

ПРОГРАМНА РЕАЛІЗАЦІЯ ПРОЦЕСІВ КОМПЛЕКСНОГО ФАКТЧЕКІНГУ ТА ВИЯВЛЕННЯ МАРКЕРІВ МАНІПУЛЯТИВНОГО ВПЛИВУ НА ОСНОВІ LLM

Вінницький національний технічний університет

Анотація

Розглянуто розробку інтелектуального веб-додатка для автоматизованого контент-аналізу та виявлення маркерів маніпулятивного впливу, клікбейту й дезінформації у новинних медіа-текстах. Запропонована система поєднує методи лексико-статистичного аналізу тексту з технологіями семантичного моделювання за допомогою двох архітектур великих мовних моделей (Llama 3 та Gemini 2.5 Flash), а також програмний модуль лінгвістичної декомпозиції та верифікації URL-адрес періоджерел за реєстрами довірених медіа-платформ. Описано функціональну структуру розробленого додатка на базі фреймворку Streamlit та наведено результати компаративного аналізу його роботи.

Ключові слова: великі мовні моделі, медіа-маніпуляції, фактчекінг, комп'ютерна лінгвістика, Streamlit, лінгвістичні маркери.

Abstract

This paper examines the development of an intelligent web application for automated content analysis and the detection of markers of manipulative influence, clickbait and disinformation in news media texts. The proposed system combines methods of lexical-statistical text analysis with semantic modelling technologies using two large language model architectures (Llama 3 and Gemini 2.5 Flash), as well as a software module for linguistic decomposition and verification of source URLs against the registries of trusted media platforms. The functional structure of the developed application, based on the Streamlit framework, is described, and the results of a comparative analysis of its performance are presented.

Keywords: large language models, media manipulation, fact-checking, computational linguistics, Streamlit, linguistic markers.

Вступ

Сучасний інформаційний простір характеризується високим рівнем динамічності та значним поширенням дезінформації, клікбейту та прихованих маніпулятивних лінгвістичних технік. Традиційні інструменти контент-аналізу часто спираються виключно на фіксовані словники маркерів, що унеможлиблює глибоке оцінювання контексту, логічних зв'язків та прихованого підтексту повідомлень. З огляду на це актуальним науково-практичним завданням є створення інтелектуальних систем моніторингу та верифікації медіа-текстів на стику прикладної лінгвістики та технологій штучного інтелекту. Особливого значення це набуває в контексті розвитку формальних засобів оцінювання сенсу текстової інформації для вдосконалення архітектур великих мовних моделей [1].

Метою роботи є розробка двосистемного лінгвістичного веб-додатка, здатного здійснювати комплексний лексичний, семантичний та фактологічний аналіз новинних заголовків та статей з метою визначення їхнього підсумкового індексу маніпулятивності й авторитетності [2].

Результати дослідження

Розроблений інтелектуальний сервіс реалізовано у вигляді веб-додатка мовою програмування Python із використанням фреймворку Streamlit [3]. Логічна структура системи складається з трьох послідовних етапів обробки вхідного текстового масиву:

Первинний лексико-статистичний аналіз. Програмні алгоритми за допомогою регулярних виразів автоматично обчислюють формальні лінгвістичні параметри заголовка. Визначається точна кількість знаків оклику, наявність маркерів емоційного нагнітання та відсоток використання верхнього регістру (Caps Lock), що дозволяє виявити первинні структурні аномалії тексту.

Аналіз авторитетності першоджерела (Фактчекінг). Програмний модуль здійснює лінгвістичну декомпозицію та парсинг URL-адреси, яку вводить користувач. Система автоматично витягує доменне ім'я другого рівня та порівнює його із завантаженими реєстрами: загальнонаціональним «білим списком» авторитетних ЗМІ України (за даними ІМІ) та реєстром офіційних регіональних порталів [5]. На основі збігів розраховується числовий індекс верифікації джерела (від 0% до 100%).

Ансамблевий семантичний аналіз (Оцінка LLM). Зібрані технічні метрики (відсоток Caps Lock, кількість знаків оклику) та індекс верифікації інтегруються у структурований промпт. Аналіз здійснюється паралельно двома мовними моделями: локальною Meta Llama 3 (через фонову службу Ollama [4]) та хмарною Google Gemini 2.5 Flash (через офіційний API). Це дозволяє реалізувати метод експертного ансамблю для мінімізації ризику лінгвістичних галюцинацій штучного інтелекту.

Експериментальна верифікація та аналіз результатів

Для підтвердження ефективності та оцінювання точності розробленого програмного комплексу було проведено тестові випробування за допомогою полярних за своєю семантико-синтаксичною структурою медіа-повідомлень.

Експеримент №1. Аналіз верифікованого інформаційного повідомлення

Вхідний заголовок: «Масштабні зміни в центрі Вінниці: де встановлять знаки та кого пускатимуть на вул. Соборну» [6]. Введено також URL-адресу джерела та текст статті.

На етапі первинної лінгво-статистичної фільтрації програмний модуль зафіксував нульову щільність знаків оклику та мінімальний коефіцієнт девіації регістру (всього 4% використання верхнього регістру, зумовленого виключно правописом власних назв). Модуль верифікації джерела успішно здійснив декомпозицію URL-адреси, ідентифікувавши домен vitatv.com.ua у реєстрі довірених регіональних медіа-платформ, унаслідок чого підсумковий індекс верифікації першоджерела склав 100%.

Під час компаративного семантичного аналізу зафіксовано повну конвергенцію (узгодженість) експертних оцінок обох досліджуваних великих мовних моделей:

- Модель Meta Llama 3 визначила рівень маніпулятивності у 20%. У ході лінгвістичного аналізу модель зафіксувала відсутність емоційного тиску чи гіперболізації, зазначивши, що окремі звернення («Вінничан та гостей міста просять бути уважними...») виступають як логічні заклики до уваги пасажирів, а не маніпулятивні елементи. Фінальний вердикт моделі: «Надійне інформаційне повідомлення».
- Модель Google Gemini 2.5 Flash встановила індекс маніпулятивності на рівні 15%. Обґрунтовуючи результат, ШІ-експерт зазначив, що використання лексем «Масштабні» та постановка запитань у заголовку є елементами залучення уваги, проте контекст статті повністю розкриває ці «зміни» як планові ремонтні роботи значного обсягу, що трансформує потенційно клікбейтну структуру на об'єктивний інформаційний анонс. Лінгвістичний аналіз моделі підтвердив використання суто нейтральної та фактологічної лексики («технічні засоби», «схему обмежень», «оновлення столітнього водогону»). Фінальний вердикт моделі: «Надійне інформаційне повідомлення».

Загальний графічний інтерфейс розробленої системи та результати проведення Експерименту №1 наведено на рисунках 1 та 2.

Первинний технічний аналіз маркерів

Кількість знаків оклику

0

Відсоток КАПСЛОКУ у тексті

4%

Етап 1: Аналіз авторитетності джерела

Проводиться лінгвістичний аналіз доменного імені: vitatv.com.ua

Джерело vitatv.com.ua успішно верифіковано. Ресурс входить до реєстру авторитетних медіа або офіційних порталів.

Індекс верифікації першоджерела

100%

Рисунок 1 – Результати роботи модуля первинного аналізу та верифікації джерела для валідного тексту

Meta Llama 3

Рівень маніпулятивності: 20%

Моя оцінка базується на врахуванні індексу верифікації джерел в Інтернеті (100/100), що свідчить про надійність інформації. Крім того, відсутність Caps Lock (<5%) дозволяє зробити висновок про відсутність маніпулятивних прийомів.

Об'єктивний лінгвістичний аналіз: новина не містить емоційного тиску, гіперболізації або закликів до дії. Автори інформації надають об'єктивну інформацію про плани робіт та їх наслідки для дорожнього руху. Втім, є кілька речень, які можуть бути сприймані як заклик до уваги (наприклад, "Вінничан та гостей міста просять бути уважними..."). Однак ці фрази не містять маніпулятивних елементів.

Остаточний вердикт: Надійне інформаційне повідомлення

У підсумку, новина є надійним інформаційним повідомленням, оскільки вона базується на достовірних даних та надає об'єктивну інформацію про плани робіт. Автори не вживали жодних маніпулятивних прийомів або емоційного тиску, що дозволяє зробити висновок про їхню хоробрість і відданість інформуванню.

Google Gemini

Звіт про контент-аналіз новинного повідомлення

1. Рівень маніпулятивності:

15%

Обґрунтування: Рівень маніпулятивності оцінюється як низький. Високий індекс верифікації джерел (100 зі 100) свідчить про фактологічну достовірність повідомлення. Відсутність знаків оклику та мінімальний відсоток використання Caps Lock (4%) вказують на відсутність сенсаційності та емоційного тиску. Заголовок "Масштабні зміни в центрі Вінниці: де встановлять знаки та кого пускатимуть на вул. Соборну" містить елемент залучення уваги через використання слова "Масштабні" та постановку запитання. Проте, зміст статті розкриває ці "зміни" як планові ремонтні роботи значного обсягу, що включають оновлення дорожнього покриття та історичного водогону, а також надає конкретну інформацію щодо організації руху. Це мінімізує потенційну маніпулятивність заголовка, перетворюючи його з гіпотетичного клікбейту на інформаційний анонс.

2. Об'єктивний лінгвістичний аналіз:

Аналіз тексту демонструє відсутність яскраво виражених маніпулятивних прийомів. Лексика повідомлення є переважно нейтральною та фактологічною ("інформаційні плани", "технічні засоби", "схему обмежень", "планові роботи", "відновлення дорожнього покриття", "комплексним", "усунення деформацій", "оновлять ділянку історичного магістрального водогону").

Емоційний тиск не спостерігається. Звернення до громадян "Вінничан та гостей міста просять бути уважними, звертати увагу на дорожні знаки та заздалегідь планувати свої поїздки через центр" є типовим для попереджень про зміни в організації руху і не містить ознак психологічного впливу чи залякування. Воно формулюється як рекомендація, а не категорична вимога.

Елементи гіперболізації присутні в заголовку через використання слова "Масштабні", проте в контексті відновлення дорожнього покриття та оновлення стоїтного водогону в центральній частині міста, це визначення можна вважати відповідним значущості робіт. Гіперболізація не переходить у деінформацію.

Заклики до дії відсутні у формі маніпулятивного впливу. Фраза "просять бути уважними, звертати увагу... та заздалегідь планувати" є м'яким проханням, спрямованим на забезпечення безпеки та зручності учасників дорожнього руху, а не на спонукання до несанкціонованих або емоційно мотивованих дій.

3. Остаточний вердикт:

Надійне інформаційне повідомлення

Рисунок 2 – Експертна оцінка та лінгвістичні звіти моделей Llama 3 та Gemini 2.5 Flash

Експеримент №2. Аналіз деструктивного (маніпулятивного) медіа-тексту

Вхідний заголовок: «УВАГА! Нова Пошта дарує безкоштовні коробки з подарунками та грошові сертифікати кожному українцю на честь свого ювілею! Кількість обмежена!». Введено також URL-адресу джерела та текст статті.

Під час первинного лексико-статистичного аналізу програмний модуль виявив явні структурні та пунктуаційні аномалії: зафіксовано 3 знаки оклику та 6% використання верхнього регістру (Caps Lock), що свідчить про штучне нагнітання емоційного тону повідомлення. На Етапі 1 модуль декомпозиції URL-адрес виділив домен стороннього ресурсу povaroshka-gift.top. Оскільки зазначена платформа повністю відсутня у реєстрах офіційних організацій та білому списку ЗМІ, додаток згенерував попереджувальний статус «Незалежний або регіональний блог. Потребує додаткової контент-перевірки» та встановив низький індекс верифікації першоджерела — 40%.

На етапі інтелектуального аналізу ансамбль великих мовних моделей продемонстрував високу конвергенцію результатів, класифікувавши повідомлення як деструктивне:

- Модель Meta Llama 3 встановила рівень маніпулятивності на позначці 80%. ШІ обґрунтував висновок тим, що заголовок використовує три знаки оклику як спробу привернути надмірну увагу, а низький коефіцієнт контролю точності інформації (40%) вказує на ймовірний намір автора спонукати користувача до швидких дій. Модель ідентифікувала такі лінгвістичні прийоми, як емоційний тиск через знаки екстреності, гіперболізацію (маніпулювання обмеженою кількістю подарунків у кількості 47 одиниць для масштабу акції) та прямий заклик до дії, що змушує жертву приймати рішення без критичного осмислення. Остаточний вердикт моделі: «Маніпуляція».
- Модель Google Gemini 2.5 Flash зафіксувала критичний індекс маніпулятивності — 98%. Модель зазначила, що комбінація низької верифікації джерела, Caps Lock та надмірної пунктуації є прямим індикатором емоційного тиску. ШІ провів декомпозицію тексту та виділив такі лінгвістичні пастки:
 - Штучний дефіцит і терміновість: вирази «Кількість обмежена!», «Поспішайте, акція діє лише до кінця сьогоднішнього дня!» та «Залишилося всього 47 подарунків!».
 - Нереалістичні обіцянки: фраза про гарантоване отримання 5000 гривень за мінімальні дії («просто опитування»), що є прямою маніпуляцією та апелюванням до соціальної довіри з метою легітимізації пропозиції.
 - Заклик до дії: імперативи «Перейти за цим посиланням» та «Перейти до опитування та забрати свій приз», спрямовані на те, щоб змусити потенційну жертву діяти негайно, оминаючи етап раціональної перевірки інформації. Остаточний вердикт моделі: «Маніпуляція».

Графічні результати фіксації маніпулятивного та фішингового контенту у веб-додатку представлено на рисунках 3 та 4.

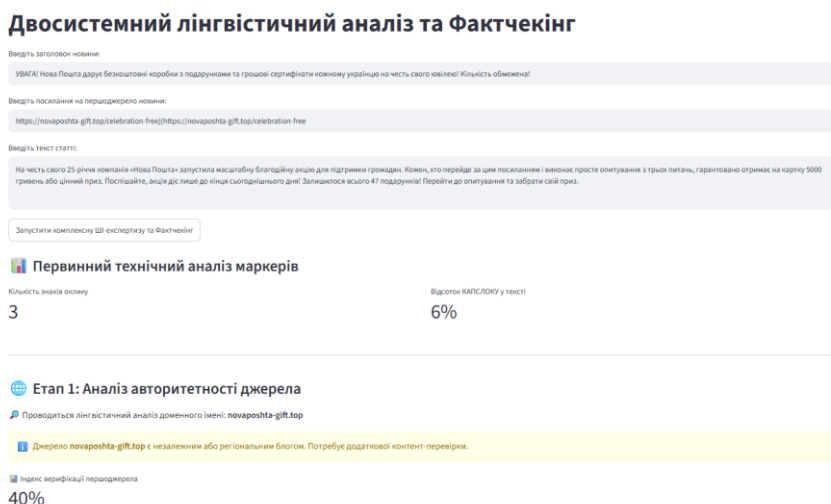


Рисунок 3 – Інтерфейс додатка під час лінгвістичного аналізу маркерів та домену фішингового ресурсу

© Етап 2: Експертна оцінка Великих Мовних Моделей

🐼 Meta Llama 3

Звіт про контент-аналіз новинного повідомлення:

1. Рівень маніпулятивності: 80%

Обґрунтування: хоча заголовок новини використовує три знаки оклику, що може бути ознакою спроби привертати увагу до інформації, індекс верифікації джерел в інтернеті становить лише 40%, що вказує на незначну кількість перевірок та контролю за точністю інформації. Крім того, Caps Lock використовується лише зрідка (6%), що є свідченням максимальної економії мовою. Враховуючи ці фактори, рівень маніпулятивності становить 80%, що вказує на можливу наміру автора новини спонукати до дії.

2. Об'єктивний лінгвістичний аналіз:

У тексті статті виявлено кілька маніпулятивних прийомів:

- Емоційний тиск: у заголовку використовуються знаки оклику, що створює відчуття екстреності та спонукання до дії.
- Гіперболізація: кількість подарунків, які залишилися, становить 47, що є досить незначною кількістю для масштабної благодійної акції.
- Заклик до дії: текст статті спонукає читачів до швидкої дії, вказуючи на обмежену кількість подарунків та необхідність їхнього отримання протягом одного дня.

3. Остаточний вердикт: Маніпуляція

У цілому, новина виглядає як маніпульоване повідомлення, спонукальне до швидкої дії з метою отримання грошового сертифіката або цінного призу. Незважаючи на обмежену кількість перевірок та контролю за точністю інформації, текст статті використовує різні маніпулятивні прийоми для привернення уваги до себе.

🌟 Google Gemini

Звіт з контент-аналізу новинного повідомлення

1. Рівень маніпулятивності: 98%. Даний показник обґрунтовується критично низьким індексом верифікації джерел (40 зі 100), що свідчить про ненадійність або фальсифікацію джерела інформації. Наявність значного відсотка тексту, написаного великими літерами (Caps Lock: 6%), та кількості знаків оклику (3) є індикаторами емоційного тиску та спроби привернути надмірну увагу, характерних для маніпулятивних повідомлень. Крім того, зміст повідомлення містить ознаки нереалістичних обіцянок та створення штучного дефіциту, що є класичними маніпулятивними прийомами.

2. Об'єктивний лінгвістичний аналіз: Повідомлення демонструє високий ступінь застосування маніпулятивних прийомів та емоційного тиску.

- **Заголовок:** Використання слова "УВАГА!" та великих літер створює ілюзію надзвичайної важливості. Обіцянка "безкоштовних коробок з подарунками та грошових сертифікатів кожному українцю" є гіперболізацією та нереалістичною обіцянкою, спрямованою на максимальне залучення аудиторії. Застереження "Кількість обмежена!" формує штучний дефіцит та підсилює терміновість.
- **Текст статті:** Фраза "масштабну благодійну акцію для підтримки громадян" апелює до соціальної відповідальності та довіри, легітимізуючи пропозицію. Обіцянка "гарантовано отримає на картку 5000 гривень або цінний приз" є прямою маніпуляцією, яка пропонує значну винагороду за мінімальні дії ("просто опитування"). Елементи терміновості та штучного дефіциту посилені фразами "Поспішайте, акція діє лише до кінця сьогоднішнього дня!" та "Залишилося всього 47 подарунків!". Ці вислови створюють відчуття невідкладності, змушуючи потенційну жертву діяти швидко, без критичного осмислення.
- **Заклик до дії:** Прямі заклики "Перейти за цим посиланням" та "Перейти до опитування та забрати свій приз" спонукають до негайної дії, мінаючи етап аналізу та перевірки інформації. Використання знаків оклику посилює емоційний тон та тиск.

3. Остаточний вердикт: Маніпуляція.

Рисунок 4 – Глибокий аналіз маніпулятивних прийомів у звітах Llama 3 та Gemini 2.5 Flash

Висновки

За результатами роботи було розроблено та протестовано інтелектуальний двосистемний веб-додаток для автоматизованого оцінювання якості медіа-контенту. Застосування технології паралельного ансамблевого аналізу моделями Llama 3 та Gemini у поєднанні з лінгво-статистичним підрахунком маркерів дозволило досягти високої точності у диференціації сухих інформаційних повідомлень від маніпулятивних та фейкових текстів.

Додавання модуля лінгвістичного аналізу та верифікації URL-адрес за реєстрами довірених доменів суттєво знизило рівень лінгвістичних галюцинацій моделей, усунуло інформаційний шум у видачі та підвищило об'єктивність експертних оцінок, що робить додаток перспективним інструментом для розгортання локальних систем автоматизованого фактчекінгу.

Подальші дослідження у напрямку виявлення фейкової інформації з метою маніпулятивного впливу планується проводити шляхом навчання великих мовних моделей за аналогією з дитячим когнітивним розвитком [7], а також на основі удосконалення великих мовних моделей формальними засобами оцінки сенсу текстової інформації [8].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Потапенко С. П. Сучасна медіа-лінгвістика: напрями та методи аналізу маніпулятивного дискурсу / С. П. Потапенко. – К. : Знання, 2022. – 184 с.
2. Дадиверін, В.; Бісікало, О. Огляд та аналіз архітектури recursive language models у контексті розпізнавання фейкової інформації. ВІТКП ВНТУ. Факультет інтелектуальних інформаційних технологій та автоматизації, Ukraine, feb. 2026. Available at: <<https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2026/paper/view/27421/22504>>. Date accessed: 03 Mar. 2026.
3. Streamlit Documentation [Електронний ресурс]. – Режим доступу: <https://docs.streamlit.io> (дата звернення: 02.06.2026).
4. Ollama Open Source Repository [Електронний ресурс]. – Режим доступу: <https://github.com/ollama/ollama> (дата звернення: 02.06.2026).
5. Мапа рекомендованих медіа та «Білий список» українських ЗМІ [Електронний ресурс] // Інститут масової інформації (ІМІ). – 2025. – Режим доступу: <https://imi.org.ua/monitorings/mapa-rekomendovanyh-media> (Дата звернення: 03.06.2026).
6. Масштабні зміни в центрі Вінниці: де встановлять знаки та кого пускатимуть на вул. Соборну [Електронний ресурс] // ТРК «ВІТА». – 2026. – Режим доступу: <https://vitatv.com.ua/misto/masshtabni-zminy-v-tsentri-vinnytsi-de> (дата звернення: 04.06.2026).
7. Дадиверін В. В., Бісікало О. В. Аналіз та імплементація ідеї навчання великих мовних моделей за аналогією з дитячим когнітивним розвитком. – Вчені записки Таврійського національного університету імені В.І. Вернадського, Серія: Технічні науки, Том 36 (75) № 4 2025, Частина 2. – СС.103-109. – ISSN 2663-5941.
8. Омельченко В., Бісікало О. Удосконалення великих мовних моделей формальними засобами оцінки сенсу текстової інформації // Матеріали XLVIII науково-технічної конференції підрозділів ВНТУ. – Вінниця : ВНТУ, 2026.

Бісікало Олег Володимирович – д-р техн. наук, професор, завідувач кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, email: obisikalo@vntu.edu.ua

Сердечна Анастасія Андріївна – студентка групи ІІСТ-24б, студентка групи ІІСТ-24б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: nastyaserdechna@gmail.com

Bisikalo Oleh V. – Dr. Sc. (Eng.), Professor, Head of the Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University, Vinnytsia, email: obisikalo@vntu.edu.ua

Serdechna Anastasia A. – student of group ІІСТ-24b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, email: nastyaserdechna@gmail.com