КІБЕРНЕТИКА та КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК 519.67 DOI:10.34229/2707-451X.25.2.5

В.В. ТРЕТИНИК, М.А. ДАВИДЕНКО

МОДЕЛІ ПОШИРЕННЯ ДЕЗІНФОРМАЦІЇ В ІНТЕРНЕТІ

Запропоновано декілька моделей поширення дезінформації в Інтернеті: модель шуму та впливу, модель вірусного поширення та гібридна модель, що поєднує концепцію теорії графів та нейронні мережі. Моделі будувалися з урахуванням їхньої ефективності щодо відображення реальних процесів поширення дезінформації.

Ключові слова: *поширення дезінформації, моделювання, гібридна модель, нейронні мережі.*

Вступ Дезінформація в Інтернеті – одна з ключових загроз сучасного інформаційного суспільства. Завдяки стрімкому розвитку соціальних мереж та цифрових платформ, поширення фейкових новин, маніпулятивного контенту та недостовірної інформації набуло глобального масштабу. Ці явища негативно впливають на суспільну думку, підривають довіру до інституцій та сприяють поляризації суспільства. Для протидії цій проблемі необхідно розробляти ефективні моделі виявлення та оцінки поширення дезінформації.

Гібридні підходи, що комбінують традиційні методи теорії графів та сучасні досягнення у галузі нейронних мереж [1], можуть стати перспективним рішенням для аналізу процесів поширення інформації. Такі моделі здатні враховувати як структуру соціальних мереж, так і семантичні особливості контенту, що розповсюджується.

Незважаючи на значний обсяг досліджень вітчизняних та зарубіжних науковців у сфері моделювання поширення дезінформації [2–6], поки що не сформовано цілісного уявлення про природу цього явища. Загалом, дезінформацію часто розглядають у контексті впливу цифрових медіа на сучасне суспільство та їхньої ролі у поширенні недостовірної інформації. Новітні комунікаційні технології зробили процес створення та розповсюдження неправдивих або шкідливих відомостей значно легшим. Моделювання розповсюдження дезінформації в Інтернет-середовищі – це актуальна задача, що дозволяє зрозуміти вплив дезінформації на поведінку людей у економічній, політичній та соціальній сферах.

Мета даної роботи – це узагальнення існуючих та розробка нових моделей поширення дезінформації в Інтернеті. Моделі повинні бути ефективними з точки зору відображення реальних процесів поширення дезінформації.

Дезінформація – це цілеспрямоване поширення неправдивої або викривленої інформації з метою маніпуляції думками аудиторії. Залежно від мети та способу поширення, дезінформацію можна класифікувати на такі типи як фейкові новини (створення повністю вигаданих подій або фактів), викривлена інформація

(спотворення реальних фактів для маніпуляції аудиторією), маніпулятивний контент (використання емоційних тригерів для поширення певної точки зору). Ці типи дезінформації відрізняються за характером, але всі вони використовують соціальні мережі як ключовий канал для досягнення цілей. Моделювання таких процесів вимагає врахування складних взаємодій між користувачами мережі та їхньої поведінки.

Соціальні мережі характеризуються складною структурою, яка може бути представленою у вигляді графів, де вузли – це користувачі, а ребра – зв'язки між ними. Поширення інформації у таких мережах відбувається через механізми, які включають репости (перепости), алгоритми рекомендацій, вірусність контенту. Ключовий фактор у поширенні дезінформації – соціальна динаміка, що залежить від так званих "інфлюенсерів" – користувачів з великою аудиторією. Моделювання таких процесів потребує інтеграції структурного аналізу соціальних мереж і аналізу семантики контенту.

У роботі [2] здійснено порівняльний аналіз різних моделей розповсюдження інформації, зокрема моделі шуму та впливу, моделі вірусного поширення, моделі залежних поширень, моделі поширення чуток, моделі впливових користувачів та моделі на основі клітинних автоматів.

Узагальнимо деякі з моделей для задачі поширення дезінформації в Інтернеті. Представимо модель шуму та впливу у контексті розповсюдження дезінформації. Платформу поширення дезінформації, а саме соціальну мережу змоделюємо як граф $G = (V, E)$, де V – множина вузлів (користувачів), E – множина ребер (зв'язків між користувачами). Кожен вузол $v \in V$ має стан: "вразливий до дезінформації", "поширює дезінформацію" або "імунний". Зв'язки Е можуть відображати міру довіри між вузлами або силу впливу одного користувача на іншого. Кожному вузлу $v \in V$ призначається порогове значення θ, яке визначає, наскільки він стійкий до впливу: якщо сума впливів сусідів перевищує θ, вузол змінює свій стан на "поширює дезінформацію". Для задачі дезінформації порогові значення можуть залежати від довіри до джерела інформації, рівня освіти та медіаграмотності користувача, суб'єктивної схильності до емоційного контенту. Кожному ребру e між вузлами u і v призначається ваговий коефіцієнт w_{uv}, що відображає силу впливу. Ваги можуть бути визначені через частоту взаємодії між користувачами (наприклад, лайки, коментарі, репости), рівень емоційності або токсичності повідомлення, історичну репутацію джерела (чи користувач раніше поширював дезінформацію).

Силу впливу будемо обчислювати за формулою:

$$I_v = \sum_{u \in neighbors(v)} w_{uv} s_u,$$

де s_u – стан вузла u (наприклад, $s_u = 1$, якщо u поширює дезінформацію, і $s_u = 0$, якщо ні).

Шум у моделі відображає випадкові або неконтрольовані фактори, які можуть впливати на поширення дезінформації. Для врахування шуму додамо випадковий компонент ε_v до порогового значення θ_v: $\theta_v' = \theta_v + \varepsilon_v$. Це дозволяє моделювати випадкову поведінку користувачів, наприклад, нераціональне поширення інформації під впливом емоцій.

Модель розповсюдження запускається у такий спосіб:

Ініціалізація: обирається підмножина початкових вузлів s_0, які є джерелами дезінформації.

Оновлення станів: на кожному кроці t:

– кожен вузол перевіряє, чи сума впливів сусідів перевищує його порогове значення: $I_v > \theta_v'$;

– якщо це так, вузол переходить до стану "поширює дезінформацію".

Зупинка: Процес завершується, коли більше жоден вузол не змінює свій стан.

Для оцінки ефективності моделі можна використовувати кількість вузлів, які опинилися у стані "поширюють дезінформацію", кількість ітерацій до завершення процесу, як поширення змінюється між різними групами вузлів (наприклад, високовпливові користувачі vs звичайні).

Дану модель можна вдосконалити ввівши додаткові стани вузлів: наприклад, ввести стан "перевіряє інформацію" , щоб моделювати користувачів, які прагнуть перевірити правдивість отриманого контенту. Також як вдосконалення моделі можна розглянути не статичні, а динамічні пороги, які змінюються з часом залежно від зовнішніх факторів, таких як новини, політичні події або попередні взаємодії.

Таким чином, модель шуму та впливу для задач поширення дезінформації в Інтернеті дозволяє врахувати ключові особливості соціальних мереж та поведінку користувачів. Основна ідея – моделювати кожного користувача як вузол із пороговою чутливістю, враховуючи вплив сусідів, рівень довіри до джерела та випадкові фактори.

Розглянемо модель вірусного поширення дезінформації. Модель включає 5 станів вузлів у графі:

1) S – сприятливі: користувачі, які ще не бачили новину або не знають про неї;
2) E – байдужі: користувачі, які бачили новину, але не реагують;
3) I – інфіковані: активні поширювачів новин;
4) R – вилікувані: користувачів, які перестали поширювати новину;
5) A – антифіковані: користувачі, що протидіють поширенню.

Модель будується як система п’яти диференціальних рівнянь наступного виду:

$$\begin{cases} \dfrac{dS}{dt} = -\beta\dfrac{SI}{N} + \gamma R + \alpha A \\ \dfrac{dE}{dt} = \beta\dfrac{SI}{N} - \xi E \\ \dfrac{dI}{dt} = \xi E - \delta I - \theta I + \alpha A, \\ \dfrac{dR}{dt} = \delta I - \gamma R \\ \dfrac{dA}{dt} = \dfrac{\theta EI}{N} - \alpha A \end{cases}$$

де, N – загальна кількість користувачів, β – ймовірність зараження дезінформацією, ε – швидкість переходу зі стану E до I, γ – ймовірність, що користувач знову буде поширювати інформацію, δ – показник, того як швидко, користувач перестає поширювати інформацію, θ – швидкість з якою користувачі переходять до антифікованих, α – швидкість з якою, антифіковані стають сприятливими.

Якщо відомі всі розглянуті параметри моделі, то можуть бути розраховані значення функцій $S(t)$, $E(t)$, $I(t)$, $R(t)$, $A(t)$ у потрібний момент часу і побудований графік зростання кількості відповідних користувачів за зазначений час дослідження.

Таке розширення системи диференційних рівнянь класичної SIR-моделі виявляється інформативнішим, оскільки це дає нам можливість враховувати розповсюдження новини різними категоріями читачів, у тому числі з повторною ймовірністю зараження дезінформацією у випадку, якщо людина піддатлива думці оточуючих чи немає ґрунтових знань для протидії тиску.

Варто зазначити деякі недоліки та обмеження запропонованих моделей. Їм може бракувати емпіричних даних для підтвердження теоретичних передумов, механізмів адаптації до нових типів загроз, і налаштування може вимагати значних ресурсів і спеціалізованих знань.

Також можна розглянути деякі удосконалення SIR-моделі за рахунок розширення її ключових параметрів. Наприклад, якщо додати фактор обізнаності аудиторії (інформаційного імунітету), що має навички критичного мислення, буде змога краще оцінити ефективність спростувань. Також, варіантом розвитку вирішення задачі, можемо розглянути виділення окремих груп користувачів, коли буде здійснений детальний аналіз даних соціальних мереж. Зазначені модифікації можуть покращити аналіз динаміки поширення фейків та зробити прогноз гнучкішим до реальних умов.

Для демонстрації такого підходу було зібрано власний набір даних з різних телеграм каналів. Як фейкова новина була використана новина, що ширилась у Телеграм каналі “Україна Онлайн: Новини“ (t.meUaOnlii) від 7 травня 2025 року. Оригінальний текст публікації наступний: “Теракт в Харкові: підірвали автівку проектувальника БПЛА Андрія Гуменного – він загинув, – РБК-Україна цитує Умєрова Він був головним конструктором авіаційного науково-технічного товариства з обмеженою відповідальністю «КНК». Цей злочин не залишиться без покарання – усі відповідні служби залучені до пошуку та затримання його виконавців. Наші співчуття“, що у свою чергу був спростований наступного дня та офіційно названо фейком. Початкові умови задані для моделювання взяті з показників приблизно на 20 хвилині від моменту публікації.

Для моделювання розповсюдження такої новини, попередньо проаналізовано новини цього ж каналу аби зрозуміти особливості поведінки підписників та читачів, і врешті підібрати найбільш вдалі коефіцієнти $\beta, \varepsilon, \gamma, \delta, \theta, \alpha$ для моделювання розповсюдження фейку підписниками цієї спільноти. Проаналізований набір даних містив близько 100 новин та відповідні показники динаміки розповсюдження (перегляди, реакції тощо) в інформаційному полі за різні проміжки часу та підібрані параметри за допомогою лінійної регресії, та було визначено такі значення коефіцієнтів: $\beta \approx 0.17810$, $\gamma \approx 0.0000013$, $\varepsilon \approx 0.0382$, $\delta \approx 0.00027$, $\theta \approx 0.01034$, $\alpha \approx 0.0201$.

Початковими умовами для моделювання є такі значення: $I_0 = 2202$, $A_0 = 100$, $R_0 = 0$, $E_0 = 57000 - I_0 - A_0 - R_0$, $S_0 = N - E_0 - I_0 - R_0 - A_0$, а кількість підписників цього каналу $N \approx 1100000$. Графічне представлення розповсюдження фейку показано на рис.1.

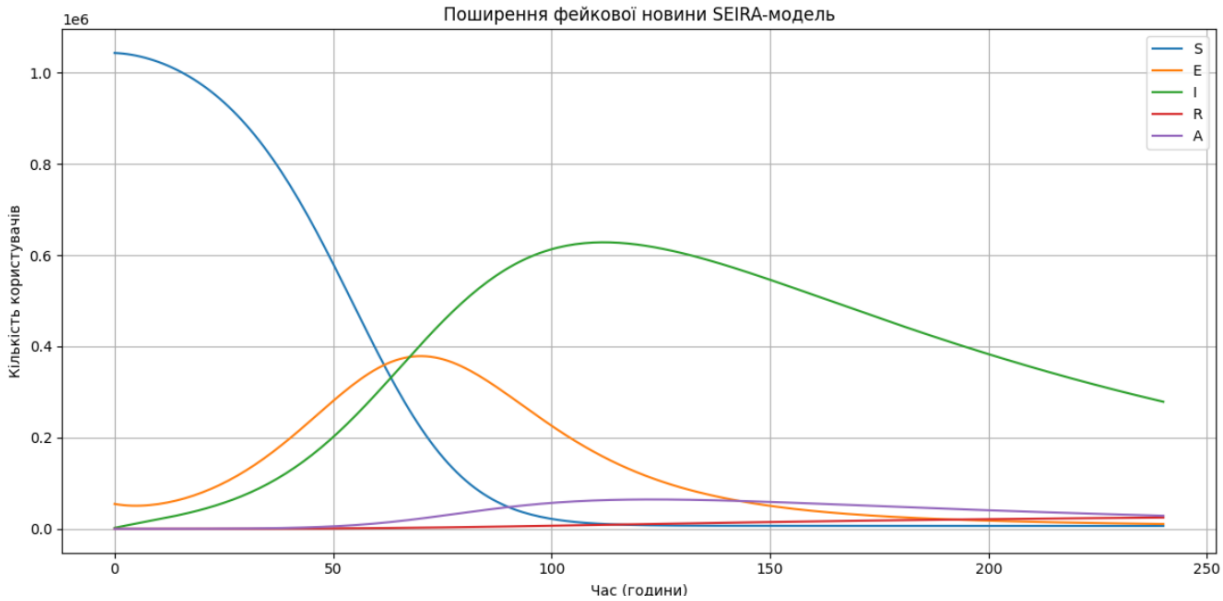

РИС. 1. Динаміка поширення фейкової новини

Розрахунки були здійснені за допомогою Python та бібліотеки scipy, що дозволяє розв`язувати диференційні рівняння чисельним методом (LSODA).

Як видно за графіком, якщо вчасно не прийняти відповідних заходів, офіційно не спростувати таку новину, вона може бути сприйнята аудиторією, як правда. Новині такого характеру, що була обрана, важко протистояти. Фейки про смерть та втрати є сильним емоційним фактором, що веде до зневіри та відчаю, і врешті переслідує головну мету – дестабілізацію.

Один з ключових етапів моделювання є попереднє виявлення фейкової новини. Для цього запропоновано використовувати власну графову нейромережу (GNN). GNN це тип нейронних мереж, призначений для роботи з даними у вигляді саме графових структур. Таке рішення обрано через специфіку даних з соціальних мереж та особливість взаємодії користувачів у Інтернеті. Алгоритм GNN враховує зв'язки між об'єктами, а не тільки окремі ознаки чи контекст. Проте варто вказати, що таке рішення вимагає значних обчислювальних ресурсів.

Одним із методів побудови графа є процес відображення вузлів або ребер графа у векторний простір низької розмірності (рис. 2) [7]. Навчання графів означає пошук значущого, потенційно малорозмірного, представлення вузлів із складних зв'язків, наявних у графі. Для цього потрібна карта від кожного вузла графа до векторного простору, який також називають латентним простором [8].

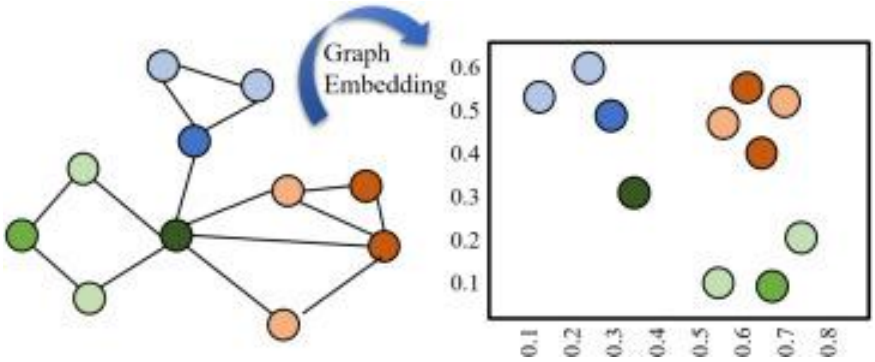

РИС. 2. Векторне представлення графової структури [7]

Застосування GNN для роботи з даними з соціальних мереж, може бути схожим до соціальних графів, але насправді – це більш складна структура порівняно з традиційними соціальними графами через їх здатність опрацьовувати складні взаємозв'язки між новинами (аудиторією, медіа, характером новини, впливом і т.д.) не тільки в евклідовому просторі [7], тобто використовує інформацію для передбачення не лише по взаємозв'язаним елементам, але із інших схожих об'єктів та контексту, зокрема.

Типова архітектура графової нейронної мережі складається з графових операторів, подібних до згортки, які виконують локальне агрегування функцій за допомогою передачі повідомлень із сусідніми вузлами, і, можливо об'єднання, що дорівнює фіксованому або навчальному укрупненню графа [9]. Обчислення здійснюються ітераційно і на кожному кроці вибудовуються зв'язки з кожним новим елементом графа. Формула пошарового поширення для графової мережі виглядає так:

$$H^{l+1} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}(A + \mathrm{I})\widetilde{D}^{\left(-\frac{1}{2}\right)}H^{(l)}W^{(l)}\right),$$

де, l – індекс шару нейронної мережі, σ – функція активації, A – матриця суміжності неорієнтовного графа, I – одинична матриця, $\widetilde{D}$ – діагональна матриця ступенів вузла, тобто кількість з'єднань вузла з іншими вузлами, H – матриця ознак вузлів, W – матриця ваг [10].

Таким чином модель поступово накопичує інформацію про структуру та характеристики графа – це підвищує її здатність до аналізу складних залежностей у даних.

Отже, комбінований підхід з графової нейронної мережі та SEIRA моделі для моделювання розповсюдження фальшивої інформації є потенційно сильним допоміжним інструментом для розробки стратегій боротьби з фейками. Акцент зроблено на гібридну модель для можливості швидкого та

точного моделювання виявлення небезпеки маніпуляціями аудиторією, оскільки такий підхід для визначення фейкових новин може допомогти отримати та змоделювати динаміку поширення дезінформації у стислі терміни.

Висновок. У роботі представлено моделі поширення дезінформації в Інтернеті. Вони є важливими для розуміння соціальних явищ та впливу в Інтернеті, а також для ефективного управління та розробки стратегій поширення неправдивого контенту та інформації. Вказано на переваги, недоліки підходу. Для подальшого вдосконалення моделювання поширення дезінформації необхідно проводити додаткові дослідження та розробляти нові підходи, які охоплювали б як технічні, так і поведінкові аспекти, що впливають на динаміку розповсюдження дезінформації у мережі.

Авторські внески: Третиник В.В. – концепція, структура, побудова моделі шуму та впливу у контексті розповсюдження дезінформації, побудова моделі вірусного поширення днзінформації, написання вступної частини. Давиденко М.В. – побудова графової нейронної мережі детекції та поширення дезінформації, розробка її архітектури, написання висновків.

Наявність даних. Дані зібрані самостійно з телеграм каналу «Україна Онлайн: Новини» (t.meUaOnlii).

Список літератури

1. Wu Z., Pan S., Chen F., Long G., Zhang C., Yu P.S. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*. 2021. Vol. 32, No. 1. P. 4–24. https://doi.org/10.1109/TNNLS.2020.2978386
2. Дмитрієнко К. Порівняльний аналіз моделей розповсюдження інформації в інтернет-середовищі. *Кібербезпека: освіта, наука, техніка*. 2023. Т. 4, № 20. С. 272–282.
3. Черній П.Д. Моделі поширення повідомлень в онлайн соціальних мережах: властивості, структура, особливості застосування. *Вісник Харківського національного університету імені В.Н. Каразіна*. 2017. С. 127–134.
4. Dennis L.A., Fu Y., Slavkovik M. Markov chain model representation of information diffusion in social networks. *Journal of Logic and Computation*. 2022. Vol. 32, Iss. 6. P. 1195–1211. https://doi.org/10.1093/logcom/exac018
5. Ulichev O.S. Research of information dissemination models and information influences in social networks. *Control, Navigation and Communication Systems*. 2018. **4** (50). P. 147–151. https://doi.org/10.26906/sunz.2018.4.147
6. Sahnoune M.Y., Akdim K., Ez-Zetouni A., Zahid M. A virus dynamics model for information diffusion in online social networks. *Communications in Mathematical Biology and Neuroscience*. 2021. Article ID 80. https://doi.org/10.28919/cmbn/6569
7. Li Y., Xie S., Wan Z., Lv H., Song H., Lv Z. Graph-powered learning methods in the Internet of Things: A survey. *Machine Learning with Applications*. 2023. Vol. 11, Article No. 100441. https://doi.org/10.1016/j.mlwa.2022.100441
8. Hetzel L., Fischer D.S., Günnemann S., Theis F. J. Graph representation learning for single-cell biology. *Current Opinion in Systems Biology*. 2021. Vol. 28, Article No. 100373. https://doi.org/10.1016/j.coisb.2021.05.008
9. Frasca F., Rossi E., Eynard D., Chamberlain B., Bronstein M., Monti F. SIGN: Scalable inception graph neural networks. 2020. https://doi.org/10.48550/arXiv.2004.11198
10. Kipf T.N., Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint*. 2016. https://doi.org/10.48550/arXiv.1609.02907

Одержано 15.05.2025

Третиник Віолета Вікентіївна,
кандидат фізико-математичних наук, доцент кафедри прикладної математики
НТУУ «КПІ ім. Ігоря Сікорського»,
https://orcid.org/0000-0002-3538-8207
viola.tret@gmail.com

Давиденко Микола Андрійович,
аспірант кафедри прикладної математики
НТУУ «КПІ ім. Ігоря Сікорського».

УДК 519.67

В.В. Третиник [*] , М.А. Давиденко

Моделі поширення дезінформації в Інтернеті

НТУУ «КПІ ім. Ігоря Сікорського»
[*] Листування: viola.tret@gmail.com

Вступ. В наш час розвиток цифрових технологій ніколи не зупиняється. Люди з кожним днем все більше отримують доступ до швидкої інформації, разом з тим все частіше залишаються обманутими, через відсутність якісної перевірки контенту на достовірність. Фейки регулярно заповнюють інформаційний простір. Головне, що небезпека криється у тому, що користувачі можуть не лише споживати новини, але й самі стати розповсюджувачами. Інформаційно-психологічні операції (ІПСО) – інструмент впливу на суспільну думку. Оманливий контент несе загрозу і через те, що може бути підкріплений емоційною складовою, доповнений правдивою інформацією та перекручуванням фактів для ускладнення розпізнання інформаційної атаки. Розповсюдження фейків може навіть призвести до дестабілізації ситуації у державі. Тобто, дезінформація це загроза для національної безпеки, частіше націлена на розсіювання паніки та підриву морального духу суспільства, особливу загрозу це несе під час війни в країні, коли достовірність інформації має критичну важливість.

Зазвичай фокус в існуючих рішеннях направлений на аналіз окремих ознак (характер тексту, динаміку поширення тощо), а також обмежується пошуком залежностей між об’єктами, оскільки більшість з них базується на аналізі даних лише в евклідовому просторі. Тож в боротьбі з дезінформацією потребуємо більш досконалих рішень, де будуть вирішені зазначені проблеми.

Мета роботи. Розробка гібридної моделі поширення дезінформації в Інтернеті шляхом комбінування нейромережевого рішення та моделі поширення інформації. Запропоноване рішення має забезпечувати високу точність виявлення фейків, а також демонструвати гнучкість і стійкість до змін середовища.

Результати. Запропоновано використати SEIRA моделі, для моделювання поширення дезінформації у соцмережах, попередньо виявленої за допомогою GNN на основі реальних даних з соціальних медіа.

Ключові слова: поширення дезінформації, моделювання, гібридна модель, нейронні мережі, соціальні мережі.

UDC 519.67

Violeta Tretynyk [*], Mykola Davydenko

Models of the Spread of Disinformation on the Internet

The National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
[*] Correspondence: viola.tret@gmail.com

Introduction. In our time, the development of digital technologies never stops. People, with each passing day, increasingly gain access to rapid information, and at the same time, they are more and more often deceived due to the lack of quality verification of content for authenticity. Fakes regularly fill the information space. The main danger lies in the fact that users can not only consume news but also become distributors themselves. Information and psychological operations (IPSOs) are a tool for influencing public opinion. Deceptive content poses a threat also because it can be supported by an emotional component, supplemented with truthful information, and involve the distortion of facts to complicate the recognition of an information attack. The spread of fakes can even lead to the destabilization of the situation in the state. That is, disinformation is also a threat to national security because it is often aimed at spreading panic and undermining the morale of society; this poses a particular threat during a war in the country, when the reliability of information is critically important.

Usually, the focus in existing solutions is directed at the analysis of individual features (nature of the text, dynamics of dissemination, etc.), and they are also limited in searching for dependencies between objects, since most of them are based on data analysis only in Euclidean space. Therefore, in the fight against disinformation, we need more advanced solutions where the mentioned problems will be addressed.

Objective of the paper. Development of a hybrid model for the spread of disinformation on the Internet by combining a neural network-based solution and an information dissemination model. The proposed solution should ensure high accuracy in fake detection, as well as demonstrate flexibility and resilience to changes in the environment.

Results. It is proposed to use SEIRA models to simulate the spread of disinformation in social networks, previously detected using GNN based on real data from social media.

Keywords: disinformation spread, modeling, hybrid model, neural networks, social networks.