

УДК 531.7

DOI: 10.18372/2073-4751.86.21282

Спірідонов А. Ю.,
orcid.org/0009-0005-4071-1733,
e-mail: 2386932@stud.nau.edu.ua

КОНЦЕПЦІЯ ТЕХНОЛОГІЇ АНАЛІЗУ ІНФОРМАЦІЙНИХ ПОТОКІВ У РАМКАХ ОПЕРАЦІЙ OSINT У СИСТЕМАХ МИТТЄВОГО ОБМІНУ ПОВІДОМЛЕННЯМИ

Національний університет «Київський авіаційний інститут»

Вступ

Сучасні збройні конфлікти дедалі частіше характеризуються ключовою роллю інформації, де здатність збирати, обробляти та аналізувати дані безпосередньо впливає на успіх операцій. Одним із ключових інструментів у цій сфері є розвідка з відкритих джерел (OSINT) – збір, аналіз та інтерпретація інформації з загальнодоступних джерел. Війна в Україні продемонструвала безпрецедентні масштаби використання OSINT як військовими структурами, так і цивільними ініціативами.

Від супутникових знімків та автоматичної геолокації фотографій до моніторингу соціальних мереж та аналізу відкритих баз даних – сьогодні OSINT дає змогу швидко збирати розвідувальну інформацію, викривати військові злочини, протидіяти дезінформації та підвищувати прозорість на полі бою. Досвід України свідчить, що систематичне використання відкритих джерел може впливати як на тактичний, так і на стратегічний рівні ведення війни, встановлюючи нові стандарти розвідки 21 століття.

Однак зростання обсягу доступних даних ставить завдання щодо обробки високоякісної інформації майже в режимі реального часу. Традиційні методи, засновані на ручному аналізі, не можуть забезпечити результати з необхідною швидкістю та масштабованістю, особливо в системах миттєвого обміну повідомленнями (наприклад, Telegram). У цьому контексті **гіперавтоматизація** – поєднання RPA, AI, ML та пов'язаних технологій – стає необхідною для комплексної, адаптивної та самонавчальної обробки даних.

У цьому дослідженні проводиться емпірична оцінка ефективності гіперавтоматизованих систем у операціях OSINT, при цьому основна увага приділяється синергії між зазначеними автоматизованими підсистемами з метою підвищення продуктивності, точності та адаптивності.

Експериментальне моделювання дозволяє оцінити запропоновану архітектуру інтелектуальної обробки інформації з точки зору економії часу, мінімізації людських помилок та підвищення надійності прийняття рішень у практиці OSINT

Основна частина

Опис запропонованої моделі обробки інформації

1. Запропонована модель інтелектуальної обробки інформації базується на принципах гіперавтоматизації та включає такі основні компоненти [3] (рис. 1):

2. Автоматизований збір даних – використання робототехнічних систем для збору інформації з різних джерел, зокрема через API, веб-скрейпінг та інтеграцію з базами даних.

3. Попередня обробка даних – очищення, нормалізація та узагальнення зібраної інформації для точного аналізу.

4. Аналітичний модуль – застосування методів штучного інтелекту (ШІ) та машинного навчання (МН) для виявлення закономірностей, прогнозування та формування висновків.

5. Система інтеграції та зворотного зв'язку – адаптація та навчання моделі на основі нових даних для підвищення довгострокової ефективності.

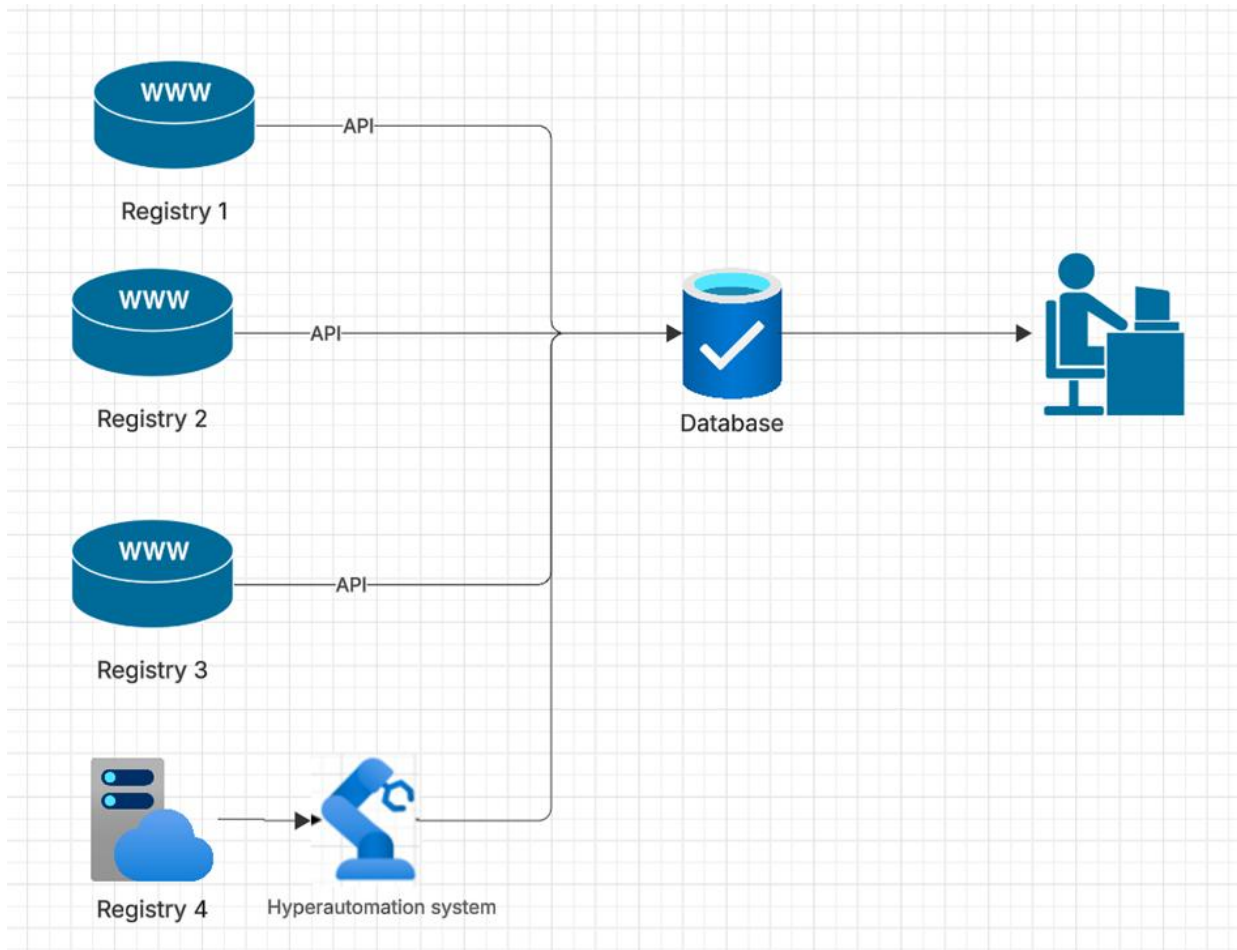


Рис. 1. Пропонована концептуальна архітектура моделі.
Архітектурна модель інформаційних технологій

Система збору

Першим етапом цієї технології є система збору даних, яка реалізує автоматизований підхід до інтеграції різнорідних джерел. Вона постійно взаємодіє з базами даних, динамічно імпортує структуровані та неструктуровані дані, а також здійснює пошук релевантної інформації у відкритих цифрових джерелах. Зібрані дані складаються з текстових масивів, числових показників, мультимедійного контенту та спеціалізованих метаданих, які підлягають подальшій верифікації та структурованій обробці.

Модуль обробки даних

Отримані дані проходять перевірку, фільтрацію, очищення, структурування та агрегацію з метою підготовки інформації до подальшого аналізу та інтелектуальної обробки, що дозволяє мінімізувати

можливі помилки та зменшити рівень шуму у вихідних даних.

Аналітичний механізм

Цей компонент, що відповідає за аналіз даних та виокремлення корисної інформації, використовує статистичні методи, алгоритми машинного навчання та моделі глибокого навчання для виявлення закономірностей, прогнозування тенденцій та визначення прихованих взаємозв'язків.

Архітектура інформаційних технологій для інтелектуальної обробки даних на основі гіперавтоматизації забезпечує комплексний підхід до аналізу даних. Вона ефективно поєднує автоматизацію процесів із застосуванням інтелектуального аналізу та прийняття рішень, що сприяє підвищенню ефективності та продуктивності обробки інформації.

Для підтвердження запропонованої архітектури було проведено експеримент, результати якого наведено в наступному розділі.

Математичний опис процесу гіперавтоматизованої обробки даних

Нехай $D = \{d_1, d_2, \dots, d_n\}$ позначає множину текстових повідомлень, отриманих із цифрових інформаційних каналів (зокрема, Telegram), де n – загальна кількість повідомлень, а кожне $d_i \in D$ представляє окремий документ у вигляді текстового масиву.

Векторизація вмісту: Кожне повідомлення $d_i \in D$ перетворюється на вектор ознак $x_i \in R^m$ за допомогою функції перетворення $T(\cdot)$ [4]:

$$x_i = T(d_i). \quad (1)$$

де m – розмірність простору ознак (лексем, лексичних одиниць, семантичних структур).

Висновок на основі нейронної моделі: де $f_\theta(x_i)$ – це функція, що моделює роботу великої мовної моделі з параметрами θ , яка генерує вектор вихідних даних $y_i \in R^k$

$$y_i = f_\theta(x_i). \quad (2)$$

Вектор y_i містить семантичні теги, тематичні ваги або інші атрибути класифікації.

Опис експериментальних досліджень

Першим етапом дослідження був вибір методології, яка б найбільш ефективно відповідала поставленим завданням. Було застосовано поєднання кількох методів, зокрема аналіз даних реальних виробничих процесів, експерименти з тестовими наборами даних та опитування фахівців у відповідних галузях.

Експеримент має на меті зібрати історичні дані з каналів месенджера Telegram, завантажити їх у базу даних, а також забезпечити, щоб система штучного інтелекту на основі лінгвістичної моделі LLAMA 3 генерувала відповіді на запити щодо вказаного діапазону дат.

З метою перевірки гіпотези щодо обробки великого обсягу новин із месенджерів було проведено експеримент[7] відповідно до процедури, показаної на рис. 2

Використані технології:

UiPath, LLAMA 3, Python та JavaScript. Вибір цих технологій зумовлений їхніми можливостями в галузі автоматизації, аналітики та обробки великих обсягів даних:

UiPath забезпечує ефективну роботизовану автоматизацію бізнес-процесів[5].

LLAMA 3 використовується як потужна лінгвістична модель для обробки текстових запитів.

Для забезпечення обчислювальної потужності було використано таке обладнання: графічний процесор NVIDIA T4[9] для ефективної роботи нейронних мереж та віртуальний сервер з 16 ГБ оперативної пам'яті для виконання складних обчислень та обробки великих обсягів даних. Така конфігурація забезпечила швидку обробку запитів та підвищила ефективність використання ресурсів.

UiPath забезпечує ефективну роботизовану автоматизацію процесів, LLAMA 3 виступає надійною лінгвістичною моделлю для обробки текстових запитів, Python надає гнучкість у роботі з даними та інтеграції алгоритмів машинного навчання, а JavaScript використовується для створення веб-інтерфейсу та інтерактивних елементів взаємодії з користувачем.

Слід зазначити, що в робочому режимі оновлення історії даних має здійснюватися регулярно. Використовується одномовна модель, яка потребує розширення до багатомовних систем; проте її застосування в експерименті дозволило оптимізувати використання ресурсів.

З огляду на сучасні загрози та виклики, цю технологію можна застосовувати для виявлення дезінформації та аналізу інформаційних потоків супротивника[6]. З цією метою під час експерименту використовувалися дані, зібрані з російськомовних інформаційних ресурсів. Таке рішення дозволило оцінити ефективність системи в контексті інформаційної безпеки та аналізу впливу.

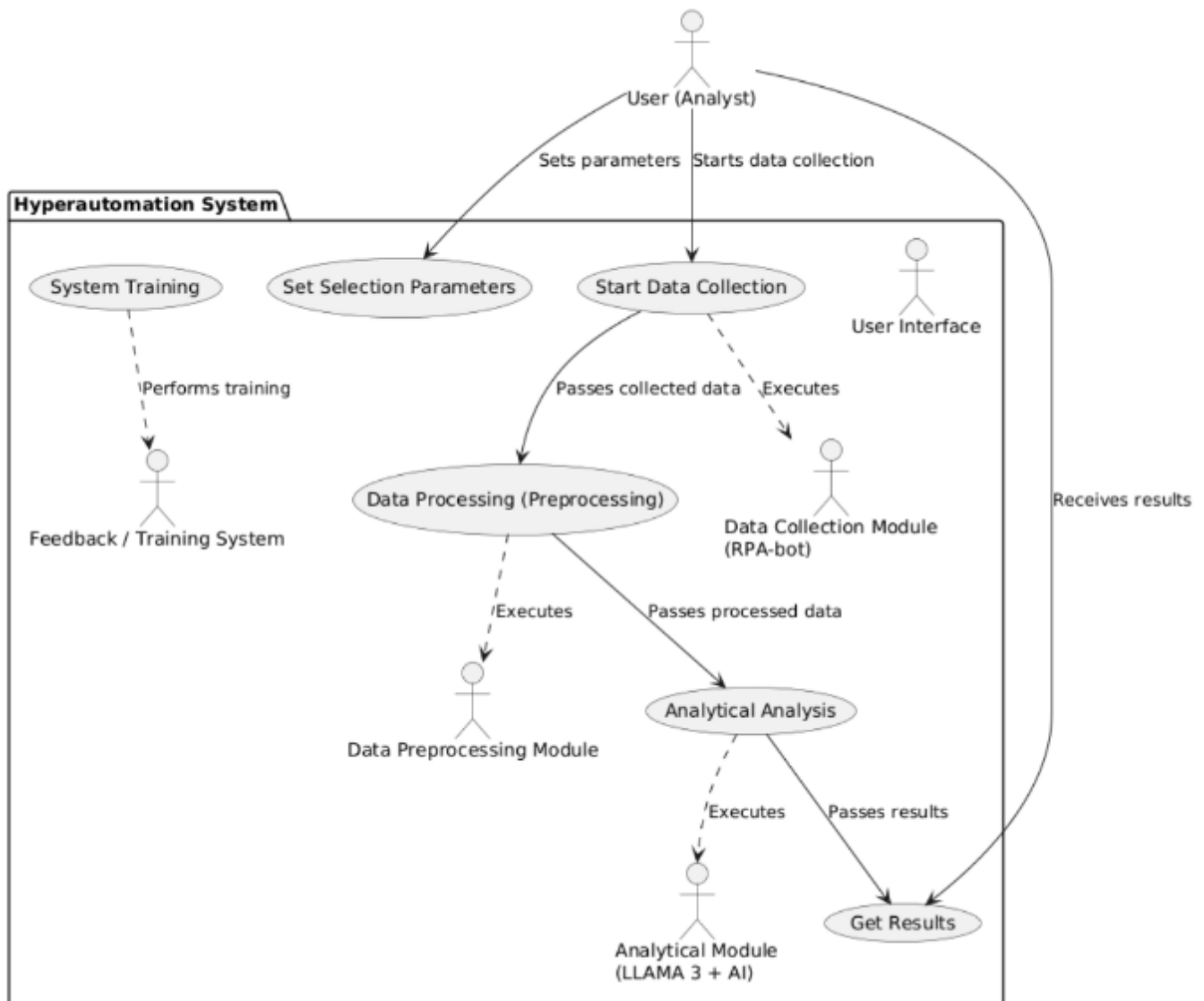


Рис. 2. Схематичний опис ходу експерименту

Опис експерименту

Вхідні параметри були реалізовані за допомогою інтерактивного веб-інтерфейсу, який давав користувачам можливість встановлювати критерії відбору даних (рис.3):

- **Період:** Для аналізу було обрано період з 01.08.2024 по 03.10.2024.

- **Джерела даних:** у ході експерименту використовувалися два новинні канали Telegram – «ТСН» та «УНІАН – новини України».

Після налаштування параметрів система автоматично розпочала процес збору даних.

Бот RPA, розроблений у UiPath, витягував інформацію із зазначених каналів Telegram і зберігав її у форматі JSON (рис.4).

Час виконання RPA-бота склав 14,7 хвилин, що значно швидше, ніж ручний збір даних.

Всього було оброблено 1472 повідомлення.

Зібрані історичні дані були збережені у структурованому форматі для подальшої обробки.


```

> Question:
Какая инфо у тебя есть по восстановлению охматдета

> Answer:
Из предоставленного текста известно следующее о восстановлении Охматдета:

1. На восстановление Охматдета было собрано более 1 млрд гривен (около $22,2 млн).
2. Средства были собраны на благотворительный счет больницы Охматдет, счет благотворительного фонда Охматдет, через платформу UNITED24 и из резервного фонда
3. Суммы, которые получило Охматдет, составляют:
    * 312 млн грн на благотворительный счет больницы Охматдет;
    * 378 млн грн на счет благотворительного фонда Охматдет;
    * 320 млн грн через платформу UNITED24;
    * 100 млн грн от правительства из резервного фонда.
4. Благодаря средствам был восстановлен реабилитационный детский бассейн в больнице.
5. Планировалось запустить программу бесплатной реабилитации детей с ампутированными конечностями, детей военных, спасателей и других.
6. Однако из-за психологического давления и хейта, возникшего из-за ситуации с тендером, на сайте фонда появилось изменение – вместо 11 человек теперь указ
7. Руководитель фонда Охматдет Евгений Вембер заявляет, что намерен обращаться к правоохранительным органам, если фонд не вернет деньги.
8. Предполагается, что тендер на восстановление Охматдета будет отменен и заменен на другой, проводимый через систему Prozorro.
-----SOURCE DOCUMENTS-----

```

Рис. 6. Приклади відповідей моделей LLM

```

** > Question:
кто имеет право на отсрочку от мобилизации

> Answer:
Из предоставленного текста следует, что следующие люди имеют право на отсрочку от мобилизации:

1. Военнообязанный, чьи близкие родственники погибли или пропали без вести во время участия в боевых действиях во время действия военного положения.
2. Мужчина, воспитывающий своего ребенка, если его супруга находится на военной службе.
3. Мужчина, который содержит своих детей.
4. Человек, состоящий в фактическом браке, может требовать раздела совместно приобретенного имущества по гражданским законам.
-----SOURCE DOCUMENTS-----

> /content/localGPT/SOURCE_DOCUMENTS/TG history.txt:
Законопроект, который обеспечивает право отсрочки от призыва на военную службу по мобилизации для военнообязанных, чьи близкие родственники погибли или проп
В рамках украинского законодательства неполнородным братом или сестрой является человек, родившийся во время пребывания матери или отца в другом браке.
13:13

```

Рис. 7. Приклади відповідей моделі LLM

Таблиця 1. Точність та релевантність контексту (запит і відповідь перекладені для цього документа, а спочатку оброблялися російською мовою

Запит	Відповідь LLAMA 3	Як
Які новини були найпопулярнішими минулого тижня?	1. Дональд Трамп: Вибори у США. 2. Ракети «Шахед» безперешкодно проникають у повітряний простір НАТО – ці новини мають найбільшу кількість переглядів.	Відповідь правильна, модель правильно визначила основні теми
Які останні новини про Охматдит?	На відновлення «Охматдиту» було зібрано понад 1 мільярд гривень (близько 22,2 мільйона доларів). – 2. Кошти надійшли на благодійний рахунок лікарні «Охматдит» та на рахунок благодійного фонду «Охматдит» через платформу UNITED24, а також 3. Суми, отримані компанією «Охматдит», становлять: 312 мільйонів гривень на благодійний рахунок лікарні «Охматдит»; 378 млн грн на рахунок благодійного фонду «Охматдит»; * 320 млн грн через платформу UNITED24; 100 мільйонів гривень від уряду з резервного фонду. 4. Завдяки цим коштам було відремонтовано дитячий басейн у лікарні.	Відповідь правильна, модель правильно визначила основні теми
Хто має право на відстрочку від призову до збройних сил?	1. Призовник, чії близькі родичі загинули або зникли безвісти під час участі у бойових діях у період дії воєнного стану. 2. Чоловік, який виховує дитину, якщо його дружина перебуває на військовій службі. 3. Чоловік, який утримує своїх дітей. 4. Особа, яка перебуває у фактичному шлюбі, може вимагати поділу спільно нажитого майна відповідно до цивільного законодавства.	Відповідь класифіковано правильно, але можна було б додати більше деталей

Оцінка продуктивності:

Оцінка продуктивності ґрунтується

на порівнянні витрат часу з показниками роботи оператора-людини (табл. 2).

Таблиця 2. Порівняння обробки одного й того самого набору даних людиною та запропонованою моделлю:

Параметри	Система (гіперавтоматизація)	Людина	Як
Час обробки	16 хв. 983 сек.	4–6 годин	Значна економія часу завдяки автоматизації.
Точність аналізу	Високий (98 %)	Залежить від кваліфікації (~70–85 %)	Алгоритми штучного інтелекту забезпечують кращу розпізнаваність образів.
Обсяг оброблених даних	До 1 мільйона повідомлень	Обмежена кількість (приблизно 500–1000 повідомлень)	Система здатна аналізувати великі масиви даних за один цикл.
Гнучкість та здатність до адаптації	Автоматична адаптація	Потрібне навчання	Машинне навчання забезпечує постійне вдосконалення.
Вартість	Одноразове впровадження, низькі експлуатаційні витрати	Значні витрати на оплату праці	Автоматизація дозволяє зменшити витрати в довгостроковій перспективі.
Імовірність помилки	Низький (1–5 %)	Середній (5–10 %)	Алгоритми мінімізують людські помилки.
Ефективність при виконанні повторюваних завдань	Висока ефективність	Продуктивність знижується через втому	Система ефективно виконує рутинні завдання.
Здатність приймати рішення	Рекомендації на основі аналізу	Залежить від досвіду	Система надає обґрунтовані висновки для прийняття рішень.

$t_h = 360$ хв – середній час ручної обробки одного запиту фахівцем;

$t_p = 14,7$ хв – постійний час попередньої обробки (операція RPA), що виконується один раз під час завантаження даних у базу даних;

$t_r = 2,2$ хв – час роботи великої мови (завантаження в базу даних та обробка моделлю), також виконується один раз;

$t_e = 5$ с = 0,083 хв – час обробки одного запиту (відображення даних оператору).

Загальний час роботи системи:

$$t_s = t_p + t_r + t_e = 16,983 \text{ хв.}$$

Економія часу:

$$\Delta t = t_h - t_s = 343,017 \text{ хв.}$$

$$E_t = \frac{1}{n} \sum \Delta t_i = 343,017 \text{ хв.}$$

Точність класифікації:

$$\text{Точність} = \frac{1}{n} \sum I[\text{argmax}(y_i) = y_i],$$

де I – індикаторна функція.

Інтегральна ефективність:

$$IE = \alpha \cdot \text{Точність} + \beta \cdot \left(\frac{E_t}{t_h}\right), \alpha + \beta = 1$$

Емпіричний приклад:

$n = 1472$, Точність = 0,98, $t_h = 360$ хв,
 $t_s = 16,983$ хв, $E_t = 343,017$ хв.

$$IE = 0,6 \cdot 0,98 + 0,4 \cdot \left(\frac{343,017}{360}\right) \approx 0,969$$

Окремий етап експерименту було проведено для перевірки адаптивності системи та її здатності класифікувати корисну інформацію. Дані було оброблено з метою пошуку потенційних систем протиповітряної оборони країни-агресора на основі новинних повідомлень.

Запит: «Проаналізувати перелік новин, доданих до файлу, та сформуванати зведення із зазначенням міст у Росії, де відбулися удари БПЛА (у тому числі уламками або ракетами). Формат відповіді: Дата – Місто – Район (або адреса, за наявності) – Тип атаки (БПЛА, ракета тощо). Включати лише новини про події на території Росії.» На основі сформованого запиту система проаналізувала набір даних, зібраний на попередньому етапі, використовуючи попередньо оброблені та збережені дані для пришвидшення обробки, та сформувала перелік відповідей у форматі JSON. (рис. 8).

```
{
  "news_summary": [
    {
      "date": "4 August 2024",
      "city": "Липецк",
      "region": "Липецкая область",
      "attack_type": "БПЛА"
    },
    {
      "date": "6 August 2024",
      "city": "Морозовск",
      "region": "Ростовская область",
      "attack_type": "БПЛА"
    },
    {
      "date": "6 August 2024",
      "city": "Белгород",
      "region": "Белгородская область",
      "attack_type": "БПЛА"
    },
    {
      "date": "6 August 2024",
      "city": "Курск",
      "region": "Курская область",
      "attack_type": "БПЛА"
    },
    {
      "date": "7 August 2024",
      "city": "Севастополь",
      "region": "Крым",
      "attack_type": "Неизвестно"
    }
  ]
}
```

Рис. 8. Приклад відповіді моделі LLM у заданому форматі виводу

Для наочності результати було відображено на карті (рис. 9), де:

- Назва позначки – дата новини
- Кольорове кодування:
 - фіолетовий – БПЛА,
 - синій – ракета,
 - червоний – невідомо

Запропонована архітектура, що поєднує роботизовану автоматизацію процесів (RPA), методи машинного навчання та великі мовні моделі (LLM), демонструє здатність істотно скоротити час виконання завдань (з 6 годин до 17 хвилин) при високій точності аналізу (98 %). Це дозволяє ефективно виявляти ключові закономірності, тенденції та приховані залежності у відкритих інформаційних потоках, а також оперативно формувати відповідні аналітичні звіти.

Порівняння з традиційними методами продемонструвало переваги гіперавтоматизованого підходу в плані масштабованості, стійкості до людських помилок, здатності обробляти багатопотокові дані та мінімізації впливу людського фактора.

Система довела свою ефективність як у вирішенні стандартних аналітичних завдань, так і в спеціалізованих сценаріях – наприклад, у виявленні ознак розташування ворожих систем ППО за даними з відкритих джерел.

Водночас результати дослідження вказують на необхідність подальшого вдосконалення технологій. Серед визначених напрямків для вдосконалення:

- розширення багатомовних можливостей моделей для обробки багатомовних інформаційних потоків;
- розширення та оптимізація багатопотокової обробки;
- розширення набору даних для навчання моделей LLM з метою виявлення та фільтрування дезінформації;
- розробка методів для підвищення стійкості системи до кіберзагроз та забезпечення етичного використання даних.

Таким чином, гіперавтоматизація у поєднанні з інструментами машинного навчання має значний потенціал для підвищення ефективності прийняття рішень у контексті інформаційної війни.

Література

1. Гудфеллоу І., Бенджо Й., Курвіль А. Глибоке навчання. MIT Press, 2016. 800 с. URL:

<https://www.deeplearningbook.org/> (дата звернення: 07.05.2026).

2. Основні стратегічні технологічні тренди 2020 року: гіперавтоматизація : Gartner Newsroom. 2020. URL: <https://www.gartner.com/en/newsroom/press-releases/2019-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2020> (дата звернення: 07.05.2026).

3. Соммервілл І. Програмна інженерія. 10-те вид. Pearson Education, 2015. 816 с. ISBN 978-0-13-394303-0.

4. Джурафські Д., Мартін Дж. Г. Обробка мовлення та мови. 3-тє вид. (чернетка). Стенфордський університет, 2021. URL: <https://web.stanford.edu/~jurafsky/slp3/> (дата звернення: 07.05.2026).

5. Приклади застосування та переваги RPA : UiPath Resources. 2022. URL: <https://www.uipath.com/resources/automation-use-cases> (дата звернення: 07.05.2026).

6. Виявлення фейкових новин та дезінформації : Центр передового досвіду НАТО зі стратегічних комунікацій. 2020. URL: <https://www.stratcomcoe.org/> (дата звернення: 07.05.2026).

7. Аналіз джерел даних Telegram у рамках розвідки з відкритих джерел : Bellingcat. 2023. URL: <https://www.bellingcat.com/resources/2023/telegram-osint-guide/> (дата звернення: 07.05.2026).

8. Технічний звіт про GPT-4 : OpenAI. 2023. URL: <https://cdn.openai.com/papers/gpt-4.pdf> (дата звернення: 07.05.2026).

9. Архітектура графічного процесора NVIDIA T4 : NVIDIA. 2021. URL: <https://www.nvidia.com/en-us/data-center/tesla-t4/> (дата звернення: 07.05.2026).

10. Бішоп К. М. Розпізнавання образів та машинне навчання. Springer, 2006. 738 с. ISBN 978-0387310732.

11. Провост Ф., Фосетт Т. Наука про дані для бізнесу. O'Reilly Media, 2013. 414 с. ISBN 978-1449361327.

Спірідонов А. Ю.

КОНЦЕПЦІЯ ТЕХНОЛОГІЇ АНАЛІЗУ ІНФОРМАЦІЙНИХ ПОТОКІВ У РАМКАХ ОПЕРАЦІЙ OSINT У СИСТЕМАХ МИТТЄВОГО ОБМІНУ ПОВІДОМЛЕННЯМИ

У цьому документі представлено результати емпіричного дослідження ефективності гіперавтоматизованих інформаційних систем при обробці великих обсягів даних під час операцій з відкритих джерел розвідки (OSINT). Основна увага приділяється розробці та експериментальній перевірці концептуальної архітектури технології OSINT нового покоління, що базується на інтеграції: роботизованої автоматизації процесів (RPA), інструментів машинного навчання та великих мовних моделей (LLM, зокрема LLAMA 3). Ця технологія застосовується для збору даних із систем миттєвого обміну повідомленнями. Описано ключові етапи конвеєра гіперавтоматизації: збір, попередня обробка, аналіз та класифікація даних із каналів Telegram. Для автоматизованого вилучення новин використовувалися боти RPA, після чого відбувалася обробка лінгвістичною моделлю для формування інформативних звітів.

Експеримент продемонстрував значне скорочення часу обробки даних (з 6 годин до 17 хвилин) при збереженні високої точності (98%). Результати було порівняно з ручною обробкою, а також оцінено потенціал запропонованого підходу в галузі інформаційної безпеки та виявлення дезінформації. Отримані дані свідчать про значний потенціал гіперавтоматизації в сфері інтелектуальної обробки інформації та прийняття рішень на основі даних.

Ключові слова: Відкриті джерела інформації (OSINT), кібербезпека, гіперавтоматизація, інтелектуальна обробка інформації, великі мовні моделі, LLAMA 3, роботизована автоматизація процесів (RPA), машинне навчання, аналіз даних, штучний інтелект

Anton Spiridonov

CONCEPT OF INFORMATION FLOW ANALYSIS TECHNOLOGY WITHIN OSINT OPERATIONS IN INSTANT MESSAGING SYSTEMS

This document presents the results of an empirical study on the efficiency of hyper-automated information systems in processing large volumes of data during Open Source Intelligence (OSINT) operations. The focus is on the development and experimental validation of a conceptual architecture for a next-generation OSINT technology based on integration: Robotic Process Automation (RPA), machine learning tools, and large language models (LLM, particularly LLAMA 3). The technology is applied to data collection from instant messaging systems. The key stages of the hyper-automation pipeline are described: collection, preprocessing, analysis, and classification of data from Telegram channels. RPA bots were used for automated news extraction, followed by linguistic model processing to generate informative reports.

The experiment demonstrated a significant reduction in data processing time (from 6 hours to 17 minutes) while maintaining high accuracy (98%). The results were compared with manual processing and the potential for the proposed approach in information security and disinformation detection was evaluated. The findings indicate the strong potential of hyper-automation in intelligent information processing and data-driven decision-making.

Keywords: Open Source Intelligence, OSINT, cyber security, hyper-automation, Intelligent Information Processing, Large Language Models, LLAMA 3, Robotic Process Automation (RPA), Machine Learning, Data Analysis, Artificial Intelligence.

Стаття подана до редакції: 07/05/2026

Стаття прийнята до опублікування: 15/05/2026

Стаття опублікована: 30/05/2026

Стаття поширюється на умовах ліцензії CC BY 4.0