Система автоматизованої протидії інформаційним впливам на основі штучного інтелекту

УДК 004.8:004.056

Євгеній Волкотруб[1], Леонід Куперштейн[2]

Вінницький національний технічний університет,
[1]zhenvolkotrub@gmail.com, [2]kupershtein@vntu.edu.ua

Гібридна війна, яку росія веде проти України, включає не лише бойові дії на фронті, але й масштабні інформаційно-психологічні операції. Їхньою метою є формування вигідного для агресора інформаційного потоку, деморалізація суспільства та підрив довіри до державних інститутів. Соціальні мережі стали ключовим майданчиком для поширення дезінформації та маніпулятивного контенту. За даними EUvsDisinfo, лише протягом березня–листопада 2023 року дослідники зафіксували 596 рекламних повідомлень із дезінформацією та пропагандою, спрямованих на українську аудиторію [1]. Організація системної протидії таким впливам передбачає координацію дій та автоматизацію рутинних операцій [2].

Особливої уваги заслуговує платформа ok.ru (Одноклассники) — підсанкційна російська соціальна мережа, офіційно заблокована в Україні з 2017 року. Незважаючи на блокування, частина аудиторії продовжує відвідувати платформу через VPN-сервіси та проксі-сервери, а сама мережа залишається активним каналом поширення кремлівських наративів. Так, лише у 2023–2024 роках дослідники зафіксували десятки пропагандистських повідомлень антимобілізаційного характеру, розміщених саме в групах Одноклассників та VKontakte [1]. Уряд України та міжнародні організації фіксують, що попри формальне блокування, доступ до цих ресурсів зберігається для значної частини користувачів [3].

Традиційні методи ручного моніторингу та контрнаративу є ресурсомісткими і не дозволяють охопити весь масив шкідливого контенту в реальному часі. Великі мовні моделі (LLM) пропонують якісно новий підхід. Завдяки навчанню на великих корпусах текстів вони здатні розуміти контекст публікацій і генерувати природнзхомовні відповіді, наближені до людського стилю спілкування [4]. Інтеграція LLM у системи з веб-автоматизацією відкриває можливість масштабувати протидію на тисячі публікацій одночасно, що суттєво знижує операційне навантаження на фахівців з інформаційної безпеки.

Метою роботи є вдосконалення процесів протидії інформаційним впливам за рахунок розробки та впровадження автоматизованої інтелектуальної системи, яка у реальному часі аналізує контент соціальних мереж та генерує контекстно релевантні спростування або альтернативні наративи на основі генеративного штучного інтелекту.

Технологія протидії інформаційним впливам побудована за конвеєрним принципом і охоплює п'ять послідовних етапів, що утворюють замкнений цикл протидії (рис. 1).

Архітектура системи, що реалізує запропоновану технологію, включає такі складові: модуль веб-взаємодії, модуль генерації відповідей на основі LLM та

модуль управління акаунтами. Загальна архітектура передбачає конвеєрне опрацювання.

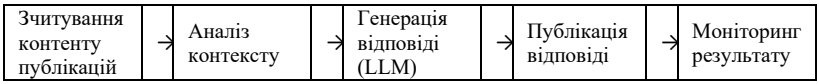

Рис. 1. Технологія застосування системи протидії інформаційним впливам

Модуль веб-взаємодії реалізовано на базі Selenium WebDriver із ChromeDriver. Для автентифікації в соціальній мережі застосовується механізм сесійних cookie, що дозволяє одночасно задіяти велику кількість акаунтів без повторного введення облікових даних. Для вилучення тексту публікацій застосовується багаторівневий парсинг за CSS-селекторами: основний текст публікації, підписи до медіа, тексти репостів та атрибути зображень. Це суттєво підвищує повноту аналізу, оскільки значна частина пропагандистського контенту поширюється у форматі зображень із текстовими підписами.

Для генерації відповідей використовується велика мовна модель Gemini від Google [5]. Gemini є мультимодальною моделлю, що підтримує розуміння тексту, зображень та структурованих даних. Доступ до моделі здійснюється через Gemini API, що забезпечує гнучку інтеграцію з автоматизованими системами без потреби у локальному розгортанні.

Ключовим елементом є спеціалізований промпт, що реалізує риторичну стратегію «Так, але…». Тобто модель спочатку демонструє поверхневу згоду з темою публікації (погода, кулінарія, свята), а потім переводить увагу читача на факти, які спростовують ворожий наратив. Така стратегія підвищує сприйнятливість аудиторії до контрнаративу, оскільки не сигналізує відразу про полемічний намір. Природність тексту додатково забезпечується повільним посимвольним введенням і випадковими затримками 3–6 с між публікаціями.

Модуль управління акаунтами забезпечує ротацію профілів із бібліотеки cookie-сесій. У разі недоступності LLM-сервісу система автоматично перемикається на резервну базу заздалегідь підготовлених відповідей, що гарантує безперервність процесу. Після кожної успішної публікації система зберігає знімок екрана для подальшого аудиту та верифікації.

Проведені випробування на платформі ok.ru підтвердили працездатність системи. Успішно опрацьовано понад 85% тестових публікацій із генерацією контекстно коректних відповідей. Середній час реакції на одну публікацію склав 8–12 секунд. Модуль моніторингу фіксує реакцію аудиторії на розміщені коментарі, що дозволяє оцінювати ефективність різних риторичних стратегій та коригувати промпт у реальному часі.

Розроблена система автоматизованої протидії інформаційним впливам поєднує можливості мовної моделі Gemini з технологіями веб-автоматизації та забезпечує протидію пропагандистських наративів у реальному часі. Конвеєрна п'ятиетапна технологія — від зчитування контенту до моніторингу результату — повністю автоматизує цикл протидії та знижує операційне навантаження на фахівців. Застосування риторичної стратегії підвищує природність та переконливість згенерованих відповідей. Масштабованість за рахунок багатоакаунтного підходу суттєво збільшує охоплення протидії. Перспективами

подальших досліджень є інтеграція мультимодальних LLM для аналізу зображень і відео, а також розробка модуля класифікації токсичного контенту.

1. EUvsDisinfo. How Russian Special Information Operations Try to Undermine Mobilisation in Ukraine. — 2024. URL: https://euvsdisinfo.eu/how-russian-special-information-operations-try-to-undermine-mobilisation-in-ukraine (дата звернення: 05.05.2026).
2. Kuperstein L. M., Lukichov V. V., Radetska A. O., Dudatyev A. V. System for Organizing Cyber Operations in the Context of Military Aggression // Science and Innovation. — 2025. — Vol. 21, № 3. — P. 86–98. https://doi.org/10.15407/scine21.03.086 (дата звернення: 05.05.2026).
3. Freedom House. Ukraine: Freedom on the Net 2023. — 2023. URL: https://freedomhouse.org/country/ukraine/freedom-net/2023 (дата звернення: 05.05.2026).
4. Baryshev Y., Kupershtein L., Maidanovych V., Voitovych O., Prokopenko S. Information System for the Fact-checker Support // CEUR Workshop Proceedings. — 2023. — Vol. 3646. — P. 127–138. URL: https://ceur-ws.org/Vol-3646/Paper_13.pdf (дата звернення: 05.05.2026).
5. Google. Gemini API Reference. — 2024. URL: https://ai.google.dev/gemini-api/docs (дата звернення: 05.05.2026).

Аналіз безпеки serverless-архітектур на основі моделювання подій

УДК 004.056:004.738.5 (043.2)

Петро Венгерський[1], Святослав Златоус[2]

Львівський національний університет імені Івана Франка,
[1]petro.venherskyy@lnu.edu.ua, [2]sviatoslav.zlatous@lnu.edu.ua

Сучасний розвиток хмарних обчислень зумовив широке впровадження serverless-архітектур, що базуються на концепції Function-as-a-Service (FaaS). Такий підхід дозволяє створювати масштабовані програмні системи без необхідності управління серверною інфраструктурою, що значно спрощує процес розробки та експлуатації додатків [1]. Водночас використання serverless-архітектур супроводжується появою нових викликів у сфері кібербезпеки, які пов'язані з динамічністю виконання функцій, складною структурою взаємодій між компонентами системи та обмеженим контролем користувачів над середовищем виконання [3]

У сучасних дослідженнях безпеки serverless-систем основна увага приділяється окремим аспектам їх функціонування, зокрема ізоляції функцій, управлінню доступом та аналізу продуктивності [3]. Водночас питання комплексного аналізу поведінки системи, що враховує взаємодії між функціями, подіями та сервісами хмарної інфраструктури, залишається недостатньо дослідженим [2]. Це ускладнює виявлення потенційних загроз безпеці, які можуть виникати у результаті нетипових сценаріїв функціонування системи.

У роботі запропоновано підхід до аналізу безпеки serverless-архітектур, що базується на дослідженні подій та взаємодій між компонентами системи. Основна ідея підходу полягає у використанні централізованого моніторингу