

РОЛЬ ІНТЕГРОВАНИХ МОДЕЛЕЙ У ПРОГНОЗУВАННІ ПОШИРЕННЯ ДЕЗІНФОРМАЦІЇ

¹ Вінницький національний технічний університет;

Анотація

Стаття присвячена аналізу ролі інтегрованих моделей у прогнозуванні поширення дезінформації. Розглянуто різні підходи до моделювання, зокрема епідеміологічні, когнітивні, агент-орієнтовані та моделі на основі машинного навчання. Обговорюються їх переваги, обмеження та сфери застосування. Окремо наголошується на важливості гібридних підходів, що інтегрують мережевий аналіз, психологічні фактори та аналіз змісту повідомлень. Автор висвітлює виклики впровадження таких моделей, включаючи доступ до даних, міжплатформну інтеграцію, етичні аспекти та необхідність постійної адаптації. У статті наведені приклади прикладних кейсів від соціальних платформ та державних ініціатив, що підтверджують ефективність інтегрованих моделей у боротьбі з дезінформацією.

Ключові слова: дезінформація, інтегровані моделі, прогнозування, епідемічні моделі, когнітивні моделі, машинне навчання, соціальні мережі, вірусність, інфодемія, агентні моделі, нейромережі, нечітка логіка, фактчекінг, моделювання інформації, медіаграмотність.

Abstract

The article examines the role of integrated models in predicting the spread of misinformation. Various modeling approaches, including epidemiological, cognitive, agent-based, and machine learning models, are reviewed. The study discusses the strengths, limitations, and application contexts of these models, highlighting the significance of hybrid approaches that integrate network analysis, psychological factors, and content analysis. Special attention is given to implementation challenges, such as data accessibility, cross-platform integration, ethical concerns, and the necessity of constant model adaptation. Practical cases from social media platforms and governmental initiatives demonstrate the effectiveness of integrated models in combating misinformation.

Keywords: disinformation, integrated models, forecasting, epidemic models, cognitive models, machine learning, social networks, virality, infodemic, agent models, neural networks, fuzzy logic, fact-checking, information modeling, media literacy.

Вступ

Дезінформація в цифровому просторі поширюється надзвичайно швидко, створюючи ризики для суспільства, політики та здоров'я. Дослідження показали, що неправдивий контент поширюється далі, швидше та глибше, ніж правдивий – зокрема, аналіз 126 тисяч випадків поширення новин у Twitter виявив, що фейкові новини охоплюють значно більшу аудиторію, ніж правдиві, і роблять це швидше [1]. Це відбувається через новизну та емоційність хибної інформації. В таких умовах особливо актуальними є моделі прогнозування поширення дезінформації, що дозволяють оцінити ризики “інфодемій” та вчасно втрутитися. Сучасні підходи прагнуть інтегрувати кілька методів – машинне навчання, аналіз соціальних мереж, нечітку логіку, епідеміологічні моделі тощо – щоб охопити всі аспекти цього складного явища. У цій доповіді наведено аналітичний огляд наукової літератури та прикладних кейсів про те, як інтегровані моделі застосовуються для прогнозування динаміки дезінформації, які з них найефективніші та які виклики стоять на шляху їх впровадження.

Результати дослідження

Одним із базових підходів є аналогія із поширенням епідемій. Ще з 1960-х років вчені адаптували епідемічні моделі (SIR, SI, SIS тощо) до “інтелектуальних епідемій” – поширення чуток і інформації [2]. У таких моделях користувачі поділяються на категорії: вразливі до дезінформації, інфіковані (ті, що повірили і поширюють її) та одужалі (ті, що перестали поширювати або стали імунними) – за аналогією із SIR-моделлю. На відміну від реальних хвороб, моделі типу Daley-Kendall або Maki-Thompson ввели специфічні для чуток поняття – наприклад, фактор забування або “цінності новини”, що зменшується з часом. Такі моделі можуть бути як детерміністичними (системи диференціальних рівнянь), так і стохастичними (агенти з випадковими контактами) [2]. Сьогодні епідеміологічні підходи розвинуто далі шляхом інтеграції з мережевим аналізом: замість припущення однорідного змішування популяції моделі запускають на реальних соціальних графах. Це дозволяє врахувати структуру мережі – хто з ким взаємодіє – і знайти “супер-розповсюджувачів” (ключових інфлюенсерів) для таргетованих інтервенцій [2]. Проте прості епідемічні моделі часто ігнорують зміст повідомлень та поведінкові відмінності користувачів, тому виникла потреба їх доповнювати іншими компонентами [2]. Дослідники з Vellore Institute (Індія) запропонували математичну епідемічну модель, що враховує сентимент повідомлень і “соціальний інтелект” користувачів [3]. Їх модель оцінює, наскільки емоційно забарвлена брехлива новина та здатність аудиторії критично її сприймати, і моделює поширення з урахуванням цих факторів. У симуляціях було показано, як дезінформація може поширюватися серед різних спільнот та коли політикам варто втрутитися, а також що ізоляція фейків (швидке видалення чи спростування) суттєво обмежує масштаб “зараження” [3]. Інтеграція психологічних чинників дозволила моделі оцінювати “емоційний і соціальний інтелект” груп, схильних до поширення фейків, та ефективніше прогнозувати точки для втручання. Інший напрям – нечітка логіка та експертні системи – використовується для моделювання невизначеності в поведінці людей. Наприклад, нечіткі множини можуть описувати рівень довіри або скептицизму аудиторії. У сфері охорони здоров'я один аналіз із застосуванням нечіткої якісної компаративної аналітики (fsQCA) показав, що під час епідемії (тroph 2022 р.) в китайських соцмережах перевірена інформація поширювалася навіть активніше, ніж дезінформація – співвідношення ~21.8:1 випадків. Дослідники виокремили три типові комбінації характеристик контенту, що привели до максимального поширення: шлях “авторитетного джерела”, “особистої ефективності” і власне “шлях дезінформації” (коли фейк стає вірусним) [4]. Цей підхід продемонстрував, що особливості контенту можуть суттєво впливати на віральність, і їх треба враховувати при моделюванні. Отже, епідемічні моделі, доповнені аналізом мережі, емоцій та змісту, дають більш повну картину ризиків поширення дезінформації.

Для точнішого прогнозування важливо врахувати, як люди сприймають інформацію. Інтегровані когнітивні моделі поєднують соціальні мережі з психологічними процесами – когнітивними упередженнями, динамікою переконань, впливом ЗМІ. Наприклад, у моделі cognitive cascade (когнітивний каскад) агенти не тільки контактують за мережею, а й мають внутрішній стан переконань, що змінюється під впливом повідомлень [5]. Дослідники з Tufts University підкреслюють: класичні “епідемічні” моделі не мають уявлення про те, що вже думає людина, тоді як агент-орієнтовані моделі з елементами когніції можуть це змодельовати [5]. Їх модель додає до мережевого каскаду прості когнітивні правила – зокрема, ефекти когнітивного дисонансу (схильність відкидати інформацію, що суперечить наявним переконанням) та вибіркової експозиції (шукаємо підтвердження своїм поглядам) [5]. Також у модель введено “медіа-агентів” – вузли, що представляють ЗМІ і запускають початкові хвилі повідомлень [5]. Результат – більш реалістичні “когнітивні каскади”: навіть спрощені функції врахування переконань суттєво підвищили пояснювальну силу моделі порівняно зі звичайними каскадними моделями [5]. За оцінкою авторів, агреговані результати симуляції узгоджуються з реальними опитуваннями громадської думки щодо COVID-19, а динаміка моделі дала уявлення, як протидіяти небезпечним переконанням. Це свідчить, що врахування когнітивних аспектів (упереджень, довіри, “імунітету” до фейків) підвищує прогностичну цінність моделей. На рисунку 1 зображено вид запропонованої моделі.

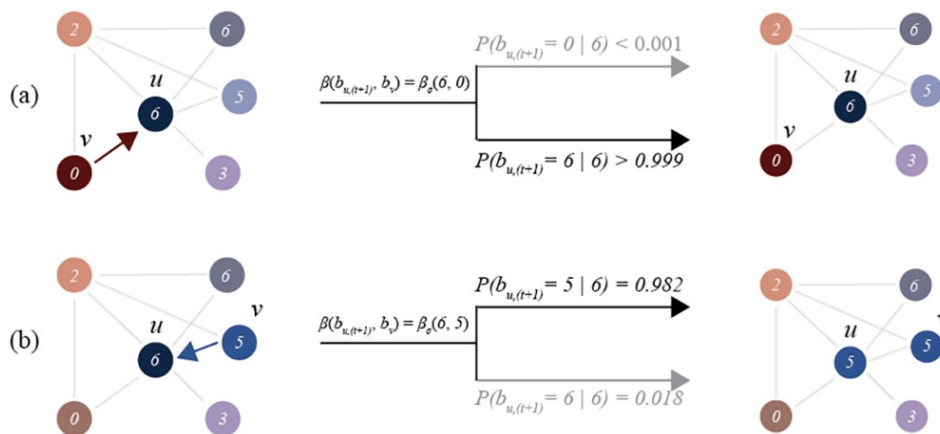


Рисунок 1 — Вид агент-орієнтованої моделі

Приклад інтеграції когнітивного фактору в модель поширення: якщо агент u з дуже сильним переконанням (рівень 6) намагається вплинути на агента v , що зовсім не поділяє цю думку (0), то ймовірність “зараження” v майже нульова ($< 0,1\%$). Натомість коли u і v вже майже однодумці (6 і 5), шанс успішної передачі переконання наближається до 98% [5]. Таким чином, модель враховує ефект підтвердження: люди значно легше сприймають дезінформацію, яка узгоджується з їхніми поточними поглядами, і пручаються тій, що їм суперечить. Агент-орієнтовані моделі також можуть інтегрувати нечітку логіку або експертні знання. Наприклад, метод *fuzzy cognitive maps* (нечітких когнітивних карт) дозволяє експертам задати фактори (довіра до джерела, емоційний фон, соціальний статус) і їх ваги, після чого модель імітує, як зміна одного фактора (наприклад, поява спростування) вплине на рішення агентів ділитися новиною. Подібні підходи поки що менш поширені, але потенційно корисні для включення якісних знань у симуляції. Загалом, когнітивно-соціальні моделі добре працюють для аналізу сценаріїв “що буде, якщо”, дозволяючи випробувати різні стратегії (наприклад, додати агентів-фактчекерів, або змінити настрої аудиторії) і подивитися, як це змінить каскад дезінформації.

Висновки

Інтегровані моделі прогнозування ризиків дезінформації стали необхідним інструментом у добу інформаційних війн та інфodemій. Комбінуючи переваги різних підходів – від математичних епідемічних моделей до глибоких нейромереж – дослідники і практики досягають більш глибокого розуміння того, як саме фейки поширюються та як цьому можна завадити. Глобальні дослідження підтверджують, що жоден окремий метод не є панацеєю: наприклад, врахування лише структури мереж без змісту повідомлень недостатньо, як і навпаки. Тому найкращі результати дають гібридні системи, які враховують і мережеві зв'язки, і поведінку користувачів, і зміст та емоції повідомлень, а подекуди – і когнітивні особливості аудиторії. Такі моделі вже демонструють успіх: від більш точного прогнозування динаміки *viral*-каскадів до визначення оптимальних точок для втручання (превентивного “щеплення” аудиторії правдою або видалення контенту). Водночас, впровадження інтегрованих моделей – це безперервний процес вдосконалення. Необхідно стандартизувати підходи, зібрати репрезентативні дані з різних країн і тем, вирішити питання етики та приватності. Виклики, такі як міжплатформна координація та адаптація до нових тактик, вимагають співпраці між дослідниками, технологічними компаніями та регуляторами. Сучасні інтегровані моделі показали свою ефективність у контрольованих умовах – наступний крок це масштабувати їх на реальну інфраструктуру, зберігаючи права користувачів і довіру суспільства. Лише поєднавши технологічні інновації з просвітою та міжнародною співпрацею, можна побудувати стійку “імунну систему” цифрового інформаційного простору проти загроз дезінформації.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. [Електронний ресурс]. URL: <https://www.science.org/doi/10.1126/science.aap9559>
2. Kühnert, M., & Beckmann, M. (2024). Are Misinformation Propagation Models Holistic Enough? Identifying Gaps and Needs. *CEUR Workshop Proceedings*, Vol-3782.
3. Kumar, R., et al. (2024). Epidemic modeling for misinformation spread in digital networks through a social intelligence approach. *PMC*.
4. Li, Y., et al. (2024). Is disinformation more likely to spread? A fuzzy-set qualitative comparative analysis of emerging infectious diseases on China's short video platform. [Електронний ресурс]. URL: <https://pmc.ncbi.nlm.nih.gov/articles/pmid/39433417>
5. Epstein, Z., Berinsky, A. J., Rand, D. G., & Pennycook, G. (2022). Cognitive cascades: How to model (and potentially counter) the spread of fake news. [Електронний ресурс]. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0261811>

Лаєров Вадим Валерійович — аспірант кафедри захисту інформації, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: vadosssss228@gmail.com

Науковий керівник: **Дудатьєв Андрій Веніамінович** — кандидат технічних наук, доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця

Lavrov Vadym V. — Department of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email : vadosssss228@gmail.com

Supervisor: **Andriy Veniaminovich Dudatyev.** — Candidate of Technical Science, Associate Professor of Information Protection, Vinnytsia National Technical University, Vinnytsia