https://doi.org/10.31891/2307-5732-2025-355-65
УДК 004.89:004

ПЕРЕГУДА ЯРОСЛАВ
Київський політехнічний інститут ім. Ігоря Сікорського
https://orcid.org/0009-0002-7292-7887
e-mail: findershtein@gmail.com
ЛЮШЕНКО ЛЕСЯ
Київський політехнічний інститут ім. Ігоря Сікорського
https://orcid.org/0000-0003-4319-5955
e-mail: lyushenkol@gmail.com

АЛГОРИТМИ КАТЕГОРИЗАЦІЇ ТА ТЕМПОРАЛЬНОГО АНАЛІЗУ БОТ-ПРОГРАМ В СОЦІАЛЬНИХ МЕРЕЖАХ

В роботі досліджується структура та динаміка активності користувачів у соціальному медіа-просторі з акцентом на виявлення автоматичної поведінки бот-програм. Для цього були створені два спеціалізовані алгоритми: категоризації користувачів за типом акаунта та темпорального аналізу їхньої активності. Застосування цих підходів до об'єднаного набору даних, що охоплює період з 2017 по 2023 рік, дозволило простежити як зміну частки шкідливих бот-програм у часі, так і особливості їх активності в різних індустріях та галузях діяльності. Результати показали, що певні індустрії є значно більш уразливими до штучного впливу. Запропоновані алгоритми можуть бути корисними для аналізу інформаційної стійкості, виявлення аномалій та удосконалення систем цифрової безпеки.

Ключові слова: *бот-програми, шкідлива автоматизація, соціальні мережі, аналіз активності, категоризація, темпоральний аналіз.*

PEREHUDA YAROSLAV
LYUSHENKO LESYA
Igor Sikorsky Kyiv Polytechnic Institute

ALGORITHMS FOR CATEGORIZATION AND TEMPORAL ANALYSIS OF BOT-PROGRAMS IN SOCIAL NETWORKS

Recent studies confirm the growing share of bot-driven activity in digital communication environments. According to Imperva's 2024 report, nearly half of web traffic is generated by automated agents, with over 30% attributed to malicious bot-programs. The accessibility of tools based on generative AI has contributed to this growth, enabling even non-specialists to launch automated scraping and spam campaigns. Simultaneously, Cloudflare data show that bot-programs target critical areas such as finance and e-commerce, causing measurable harm to businesses and undermining user trust. Academic research highlights both the increasing sophistication of bot-programs and the emergence of new detection strategies, including behavior-based modeling and machine learning methods. However, many existing approaches remain limited in their ability to integrate structural and temporal dimensions of user activity.

The study aims to analyze behavioral characteristics of users in digital communication platforms, with a focus on the distinction between human-operated and automated accounts. The goal is to trace the distribution and evolution of various account types across time and thematic sectors, as well as to identify sectors most vulnerable to the influence of coordinated bot-program activity.

To achieve this, the research introduces two purpose-built algorithms: one for user categorization and another for temporal activity analysis. The categorization algorithm accommodates both pre-labeled and raw user data, using heuristic thresholds and public bot-likelihood metrics to assign accounts to one of three groups: humans, benign bot-programs, and malicious bot-programs. The temporal analysis algorithm models user activity over time by comparing expected versus actual interaction levels per group, enabling the identification of overstimulated or understimulated behavior patterns. These algorithms were applied to a dataset comprising the TwiBot-2022 benchmark and a manually compiled corpus of tweets from 2023. The resulting analysis tracks the changing distribution of user types between 2017 and 2023 and maps their activity across multiple thematic sectors.

The findings reveal a noticeable rise in malicious bot-programs activity between late 2019 and 2020, coinciding with the outbreak of the COVID-19 pandemic and the U.S. presidential election—both of which are known triggers for disinformation campaigns. Furthermore, sectoral analysis shows disproportionate bot-programs involvement in domains such as politics, marketing, and financial services. Meanwhile, categories such as entertainment and lifestyle remain dominated by human activity. The study demonstrates the effectiveness of the proposed algorithms in capturing dynamic and structural aspects of online behavior and contributes new tools for the monitoring of digital ecosystems and the mitigation of information threats.

Keywords*: bot-programs, malicious automation, social networks, activity analysis, categorization, temporal analysis.*

Стаття надійшла до редакції / Received 11.05.2025
Прийнята до друку / Accepted 26.06.2025

Постановка проблеми

В умовах стрімкої цифровізації суспільного життя інформаційний простір стає дедалі вразливішим до впливу автоматизованих систем, зокрема бот-програм (ботів), здатних масштабно розповсюджувати маніпулятивний або деструктивний контент. Це ускладнює підтримання інформаційної прозорості, особливо в періоди суспільно-політичної напруги чи глобальних криз. Соціальні мережі, як ключові платформи комунікації, є особливо чутливими до такого впливу. При цьому, існує брак інструментів, що дозволяють не лише відокремлювати поведінку людей-користувачів від бот-програм, але й системно аналізувати часову та тематичну структуру активності користувачів. Це зумовлює необхідність розробки спеціалізованих методів для глибшого розуміння поведінкових закономірностей у цифрових середовищах.

Аналіз останніх досліджень і публікацій

Останні дослідження підтверджують зростання частки бот-програм в інтернет-трафіку. Згідно зі звітом Imperva за 2024 рік, майже половина веб-трафіку є автоматичною, зокрема 32% становлять шкідливі

бот-програми [1]. Це зростання триває з 2013 року, і його спричиняє доступність інструментів для створення бот-програм, зокрема на базі генеративного штучного інтелекту (ШІ), який дозволяє навіть некваліфікованим користувачам запускати скрапери чи спам-ботів. Частка "простих" шкідливих ботів зросла з 33,4% до 39,6% у 2022–2023 роках, що свідчить про роль ШІ в автоматизації [1].

Технічні наслідки діяльності ботів різноманітні. Згідно з даними Cloudflare (2024), 93% ботів мають потенційно шкідливий характер, особливо в галузях фінансів і електронної комерції [2]. Вони експлуатують вразливості бізнес-логіки (наприклад, API без обмежень входу), спричиняючи резервування товарів, збирання цін, зловживання API — усе це шкодить бізнесу та підриває довіру користувачів. У 2023 році кількість атак з підбору облікових даних зросла на 10%, причому 36,8% були спрямовані на фінансові послуги [1]. Крім того, IoT-ботнети поглиблюють загрозу DDoS, про що свідчить рекордна атака 2024 року на 5,6 Тбіт/с [3].

Новітні дослідження досліджують вплив ШІ на бот-програми. Університет Вашингтона (2024) встановив, що LLM допомагають ботам уникати виявлення на 30%, але також можуть підвищити точність визначення на 9%, якщо застосовуються у захисті [4]. Інше дослідження (IEEE Transactions) виявило ознаки ботнетів — синхронні публікації та аномальні спалахи активності [5], що вказує на постійну "гонку озброєнь" між ботами і системами виявлення [4, 5].

Звіти Imperva та Cloudflare підкреслюють економічні наслідки: 84 млрд доларів щорічних втрат через рекламне шахрайство та зростаючі витрати на захист від DDoS [1, 2]. ЗМІ, такі як Forbes та VOA, також звертають увагу на труднощі виявлення спаму й шахрайства, створених за допомогою ШІ [6]. У відповідь експерти закликають до розвитку поведінкової аналітики, безпеки API та нормативних заходів для стримування шкідливої автоматизації, не шкодячи легальним ботам [1, 2, 5].

Формулювання цілей статті

Метою дослідження є вивчення поведінкових характеристик користувачів Twitter з урахуванням типу акаунта (людина, корисна або шкідлива бот-програма) та галузевого контексту створеного контенту. Дослідження передбачає аналіз змін у кількості користувачів протягом часового періоду з 2017 по 2023, а також виявлення індустрій із підвищеною активністю бот-програм. Для досягнення цієї мети були розроблені спеціалізовані програмні алгоритми, які забезпечили як категоризацію користувачів, так і оцінку їхньої активності у часовому вимірі.

Викладення матеріалу дослідження

Використані набори даних. У дослідженні було використано два основні набори даних. Перший — TwiBot-22, репрезентативний набір даних, що містить інформацію про користувачів Twitter із зазначеними типами акаунтів, зібраний у рамках масштабної ініціативи з вивчення бот-активності в соціальних мережах. Цей набір забезпечує широке покриття та охоплює різні типи взаємодій у мережі. Другий набір — власноруч зібраний корпус аккаунтів користувачів та їх твітів, що охоплює період за 2023 рік, зі спеціальним фокусом на теми, що активно обговорювались у зазначений проміжок часу.

Алгоритм категоризації. Для аналізу зазначених наборів даних був використаний розроблений алгоритм категоризації груп користувачів, представлений на рис. 1.

Data: $\{u \mid u \in \mathcal{N}, \text{ and } u \text{ is not } categorized\}$
Result: $\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_n, \{u^{\text{group}} \mid u \in \mathcal{N}\}$

$\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_n \leftarrow \emptyset$;
for $u \in \mathcal{N}$ **do**
  $c \leftarrow f(u)$;
  if c *is categorical* **then**
    $u^{\text{group}} \leftarrow c$;
    add u to the subset $\mathcal{N}_s \subset \mathcal{N}$ associated with c;
  else if c *is numerical* **then**
    $h \leftarrow heuristic(c)$;
    $u^{\text{group}} \leftarrow h$;
    add u to the subset $\mathcal{N}_s \subset \mathcal{N}$ associated with h;
  end
end

Рис. 1. Алгоритм категоризації груп користувачів

Алгоритм здійснює класифікацію кожного користувача $u \in N$ у відповідну групу на основі значення, що повертається функцією категоризації $f(u)$. Результатом виконання цього алгоритму є формування множин $N_1, N_2, \ldots, N_n$, які представляють кластери користувачів, згрупованих за спільними характеристиками.

На вхід алгоритму подається сукупність не категоризованих користувачів $\{u \mid u \in N, \text{and } u \text{ is not } categorized\}$. Для кожного користувача обчислюється значення категоризації $c = f(u)$. Якщо значення с є категоріальним (наприклад, "Людина", "Корисна бот-програма", "Шкідлива бот-програма"), користувачі одразу приєднуються до відповідних груп за ознакою:

$$u^{group} = c, \text{де } u \in N_s \subset N \quad (1)$$

де N_s— підмножина, що відповідає категорії c.

У випадку, коли значення c є числовим, алгоритм використовує додаткову евристику $h = heuristic(c)$, яка відносить числове значення до відповідного інтервалу, тобто категорії:

$$u^{group} = h(c), \text{де } u \in N_s \text{ для } h(c) \quad (2)$$

Евристика передбачала розбиття значень у межах від 0 до 1 на три категорії згідно з порогами, що апроксимують категорії "Людина", "Корисна бот-програма", "Шкідлива бот-програма". Для цього застосовувались квантильні пороги, що відповідали спостережуваним порівнянням із TwiBot-2022.

Для числової оцінки ботоподібності використовувалися два інструменти:

− Botometer — інструмент, розроблений Університетом Індіани, який аналізує понад 1 200 ознак акаунта, включаючи метадані, мережеву структуру та поведінкові патерни, щоб оцінити ймовірність того, що акаунт є ботом. Botometer надає оцінку від 0 до 5, де вищі значення свідчать про більшу ймовірність ботоподібної поведінки;

− Bot Sentinel — платформа, яка використовує машинне навчання та штучний інтелект для класифікації акаунтів Twitter. Вона оцінює акаунти за шкалою від 0% до 100%, де вищі відсотки вказують на більшу ймовірність того, що акаунт займається токсичною поведінкою або поширює дезінформацію. Bot Sentinel також надає категорії, такі як "нормальний", "задовільний", "деструктивний" та "проблематичний", для подальшої класифікації акаунтів.

Таким чином, весь набір користувачів N був поділений на три основні підмножини $N_{\text{люди}}$, $N_{\text{хороші боти}}$, $N_{\text{погані боти}}$, з урахуванням як наперед заданої категоризації, так і числового аналізу.

Результати роботи алгоритму категоризації. На рис. 2 та в табл. 1 подано відсоткову динаміку розподілу користувачів по групах у період 2017–2023 років, отриманих в результаті роботи алгоритму.

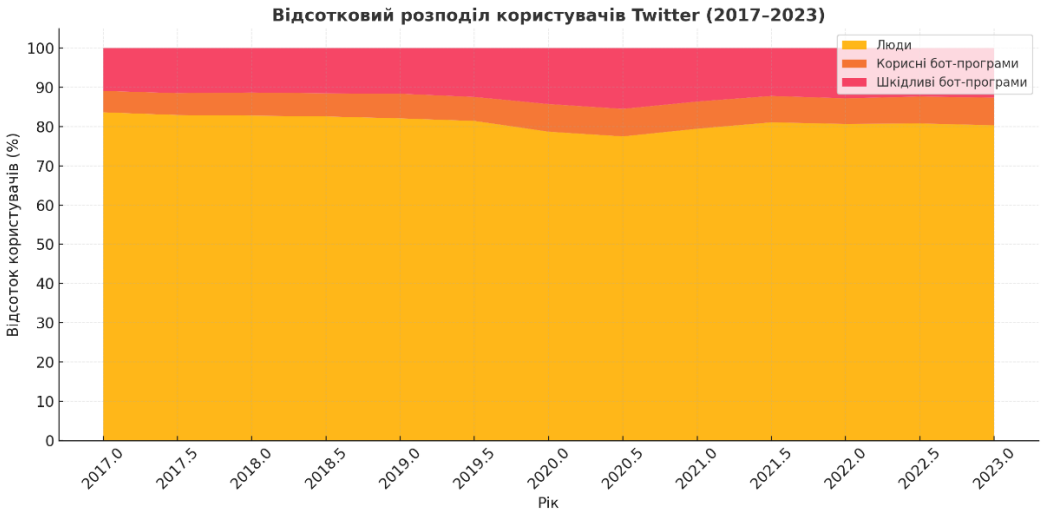

Рис. 2. Відсотковий розподіл користувачів Twitter (2017-2023)

Як видно з графіку (рис. 2), протягом 2017–2019 років частка поганих ботів була відносно стабільною (~11–12%), однак з середини 2019 року спостерігається стійке зростання, яке досягає піку у середині 2020 року (15.5%).

Цей стрибок частки зловмисних акаунтів збігається в часі з двома ключовими глобальними подіями:

Таблиця 1

Відсотковий розподіл користувачів Twitter (2017-2023)

Рік	Люди (%)	Хороші боти (%)	Погані боти (%)
2017.0	83.63	5.45	10.92
2017.5	82.94	5.57	11.49
2018.0	82.78	5.89	11.33
2018.5	82.57	5.89	11.54
2019.0	82.08	6.27	11.64
2019.5	81.43	6.1	12.47
2020.0	78.68	7.04	14.28
2020.5	77.48	7.02	15.5
2021.0	79.43	6.93	13.64
2021.5	81.07	6.73	12.21
2022.0	80.65	6.54	12.81
2022.5	80.79	6.85	12.37
2023.0	80.32	7.04	12.64

− початком пандемії COVID-19 — з моменту оголошення надзвичайної ситуації у січні 2020 та визнання пандемії у березні 2020 року. Попередні дослідження показують, що соціальні мережі в цей час стали основним каналом розповсюдження дезінформації про вірус [7, 8];

− президентськими виборами у США 2020 року, що супроводжувались хвилею дезінформації, поширенням теорій змови, антидемократичними кампаніями, до яких залучались автоматизовані облікові записи [9].

Таким чином, використаний підхід до категоризації не лише забезпечив структуроване представлення динаміки змін у поведінковому складі користувачів, але й дозволив провести ретроспективний аналіз інформаційних загроз на тлі суспільно-політичних криз.

Алгоритм аналізу у часі. У межах дослідження було розроблено алгоритм аналізу у часі, або темпорального аналізу, представлений на рис. 3.

Даний алгоритм створений для оцінки активності користувачів у Twitter за галузевим поділом. Аналіз базувався на попередньо категоризованих користувачах згідно, отриманих з наборів даних TwiBot-2022 та вручну зібраного набору даних за 2023 рік.

Алгоритм працює на основі графу взаємодій $G = (N, E)$, де:

− N — множина користувачів;

− E — множина взаємодій (наприклад, ретвіти).

Algorithm 8: Temporal analysis

Data: $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, Dataset of interactions
Result: Temporal activity of $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3, \ldots, \mathcal{N}_n$

for $g \in \{1, 2, 3, \ldots, n\}$ **do**
  $p_g \leftarrow |\mathcal{N}_g|/|\mathcal{N}|$;
  for $d \leftarrow d_1$ *to* d_m **do**
    $p_{g,d} \leftarrow \frac{|\{t \mid t^{\text{retweeter}} \in \mathcal{N}_g \wedge t^{\text{timestamp}} = d\}|}{|\{t \mid t^{\text{timestamp}} = d\}|}$;
    if $p_{g,d} \approx p_g$ **then**
      g is *normally stimulated* on d;
    else if $p_{g,d} < p_g$ **then**
      g is *understimulated* on d;
    else
      g is *overstimulated* on d;
    end
  end
end

Рис. 3. Алгоритм аналізу активності користувачів у часі

Для кожної групи користувачів $N_g \subset N$ з індустрії g, алгоритм обчислює очікувану частку активності:

$$p_g = \frac{|N_g|}{|N|} \tag{3}$$

Це базова пропорція групи відносно всієї мережі.

Далі, для кожного моменту часу $d \in \{d_1, d_1, \ldots, d_m\}$, тобто певного дня або тижня, обчислюється фактична частка активності групи:

$$p_{g,d} = \frac{|\{t \in E | t^{retweeter} \in N_g \wedge t^{timestamp} = d\}|}{|\{t \in E | t^{timestamp} = d\}|} \tag{4}$$

Тобто, відношення кількості твітів/ретвітів, створених користувачами з групи g у день d, до загальної кількості взаємодій у той самий день.

Порівнюючи $p_{g,d}$ та p_g, алгоритм дозволяє класифікувати ступінь активності:

− якщо $p_{g,d} \approx p_g$ — група активна у нормальному режимі (normally stimulated);

− якщо $p_{g,d} < p_g$ — група є недостимульованою (understimulated);

− якщо $p_{g,d} > p_g$ — група є перестимульованою (overstimulated).

Це дозволяє виявляти часові аномалії або пікову поведінку окремих груп, яка може бути спричинена зовнішніми подіями, наприклад дезінформаційними кампаніями або маркетинговою активністю.

Результати роботи алгоритму аналізу у часі. Результати цього аналізу подано у табл. 2 та рис. 4, де для кожної індустрії представлено середній відсотковий розподіл активності користувачів за трьома групами.

Таблиця 2

Середній розподіл активності в Twitter за індустріями

Індустрія	**Активність від людей (%)**	**Активність від корисних бот-програм (%)**	**Активність від шкідливих бот-програм (%)**
Ігри	75.98	13.33	10.7
Телекомунікації	79.3	12.78	7.91
Обчислення та IT	77.06	13.7	9.24
Спільнота та суспільство	75.9	14.79	9.32
Бізнес-послуги	73.47	12.99	13.54
Охорона здоров'я	77.92	9.61	12.47
Новини	73.75	12.81	13.44
Розваги	82.84	6.18	10.98
Азартні ігри	84.27	11.4	4.33
Фінансові послуги	72.67	6.43	20.9
Роздрібна торгівля	80.83	14.45	4.72
Освіта	75.58	10.22	14.2
Маркетинг	76.36	9.15	14.49
Стиль життя	83.51	7.65	8.84
Законодавство та уряд	66.42	12.74	20.84
Спорт	66.74	9.56	23.7
Їжа та продукти	65.4	10.68	23.91

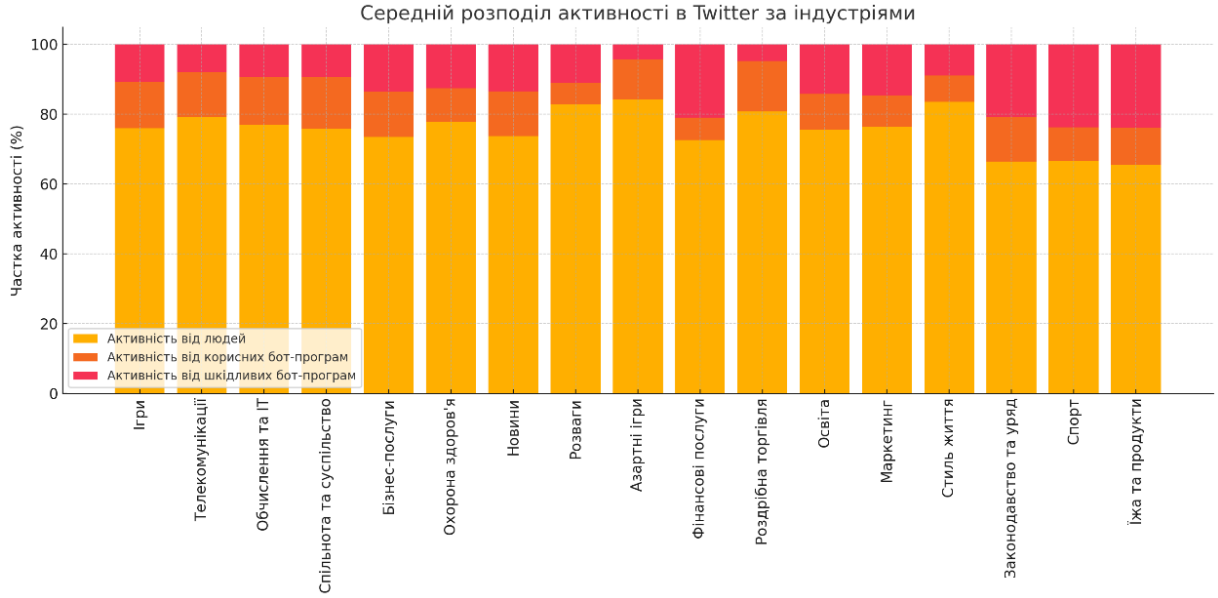

Рис. 4. Середній розподіл активності в Twitter за індустріями

Виявлено, що деякі індустрії мають аномально високий рівень активності шкідливих ботів, зокрема:

− Законотворчість та уряд — 28.07%;

− Маркетинг — 22.89%;

− Азартні ігри — 17.11%.

Ці числа можуть свідчити про навмисне використання автоматизованих облікових записів для інформаційного впливу, поширення маніпулятивного контенту або спам-кампаній.

У той же час індустрії на зразок Розваги, Фінансові послуги та Стиль життя демонструють переважно органічну поведінку — частка контенту від людей перевищує 80%, без суттєвого втручання з боку шкідливих бот-мереж.

Висновки

У ході дослідження було проаналізовано динаміку активності користувачів Twitter на основі двох наборів даних, що охоплюють як історичний, так і сучасний період. Розроблені програмні алгоритми — категоризації та темпорального аналізу — дозволили ефективно структурувати користувачів за поведінковими характеристиками та простежити часові закономірності їхньої активності в різних тематичних індустріях. Результати показали як загальні тенденції (зростання частки шкідливих ботів у період пандемії), так і галузеву специфіку — підвищену активність автоматизованих акаунтів у політичному, маркетинговому та фінансовому сегментах. Отримані результати підтверджують доцільність використання спеціалізованих

програмних алгоритмічних підходів для аналізу складних соціальних даних, а також відкривають перспективи для подальших досліджень інформаційних впливів і цифрової безпеки.

Література

1. Imperva Threat Research. 2024. 2024 Bad Bot Report. Imperva. [Електронний ресурс]. – Режим доступу: https://palai.media/wp-content/uploads/2024/04/imperva-bad-bot-report-2024.pdf

2. Tremante M., Zejnilovic S., Newcomb C. Application Security Report: 2024 Update // Cloudflare Blog. – 2024. – 11 липня. [Електронний ресурс]. – Режим доступу: https://blog.cloudflare.com/application-security-report-2024-update/

3. Lakshmanan R. Mirai Botnet Launches Record 5.6 Tbps DDoS Attack with 13,000+ IoT Devices // The Hacker News. – 2025. – 22 січня. [Електронний ресурс]. – Режим доступу: https://thehackernews.com/2025/01/mirai-botnet-launches-record-56-tbps.html

4. Milne S. Large language models can help detect social media bots — but can also make the problem worse // University of Washington News. – 2024. – 28 серпня. [Електронний ресурс]. – Режим доступу: https://www.washington.edu/news/2024/08/28/large-language-models-social-media-bots-twitter-ai/

5. Hickey D., Fessler D. M. T., Lerman K., Burghardt K. X under Musk's leadership: Substantial hate and no reduction in inauthentic activity // PLOS ONE. – 2025. – Т. 20, № 2. – С. e0313293. – DOI: https://doi.org/10.1371/journal.pone.0313293

6. Voice of America. Report: Bots Make up Nearly Half of Worldwide Internet Traffic // VOA Learning English. – 2024. – 17 квітня. [Електронний ресурс]. – Режим доступу: https://learningenglish.voanews.com/a/report-bots-make-up-nearly-half-of-worldwide-internet-traffic/7573861.html

7. Ferrara E., Cresci S., Luceri L. Misinformation, manipulation, and abuse on social media in the era of COVID-19 // Journal of Computational Social Science. – 2020. – Т. 3. – С. 271–277.

8. Kouzy R., Abi Jaoude J., Kraitem A., El Alam M. B., Karam B., Adib E., та ін. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter // Cureus. – 2020. – Т. 12, № 3.

9. Ferrara E., Chang H., Chen E., Muric G., Patel J. Characterizing social media manipulation in the 2020 US presidential election // First Monday. – 2020.