Гой Катерина Володимирівна
студентка 202_СПС н.гр. ННІ права та психології НАВС

Науковий керівник:
Пакриш Олександр Євгенійович
кандидат технічних наук, доцент, доцент кафедри інформаційних технологій ННІ права та психології НАВС

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ АВТОМАТИЗАЦІЇ КІБЕРАТАК ТА СТВОРЕННЯ ДІПФЕЙКІВ

У сучасному світі стрімкий розвиток технологій суттєво змінює всі сфери людської діяльності. Особливе місце серед інновацій займає штучний інтелект (ШІ), який відкриває нові можливості для науки, медицини, освіти та бізнесу. Проте разом із перевагами виникають і нові ризики. Однією з найсерйозніших проблем є використання ШІ в кіберзлочинності, зокрема для автоматизації кібератак та створення діпфейків [1].

Актуальність цієї теми зумовлена тим, що кіберзагрози стають дедалі складнішими, а їх наслідки – масштабнішими. Зловмисники активно використовують сучасні технології для досягнення своїх цілей, що створює серйозну небезпеку для окремих осіб, організацій і навіть держав.

Штучний інтелект – це галузь комп'ютерних наук, яка займається створенням систем, здатних виконувати завдання, що зазвичай потребують людського інтелекту. До таких завдань належать навчання, аналіз інформації, прийняття рішень та розпізнавання образів.

Основними технологіями ШІ є: машинне навчання, нейронні мережі та обробка природної мови. Завдяки цим інструментам ШІ може аналізувати великі обсяги даних, знаходити приховані закономірності та адаптуватися до нових умов. Саме ці властивості роблять його потужним інструментом як для захисту інформації, так і для її порушення [2].

Однією з найбільш небезпечних сфер застосування штучного інтелекту є кіберзлочинність [3]. ШІ дозволяє автоматизувати багато процесів, які раніше потребували значних зусиль з боку людини.

По-перше, штучний інтелект здатний самостійно знаходити вразливості в комп'ютерних системах. Він аналізує програмний код і мережеву активність, виявляючи слабкі місця, які можуть бути використані для атак.

По-друге, ШІ використовується для створення шкідливого програмного забезпечення. Сучасні віруси можуть змінювати свою структуру, щоб уникати виявлення антивірусами. Такі програми здатні навчатися і адаптуватися до нових умов.

По-третє, значно вдосконалилися фішингові атаки. Завдяки ШІ зловмисники можуть створювати персоналізовані повідомлення, які виглядають максимально правдоподібно. Вони враховують інтереси, стиль спілкування та поведінку жертви, що значно підвищує ефективність атак.

Крім того, ШІ використовується в соціальній інженерії – методі психологічного впливу на людину з метою отримання конфіденційної інформації. Наприклад, зловмисник може імітувати стиль листування керівника компанії або знайомої особи.

Також варто згадати про автоматизовані DDoS-атаки, які за допомогою ШІ можуть змінювати свою тактику в режимі реального часу, ускладнюючи захист від них.

Окрему загрозу становлять діпфейки – штучно створені відео, аудіо або зображення, які виглядають як справжні. Вони створюються за допомогою складних алгоритмів, зокрема генеративних нейронних мереж. Діпфейки дозволяють підробляти обличчя людей у відео, змінювати голос або створювати повністю вигадані сцени. Сучасні технології настільки вдосконалені, що відрізнити підробку від реальності стає дуже складно [4].

Існують різні види діпфейків:

- відео (підміна обличчя);
- аудіо (імітація голосу);
- зображення;
- текстові фальсифікації.

Діпфейки становлять серйозну небезпеку для суспільства. Однією з головних загроз є політичні маніпуляції. За допомогою підроблених відео можна поширювати неправдиву інформацію про політичних діячів, впливаючи на громадську думку.

Ще однією проблемою є шахрайство. Наприклад, зловмисники можуть імітувати голос керівника компанії та віддавати фальшиві розпорядження про переказ коштів.

Діпфейки також використовуються для дискредитації людей, що може завдати значної шкоди їхній репутації. Крім того, вони сприяють поширенню дезінформації, викликаючи паніку та недовіру до медіа.

Не менш важливим є питання порушення приватності, оскільки зображення людини можуть використовуватися без її дозволу.

Сучасні кіберзлочинці часто поєднують різні технології для досягнення максимального ефекту. Наприклад, діпфейк може використовуватися разом із фішингом або соціальною інженерією.

Уявімо ситуацію, коли працівник компанії отримує відеозвернення від «керівника» з проханням терміново виконати певну дію. Якщо це відео є діпфейком, то ймовірність того, що працівник повірить і виконає вказівку, значно зростає.

Такі комбіновані атаки є особливо небезпечними, оскільки вони впливають не лише на технічні системи, а й на людську психологію.

Для боротьби з використанням ШІ у кіберзлочинності необхідно застосовувати комплексний підхід.

По-перше, важливо розвивати технології виявлення діпфейків. Це можуть бути спеціальні алгоритми, які аналізують відео та аудіо на наявність ознак підробки.

По-друге, необхідно підвищувати рівень кіберграмотності населення. Люди повинні вміти розпізнавати фішингові атаки та критично оцінювати інформацію.

По-третє, важливу роль відіграє законодавче регулювання. Необхідно розробляти закони, які обмежують незаконне використання ШІ.

Також слід використовувати сам штучний інтелект для захисту – наприклад, для виявлення аномалій у мережі та запобігання атакам.

Отже, штучний інтелект є потужним інструментом, який може використовуватися як на благо людства, так і на шкоду. Його застосування у сфері кіберзлочинності відкриває нові можливості для зловмисників, роблячи атаки більш складними, масштабними та небезпечними.

Особливу загрозу становлять діпфейки, які здатні впливати на свідомість людей, поширювати дезінформацію та підривати довіру до інформаційних джерел.

У зв’язку з цим надзвичайно важливо розвивати засоби захисту, підвищувати обізнаність суспільства та вдосконалювати законодавство. Лише комплексний підхід дозволить мінімізувати ризики та забезпечити безпечне використання штучного інтелекту.

Список використаних джерел:

1. Kietzmann J., Lee L. Deepfakes: Trick or treat? Understanding the rise of synthetic media. *Business Horizons*. 2020. Vol. 63(2). P. 135–146.
2. Goodfellow I., Pouget-Abadie J., Mirza M. Generative Adversarial Networks. *Communications of the ACM*. 2014. Vol. 63(1). P. 139–144.
3. Brundage M., Avin S., Clark J. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *arXiv preprint*. 2018.
4. Tolosana R., Vera-Rodriguez R., Fierrez J. Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection. *Information Fusion*. 2020. Vol. 64. P. 131–148.