Градюк Іванна Миколаївна
Студентка н.гр. 205_СПД ННІ права та психології НАВС

Науковий керівник:
Тарасенко Володимир Петрович
кандидат фізико-математичних наук, доцент кафедри інформаційних технологій ННІ права та психології НАВС

ВИКОРИСТАННЯ «ШІ» ДЛЯ ВИЯВЛЕННЯ ФЕЙКІВ І ДЕЗІНФОРМАЦІЇ В СИСТЕМАХ ДЕРЖАВНОЇ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

У сучасному світі інформаційні потоки стали однією з головних арен боротьби за вплив. Фейки, дезінформація та інформаційні маніпуляції використовуються як інструменти гібридної війни, що становить серйозну загрозу національній безпеці. Особливо це актуально для України, яка вже понад десятиліття перебуває під інформаційним тиском з боку агресора. У таких умовах критично важливо мати ефективні інструменти для протидії дезінформації. Штучний інтелект (ШІ) відкриває нові можливості для автоматизованого виявлення неправдивих повідомлень, фейкових джерел та координованих інформаційних атак.

Безперечно, впровадження технологій штучного інтелекту знаменує собою якісно новий етап у боротьбі з загрозами інформаційній безпеці. Інтелектуальні системи отримують можливість у режимі реального часу здійснювати глибокий аналіз величезних масивів даних, що надходять з різних інформаційних джерел – від соціальних мереж до новинних агрегаторів та медіа-платформ. Завдяки цьому стає можливим оперативно виявляти стрімко розповсюджувані фейкові повідомлення, розпізнавати маніпулятивні інформаційні конструкції, спрямовані на руйнування суспільної єдності та довіри до інститутів, а також розкривати організовані дезінформаційні атаки, метою яких є дестабілізація внутрішньої ситуації в країні.

Штучний інтелект здатний не лише реагувати на вже виявлені випадки дезінформації, а й аналізувати потоки інформації для виявлення потенційних загроз на ранніх стадіях їхнього формування. Це досягається шляхом виявлення певних лінгвістичних патернів, різких змін у тематиці обговорень або активності певних груп користувачів, що може свідчити про підготовку або початок інформаційної атаки.

Штучний інтелект може використовуватися:

1) для автоматичного аналізу великої кількості даних, що надходять з різних джерел,

2) для виявлення неправдивої інформації,

3) для розпізнавання змінених (підроблених) зображень шляхом порівняння їх з оригінальними,

4) для встановлення принципів, схем і способів поширення дезінформації,

5) для блокування виявленої дезінформації [1].

В умовах глобального інформаційного простору дезінформація часто поширюється різними мовами. Системи штучного інтелекту, навчені на багатомовних даних, здатні аналізувати та виявляти неправдиву інформацію незалежно від мови її поширення, що є критично важливим для протидії зовнішньому інформаційному впливу. Розробка передових багатомовних моделей ШІ дозволяє здійснювати раннє виявлення фейків та надавати прозорі пояснення щодо прийнятих рішень, що сприяє ефективній перевірці інформації людиною. Важливо, щоб такі системи ШІ розроблялися та використовувалися з дотриманням принципів справедливості, прозорості, підзвітності та безпеки, що забезпечує довіру суспільства та ефективність контрзаходів проти дезінформації [2].

Одним із найнебезпечніших інструментів поширення дезінформації є синтетичний медіаконтент, зокрема фальшиві зображення, відео та аудіофайли, створені за допомогою технологій штучного інтелекту, таких як deepfake. Ці технології дозволяють маніпулювати зовнішністю або голосом людини, створюючи реалістичні, але неправдиві матеріали, які можуть дискредитувати публічних осіб, викривляти факти або сіяти паніку в суспільстві.

У відповідь на ці загрози, сучасні алгоритми ШІ розвиваються у напрямку детектування ознак підробки та маніпуляцій. Це включає аналіз метаданих, виявлення невідповідностей у тінях, звукових частотах або глибинах зображення. Наприклад, нейронні мережі, спеціально навчені на великій кількості справжніх та фальшивих зразків, можуть з високою точністю виявляти використання штучно згенерованих облич або голосів. Це відкриває нові можливості для протидії аудіовізуальній дезінформації, знижуючи ймовірність успішного поширення фейкових матеріалів через соціальні мережі та месенджери.

В Україні вже впроваджуються практичні рішення для боротьби з дезінформацією. Зокрема, платформа Mantis Analytics, розроблена українськими фахівцями, використовує штучний інтелект для моніторингу інформаційного простору в режимі реального часу. Ця система аналізує тисячі повідомлень з медіа та соціальних мереж, виявляючи фейки та інформаційно-психологічні операції (ІПСО), що дозволяє оперативно реагувати на загрози інформаційній безпеці держави. Сучасні виклики інформаційної безпеки вимагають нових технологічних рішень, і штучний інтелект уже сьогодні доводить свою ефективність у боротьбі з дезінформацією.

Його здатність швидко обробляти великі обсяги даних, виявляти приховані зв’язки, аналізувати медіаконтент та розпізнавати фейки значно підсилює можливості держави в інформаційній протидії.

Запровадження ШІ в системи державної інформаційної безпеки дозволяє не лише оперативно реагувати на загрози, а й діяти на випередження – виявляючи інформаційні атаки ще на етапі їхнього поширення. Водночас важливо враховувати етичні, правові та технічні аспекти використання таких технологій, щоб забезпечити баланс між безпекою та правами громадян.

Список використаних джерел:

1. Штучний інтелект в системі інформаційної безпеки України в умовах російсько - української війни//Воропаєва Т. С., Авер'янова Н. М.//ст. 62-65 // URL: https://previous.scientia.report/index.php/archive/article/view/2026/2042