УДК 004.7:004.8:004.9
Л.П. Дмитроца, канд.техн.наук, С.В.Дацик
(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ЗАСТОСУВАННЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЯВЛЕННЯ ТА ПРОТИДІЇ ДЕЗІНФОРМАЦІЇ У FACEBOOK

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

APPLICATION OF ARTIFICIAL INTELLIGENCE METHODS TO DETECT AND COUNTERACT DISINFORMATION ON FACEBOOK

В останні роки проблема дезінформації на платформах соціальних мереж стала серйозною проблемою. Як одну з найбільших платформ соціальних медіа, Facebook зазнав критики за нездатність ефективно виявляти та запобігати поширенню дезінформації. Щоб вирішити цю проблему, Facebook запровадив методи на основі ШІ для виявлення та позначення неправдивої інформації. Однак методи не позбавлені проблем. У цьому досліджені виконаємо огляд методів на основі штучного інтелекту для виявлення дезінформації у Facebook, обговоримо проблеми виявлення дезінформації за допомогою розглянутих методів і дослідимо стратегії підвищення їх точності та ефективності.

Штучний інтелект (AI) став критично важливим інструментом для виявлення та запобігання поширенню шкідливого контенту у Facebook. Згідно із заявою Facebook у листопаді 2020 року, штучний інтелект допомагає масштабувати роботу експертів-людей і підвищити ефективність модерації контенту. Facebook використовує поєднання технологій примусового контролю, перевірки людьми та методів на основі ШІ для виявлення та видалення неправдивої інформації. Використання методів штучного інтелекту для виявлення дезінформації на платформах соціальних мереж в останні роки набирає обертів. Дослідження, проведене Santos et al. у 2023 році [1] проаналізував потенційні переваги автоматизованого виявлення дезінформації з точки зору інформаційних наук. Однак, незважаючи на переваги ШІ у виявленні дезінформації, існують також значні проблеми, які необхідно вирішити.

Однією з основних проблем у виявленні дезінформації за допомогою методів на основі штучного інтелекту є поширення фейкових новин, що досі є складною невирішеною проблемою. Пандемія COVID-19 також висвітлила проблему дезінформації на платформах соціальних мереж із поширенням неправдивої інформації про вірус та його лікування. Крім того, використання змагальних прикладів, які є спеціально створеними вхідними даними, призначеними для обману моделей машинного навчання, також може стати проблемою для методів на основі ШІ для виявлення дезінформації [2]. Виклики підкреслюють необхідність стратегій для підвищення точності та ефективності методів на основі ШІ для виявлення дезінформації на платформах соціальних мереж.

Щоб підвищити точність і ефективність методів на основі ШІ для виявлення дезінформації у Facebook, були запропоновані різні стратегії. Сантос та ін. у 2023 році [2] проаналізував низку підходів, включаючи перевірку фактів, лінгвістичний аналіз, аналіз настроїв та використання систем людського циклу. Баласубраманіам та ін. у 2023 році [3] запропонував систематичний і структурований спосіб визначення вимог пояснювання систем штучного інтелекту, що може покращити прозорість та інтерпретацію методів на

основі штучного інтелекту. Лі та ін. [4] провели дослідження наслідків коментування посту дезінформації у Facebook, яке показало, що втручання агентства користувачів може бути ефективним у зниженні поширення неправдивої інформації. Попередні стратегії можуть допомогти вирішити проблеми з виявленням дезінформації за допомогою методів на основі штучного інтелекту та підвищити їх точність і ефективність.

Хоча методи на основі штучного інтелекту пропонують багатообіцяюче рішення для виявлення та запобігання поширенню дезінформації у Facebook, є також етичні міркування, які слід брати до уваги. Однією з головних проблем є можливість зміщення в алгоритмах, які використовуються в цих методах. Як зазначив Діпак у 2021 році [5], використання автоматизації на основі даних для виявлення фейкових новин може викликати етичні та нормативні міркування. Флорес у 2022 році [6] далі досліджує етичні міркування використання штучного інтелекту, наголошуючи на необхідності прозорості та підзвітності в застосовуваних методах ШІ. Крім того, Лауер у 2021 році [7] підкреслює питання свободи слова та потенціал цензури при виявленні дезінформації. Етичні міркування необхідно ретельно розглянути та розглянути, щоб переконатися, що методи на основі ШІ використовуються відповідально та етично.

Висновок. методи на основі ШІ пропонують багатообіцяюче рішення для виявлення та запобігання поширенню дезінформації у Facebook. Однак існують значні проблеми та етичні міркування, які необхідно брати до уваги. Поширення фейкових новин, використання суперечливих прикладів і потенціал упередженості в алгоритмах – все це виклики, які потребують вирішення. Такі стратегії, як перевірка фактів, лінгвістичний аналіз і втручання на основі користувачів, можуть допомогти підвищити точність і ефективність методів на основі ШІ. Крім того, прозорість і підзвітність мають бути пріоритетними для вирішення етичних міркувань, таких як потенційна упередженість і цензура.

Література

1. Artificial Intelligence in Automated Detection of Disinformation: A Thematic Analysis – [Електронний ресурс] – Режим доступу: https://www.mdpi.com/2673-5172/4/2/43

2. An Adversarial Benchmark for Fake News Detection Models – [Електронний ресурс] – Режим доступу: https://arxiv.org/pdf/2201.00912.pdf

3. Transparency and explainability of AI systems – [Електронний ресурс] – Режим доступу: https://www.sciencedirect.com/science/article/pii/S0950584923000514

4. User agency–based versus machine agency–based misinformation interventions – [Електронний ресурс] – Режим доступу: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10113910/

5. Ethical Considerations in Data-Driven Fake News Detection – [Електронний ресурс] – Режим доступу: https://link.springer.com/chapter/10.1007/978-3-030-62696-9_10

6. Ethical Considerations in the Application of Artificial Intelligence – [Електронний ресурс] – Режим доступу: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9406274/

7. Facebook’s ethical failures are not accidental; they are part of the business model – [Електронний ресурс] – Режим доступу: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8179701/