***Multifactor Model for Propaganda Detection in Conditions of Information Warfare***

|  |
| --- |
| Kyryl Feshchenko, Olena Havrylenko  Department of Information Systems and Technologies  Igor Sikorsky Kyiv Polytechnic Institute  Kyiv, Ukraine |
| ***Abstract.* The paper presents a multifactor model for the detection of propaganda in the context of information warfare. The model integrates data mining techniques and machine learning algorithms to evaluate the likelihood of propagandistic content in textual data. It addresses the complexity and adaptability of modern disinformation strategies by incorporating multiple linguistic, semantic, and emotional indicators. A prototype system has been developed to automate the analysis of online content, enabling more efficient identification and filtering of potentially harmful information.**  ***Keywords*: *propaganda detection, information warfare, data mining, machine learning, text analysis, multifactor model.*** |

***Багатофакторна модель виявлення пропаганди в умовах інформаційної війни***

Фещенко Кирил Юрійович, Гавриленко Олена Валеріївна

Кафедра інформаційних систем та технологій,

КПІ ім. Ігоря Сікорського

м. Київ, Україна

fkirill440@gmail.com, gelena1980@gmail.com

*Анотація.* У роботі представлено багатофакторну модель виявлення пропаганди в умовах інформаційної війни. Модель поєднує методи інтелектуального аналізу даних та алгоритми машинного навчання для оцінки ймовірності пропагандистського характеру текстового контенту. Вона враховує складність і адаптивність сучасних стратегій дезінформації шляхом використання низки лінгвістичних, семантичних та емоційних ознак. Розроблено прототип системи, що автоматизує аналіз онлайн-контенту та забезпечує ефективніше виявлення і фільтрацію потенційно шкідливої інформації.

# *Ключові слова: виявлення пропаганди, інформаційна війна, інтелектуальний аналіз даних, машинне навчання, аналіз тексту, багатофакторна модель.*

Вступ

У світлі повномасштабної агресії Російської Федерації проти України, питання інформаційної безпеки набуло стратегічної ваги. Інформаційна війна, як складова гібридної агресії, спрямована на свідомість громадян, формування маніпулятивних наративів, деморалізацію суспільства та підрив довіри до державних інституцій.

Одним з ключових інструментів є **пропаганда** — систематичне поширення повідомлень, що містять емоційно забарвлені оцінки, маніпулятивні заклики, викривлені факти та тригерні формулювання[2].

Особливу загрозу становить **цифрова пропаганда**, що поширюється через соціальні мережі, месенджери та інші інформаційні платформи, з високою швидкістю, анонімно та безконтрольно[3]. Традиційні методи виявлення дезінформації виявилися малоефективними в умовах зростаючої складності контенту.

У зв’язку з цим запропоновано **багатофакторну модель автоматизованого виявлення пропаганди**, що поєднує методи **інтелектуального аналізу даних (data mining)** та **машинного навчання**, з метою оцінки ймовірності пропагандистського характеру текстового повідомлення[4].

Опис Моделі

### Вхідні параметри:

1. 𝑇 — текст повідомлення;
2. 𝐻 — заголовок;
3. 𝑆 — джерело публікації.

### Вихідний параметр:

1. **Total Score** (TS) — інтегральна оцінка пропагандистського навантаження, що розраховується за формулою:

TS=i=1∑10​wi​⋅Pi​

де:

1. Pi​∈[0;1] — нормалізоване значення i-го параметра;
2. wi​∈[0;1] — ваговий коефіцієнт i-го параметра.

### Оцінювані параметри:

| № | Позначення | Назва показника | Короткий опис |
| --- | --- | --- | --- |
| 1 | P1​ | Сентимент-аналіз | Емоційне забарвлення тексту |
| 2 | P2​ | Тригерні слова | Наявність провокаційних термінів |
| 3 | P3​ | Простота тексту | Лексична та синтаксична складність |
| 4 | P4​ | Надійність джерела | Репутаційна оцінка джерела |
| 5 | P5​ | Тригерні теми | Соціально-чутливі наративи |
| 6 | P6​ | Клікбейтний заголовок | Типові шаблони клікбейту |
| 7 | P7​ | Суб’єктивність | Частота суджень замість фактів |
| 8 | P8​ | Заклик до дії | Прямі чи непрямі мотивації до дії |
| 9 | P9​ | Повторювані тези | Частота повторення в межах тексту |
| 10 | P10​ | Повторювані тексти у джерелі | Виявлення кампаній повторного публікування |

### Розрахунок вагових коефіцієнтів:

wi​=∑j=110​fj​fi​​

де:

1. fi​ — частка випадків, коли Pi​>0.3, серед усіх текстів у вибірці.

Це дозволяє адаптивно враховувати релевантність кожного параметру на основі емпіричних даних[1].

### Інтерпретація показника Total Score:

| Значення TS | Інтерпретація |
| --- | --- |
| < 0.1 | Відсутні ознаки пропаганди |
| 0.1–0.2 | Низький рівень |
| 0.2–0.3 | Помітний рівень |
| 0.3–0.5 | Помірний рівень |
| 0.5–0.7 | Високий рівень |
| > 0.7 | Вельми високий рівень |

Значення TS ≥ 0.3 є **пороговим** і потребує глибшого аналізу.

Результати Експерименту

З метою апробації ефективності багатофакторної моделі виявлення пропаганди було проведено експериментальне дослідження, в рамках якого проаналізовано 10 текстових повідомлень, відібраних за принципом різноманітності пропагандистського навантаження. До вибірки увійшли як тексти, що є зразками відкритої політичної пропаганди (зокрема, публічні виступи тоталітарних лідерів), так і нейтральні за змістом матеріали науково-популярного та освітнього характеру. Оцінка проводилася шляхом обчислення інтегрального показника total score, який відображає рівень ймовірної пропагандистської насиченості тексту.

Таблиця 1

|  |  |  |
| --- | --- | --- |
| **Назва повідомлення** | **Total score** | **Посилання** |
| Звернення путіна до Генеральної Асамблеї перед оголошенням СВО | 0.64 | Офіційний сайт кремля |
| Звернення Емануеля Макрона про стратегічне бачення Європи | 0.41 | https://www.rfi.fr/uk |
| Промова Трампа під час раллі 2020 року у Тулзі | 0.63 | https://www.rev.com/transcripts/donald-trump-tulsa-oklahoma-rally-speech-transcript |
| Промова Трампа під час раллі 2020 року перед закінченням підрахунку голосів | 0.78 | https://www.rev.com/blog/transcripts/donald-trump-speech-save-america-rally-transcript-january-6 |
| Звернення Гітлера до нації (1 вересня 1939 року) | 0.61 | https://alphahistory.com/nazigermany/hitler-declaration-of-war-against-poland-1939 |
| Промова Гітлера на Спортпалаці | 0.68 | https://research.calvin.edu/german-propaganda-archive/hitler1.html |
| Введення до лійної геометрії | 0.24 | https://www.khanacademy.org/math/geometry/hs-geo-congruence/hs-geo-linear-segments/a/geometry-introduction-to-lines |
| Магія математики | 0.2 | https://www.networkpages.nl/the-magic-of-mathematics/ |
| Дослідження океанів | 0.23 | https://www.noaa.gov/education/resource-collections/ocean-coasts/ocean-exploration |
| Вирощування яблук | 0.25 | https://www.nifa.usda.gov/topic/apple-production |

Для візуальної інтерпретації результатів з табл. 1 класифікувалися за шкалою:  
  
**Зелений**: total score < 0.3 — тексти без ознак пропаганди або з незначним впливом

**Жовтий**: 0.3 ≤ total score < 0.5 — помірний рівень пропаганди

**Червоний**: total score ≥ 0.5 — високий та вельми високий рівень пропаганди

Аналіз даних з табл 1. показав чітку кореляцію між тематичним змістом текстів і рівнем виявленого пропагандистського впливу. Так, усі повідомлення, що містять політичну риторику з елементами ідеологічного навантаження (виступи Путіна, Трампа, Гітлера), були класифіковані як високопропагандистські (червона зона), що підтверджується високими показниками total score у межах від 0.61 до 0.78. Це свідчить про здатність моделі фіксувати маніпулятивні структури, такі як апеляція до страху, нав’язування дихотомій, використання категоричних суджень та інші риторичні прийоми, характерні для пропаганди.

Водночас тексти з науково-популярних ресурсів, присвячені темам освіти, досліджень і побутової інформації (наприклад, «Магія математики» чи «Вирощування яблук»), стабільно потрапляють у зелену зону — їхні показники не перевищують 0.25. Це демонструє здатність моделі ігнорувати інформацію, яка не несе ідеологічного чи маніпулятивного підтексту, що важливо для зменшення кількості хибнопозитивних результатів.

Особливу увагу заслуговують тексти з жовтої зони (наприклад, виступ Макрона), які відзначаються помірним рівнем total score. Вони, як правило, містять елементи стратегічного чи політичного дискурсу, але подаються в більш нейтральному та раціоналізованому тоні. Це свідчить про чутливість моделі до відтінків політичної комунікації, дозволяючи не лише виявляти відверту пропаганду, а й диференціювати проміжні випадки.

Таким чином, результати підтверджують здатність моделі не лише правильно ідентифікувати високопропагандистські тексти, а й ефективно розрізняти ступінь пропагандистської насиченості повідомлень. Це відкриває можливості для її використання у системах попереднього фільтрування контенту в інформаційних просторах, де важливо зберігати баланс між виявленням шкідливого впливу та збереженням свободи слова.

# Висновки

Результати проведеного експерименту переконливо демонструють здатність запропонованої багатофакторної моделі з високим ступенем достовірності диференціювати тексти за рівнем пропагандистського впливу. Аналіз показав, що модель ефективно розпізнає збільшену концентрацію пропагандистських елементів у промовах політичних лідерів, які часто містять ознаки маніпулятивного дискурсу та цілеспрямованого впливу на аудиторію. Водночас система демонструє стабільно низькі показники total score при аналізі нейтральних науково-популярних текстів, що свідчить про її здатність уникати хибнопозитивних результатів. Це вказує на високий рівень точності класифікації та релевантності визначених критеріїв оцінювання.

Застосування динамічного алгоритму вагових коефіцієнтів забезпечує виняткову гнучкість системи, дозволяючи їй адаптуватися до широкого спектра стилістичних, жанрових та тематичних особливостей текстів. Такий підхід значною мірою підвищує ефективність роботи моделі в умовах реального інформаційного середовища, де характер повідомлень може варіюватися залежно від джерела, контексту та цільової аудиторії. Отримані емпіричні дані підтверджують не лише наукову обґрунтованість та логічну стрункість запропонованої моделі, але й її практичну придатність до попередньої автоматизованої фільтрації контенту з ознаками пропаганди. Це відкриває перспективи її подальшого розвитку, вдосконалення точності та інтеграції у комплексні системи інформаційної безпеки в умовах гібридної війни та поширення дезінформації.

##### Література

**1. Гавриленко, О., & Фещенко, К.** (2025). Виявлення пропаганди в потоках новин. Адаптивні Системи Автоматичного Управління 1(46), [164-174]. https://doi.org/10.20535/1560-8956.46.2025.323759

2. Черниш, Н. О., & Мусієнко, І. О. (2021). Інформаційна війна та пропаганда в умовах гібридної агресії проти України. Науковий вісник Херсонського державного університету. Серія: Політологія, 2(15), 102–109.

3. Barro, R. J., & Zussman, A. (2022). Propaganda and manipulation in political discourse: Theory and evidence. Journal of Political Economy, 130(3), 677–712. https://doi.org/10.1086/717203

4. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2018). A stylometric inquiry into hyperpartisan and fake news. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (pp. 231–240). https://doi.org/10.18653/v1/P18-1022