конкретним людям, так і збірному поняттю "експерти" чи взагалі не мати автора. Алгоритми HC-CB-3 і UFD будуть порівнюватися за базою даних BuzzFeed, а алгоритм FAKEDETECTOR – за даними PolitiFact.

У якості метрик застосовувалися [10]:

- достовірність (accuracy) - відношення кількості правильних передбачень до загальної кількості передбачень;
- точність (precision) - відношення правильно передбачених позитивних спостережень до загальної кількості передбачуваних позитивних спостережень;
- чутливість (recall) - відношення правильно передбачених позитивних спостережень до всіх спостережень у реальному класі;
- критерій F1 - зважене середнє між точністю і чутливістю.

Автори цих алгоритмів навели дані щодо ефективності інших популярних алгоритмів та фреймворків для порівняння, зокрема:

- TRIFN [11]: використовує концепції користувача, новини та публікатора для виявлення фейкових новин;
- DEEPWALK [12] – модель вбудованої мережі (network embedding model). Базуючись на структурі мережі фейкових новин, DEEPWALK вбудовує статті, їхніх авторів та заголовки в латентний простір ознак (latent feature space);
- RST [13] – базується лише на контенті та використовує SVM класифікатор;
- LIWC [14] – базується лише на контенті та використовує психолінгвістичні категорії;
- Castillo [15] – базується лише на соціальних ознаках та враховує профілі користувачів і мережу друзів;
- RST+Castillo [15] – комбінує підходи контенту та соціальних ознак.

Проте, через порівняно невисоку ефективність, вони не розглядалися в дослідженні. Для порівняння ефективності трьох алгоритмів, були виокремлені необхідні дані з досліджень Fakedetector, HC-CB-3 та UFD. Зокрема:

- для алгоритму Fakedetector бралися результати класифікації за категорією контенту, оскільки для протидії інформаційній пропаганді найбільш важливим є спершу визначити правдивість саме контенту новини;
- для алгоритму HC-CB-3 в якості кінцевих результатів бралося усереднене значення ефективності для груп завідомо правдивих та завідомо фейкових груп новин.

Результати ефективності алгоритмів Fakedetector, HC-CB-3 та UFD наведені на рис. 1. Як можна побачити, найбільш ефективним є алгоритм HC-CB-3. Він переважає алгоритми UFD та Fakedetector згідно з усіх застосованих метрик. Окрім того, HC-CB-3 імплементований авторами у вигляді чатботу для соціальної мережі Facebook і вже протестований на реальних даних. Отже, підхід з комплексним поєднанням особливостей контенту та соціаль-

ної взаємодії може бути ефективним для виявлення інформаційної пропаганди в соціальних мережах. Він задовольняє всім критеріям, які автори визначили вище: забезпечує виявлення неправдивої інформації в дописах на ранніх етапах, має високий коефіцієнт виявлення фейків, а також враховує як особливості контенту допису, так і соціальні реакції користувачів на нього.

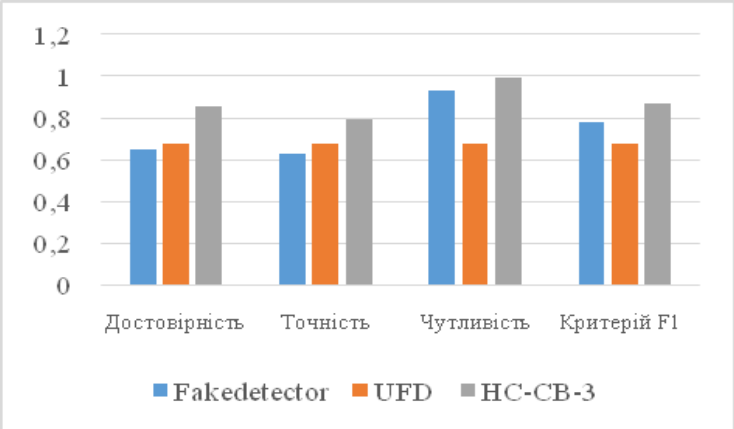

Рис. 1. Результати ефективності алгоритмів Fakedetector, HC-CB-3 та UFD

Проте, він вимагає тренувальної вибірки для навчання, тому автори вважають, що цей підхід можливо вдосконалити для роботи саме в умовах інформаційного впливу. Для цього вони пропонують динамічно збирати дані стосовно авторів вже проаналізованих дописів та зберігати їх в окрему базу даних. Це дозволить побудувати систему оцінки автору новини і в майбутньому додати в алгоритм можливість врахування цієї оцінки при формуванні остаточного рішення. Таке вдосконалення може зробити HC-CB-3 більш ефективним в умовах інформаційної пропаганди, оскільки багато фейкових новин поширюється спеціально створеними акаунтами-ботами.

Висновки

В ході дослідження були проаналізовані особливості інформаційної пропаганди з деяких країн пострадянського простору; розглянуті сучасні підходи до розпізнавання фейкових новин у соціальних мережах та їхня ефективність; обраний найбільш перспективний підхід для ефективного виявлення неправдивої інформації, розповсюджуваної в соціальних мережах у рамках ведення інформаційної пропаганди, який полягає в комбінованому використанні двох алгоритмів, які враховують різні сторони інформаційного контенту та особливостей його поширення в соціальних мережах - HC-CB-3. Його ефективність на базі даних фейкових новин BuzzFeed є вищою за ефективність розглянутих алгоритмів UFD та Fakedetector за метриками accuracy, precision, recall та F1. Проте, автори вважають, що роботу алгоритму в умовах інформаційної пропаганди можна покращити, якщо сформувати базу даних з оцінками довіри до авторів дописів і враховувати її при прийнятті рішень щодо нових дописів.

Література

[1] D. Katsaros, G. Stavropoulos, D. Papakostas, "Which machine learning paradigm for fake news detection?," *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Thessaloniki, Greece, 2019, pp. 383-387.

[2] М.М. Браіловський, І.С. Іванченко, І.Р. Опірський, В.О. Хорошко. Інформаційно-психологічне протиборство в Україні// *Науковий журнал «Безпека інформації»*, 2019 , том 25 №3. – С. 144-149.

[3] H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 2017. – pp. 211-236.

[4] Daniel Boffey. EU raises funds to fight 'disinformation war' with Russia. *The Guardian*, [Електронний ресурс]. Режим доступу: https://www.theguardian.com/profile/daniel-boffey? page=40.

[5] Paul, Christopher and Miriam Matthews, The Russian "Firehose of Falsehood" Propaganda Model: Why It Might Work and Options to Counter It, Santa Monica, Calif.: *RAND Corporation*, PE-198-OSD, 2016. As of September 29, 2020 [Електронний ресурс]. Режим доступу: https://www.rand.org/pubs/perspectives/PE198.html.

[6] M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," *22nd Conference of Open Innovations Association (FRUCT)*, Jyvaskyla, 2018. - pp. 272-279.

[7] J. Zhang, B. Dong and P. S. Yu, "Fakedetector: Effective Fake News Detection with Deep Diffusive Neural Network," *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, Dallas, TX, USA, 2020, pp. 1826-1829.

[8] Yang, Shuo & Shu, Kai & Wang, Suhang & Gu, Renjie & Wu, Fan & Liu, Huan. (2019). Unsupervised Fake News Detection on Social Media: A Generative Approach. *Proceedings of the AAAI Conference on Artificial Intelligence*. 33, 2019. – pp. 5644-5651.

[9] Yingzhao Ouyang. Identifying fake news: The LIAR dataset and its limitations. *Towards Data Science, June 29*, 2020. [Електронний ресурс]. Режим доступу: https://towardsdatascience.com/identifying-fake-news-the-liar-dataset-713eca8af6ac.

[10] Jason Brownlee. *How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification - Machine Learning Mastery*, January 3, 2020. [Електрон-

ний ресурс]. Режим доступу: https:// machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/.

[11] K. Shu, S. Wang, and H. Liu. *Exploiting tri-relationship for fake news detection*. CoRR, abs/1712.07709, 2017 [Електронний ресурс]. Режим доступу: https:// export. arxiv.org/pdf/1712.07709.

[12] B. Perozzi, R. Al-Rfou, and S. Skiena. *Deepwalk: Online learning of social representations*. In KDD, 2014.

[13] V. L. Rubin, N. J. Conroy, Yimin Chen, *"Towards News Verification: Deception Detection Methods for News Discourse,"* 2015.

[14] J. W. Pennebaker, R. L. Boyd, K. Jordan, K. Blackburn, *"The Development and Psycho-metric Properties of LIWC2015,"* 2015.

[15] C. Castillo, M. Mendoza, B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th International Conference on World Wide Web*. ACM, 2011, pp. 675–684.

УДК 654.071

Shtefaniuk Y.F., Opirskyy I.R., Harasymchuk O.I. Analysis of application of existing fake news recognition techniques to counter information propaganda

Abstract. *The problem of detecting false (fake) information transmitted through various channels on the Internet is becoming increasingly important. One type of such information is targeted propaganda, which has a specific purpose and uses specially created resources. To combat such informational influence, one can use already developed tools to detect fake news. This article considers the features of information propaganda and approaches to combating it; the effectiveness of several well-known techniques for recognizing fake news; the analysis of possible efficiency of these techniques in the context of possibility of their application for counteraction to purposeful information influences was carried out. Based on the study, the most promising algorithm for recognizing information propaganda in social networks was selected.*

Key words: *fake news, neural networks, information influence counteraction, information propaganda, Fakedetector, UDF, HC-CB-3.*

Штефанюк Е.Ф., Опирский И.Р., Гарасимчук О.И. Анализ применения существующих техник распознавания фейковых новостей для противодействия информационной пропаганде

Аннотация. *Проблема выявления ложной (фейковой) информации, передаваемой через различные каналы в сети Интернет, становится все более актуальной. Одной из разновидностей такой информации является целевая пропаганда, которая имеет конкретную цель и использует специально созданные ресурсы. Для борьбы с таким информационным воздействием можно использовать уже разработанные средства обнаружения фейковых новостей. В данной статье рассмотрены особенности информационной пропаганды и подходов к борьбе с ней; эффективность работы нескольких известных техник распознавания фейковых новостей; проведен анализ возможной эффективности этих техник в контексте их применения для противодействия целенаправленным информационным воздействиям. На основании проведенного исследования выбран наиболее перспективный метод для распознавания информационной пропаганды в социальных сетях.*

Ключевые слова: *фейковые новости, нейронные сети, противодействие информационному воздействию, пропаганда, Fakedetector, UDF, HC-CB-3.*

Shtefaniuk Yevheniy Fedorovych, post-graduate student of the Department of Information Protection of the National University "Lviv Polytechnic".
Штефанюк Євгеній Федорович, аспірант кафедри захисту інформації Національного університету «Львівська політехніка».
Штефанюк Евгений Федорович, аспирант кафедры защиты информации Национального университета «Львовская политехника».

Opirskyy Ivan Romanovych, Dc.S, Associate Professor, Professor of the Department of Information Protection of the National University "Lviv Polytechnic".
Опірський Іван Романович, доктор технічних наук, доцент, професор кафедри захисту інформації Національного університету «Львівська політехніка».
Опирский Иван Романович, доктор технических наук, доцент, профессор кафедры защиты информации Национального университета «Львовская политехника».

Harasymchuk Oleh Igorovych, Ph.D., Associate Professor, Associate Professor of Information Protection, National University "Lviv Polytechnic".
Гарасимчук Олег Ігорович, кандидат технічних наук, доцент, доцент кафедри захисту інформації Національного університету «Львівська політехніка».
Гарасимчук Олег Игоревич, кандидат технических наук, доцент, доцент кафедры защиты информации Национального университета «Львовская политехника».

Отримано *19 листопада 2020 року*, затверджено редколегією *15 грудня 2020 року*