

УДОСКОНАЛЕНИЙ МЕТОД ВИЯВЛЕННЯ НЕПРАВДИВОЇ ІНФОРМАЦІЇ

За умов швидкого розвитку глобального інформаційного суспільства, широкого використання інформаційно-комунікаційних технологій у всіх сферах життя особливе значення набуває проблема інформаційної безпеки. Технології - двосічний меч у цій битві - також мають потенціал для вирішення проблеми. Алгоритми перевірки фактів, хоч і недосконалі, можуть діяти як вартіві, позначаючи підозрілий контент. Прозорість і співпраця між платформами, дослідниками та незалежними фактчекерами мають вирішальне значення для вдосконалення цих інструментів і забезпечення відповідального застосування. Однак самі по собі технології не можуть виграти цю війну. Поле бою виходить за межі алгоритмів і коду, сягаючи самої тканини наших суспільств. Сучасні методи виявлення неправдивої інформації враховують лише зміст джерела даних, тобто самих даних, не звертаючи уваги на контекст даних. Беручи до уваги, що дані стають складнішими, їх стає все більше та більше, стає важливим розробка та удосконалення алгоритмів та методів виявлення неправдивої інформації з урахуванням сучасних особливостей розповсюдження інформації. Тому метою роботи є розробка удосконаленого методу виявлення неправдивої інформації на основі синергії запропонованих додаткових методів. Удосконалення методу полягає у синергії п'яти складових. Кожна з котрих має можливість виявляти неправдиву інформацію з де якою ймовірністю. Саме синергія та комбінаторика цих методів дає удосконалення загального методу та виявляти неправдиву інформацію з більший ефективністю ніж існуючі методи.

Ключові слова: неправдива інформація, метод, синергія, розповсюдження, припинення, алгоритм.

Вступ

Інформація – це джерело влади. Доступ до інформації – фундамент соціальності, основа громадянського суспільства. Це виявляється у володінні світовим контентом, свободі знати, оцінювати, голосувати, керувати, приймати рішення. Сучасна людина отримала більш широкі права й реальну можливість заволодіти світовою увагою, втрутитися у світові сценарії, оцінювати їх і транслювати світові свою оцінку того, що відбувається, створювати власні версії, діяти і закликати до соціальної дії. Вплив інформації на соціум, виявлення та блокування неправдивої інформації є складовою забезпечення інформаційної безпеки держави.

Основна моральна проблема інформаційного суспільства полягає в тому, що комунікація перестала бути справжньою. Інтенсивність інформаційних потоків, швидка зміна ціннісних і ідеологічних пріоритетів, ставка на фактичність і сенсаційність, байдужість до духовних цінностей призводять до того, що комунікація стає формальною і вихолощеною, позбавленою людського начала, а інформація губить правдивість.

Сучасний світ переповнений інформацією. Ця оцінка правдивості інформації потрібно враховувати багато факторів, це є неможливим без обробки великих як структурованих так і не структурованих даних, з різною точністю та повнотою і звичайно з наявними помилками. Виконання прогностичного моделювання, наприклад виявлення аномалій, у Big Data є складним завданням.

Сприймання медійних продуктів є невіддільною частиною повсякденного життя сучасних людей, яка за кількістю витраченого на неї часу перебуває на другому місці серед усіх видів активності, поступаючись лише праці. Недостатній контроль з боку держави за дотриманням законів України політичними силами, ЗМІ та окремими особами, які займаються підприємницькою діяльністю в інформаційній сфері, призводить до того, що нині трапляються непоодинокі випадки надання ефірного часу телепрограмам та радіопрограмам, спрямованим на руйнування моральних цінностей, свідомості української нації, підривання морального і фізичного здоров'я громадян. Певна залежність ЗМІ від держави, контроль з боку фінансових чи політичних груп перетворюють вітчизняні ЗМІ на знаряддя маніпуляції суспільною свідомістю, провідників певної ідеології і, що є найнебезпечнішим, часто сприяють упровадженню чужих, не властивих суспільству духовно-моральних і політичних цінностей, що руйнує духовний фундамент її існування. Час від часу відвертаючи увагу споживачів від реальності, ЗМІ створюють для них специфічний інформаційний світ, формують певні

ціннісно- смислові моделі для засвоєння суспільством і таким чином змінюють аксіологічну картину світу соціуму.

Достовірність інформації життєво необхідна здоровій демократії. Хибна чи неточна інформація може негативно вплинути на громадське обговорення питань та політичні рішення громадян призводячи до порушення коректної політичної дискусії та перешкоджаючи досягненню угод. Можливість поінформованого та поважного обговорення громадянами політичних ідей та суспільно-політичних питань є ключовим аспектом збереження демократії у довгостроковій перспективі. Це повною мірою відноситься також до діалогу всередині уряду та між політиками.

Аналогічним чином, громадяни повинні розуміти роботу уряду та мати у своєму розпорядженні необхідну інформацію, щоб урядовці несли відповідальність перед ними за рішення. Поширення помилкової інформації, що вводить в оману та її використання для підриву довіри суспільства, посилення розколу та обмеження можливості громадян діяти за окремо чи спільно можна розглядати як загрозу для демократії. Дезінформація може мати особливу руйнівну силу в період виборів, за наявності значних, що укорінилися протиріч у пріоритетах та політичних принципах. У такі періоди дезінформація може маніпулювати перевагами виборців, порушувати нормальний перебіг виборного процесу, підживлювати невдоволення та розчарування громадськості. Зрозуміло, не кожна спроба впровадження дезінформації пов'язано з особливою подією, такою як вибори. Дезінформація також може застосовуватись для зміни загального інформаційного поля, в якому люди обговорюють питання, формують погляди та приймають політичні рішення. У деяких випадках метою дезінформації є поступове формування ширшої інформаційної картини чи перешкоджання суспільної дискусії за рахунок розбіжностей чи цинізму.

Зрозуміло, що маніпуляція інформацією не є чимось новим для демократичних суспільств, але цифрові технології збільшили масштаб проблеми через те, що зловмисні особи отримали можливість анонімно впливати на громадську думку та загрожувати достовірності інформації. Соціальні мережі посилюють наслідки через відносно низьку вартості та високої швидкості донесення інформації до великої аудиторії. Процесу часто сприяють автоматизовані системи, наприклад боти, просувають матеріали користувачам згідно з даними про їх демографію та особисті переваги. Фальшиві новини - це термін, який використовують взаємозаміно з терміном «дезінформація» та, іншими словами, що позначають інші порушення інформаційної екосистеми. Зараз його застосовують у загальному сенсі щодо неточних чи сфабрикованих новин. В той же час термін «фальшиві новини» недостатньо точно відображає складність дезінформації, недостовірної інформації та шкідливої інформації. Його часто використовують авторитарні та інші політичні діячі у тому, щоб девальвувати неугодні факти, вплітаючи у яких хибні установки.

Ключовою відмінністю дезінформації від повідомлення недостовірної інформації є намір. Мотиви, якими керуються дійові особи при розробці, створенні та розповсюдженні дезінформації, дозволяють вивчити це ще глибше. Мотиви поділяються на чотири категорії: фінансові, політичні, соціальні та психологічні. Дезінформацію можуть використовувати з метою маніпулювання думкою чи поглядами цільової аудиторії як державні, так та недержавні політичні діячі. Політики можуть поширювати дезінформацію про установи чи політичні супротивники як усередині своєї країни, так і за кордоном з метою заглушити їх голоси і відвести дискурс у потрібне русло. Подібні політичні діячі можуть як мати зв'язок з урядами, так і діяти самостійно та координувати свої дії з іншими особами на підтримку загальної ідеологічної концепції. Інші діячі, які займаються дезінформацією, можуть керуватися неполітичними мотивами, наприклад, бажанням розважитися або збільшити прибуток. На сьогоднішній день реклама в Інтернеті виступає фінансовим стимулом для дезінформації, здатної швидко поширюватися та залучати відвідувачів на певний веб-сайт. Корпоративні та незалежні діячі, які прагнуть збільшити прибуток за рахунок нарощування відвідуваності ресурсу можуть маніпулювати внутрішніми механізмами (алгоритмами) соціальних мереж,

призначеними для надання інформації, а також самої інформацією. Оскільки в соціальних мережах розваги та новини співіснують у тісному сусідстві один з одним, введення в оману споживачів у мережі може бути побічним результатом основний мети — отримання прибутку.

Таким чином виявлення неправдивої інформації відіграє величезну роль у інформаційній безпеці держави. Тому розробка та удосконалення науково-методичного апарату у напрямку виявлення та блокування неправдивої інформації є актуальним науковим завданням.

Постановка проблеми

Сучасні методи виявлення неправдивої інформації враховують лише зміст джерела даних, тобто самих даних, не звертаючи уваги на контекст даних. Беручи до уваги, що дані стають складнішими, їх стає все більше та більше, стає важливим розробка та удосконалення алгоритмів та методів виявлення неправдивої інформації з урахуванням сучасних особливостей розповсюдження інформації. Тому метою роботи є розробка удосконаленого методу виявлення неправдивої інформації на основі синергії запропонованих додаткових методів.

Аналіз останніх досліджень і публікацій

Серед пропагандистським прийомів по просуванню неправдивої інформації, виокремлених науковцями, найпоширенішими є спрощення, перебільшення, навіювання страху, культ особи, демонізація іншої групи, деморалізація, дезінформація, навішування ярликів, фальшиве звинувачення, «розмахування прапорами» (flag-waving), фреймування, заперечення очевидного, використання протиріч, відверта брехня і напівправа, обзивання, «надузагальнення», перенос відповідальності з інверсією ролей «жертва-переслідувач», «жертва-злочинець», «агресор – жертва агресії», апелювання до «очевидців», «перетасовування карт» (Card Stacking») як непропорційна кількість інформації під виглядом збалансованої аргументації та багато інших. Проблеми виявлення та блокування неправдивої інформації присвячено багато наукових робіт. Так у роботі [1] розглянуто метод словотвірного аналізу і метод, базований на теорії концептуальної інтеграції. Словотвірний аналіз залучався для виявлення способів і засобів словотвору українських мемів-неологізмів і визначення найбільш продуктивних словотвірних моделей. Але узагальнене рішення не запропоновано.

У роботах [2–5] використовуються елементи методу аналізу когнітивної метафори, що ґрунтується на моделі чотирьох просторів – вихідних джерельного і цільового просторів, просторів спільних ознак і бленду, в якому відбувається змішування всіх просторів на основі процесів композиції – встановлення зв'язків між вихідними просторами, завершення як узгодження змішаної структури з базовими знаннями та розробки як подальшого розвитку «сценарію» бленду. Але узагальнення у роботі не приведено, наукового результату не отримано.

У роботах [6–9] загальні положення виявлення аномалій. Надається поняття аномалій. Аномалія – це виявлення аномальних подій або моделей, які не відповідають очікувані події або закономірності. Виявлення аномалій важливо в широкому діапазоні різнорідних сфер, таких як діагностика медичних проблем, банківське та страхове шахрайство, вторгнення в мережу та інше. Нажаль застосування методології для виявлення неправдивої інформації не розглядається.

З проведеного аналізу наукової літератури можливо зробити висновок, що універсальних пристроїв або автоматизованих програмних комплексів для забезпечення виявлення неправдивої інформації не існує. Тому тема розробки нових та удосконалення існуючих методів виявлення неправдивої інформації є актуальною і дуже важливою.

Виклад основного матеріалу

Інформаційна безпека відіграє важливу роль у забезпеченні життєво важливих інтересів будь-якої держави. Створення розвиненого та захищеного інформаційного середовища є обов'язковою умовою розвитку суспільства та держави. Інформаційна безпека держави досягається шляхом балансу між інформаційними правами та свободами різноманітних суб'єктів права та захистом національного інформаційного суверенітету. Питання інформаційної безпеки та національної безпеки держави взагалі – питання балансу між правами та інтересами людини та компетенцією та інтересами державної влади, балансу, який

можна встановити лише з допомогою правових норм. Усі країни зацікавлені у розвитку глобального інформаційного суспільства та використанні нових можливостей, які відкриваються завдяки покращенню доступу до інформації та кращому забезпеченню інформацією. Інформаційна безпека розглядається як глобальна проблема захисту інформації, інформаційного простору та інформаційного суверенітету.

Сучасні методи виявлення неправдивої інформації враховують лише зміст джерела даних, тобто самих даних, не звертаючи уваги на контекст даних. Беручи до уваги, що дані стають складнішими, їх стає все більше та більше, стає важливим розробка методів виявлення аномалій упередженості для контексту, будь то просторовий, часовий, або семантичні методи. Саме тому стає необхідним ретельне дослідження теоретичних та практичних проблем інформаційної безпеки у сучасному глобалізованому світі.

Причини поширення неправдивої інформації

Розглянемо причини поширення неправдивої інформації, з яких складні та можливо виділити кілька ключових:

Індивідуальні фактори:

1) *Когнітивні упередження*: Люди схильні до когнітивних упереджень, таких як упередження підтвердження, коли ми шукаємо інформацію, що підтверджує наші існуючі переконання, та ефект Даннінга-Крюгера, коли ми переоцінюємо наші знання. Упередження підтвердження може зробити людей більш сприйнятливими до дезінформації, яка відповідає їхнім упередженням уявленням, що призводить до того, що вони приймають і поширюють неправдиву інформацію, не оцінюючи її достовірність критично.

2) *Емоційна привабливість*: Неправдива інформація часто використовує емоційну мову і викликає страх, гнів або обурення, що підвищує ймовірність привернути увагу і поширити її.

3) *Обмежена цифрова грамотність*: Відсутність навичок оцінювання джерел інформації та виявлення "червоних прапорців" може зробити людей більш вразливими до віри та поширення неправдивої інформації.

Фактори інформаційного середовища:

1) *Алгоритми соціальних мереж*: Платформи надають пріоритет залученню, що може посилювати сенсаційний і суперечливий контент, навіть якщо він є неправдивим.

2) *Ехо-камери*: Бульбашки фільтрів, створені алгоритмами та персоналізованими стрічками новин, можуть показувати людям насамперед інформацію, яка підтверджує їхні наявні переконання, що зменшує ймовірність зіткнутися з суперечливою інформацією.

3) *Економічні стимули*: Деякі суб'єкти створюють і поширюють неправдиву інформацію заради фінансової вигоди, наприклад, за допомогою клікбейтів або доходів від реклами.

4) *Брак медіаграмотності*: Багатьом людям бракує необхідних навичок, щоб критично оцінювати інформацію та відрізнити достовірні джерела від ненадійних. Без належної медіаграмотності люди можуть бути більш вразливими до дезінформації та ненавмисного поширення неправдивої інформації серед інших.

5) *Швидкість і вірусність*: В епоху миттєвої комунікації та соціальних мереж неправдива інформація може поширюватися швидко і широко, охоплюючи мільйони людей протягом декількох хвилин або годин. Швидкість і вірусність дезінформації можуть ускладнити виправлення неправдивих наративів після того, як вони вкоренилися, посилюючи їхній вплив і увічнюючи їхнє поширення.

Суспільні фактори:

1) *Політична поляризація*: Посилення політичного поділу може створити середовище, в якому люди більш схильні довіряти інформації, яка відповідає їхнім політичним поглядам, незалежно від її достовірності.

2) *Недовіра до традиційних інститутів*: Якщо люди втрачають довіру до усталених джерел інформації, таких як уряди та ЗМІ, вони можуть з більшою ймовірністю звертатися до альтернативних джерел, навіть якщо вони ненадійні.

3) *Великі події та кризи*: Періоди невизначеності та високих емоцій, такі як вибори або стихійні лиха, можуть створити сприятливий ґрунт для поширення неправдивої інформації.

Протидія поширенню неправдивої інформації вимагає узгоджених зусиль багатьох зацікавлених сторін, зокрема технологічних компаній, урядів, освітян, журналістів та організацій громадянського суспільства. Впроваджуючи стратегії, спрямовані на підвищення медіаграмотності, покращення модерації контенту, підвищення прозорості та розвиток навичок критичного мислення, суспільство може працювати над пом'якшенням впливу дезінформації та побудовою більш стійкої інформаційної екосистеми.

Удосконалений метод виявлення неправдивої інформації

Виявлення неправдивої інформації вимагає поєднання технологічних інструментів, навичок критичного мислення та методів аналізу. Для пояснення удосконалення методу розглянемо більш ретельно його складові. А саме з яких додаткових технік, алгоритмів та методів він складається. Розглянемо саме ті, які будемо використовувати у нашому методі:

Індивідуальні техніки:

1) *Критичне мислення*: Це основа виявлення неправдивої інформації. Треба задавати собі такі питання, як: Хто є джерелом? Яка його мета? Чи не звучить це занадто добре, щоб бути правдою? Чи узгоджується інформація з іншими, надійними джерелами?

2) *Перевірка фактів*: Не треба покладатися лише на саму інформацію. Веб-сайти для перевірки фактів, такі як Snopes, PolitiFact або FactCheck.org, тощо підходять для перевірки твердження та джерела.

3) *Оцінка джерела*: пошук авторитетних джерел з визнаним досвідом у даній темі. Перевірка наявності ознак упередженості, кваліфікації автора, а також прозорості фінансування чи афілійованості.

4) *Зворотний пошук зображень*: Це може допомогти розвінчати сфабриковані або маніпульовані візуальні матеріали. Такі інструменти, як Google Lens, TinEye, тощо можуть допомогти знайти першоджерело зображення.

5) *Перехресні посилання*: Не треба покладатися лише на одне джерело. Підтвердження інформації з інших достовірних джерел, щоб перевірити, чи відповідає вона дійсності; пошук невідповідності або пропущеної інформації.

Алгоритми з технологічною підтримкою:

1) *Алгоритми перевірки фактів*: Платформи соціальних мереж використовують алгоритми для позначення підозрілого контенту та з'єднання користувачів з ресурсами для перевірки фактів. Однак ці алгоритми не є досконалими і потребують людського нагляду.

2) *Сторонні фактчекери*: Платформи співпрацюють з незалежними фактчекерами, які аналізують контент і надають рейтинги або позначки. Слід шукати значки або індикатори відомих організацій, що займаються перевіркою фактів.

3) *Інструменти для виявлення депфейків*: Оскільки "глибокі фейки" стають все більш витонченими, розробляються інструменти для їх виявлення. Ці інструменти аналізують відео- та аудіоматеріали на наявність ознак маніпуляцій.

4) *Краудсорсингові платформи*: Такі платформи, як Вікіпедія, використовують колективні знання користувачів для виявлення та усунення неточностей. Слід остерігатися платформ із низькими редакційними стандартами або анонімними дописувачами.

Додатковий розгляд:

1) *Усвідомлення власних упереджень*: Кожен має упередження, які можуть впливати на те, як сприймається інформація. Варто пам'ятати про власні упередження і намагатися підходити до інформації об'єктивно.

2) *Пошук червоні прапорці*: Емоційна мова, сенсаційні заголовки та твердження, які здаються занадто хорошими, щоб бути правдою, часто є ознаками дезінформації.

3) *Слід бути скептичним, але не зневажливим*: Не треба одразу відкидати інформацію, яка кидає виклик вашим переконанням. Однак слід підходити до неї зі здоровою дозою скептицизму і ретельно дослідити її, перш ніж прийняти.

4) *Повідомляйте про дезінформацію*: Якщо виявилось, що поширюється неправдива інформація, треба повідомити про це платформу, використовуючи її механізми повідомлення. Це допоможе платформі виявити та видалити шкідливий контент.

5) *Використання штучного інтелекту та машинного навчання*: Вдосконалені алгоритми та моделі машинного навчання можуть аналізувати величезні обсяги даних для виявлення закономірностей, аномалій та ознак дезінформації. Методи обробки природної мови можуть оцінювати достовірність текстового контенту на основі лінгвістичних сигналів і контексту.

6) *Залишення поінформованими та критичними*: Оволодіння навичками медіаграмотності та збереження здорового скептицизму щодо інформації, яка зустрічається в Інтернеті, є важливими для навігації в сучасному інформаційному ландшафті. Критичне мислення, скептицизм і готовність перевіряти інформацію можуть допомогти людям виявляти та уникали неправдивої інформації.

7) *Звернення до експертної думки*: Звернення до експертів або професіоналів у відповідних галузях може дати цінну інформацію про складні або технічні теми. Експерти можуть запропонувати нюансовані перспективи та виявити неточності або хибні уявлення.

На рис.1. наведено блок схема удосконаленого методу виявлення неправдивої інформації.

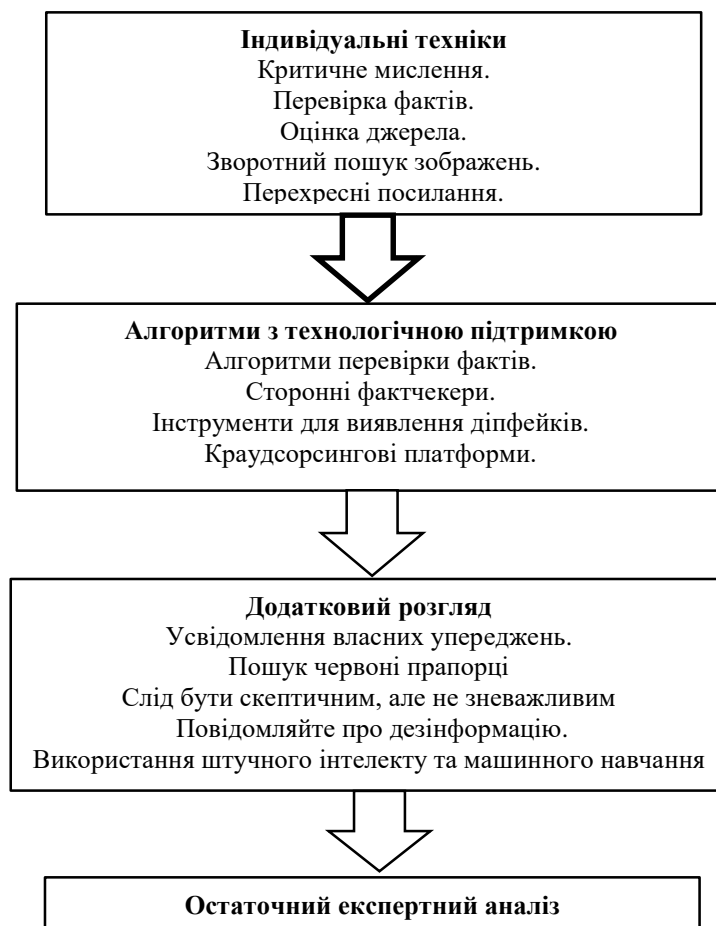


Рис.1. Блок схема удосконаленого методу виявлення неправдивої інформації

Треба відмітити, що жодний крок не є надійним, тільки синергія запропонованих складових розробленого методу дає змогу виявляти неправдиву інформації з необхідної достовірністю. Саме проведене удосконалення полягає у синергії та комбінаторики запропонованих складових. Такий варіант значно покращить можливості виявлення неправдивої інформації у Big Data – світі інформації.

Висновок

Проведені дослідження показали, що в сучасний період військових дій з росією боротьба з неправдивою інформацією вимагає не просто захисту, а багатогранного арсеналу захисту. Триває інформаційна війна. Емоційна привабливість сфабрикованих наративів сприяє поширенню неправдивої інформації і відповідь полягає не в цензурі, а в розбудові суспільства, здатного володіти інструментами для розпізнавання правди. Існуючі алгоритми перевірки фактів, хоч і недосконалі, можуть діяти як вартові, позначаючи підозрілий контент. Розроблений удосконалений метод виявлення неправдивої інформації є черговим кроком у виявленні та припиненні розповсюдження ворожої пропаганди. Він відрізняється від існуючих у синергії та комбінаторики запропонованих складових виявлення неправдивої інформації, що дає змогу застосовувати його для аналізу будь якої інформації з високою ефективністю.

Перелік посилань

1. Лаптева, Т. О., Лукова-Чуйко, Н. В. Удосконалення методу виявлення неправдивої інформації на основі методу експертної оцінки «Дельфі». Наукоємні технології. Том 55 № 3 (2022) стр.193 – 199. DOI: 10.18372/2310-5461.55.16901.
 2. Гнатієнко, Г. М., Снитюк, В. Є. Експертні технології прийняття рішень. – Київ: McLaut, 2008. – 444 с.
 3. Schefer-Wenzl, S., Strembeck, M. Modeling support for role-based delegation in process-aware information systems. *Business and Information Systems Engineering*. 6 (4). 2014. pp. 215-237. DOI: 10.1007/s12599-014-0343-3
 4. Лаптева, Т. Алгоритм визначення міри існування недостовірної інформації в умовах інформаційного протистояння. *Кибербезпека: освіта, наука, техніка*. No 2 (14), 2021, с. 15-25. DOI 10.28925/2663-4023.2021.14.1525, ISSN 2663-4023.
 5. Лукова-Чуйко, Н., Лаптева, Т. Виділення та відбір ознак для визначення неправдивої інформації. V Міжнародна науково-практична конференція. “Проблеми кібербезпеки інформаційно-телекомунікаційних систем” (PCSITS) 27-28 жовтня 2022 р. Київ, Україна. Збірник матеріалів доповідей та тез. С 13-15.
 6. Лаптева, Т. О. Методика виявлення неправдивої інформації для безпеки Держави. Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «Об’єднати наукою: перспективи міждисциплінарних досліджень» 23-24 листопада 2023 р. Київ, Україна. С.131–132.
 7. Yevseiev, S., Laptiev, O., Korol, O., Pohasii, S., Milevskyi, S., Khmelevsky, R. Analysis of information security threat assessment of the objects of information activity. *International independent scientific journal. Poland*. Vol. 1, №34, 2021, pp.33 – 39. ISSN 3547-2340.
 8. Yevseiev, S., Laptiev, O., Korol, O., Pohasii, S., Milevskyi, S. The methodology of automatical detection of digital illegal obtaining means of information. *Scientific discussion. Praha, Czech Republic*. Vol. 1, No 62, 2021. pp. 16–22. ISSN 3041-4245.
- Laptiev, O., Pohasii, S., Milevskyi, S., Khmelevsky, R., Barabash, A., Ponomarenko, V. Information security of the eGovernment. *Journal of science. Lyon*. №27, 2021, pp.49-54. ISSN 3475-3281.

Надійшла 17.02.2024