

УДК 316.77:004.896:17

Л.П. Дмитроца, канд.техн.наук; С.В.Дацик

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

АНАЛІЗ ІНСТРУМЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЯВЛЕННЯ ДЕЗІНФОРМАЦІЇ В НОВИНАХ FACEBOOK

L.P. Dmytrotsa Ph.D, S.V. Datsyk

ANALYSIS OF ARTIFICIAL INTELLIGENCE TOOLS TO DETECT DISINFORMATION IN FACEBOOK NEWS

Поширення дезінформації на платформах соціальних мереж стало серйозною проблемою в епоху цифрових технологій. Facebook, будучи однією з найбільш поширених платформ соціальних медіа, перебуває під пильною увагою через його роль у поширенні дезінформації. Для вирішення цієї проблеми існує кілька інструментів для виявлення дезінформації в новинах Facebook. У цьому дослідженні здійснено порівняння та зіставлено чотири із них, а саме: Sphere, CrowdTangle, Factmata та NewsGuard. Досліджуючи їхні переваги та недоліки, можемо краще зрозуміти ефективність цих інструментів у виявленні дезінформації в новинах Facebook.

Sphere – інструмент, розроблений дослідниками з Кембриджського університету, який використовує штучний інтелект (AI) і машинне навчання (ML) для виявлення та пом'якшення поширення дезінформації на платформах соціальних мереж, таких як Facebook [1]. Однією з переваг Sphere є його здатність точно виявляти дезінформацію та запобігати її поширенню серед широкої аудиторії. Однак покладатися виключно на штучний інтелект і машинне навчання для виявлення дезінформації може бути недостатньо. У міру розвитку тактики дезінформації може знадобитися включення інших методів, наприклад, перевірка фактів людьми, щоб забезпечити точне виявлення дезінформації [2].

CrowdTangle – інструмент публічного аналізу, розроблений Facebook, який допомагає видавцям, журналістам, дослідникам, перевіряючим факти та іншим слідкувати, аналізувати та повідомляти про те, що відбувається в соціальних мережах [3]. Однією з переваг CrowdTangle є його здатність надавати дані про поширення дезінформації в режимі реального часу, що може допомогти ідентифікувати джерело та запобігти її подальшому поширенню. Однак CrowdTangle може бути неефективним у виявленні дезінформації, оскільки він значною мірою покладається на створений користувачами контент, який не завжди може бути надійним [4].

Factmata – інструмент, який використовує III та обробку природної мови (NLP) для ідентифікації та позначення дезінформації на платформах соціальних мереж, таких як Facebook [5]. Однією з переваг Factmata є його здатність аналізувати контекст і настрої публікацій, щоб точно виявляти дезінформацію. Крім того, Factmata розробила продукт Narrative Monitoring, який поєднує дві унікальні технології: кластеризацію тем і оцінку вмісту. Однак алгоритми Factmata можуть бути не в змозі виявити більш складні форми дезінформації, такі як глибокі фейки [6].

NewsGuard – інструмент, який надає незалежні, аполітичні рейтинги довіри для онлайн-джерел новин, що може допомогти користувачам визначити надійні джерела інформації [7]. Однією з переваг NewsGuard є його здатність надавати користувачам

простий і легкий у використанні інструмент для оцінки довіри до джерел новин у Facebook. Проте рейтинги NewsGuard можуть бути не зовсім точними, оскільки вони значною мірою покладаються на людське судження та можуть залежати від особистих упереджень [8].

Хоча кожен із інструментів має свої переваги та недоліки, очевидно, що жоден інструмент не може повністю усунути дезінформацію у Facebook. Натомість для ефективної боротьби з дезінформацією на цій платформі може знадобитися поєднання інструментів і стратегій [9]. Алгоритми машинного навчання, класифікаційний аналіз, регресія, кластеризація даних, розробка функцій і зменшення розмірності – інструменти, які можна використовувати для виявлення дезінформації у Facebook [10]. Зрештою, найефективнішим інструментом буде той, який зможе точно виявляти дезінформацію, а також буде зручним і доступним для широкого кола користувачів.

Висновок. Боротьба з дезінформацією в соціальних мережах є постійною проблемою, і такі інструменти, як Sphere, CrowdTangle, Factmata та NewsGuard, є важливими для боротьби з цією проблемою. Хоча кожен інструмент має свої сильні сторони та обмеження, усі вони відіграють вирішальну роль у виявленні та усуненні дезінформації у Facebook.

Література

1. Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions – [Електронний ресурс] – Режим доступу: <https://link.springer.com/article/10.1007/s10479-022-05015-5>
2. Why the Government Should Not Regulate Content Moderation of Social Media – [Електронний ресурс] – Режим доступу: <https://www.cato.org/policy-analysis/why-government-should-not-regulate-content-moderation-social-media>
3. A tool from Meta to help follow, analyze, and report on what’s happening across social media – [Електронний ресурс] – Режим доступу: <https://www.crowdtangle.com/>
4. What data is CrowdTangle tracking? – [Електронний ресурс] – Режим доступу: <https://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking>
5. Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis – [Електронний ресурс] – Режим доступу: <https://www.mdpi.com/2673-5172/4/2/43>
6. Factmata Narrative Monitoring. – [Електронний ресурс] – Режим доступу: <https://ircai.org/top100/entry/factmata-narrative-monitoring/>
7. NewsGuard Ratings. – [Електронний ресурс] – Режим доступу: <https://www.newsguardtech.com/solutions/newsguard/>
8. Number of Unreliable AI-Generated Sites, Labeled 'UAINs – [Електронний ресурс] – Режим доступу: <https://www.newsguardtech.com/press/newsguard-now-identifies-125-news-and-information-websites-generated-by-ai-develops-framework-for-defining-unreliable-ai-generated-news-and-information-sources/>
9. The Future of Truth and Misinformation Online – [Електронний ресурс] – Режим доступу: <https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/>
10. Machine Learning: Algorithms, Real-World Applications – [Електронний ресурс] – Режим доступу: <https://link.springer.com/article/10.1007/s42979-021-00592-x>