

УДК 519.688; 004.89; 004.9

**О. Гавриленко, К. Фещенко**

## **ВИЯВЛЕННЯ ПРОПАГАНДИ В ПОТОКАХ НОВИН**

*Анотація:* Виявлення елементів пропаганди в масивах текстових даних наразі є одним із основних засобів боротьби в інформаційній війні, яка відбувається в світі. В даній роботі представлено багатофакторну модель для визначення рівня пропаганди в публікації.

*Ключові слова:* пропаганда, публікація, новини, дописи в соціальних мережах, багатофакторна модель, статистичний аналіз, інформаційна технологія інтелектуального аналізу, машинне навчання.

### **Вступ**

Пропаганда — це форма комунікації, спрямована на вплив на думки, ставлення або дії великої групи людей. Вона часто використовується урядами, політичними партіями, корпораціями або іншими організаціями для просування певної ідеології, політики або товару. Пропаганда може бути як позитивною, так і негативною, але часто асоціюється з маніпуляцією інформацією та дезінформацією для досягнення своїх цілей.

Актуальність виявлення пропаганди в новинах є дуже високою в сучасному інформаційному середовищі. Існує ряд причин, чому це питання є важливим:

1. Захист демократії. Демократичні суспільства залежать від інформованих громадян, які можуть робити свідомий вибір. Пропаганда може викривлювати реальність та маніпулювати громадською думкою, що підриває демократичні процеси.

2. Поширення дезінформації. Пропаганда часто використовує дезінформацію або викривлені факти для досягнення своїх цілей. Це може призвести до того, що громадськість отримує неправдиву інформацію, що в свою чергу може викликати паніку, невиправдані страхи або ненависть.

3. Політичні маніпуляції. Уряди або політичні групи можуть використовувати пропаганду для впливу на громадську думку з метою підтримки своїх політичних цілей, що може призвести до несправедливих виборів або інших політичних переваг.

4. Соціальні наслідки. Пропаганда може посилювати розподіли у суспільстві, поширювати ненависть, дискримінацію та підвищувати рівень насильства. Це може призвести до соціальної нестабільності та конфліктів.

5. Економічні наслідки. Дезінформація може впливати на фінансові ринки, викликати економічні потрясіння та підривати довіру до бізнесу та економічних інституцій.

6. Здоров'я громадян. Під час пандемій або інших кризових ситуацій пропаганда та дезінформація можуть мати серйозні наслідки для здоров'я громадян, спричиняючи недовіру до медичних порад, вакцинацій та інших заходів охорони здоров'я.

7. Міжнародні відносини. Пропаганда може впливати на міжнародні відносини, викликати напругу між країнами та сприяти конфліктам. Це може мати далекосяжні наслідки для глобальної безпеки та стабільності.

#### *Методи виявлення пропаганди*

На сьогоднішній день існують наступні методи виявлення пропаганди:

1. *Аналіз контенту* – використання алгоритмів для аналізу тексту та ідентифікації пропагандистських технік, таких як емоційно забарвлена мова, перебільшення, використання стереотипів тощо.

2. *Перевірка фактів* – створення платформ для перевірки фактів, які можуть аналізувати інформацію та вказувати на її достовірність або фальшивість.

3. *Освітні програми* – підвищення медіаграмотності серед громадян, щоб вони могли самостійно ідентифікувати пропаганду та дезінформацію.

4. *Співпраця з технологічними компаніями* – розробка інструментів та алгоритмів для соціальних медіа платформ, які можуть виявляти та позначати пропаганду та дезінформацію.

5. *Дослідження мереж поширення* – аналіз того, як пропагандистські матеріали поширюються в Інтернеті, включаючи ідентифікацію ключових гравців та каналів поширення.

Виявлення та протидія пропаганді в новинах є ключовими елементами для забезпечення здорового інформаційного середовища, яке сприяє свідомому прийняттю рішень громадянами та підтримці демократичних цінностей.

### **Аналіз літературних джерел та постановка проблеми**

Розуміння пропаганди через науковий аналіз вимагає міждисциплінарного підходу, що включає історію, психологію, соціологію, політологію та технології. Вивчаючи наявну літературу та тематичні дослідження, можна отримати комплексне уявлення про те, як працює пропаганда та її далекосяжні наслідки.

Аналіз наукових статей на тему пропаганди виявляє різні підходи та висновки в цій галузі. Дослідники все частіше застосовують передові методи для виявлення та аналізу пропаганди в новинних статтях, включаючи використання моделей машинного навчання, таких як BERT і GPT-4. Ці моделі навчені виявляти і класифікувати різні пропагандистські технології в текстових даних.

В роботі [1] наведено результати дослідження з використанням попередньо навченої моделі BERT, з метою покращити виявлення пропагандистських матеріалів

у новинних статтях. Дана модель обробляє текст на рівні слів та інтегрує ознаки на рівні речень, що дозволяє їй ефективно розрізняти пропагандистський і непропагандистський контент. Дослідження висвітлює такі проблеми, як незбалансованість даних, і запропонувало такі методи, як надмірна вибірка та доповнення даних для вирішення цих проблем

В роботі [2] представлено дослідження, яке полягало у анотуванні та виявленні пропагандистських матеріалів за допомогою GPT-4. Це передбачало комплексний процес анотування, що складався з кількох етапів для забезпечення високої якості даних. Створений набір даних включав анотовані абзаци з різних джерел новин, що дало змогу детально проаналізувати методи пропаганди в різних темах і сферах. Це дослідження також підкреслило важливість високої узгодженості між анотаторами з метою збереження надійності анотацій.

В публікації [3] наведено результати дослідження впливу пропаганди на політичну ситуацію в США. Було встановлено, що поширення неправдивої інформації (дезінформації) в засобах масової інформації мало глибокий вплив на соціальний дискурс і політику в Сполучених Штатах. У зв'язку з цим, виявлення «фейкових новин» привернуло багато уваги з деякими вражаючими результатами, однак процес виявлення пропаганди залишається складним. Наведена в даній публікації пропозиція, спрямована на подальше дослідження виявлення пропаганди шляхом побудови онтології з цією конкретною метою на увазі, спираючись при цьому на численні дисципліни в межах інформатики та соціальних наук.

В роботі [4] наведено результати наступного дослідження: виконано детальний аналіз текстів шляхом виявлення всіх фрагментів, які містять пропагандистські методи, а також їх тип. Зокрема, створено корпус новинних статей, анотованих вручну на рівні фрагментів за допомогою вісімнадцяти пропагандистських прийомів, і пропонуємо відповідну міру оцінки. Крім того, розроблено нову нейронну мережу з деталізацією й показали, що вона перевершує кілька потужних базових ліній на основі BERT.

В публікації [5] представлена методика оцінки достовірності сумнівної інформації за допомогою обчислення найкоротшого шляху між концептуальними вузлами, відповідно до правильно визначених метрик семантичної близькості на графах знань.

У роботі [6] системно викладено історію виникнення і розвитку методології інформаційних воєн та розкрито різницю між американською, британською і російською моделями. Вперше вводиться поняття смислової війни, розглядається її роль в сучасному світі.

Варто підкреслити, що перераховані моделі машинного навчання використовують принцип «навчання з учителем», тобто потребують людського втручання у формування початкових вибірок. Це може внести певний елемент суб'єктивізму у процес прийняття рішення щодо того, чи має певна стаття елементи пропаганди чи ні.

Слід також зазначити, що на сьогоднішній день велике значення у розповсюдженні пропаганди мають соціальні мережі [7].

Зокрема, у статті [8] пропонується нова модель під назвою CatRevenge, яка ідентифікує такий аспект нездорового спілкування в соціальних мережах, як активну, так і пасивну помсту, що також стосується поняття пропаганди. Ця модель попередньо обробляється зі словником значень інтернет-сленгу Slangzy, щоб ефективніше виявляти текст помсти. CatRevenge призначає вагу впливу кожній частині мови в реченнях на основі її релевантності та оцінки TF-IDF слів. Нова модель CatRevenge також розглядає модель вбудовування абзаців для контекстного семантичного аналізу тексту про помсту. Крім того, у цьому дослідженні застосовано класифікатор CATBoost із підсиленням градієнта з категоріальними функціями, щоб зменшити переобладнання моделі.

В результаті дослідження, представленого в роботі [9], було виявлено осіб, які впливають на процеси обміну знаннями через внутрішню соціальну мережу, і спрогнозувати майбутні потоки знань. Інакше кажучи, в роботі досліджується вплив пропаганди на ідеологічному рівні. В результаті, представлено чотирифазову методологію, яка поєднує аналіз соціальних мереж із структурним моделюванням.

В роботі [10] досліджено рівень впливу на формування курсу криптовалют такого фактору, як дописи відомих людей в соціальних мережах. Дослідження проводилося на основі статистичного аналізу. В результаті, було встановлено, що дописи відомих людей в соціальних мережах істотно впливають на курси криптовалют. Таким чином, дана робота є прикладом пропаганди в галузі комерції.

Ці дослідження підкреслюють важливість використання складних методів машинного навчання, статистичного аналізу, Data Mining, а також ретельних процесів анотування даних для виявлення та аналізу пропаганди. Вони дають цінну інформацію про методології, які можуть підвищити точність і надійність систем виявлення пропаганди, що має вирішальне значення для розуміння і пом'якшення впливу пропаганди.

Слід зазначити, що процес виявлення пропаганди продовжує потребувати створення різноманітних математичних моделей для ще більш якіснішої ідентифікації цієї форми комунікації.

Авторами даної роботи запропоновано багатофакторну модель оцінки рівня пропаганди в новинах або дописах в соціальних мережах, яка ґрунтується лише на статистичних даних та математичних розрахунках, зроблених за допомогою алгоритмів Data Mining та Теорії прийняття рішень.

### Мета і задачі дослідження

Метою дослідження є створення багатофакторної моделі, за допомогою якої буде розраховуватися метрика, яка показуватиме рівень вмісту пропагандистського контенту в новинах чи дописах в соціальних мереж (надалі – публікаціях). Використання цієї метрики дозволяє сформуванню рекомендації щодо того, чи є публікація пропагандистською чи ні.

Для досягнення мети були поставлені такі завдання:

- сформувати датасет з публікацій;
- визначити ознаки, які є характерними при визначенні пропаганди в публікаціях;
- на основі обраних факторів створити багатофакторну модель для обчислення рівня пропаганди в публікаціях;
- сформувати рекомендації, щодо того, чи є публікація пропагандистською чи ні.

### Матеріали та методи дослідження

Об'єктом дослідження є процес визначення пропаганди в публікаціях, на основі аналізу статистичної інформації про неї, а саме:

- час публікації;
- кількість слів у публікації;
- кількість речень у публікації;
- кількість складів;
- першоджерело публікації;
- кількість перепостів публікації;
- джерела, де було перепощено дану публікацію.

Інформація, яка потрібна для аналізу наявності пропаганди, є простою, не потребує додаткової обробки і є доступною.

Множину публікацій та всю необхідну статистичну інформацію для даного дослідження може бути взято з [11].

Таблиця 1. Зразок вхідних статистичних даних

№	Назва публікації	Час публікації	Першоджерело публікації	Кількість перепостів публікації	Джерела перепостів	Кількість слів у публікації	Кількість речень у публікації	Кількість складів у публікації
1	$P_1$	$t_1$	$L_1$	$m_1$	$Q_1^1; Q_2^1; \dots; Q_{m_1}^1$	$a_1$	$b_1$	$c_1$
2	$P_2$	$t_2$	$L_2$	$m_2$	$Q_1^1; Q_2^1; \dots; Q_{m_2}^1$	$a_2$	$b_2$	$c_2$
...	...	...	...	...	...	...	...	...
$l$	$P_l$	$t_l$	$L_l$	$m_l$	$Q_1^1; Q_2^1; \dots; Q_{m_l}^1$	$a_l$	$b_l$	$c_l$

У табл. 1 у стовпці 2 наведено  $P_j$  – назва публікацій, у стовпці 3  $t_j$  – час, коли було зроблено дану публікацію, у стовпці 4  $L_j$  – назви її першоджерела, у стовпці 5  $m_j$  – кількості її перепостів в інших новосних джерелах чи соціальних мережах, у стовпці 6  $Q_1^j; Q_2^j; \dots; Q_{m_j}^j$  – перелік джерел, в яких було перепощено публікацію, у стовпці 7  $a_j$  – кількість слів у публікації, у стовпці 8  $b_j$  – кількість речень у публікації, у стовпці 9  $c_j$  – кількість складів у публікації ( $j = 1, 2, \dots, l$ ).

Сформований таким чином датасет є вхідною інформацією для алгоритму визначення рівня пропаганди в наборі публікацій.

Але слід зазначити, що процес виявлення пропаганди потребує також додаткових даних, які можна отримати після додаткового аналізу заголовків та тексту публікації  $P_j$ . Зокрема:

- визначення тематики публікації ( $S_k$ );
- множини фактів, про які йдеться у публікації ( $F^j$ );
- емоційне забарвлення публікації;
- наявність в публікації посилань на авторитетні видання, статистичні дані чи нормативні акти ( $F_1^j$ );
- події, які пов'язані з публікацією ( $D$ );
- наявність клікабельних виразів у заголовках та тексті публікації;
- відносна частота появи пропагандистських публікацій у першоджерелі даної публікації;
- особи, зацікавлені в даній публікації ( $M$ ).

Таким чином, успішне проведення дослідження потребує як «простих» статистичних даних, так і даних, отриманих за допомогою засобів Data Mining.

На рис. 1 наведено модель проведення досліджень.

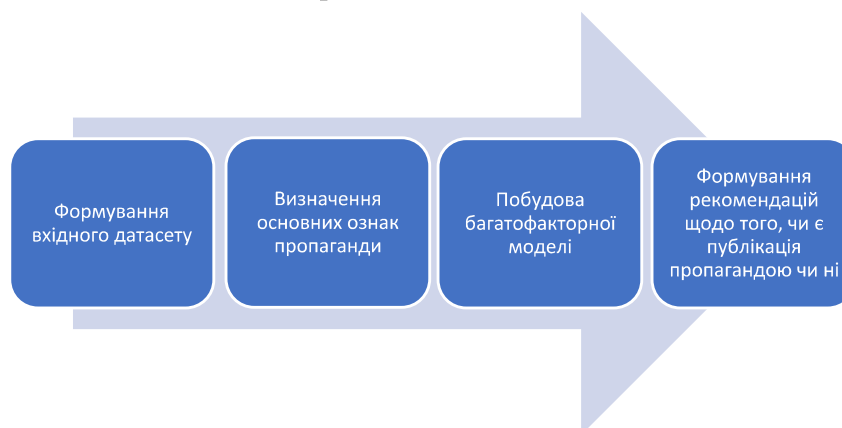


Рисунок 1. Модель досліджень для визначення рівня пропаганди в публікаціях

В основі запропонованого принципу визначення пропаганди лежить розрахунок метрики, яка показує рівень відповідності кожної з обраних публікацій попередньо відібраним ознакам пропаганди. Для цього використовується метод лінійної згортки. Для розрахунків значень показників використовувалися методи статистичного аналізу та методи Data Mining. Для визначення рівня впливу кожного показника на загальне значення рівня довіри до публікації використовується коефіцієнт подібності Жаккара. На основі розрахованої метрики формуватимуться рекомендації щодо вмісту пропаганди в даній публікації і складається рейтинг довіри до видань чи соціальних мереж, звідки її було взято.

Для проведення інтелектуального аналізу та отримання результатів на основі вказаних методів інтелектуального аналізу було розроблено програмне забезпечення.

### **Алгоритм визначення пропаганди в публікації**

#### **Формування датасету з публікацій**

На вхід подається набір публікацій  $P_1, \dots, P_l$ , які відповідатимуть наступним критеріям:

- публікації відбираються з різноманітних новосних видань світу чи постів в соціальних мережах;
- публікації мають відноситися до різної тематики;
- тексти публікацій повинні бути написані різними мовами;
- публікації повинні містити статистичні дані, наведені в пункті 4.

На виході отримується множина публікацій  $P = (P_1, \dots, P_l)$  зі статистичними даними щодо них (табл. 1).

Сформований таким чином датасет дає змогу охопити більшу кількість різноманітних інформаційних джерел світу, що, в свою чергу, зробить більш якісним експеримент з розпізнання елементів пропаганди. Зазначені статистичні дані будуть враховані при обчислень деяких значень показників відповідності ознакам пропаганди.

#### **Визначення ознак, які є характерними при визначенні пропаганди в публікаціях**

На вхід подаються наступні ознаки пропаганди:

$x_1 = \{\text{спроби маніпулювати аудиторією}\};$

$x_2 = \{\text{спрямованість публікації викликати емоції}\};$

$x_3 = \{\text{часті повторення певної думки в публікації}\};$

$x_4 = \{\text{часті перепости публікації}\};$

$x_5 = \{\text{простота читання тексту публікації}\};$

$x_6 = \{\text{високий рівень пропагандистських публікацій в першоджерелі}\};$

$x_7 = \{\text{приналежність публікації до певної тематики, яка є особливо підвладною пропаганді}\};$

$x_8 = \{\text{дана публікація може мати вплив на настання певної події}\};$

$x_9 = \{\text{наявність осіб, яким вигідне розповсюдження даної публікації}\};$

$x_{10} = \{\text{наявність клікабельних виразів в заголовках чи тексті публікації}\}.$

На виході отримується множина ознак  $X = (x_1, \dots, x_{10})$ .

Вибір даних ознак обумовлений тим, що, як відомо, пропаганда використовує наступні засоби:

- спроба впливу на переконання аудиторії без обґрунтування фактів чи раціональних аргументів;
- використання сильних емоцій, таких як страх, гнів, чи патріотизм, для мобілізації аудиторії;
- використання частих повторів однієї і тієї ж думки, з метою кращого її запам'ятовування аудиторією;
- часті перепости однієї й тієї ж публікації, з метою охоплення більшої аудиторії;
- написання текстів публікації максимально простою мовою, легкою для читання, з метою покращення сприйняття та запам'ятовування інформації аудиторією;
- в разі, якщо першоджерело публікації досить часто публікує матеріали пропагандистського характеру, то з високою долею ймовірності чергова його публікація також може виявитись пропагандою;
- написання публікацій певної тематики (зокрема: політика, релігія, комерція, екологія, військова сфера, соціальна сфера, певна ідеологія), оскільки серед саме серед цих тем зустрічаються пропагандистські публікації;
- пропагандистська публікація має на меті підготувати аудиторію до настання якоїсь події;
- написання подібних публікацій завжди має бути вигідним якійсь особі чи колу осіб;
- часто пропагандистські публікації використовують клікабельні вирази в заголовках та текстах, з метою привертання уваги аудиторії.

### **Побудова багатofакторної моделі для обчислення рівня пропаганди в публікаціях**

Процес побудови багатofакторної моделі для обчислення рівня пропаганди в публікаціях, згідно методу лінійної згортки, можна представити наступними кроками [12, 13].

КРОК 0: Передобробка тексту публікації;

КРОК 1: Обчислення числових показників моделі;



КРОК 2: Обчислення коефіцієнтів важливості кожного показника;

КРОК 3: Обчислення функції цінності.

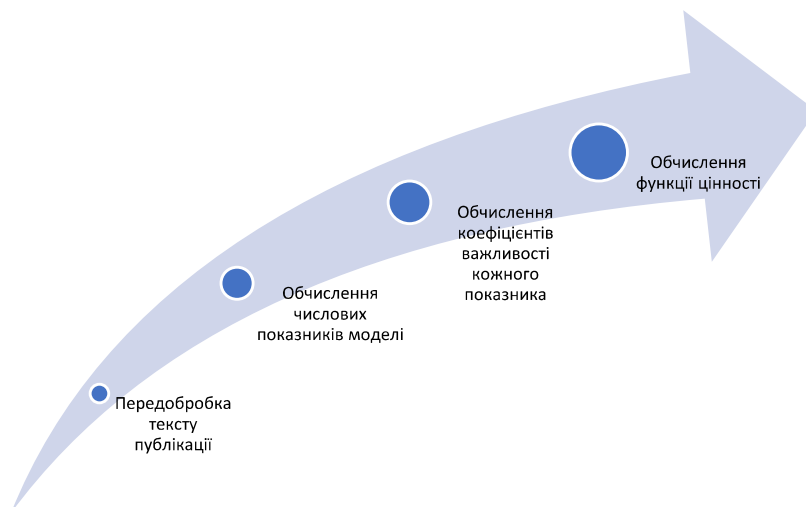


Рисунок 2. Процес побудови багатофакторної моделі

#### КРОК 0

На вхід подаються: множина публікацій  $P = (P_1, \dots, P_l)$ .

На виході отримуються множини  $A_1, A_2, \dots, A_l$  – множини вживаних слів в публікаціях  $P_1, \dots, P_l$  відповідно.

Для формування кожної з множини  $A_j$  ( $j = 1, \dots, l$ ) рекомендується провести передобробку тексту публікації за допомогою процесів лематизації та стемінгу з метою зменшення потужності даної множини за рахунок відкидання спільнокореневих слів та допоміжних частин мови.

#### КРОК 1

На вхід подаються: множина публікацій  $P = (P_1, \dots, P_l)$ , множина ознак  $X = (x_1, \dots, x_{10})$  (див. пункт 5.2) та статистичні дані про публікації (табл. 1) [14, 15].

Необхідно розрахувати рівні пропаганди за кожним з наведених факторів.

На виході отримується множина  $K^j = (K_1^j, \dots, K_{10}^j)$ , де  $K_i^j, i = 1, \dots, 10; j = 1, \dots, l$  – значення метрик, які показують рівень пропаганди в публікації  $P_j$  згідно ознаки  $x_i$ .

1. Обчислення метрики  $K_+^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$ , в якій йдеться про деякі факти (події), які необхідно попередньо виділити з публікації. Множину фактів  $F^j$  представимо у вигляді:

$$F^j = F_1^j \cup F_2^j, F_1^j \cap F_2^j = \emptyset,$$

де  $F_1^j, F_2^j$  – підмножини фактів, відповідно підтверджених та непідтвердженими достовірними джерелами, статистичними даними та нормативними актами.

Метрика  $K_1^j$  публікації  $P_j$  вказує на те, яка частка фактів викладених в публікації лишаються без підтвердження з достовірних джерел, статистичних даних чи нормативних актів.

$$K_1^j = \frac{|F_2^j|}{|F^j|}, \quad (1)$$

де  $|F_2^j|$  – кількість непідтверджених фактів у публікації  $P_j$ ,  $|F^j| = |F_1^j| + |F_2^j|$  – загальна кількість фактів у публікації  $P_j$ .

Очевидно, що  $0 \leq K_1^j \leq 1$  і чим ближче його значення до одиниці, тим більше непідтверджених фактів має дана публікація. Отже, за ознакою  $x_1$  її можна рекомендувати як пропагандистську.

2. Обчислення метрики  $K_2^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$ . Як відомо, будь-який текст може викликати в людини три типи емоцій: позитивні, негативні та нейтральні. Для їх визначення у текстах створено спеціальні словники.

Метрика  $K_2^j$  вказує на рівень емоційного забарвлення публікації, тобто позитивних або негативних емоцій [16]. Нейтральні емоції не слід розглядати, оскільки пропагандистські публікації не повинні лишати байдужим аудиторію.

$$K_2^j = \frac{K_{2p}^j + K_{2n}^j}{2}, K_{2p}^j = \frac{|A_j \cap B|}{|A_j|}, K_{2n}^j = \frac{|A_j \cap C|}{|A_j|}, \quad (2)$$

де  $A_j$  – множина вживаних слів в публікації  $P_j$ , а  $|A_j| = a_j$  (табл. 1);  $B$  – множина слів у словнику для визначення позитивних емоцій;  $C$  – множина слів у словнику для визначення негативних емоцій;  $K_{2p}^j$  – рівень позитивного забарвлення публікації;  $K_{2n}^j$  – рівень негативного забарвлення публікації.

Очевидно, що  $0 \leq K_2^j \leq 1$  і чим ближче його значення до одиниці, тим вищий рівень емоційного забарвлення має дана публікація. Отже, за ознакою  $x_2$  її можна рекомендувати як пропагандистську.

3. Обчислення метрики  $K_3^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$ . Нехай  $A_j$  – множина вживаних слів. Ці слова формують певну думку в даній публікації, має бути попередньо визначена.

Метрика  $K_3^j$  вказує на частоту повторів цієї думки в публікації.

$$K_3^j = \frac{m_{A_j}}{n_j}, \quad (3)$$

де  $m_{A_j}$  – кількість повторів слів, що входять в множину  $A_j$ ;  $n_j$  – загальна кількість слів в публікації  $P_j$ .

Очевидно, що  $0 \leq K_3^j \leq 1$  і чим ближче його значення до одиниці, тим частіше повторюється певна думка в даній публікації. Отже, за ознакою  $x_3$  її можна рекомендувати як пропагандистську.

4. Обчислення метрики  $K_4^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$  та множина  $Q_j = (Q_1^j; Q_2^j; \dots; Q_{m_j}^j)$  джерел, в яких зроблено перепости публікації  $P_j$  (табл. 1).

Метрика  $K_4^j$  вказує на частоту перепостів даної публікації.

$$K_4^j = \frac{m_j}{n_{Q_j}}, \quad (4)$$

де  $m_j$  – кількість перепостів публікації  $P_j$  (табл.1);  $n_{Q_j}$  – середня кількість перепостів публікацій в множині  $Q_j$  – має бути визначена додатково.

Очевидно, що  $0 \leq K_4^j \leq 1$  і чим ближче його значення до одиниці, тим частіше перепощується дана публікація. Отже, за ознакою  $x_4$  її можна рекомендувати як пропагандистську.

5. Обчислення метрики  $K_5^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$ . Нехай вона містить: кількість слів –  $a_j$ ; кількість речень –  $b_j$ ; кількість складів –  $c_j$  (табл.1).

Метрика  $K_5^j$  вказує на легкість читання тексту даної публікації.

$$K_5^j = \left( 206,835 - 1,015 \cdot \frac{a_j}{b_j} - 84,6 \cdot \frac{c_j}{a_j} \right) \cdot 0,01. \quad (5)$$

Метрика  $K_5^j$  називається індексом читабельності Флеша [17].

Значення даної метрики можна інтерпретувати, як показано в табл. 2

Очевидно, що  $0 \leq K_5^j \leq 1$  і чим ближче його значення до одиниці, тим більш простою для читання є дана публікація. Отже, за ознакою  $x_5$  її можна рекомендувати як пропагандистську.

6. Обчислення метрики  $K_6^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$  та її першоджерело  $L_j$ .

Метрика  $K_6^j$  вказує на рівень недовіри читачів до джерела  $L_j$ .

$$K_6^j = \frac{m_{L_j}}{n_{L_j}}, \quad (6)$$

де  $m_{L_j}$  – кількість публікацій пропагандистського характеру в інформаційному джерелі  $L_j$ , а  $n_{L_j}$  – загальна кількість публікацій джерела  $L_j$ .

Очевидно, що  $0 \leq K_6^j \leq 1$  і чим ближче його значення до одиниці, тим вищою є ймовірність того, що дана публікація носитиме пропагандистський характер. Отже, за ознакою  $x_6$  її можна рекомендувати як пропагандистську.

Слід зазначити, що з джерел  $L_j$ , публікації яких було проаналізовано на наявність пропаганди, необхідно скласти ранжований список  $L$  щодо рівня пропаганди, який буде весь час оновлюватися. Цей список міститиме всю необхідну статистичну інформацію для подальших досліджень (табл. 3).

Таблиця 2. Інтерпретація значень індексу читабельності Флеша

Оцінка	Шкільний рівень	Примітки
0,9–1,0	5 клас	Дуже легко читається. Легко зрозумілий середньому 11-річному учневі.
0,9–0,8	6 клас	Легко читається. Розмовна мова для споживачів.
0,8–0,7	7 клас	Досить легко читається.
0,7–0,6	8 і 9 клас	Звичайна мова. Легко сприймається учнями 13-15 років.
0,6–0,5	10-12 клас	Досить важко читати.
0,5–0,3	Коледж	Важко читати.
0,3–0,1	Випускник технікуму	Дуже важко читати. Найкраще розуміють випускники університетів.
0,1–0,0	Професійний	Надзвичайно важко читати. Найкраще розуміють випускники університетів.

Таблиця 3. Зразок списку  $L$

№	Назва інформаційного джерела	Кількість пропагандистських публікацій $m_{L_j}$	Загальна кількість публікацій $n_{L_j}$
1	$L_1$	$m_{L_1}$	$n_{L_1}$
2	$L_2$	$m_{L_2}$	$n_{L_2}$
...	...	...	...
$l$	$L_l$	$m_{L_l}$	$n_{L_l}$

Окрім того, в разі відсутності такого списку можна скористатися рейтинговими списками ЗМІ, складеними журналістами, зокрема видання Forbes [18]. Це вирішить проблему «холодного старту» на початкових етапах дослідження. Але в подальшому слід використовувати список  $L$ , що унеможливить суб'єктивний вплив людини (журналіста) на прийняття рішення щодо рівня довіри тому чи іншому інформаційному джерелу.

7. Обчислення метрики  $K_{\tau}^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$ ; множину вживаних слів у публікації  $A_j$ ; множина тем  $S = (s_1; s_2; \dots; s_r)$ , в яких найчастіше зустрічаються пропагандистські публікації, та словники характерних слів для даних тем,  $T_1; T_2; \dots; T_r$ . Теми і відповідні словники мають бути сформовані попередньо.

Метрика  $K_{\tau}^j$  вказує на те, чи належить публікація  $P_j$  до якоїсь із тем з множини  $S$ . Вона обчислюється наступним чином:

1. обчислюємо коефіцієнти подібності Жаккара між множиною  $A_j$  та кожним з словників  $T_k, k = 1, 2, \dots, r$  [19].

$$J(A_j, T_k) = \frac{|A_j \cap T_k|}{|A_j \cup T_k|}, \quad (7)$$

2. обираємо максимальне значення коефіцієнту Жаккара:

$$J_{max}(A_j, T_k) = \max_{k=1,2,\dots,r} J(A_j, T_k), j = 1, \dots, l$$

та встановлюємо тему  $s_k$ , якій відповідає дане значення;

3. робимо висновок, що з ймовірністю  $p_j = J_{max}(A_j, T_k)$  публікація  $P_j$  належить до теми  $s_k$ .

4. метрика  $K_7^j$  дорівнює:

$$K_7^j = \begin{cases} 1, & \text{якщо існує така тема } s_k \in S; \\ 0, & \text{якщо не існує такої теми } s_k \in S. \end{cases} \quad (8)$$

Очевидно, що якщо для публікації  $P_j$  значення  $K_7^j = 1$ , то за ознакою  $x_7$  її можна рекомендувати як пропагандистську.

8. Обчислення метрики  $K_8^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$  та множину вживаних слів у публікації  $A_j$ .

Метрика  $K_8^j$  вказує на те, чи матиме вплив публікація  $P_j$  на настання деякої події  $D$ .

Під впливом публікації  $P_j$  на подію  $D$  розуміються прямі або опосередковані згадки про неї в тексті публікації.

Таким чином, з множини  $A_j$  необхідно виділити підмножину  $D_j \subset A_j$  слів, які описують подію  $D$ .

Тоді:

$$K_8^j = \begin{cases} 1, & \text{якщо } D_j \neq \{\emptyset\}; \\ 0, & \text{якщо } D_j = \{\emptyset\}. \end{cases} \quad (9)$$

Очевидно, що якщо для публікації  $P_j$  значення  $K_8^j = 1$ , то за ознакою  $x_8$  її можна рекомендувати як пропагандистську.

9. Обчислення метрики  $K_9^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$  та множину вживаних слів у публікації  $A_j$ .

Метрика  $K_9^j$  вказує на те, чи існують особи (група осіб  $M$ ), які зацікавлені у публікації  $P_j$ .

Під зацікавленістю особи чи групи осіб  $M$  у публікації  $P_j$  розуміються прямі або опосередковані згадки про неї (них) в тексті публікації.

Таким чином, з множини  $A_j$  необхідно виділити підмножину  $M_j \subset A_j$  слів, які описують осіб  $M$ , що згадуються в публікації  $P_j$ .

Тобто:

$$K_9^j = \begin{cases} 1, & \text{якщо } M_j \neq \{\emptyset\}; \\ 0, & \text{якщо } M_j = \{\emptyset\}. \end{cases} \quad (10)$$

Очевидно, що якщо для публікації  $P_j$  значення  $K_9^j = 1$ , то за ознакою  $x_9$  її можна рекомендувати як пропагандистську.

10. Обчислення метрики  $K_{10}^j$ . Розглянемо деяку публікацію  $P_j, j = 1, \dots, l$ ; множину вживаних слів у публікації  $A_j$  та словник клікабельних слів  $N$ , який також необхідно створювати попередньо.

Метрика  $K_{10}^j$  вказує на те, чи використовує публікація  $P_j$  слова з множини  $N$ .

$$K_{10}^j = \begin{cases} 1, \text{ якщо } A_j \cap N \neq \emptyset; \\ 0, \text{ якщо } A_j \cap N = \emptyset. \end{cases} \quad (11)$$

Очевидно, що якщо для публікації  $P_j$  значення  $K_{10}^j = 1$ , то за ознакою  $x_{10}$  її можна рекомендувати як пропагандистську.

Слід зазначити, що розглянуті метрики можна інтерпретувати як ймовірності того, що публікація  $P_j$  буде пропагандою за ознакою  $x_i$ .

### КРОК 2

На вхід подаються: показники  $K_i^j, i = 1, \dots, 10; j = 1, \dots, l$  (формули (1)-(11)).

Необхідно для кожного критерія розрахувати коефіцієнти їх важливості для обчислення функції цінності.

На виході коефіцієнти  $\omega_i^j$ .

Для обчислення коефіцієнтів  $\omega_i^j$  необхідно сформулювати наступну задачу [14, 15]:

При перевірці публікації на наявність пропаганди настала подія  $W = \{\text{публікація } P_j \text{ є пропагандою}\}$ . Сформулюємо наступні гіпотези:

$$H_i^j = \{\text{публікація } P_j \text{ має ознаку } x_i\}, i = 1, \dots, 10; j = 1, \dots, l.$$

Необхідно обчислити ймовірності  $P(H_i^j/W)$ , тобто ймовірність того, що публікація має ознаку  $x_i$  за умови, що вона є пропагандою.

Вважатимемо, що  $P(H_i^j) = \frac{1}{10}$ , тобто події  $H_i^j$  – рівноймовірні, а  $P(W/H_i^j) = K_i^j$ . Тоді за формулою Баєса:

$$P(H_i^j/W) = \frac{P(H_i^j) \cdot P(W/H_i^j)}{\sum_{i=1}^{10} P(H_i^j) \cdot P(W/H_i^j)} = \frac{K_i^j}{\sum_{i=1}^{10} K_i^j}.$$

Таким чином, коефіцієнт важливості показника  $K_i^j$  обчислюватиметься за формулою (12):

$$\omega_i^j = \frac{K_i^j}{\sum_{i=1}^{10} K_i^j}. \quad (12)$$

### КРОК 3

На вхід подаються: множина публікацій  $P = (P_1, \dots, P_l)$ , показники  $K_i^j, i = 1, \dots, 10; j = 1, \dots, l$  та коефіцієнти  $\omega_i^j$ .

Необхідно для кожної публікації розрахувати значення функції цінності з точки зору наявності в ній ознак пропаганди.

На виході значення функції  $V_j$ .

Значення функції цінності  $V_j$ , згідно методу лінійної згортки, обчислюються за формулою (13) [12, 13]:

$$V_j = \sum_{i=1}^{10} (\omega_i^j \cdot K_i^j). \quad (13)$$

Результати розрахунків можна представити у вигляді таблиці (табл. 4). Сформуємо статистичну вибірку зі значень функції цінності (13):

$$V = (V_1, V_2, \dots, V_l).$$

**Таблиця 4. Розрахункова таблиця**

	$K_1^j$	$K_2^j$	$K_3^j$	$K_4^j$	$K_5^j$	$K_6^j$	$K_7^j$	$K_8^j$	$K_9^j$	$K_{10}^j$	$V_j$
$P_1$	$K_1^1$	$K_2^1$	$K_3^1$	$K_4^1$	$K_5^1$	$K_6^1$	$K_7^1$	$K_8^1$	$K_9^1$	$K_{10}^1$	$V_1$
$P_2$	$K_1^2$	$K_2^2$	$K_3^2$	$K_4^2$	$K_5^2$	$K_6^2$	$K_7^2$	$K_8^2$	$K_9^2$	$K_{10}^2$	$V_2$
$P_l$	$K_1^l$	$K_2^l$	$K_3^l$	$K_4^l$	$K_5^l$	$K_6^l$	$K_7^l$	$K_8^l$	$K_9^l$	$K_{10}^l$	$V_l$

**Формування рекомендації, щодо того, чи є публікація пропагандистською чи ні**

На вхід подається набір публікацій  $P = (P_1, \dots, P_l)$  та вибірка  $V = (V_1, V_2, \dots, V_l)$  (див. пункт 5.3).

На виході формуються рекомендації щодо того, які з публікацій  $P = (P_1, \dots, P_l)$  є пропагандою.

Рекомендації формуються за наступним правилом [14, 15]:

Якщо  $V_j \geq \bar{V}_B$ , ( $j = 1, \dots, l$ ), то публікація  $P_j$  рекомендується як пропагандистська. Якщо  $V_j < \bar{V}_B$ , ( $j = 1, \dots, l$ ), то публікація  $P_j$  не рекомендується як пропагандистська.

В даному правилі  $\bar{V}_B$  – вибіркове середнє для вибірки  $V$ .

$$P_j = \begin{cases} 1, & \text{якщо } V_j \geq \bar{V}_B; \\ 0, & \text{якщо } V_j < \bar{V}_B. \end{cases} \quad (14)$$

Тобто, публікації ставиться у відповідність значення «1», якщо вона є пропагандою і значення «0» – в протилежному випадку.

Релевантність наданих рекомендацій оцінюється за допомогою метрик *recall* та *precision*.

## Висновки

В даній роботі представлено багатофакторну модель, на основі якої можна визначити рівень пропаганди в публікації. Дана модель створена на основі методу лінійної згортки, для реалізації якого було відібрано 10 показників, які є найбільш ефективними при виявленні елементів пропаганди в тексті. Кожен з відібраних показників розраховувався засобами статистичного та інтелектуального аналізів. Їх коефіцієнти важливості обчислювалися за допомогою підходу, в основу якого покладено формулу Баєса. В результаті запропонованого методу, для кожної публікації розраховується значення функції цінності, на основі якого формується рекомендація, щодо того, чи містить вона елементи пропаганди, чи ні.

Перевагами запропонованої моделі є:

1. Відсутність людського (суб'єктивного) впливу на процес обчислення показників пропаганди, оскільки на вхід моделі подаються лише статистичні дані чи дані, отримані засобами інтелектуального аналізу.
2. Легкий процес масштабування моделі за рахунок додавання нових показників чи видалення показників, які вже втратили свою актуальність.
3. Подолання проблеми «холодного старту» за рахунок використання на першому етапі публікацій зі списку Forbs найбільш поважних видань.
4. Створення в процесі роботи моделі власного рейтингу довіри до джерел публікації.
5. Коректність отриманих результатів гарантується використанням класичних математичних методів.

Недоліками запропонованої моделі є:

1. Необхідність накопичувати та зберігати великі масиви статистичних даних.
2. Постійна потреба у контролі рівня релевантності наданих рекомендацій.

Отримані результати можна використовувати у інформаційній війні, яка точиться в Україні та в Світі, а також як дієвий елемент інтен-аналізу.

В якості вдосконалення моделі з метою підвищення рівня релевантності наданих рекомендацій, в подальшому планується проведення ґрунтового дослідження джерел публікацій (як новин так і соціальних мереж) для вибору таких, що найбільше відповідають специфіці дослідження. Також можливе не тільки виявлення пропаганди в публікації, а й встановлення її типу.



## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Li, W., Li, S., Liu, C. et al. (2022). «Span identification and technique classification of propaganda in news articles». *Complex Intell. Syst.* 8, 3603–3612. DOI: <https://doi.org/10.1007/s40747-021-00393-y>.
2. Maram Hasanain, Fatema Ahmed, Firoj Alam. (2024). «Can GPT-4 Identify Propaganda? Annotation and Detection of Propaganda Spans in News Articles». *Computation and Language (cs.CL)*. DOI: <https://doi.org/10.48550/arXiv.2402.17478>.
3. Hamilton, K. (2021). «Towards an Ontology for Propaganda Detection in News Articles». In: Verborgh, R., et al. *The Semantic Web: ESWC 2021 Satellite Events. ESWC 2021. Lecture Notes in Computer Science*, vol 12739. Springer, Cham. DOI: [https://doi.org/10.1007/978-3-030-80418-3\\_35](https://doi.org/10.1007/978-3-030-80418-3_35).
4. Da San Martino, G., Yu, S., Barrón-Cedeño, A., Petrov, R., Nakov, P. (2019). «Fine-grained analysis of propaganda in news article.» In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5635–5645. Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/D19-1565>.
5. Ciampaglia, G.L., Shiralkar, P., Rocha, L.M., Bollen, J., Menczer, F., Flammini, A. (2015). «Computational fact checking from knowledge networks». *PLoS One* 10(6), 15. DOI: <https://doi.org/10.1371/journal.pone.0128193>.
6. Pocheptsov, G. (2015). «Modern information wars». Kyiv: Kyiv- Mogylianska Academy, pp. 497. ISBN: 9789665186748.
7. Ghosal, S., Jain, A. (2024). «CatRevenge: towards effective revenge text detection in online social media with paragraph embedding and CATBoost». *Multimed Tools Appl.* DOI: <https://doi.org/10.1007/s11042-024-18791-y>.
8. Alhajj, R., Rokne, J. (eds). (2018) «Encyclopedia of Social Network Analysis and Mining». Analysis and Mining. Springer, New York, NY. P. 2699. <https://doi.org/10.1007/978-1-4614-7163-9>.
9. Ramona-Diana Leon, Raúl Rodríguez-Rodríguez, Pedro Gómez-Gasquet, Josefa Mula. (2017) «Social network analysis: A tool for evaluating and predicting future knowledge flows from an insurance organization». *Technological Forecasting and Social Change*. Vol. 114, pp. 103-118. <https://doi.org/10.1016/j.techfore.2016.07.032>.
10. Sergii Telenyk, Grzegorz Nowakowski, Olena Gavrilenko, Mykhailo Miahkyi, Olena Khalus. (2024). «Analysis of the influence of posts of famous people in social networks on the cryptocurrency course». *Bulletin of the Polish Academy of Sciences Technical Sciences*, Vol. 72(4). DOI: [10.24425/bpasts.2024.150117](https://doi.org/10.24425/bpasts.2024.150117).
11. Ukrainian News Dataset. URL: <https://huggingface.co/datasets/zeusfsx/ukrainian-news>. (accessed: 14.09.2024).

12. *Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, Roman Słowiński.* (2008). «Multiobjective Optimization». Springer-Verlag Berlin Heidelberg, P. 470. DOI: <https://doi.org/10.1007/978-3-540-88908-3>.
13. *Kalyanmoy Deb.* (2001). «Multi-Objective Optimization using Evolutionary Algorithms». Wiley, P. 536, ISBN: 978-0-471-87339-6.
14. *Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying E. Ye.* (2016). «Probability and Statistics for Engineers and Scientists». 9th ed. Pearson, P. 816. ISBN-13: 978-0134115856.
15. *Sheldon Ross.* (2018). «A First Course in Probability». 10th ed. Pearson, P. 528. ISBN-13: 978-0134753119.
16. *P.G. Preethi, V. Uma, A. Kumar.* (2015). «Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction». Procedia Computer Science. №48. pp. 84–89. <https://doi.org/10.1016/j.procs.2015.04.154>.
17. *Rudolf Flesch.* (1979). «How to Write Plain English: A Book for Lawyers and Consumers». Harper & Row, P. 126, ISBN: 9780060112783.
18. Ranked: America's Most Trustworthy News Organizations in 2024. URL: <https://www.visualcapitalist.com/ranked-americas-most-trustworthy-news-organizations-in-2024/>. (accessed: 14.09.2024).
19. *Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman.* (2014). «Mining of Massive Datasets». Cambridge University Press, P. 326, ISBN: 9781107015357.