SUN'IY INTELLEKT SHAROITIDA MA'LUMOTLAR HIMOYASI: TAHDIDLAR VA YECHIMLAR

Panjiyeva Gulhayo
Chirchiq davlat pedagogika universiteti

Annotatsiya: Sun'iy intellekt (SI) texnologiyalarining keng tarqalishi kiberxavfsizlik sohasida yangi imkoniyatlar bilan bir qatorda, jiddiy tahdidlarni ham yuzaga keltirmoqda. Ushbu maqolada SI bilan bog'liq kiberxavfsizlik muammolari chuqur tahlil qilinadi. Jumladan, deepfake texnologiyalarining xavfi, shaxsiy ma'lumotlar bilan bog'liq hujumlar va differential privacy tamoyili asosidagi himoya mexanizmlari ko'rib chiqiladi. Maqolada global statistika, real tadqiqotlar va amaliyotda qo'llanilayotgan texnologik yondashuvlar asosida ushbu muammolarning dolzarbligi va ularni bartaraf etish yo'llari tahlil qilinadi. Tadqiqot natijalariga ko'ra, SI asosidagi himoya usullari (masalan, avtomatlashtirilgan tahdid aniqlash, differensial maxfiylik) kiberxavfsizlikni oshirishda katta salohiyatga ega, biroq ularning o'zi yetarli emas — yuridik va etik me'yorlar bilan birga ishlashi muhim.

Kirish so'z. kiberxavfsizlik, ma'lumotlar tahlili, deepfake, aniqlash algoritmlari, shaxsiy ma'lumotlar

Annotation: The widespread adoption of artificial intelligence (AI) technologies is bringing both new opportunities and serious challenges in the field of cybersecurity. This paper explores key security issues related to AI, including the risks posed by deepfake technologies, personal data breaches, and privacy protection through differential privacy principles. Based on global statistics, recent academic research, and practical technological approaches, the paper analyzes the relevance of these threats and the effectiveness of current solutions. The findings suggest that AI-based defense mechanisms — such as automated threat detection and differential privacy — have great potential to enhance cyber security. However, they are not sufficient on their own and must be complemented by legal and ethical frameworks.

Keywords: cybersecurity, data analysis, deepfake, detection algorithms, personal data

Аннотация: Широкое распространение технологий искусственного интеллекта (ИИ) приносит как новые возможности, так и серьёзные угрозы в сфере кибербезопасности. В данной статье рассматриваются ключевые проблемы безопасности, связанные с ИИ, включая риски, связанные с технологией дипфейков, утечки персональных данных и механизмы защиты конфиденциальности на основе дифференциальной приватности. На основе глобальной статистики, современных научных исследований и практических подходов анализируется актуальность указанных угроз и эффективность

существующих решений. Результаты показывают, что методы защиты, основанные на ИИ — такие как автоматическое обнаружение угроз и дифференциальная приватность — обладают высоким потенциалом для повышения уровня кибербезопасности, однако требуют дополнения нормативно-правовой и этической поддержкой.

Ключевые слова: кибербезопасность, анализ данных, дипфейк, алгоритмы обнаружения, персональные данные

KIRISH

So‘nggi yillarda sun’iy intellekt (SI) texnologiyalarining jadal rivojlanishi hayotning barcha jabhalariga, xususan, kiberxavfsizlik sohasiga ham sezilarli ta’sir ko‘rsatmoqda. Xususan, ma’lumotlarga asoslangan qarorlar qabul qilish, foydalanuvchi xatti-harakatlarini kuzatish va avtomatlashtirilgan xavf tahlilini amalga oshirishda SI vositalarining ahamiyati ortib bormoqda. Biroq, bu texnologiyalar bilan birga **kiberxavfsizlikka doir yangi tahdidlar** ham yuzaga chiqmoqda. Xususan, **deepfake** texnologiyalarining ommalashuvi, **shaxsiy ma’lumotlarni himoya qilishdagi murakkabliklar**, va **qonuniy bo‘lmagan ma’lumot sızdirilishi** bugungi kunda dolzarb muammolar sirasiga kiradi.

Statistik ma’lumotlarga ko‘ra, 2023-yilda butun dunyo bo‘yicha **ma’lumotlar buzilishi (data breach)** holatlari soni 8,2 milliarddan ortiq foydalanuvchiga ta’sir qilgan bo‘lib, bu 2021-yildagidan qariyb 45% ko‘pdir [1]. Ayniqsa, tibbiyot, ta’lim va moliyaviy sektorlardagi ma’lumotlar, nisbatan himoyasiz ekani sababli, SI asosida uyushtirilgan hujumlar uchun asosiy nishonga aylanmoqda. Shu bilan birga, ma’lumotlarni yashirin holda yig‘uvchi AI vositalarining ko‘payishi shaxsiy hayot daxlsizligiga tahdid solmoqda.

Sun’iy intellekt asosida yaratilgan **deepfake texnologiyalari** esa eng xavfli tahdidlar qatoriga kiradi. Ushbu texnologiya yordamida video yoki audio kontentni realdek ko‘rsatib, yolg‘on axborot tarqatish mumkin. 2019-yilda Facebook, Microsoft, va Amazon kabi kompaniyalar deepfake’ga qarshi kurashish uchun **Deepfake Detection Challenge** musobaqasini e’lon qildi [2]. Bu loyiha doirasida yuzlab ilmiy guruhlar sun’iy intellekt yordamida deepfake’larni aniqlashning samarali usullarini izlashdi. Ushbu musobaqa orqali soha olimlari orasida bu tahdidga qarshi global ilmiy harakat boshlandi.

Kiberxavfsizlik masalasida yana bir dolzarb yondashuv — **Differential Privacy** (farqlovchan maxfiylik) tamoyilidir. Bu uslub foydalanuvchi ma’lumotlarining tahlil qilinishiga imkon berar ekan, ularning shaxsiylik darajasini saqlab qoladi. Ushbu yondashuv hozirgi kunda Google, Apple kabi kompaniyalar tomonidan amaliyotda qo‘llanilmoqda. Maqolada biz aynan ushbu uch yo‘nalishga — **ma’lumotlar**

himoyasi, deepfake texnologiyalari xavfi va differential privacy asosidagi yechimlarga alohida e’tibor qaratamiz.

Ma’lumotlar himoyasi va sun’iy intellekt

AI orqali ma’lumotlar tahlili: xavf va foyda: Sun’iy intellekt (AI) bugungi kunda katta hajmdagi ma’lumotlarni tahlil qilishda beqiyos imkoniyatlar yaratmoqda. Kompaniyalar foydalanuvchi xatti-harakatlarini aniqlash, xavfli tendensiyalarni oldindan bashorat qilish va real vaqtli monitoringni amalga oshirishda AI algoritmlaridan keng foydalanmoqda. Masalan, bank tizimlarida AI yordamida firibgarliklarni (fraud detection) aniqlash darajasi 90% dan yuqoriga ko‘tarilgan (McKinsey Global Institute, 2020).

Biroq, ushbu texnologiyalar noto‘g‘ri maqsadlarda ham qo‘llanilishi mumkin. Ma’lumotlarga asoslangan qarorlar diskriminatsion natijalarga olib kelishi yoki foydalanuvchining maxfiyligini buzishi ehtimoli yuqori. Shuningdek, ma’lumotlar noto‘g‘ri tuzilgan yoki etarli darajada anonimlashtirilmagan bo‘lsa, AI model orqali shaxsni qayta aniqlash mumkin. Bu esa **GDPR** (General Data Protection Regulation) singari xalqaro qonunlar bilan zid keladi.

AI yordamida hujumlarni aniqlash (IDS/IPS tizimlari)

Intrusion Detection System (IDS) va Intrusion Prevention System (IPS) tizimlari — kiberxavfsizlikdagi muhim elementlardir. An’anaviy IDS/IPS tizimlari belgilangan qoidalar asosida ishlagan bo‘lsa, hozirda ular **AI asosida o‘z-o‘zini o‘rganuvchi** tizimlarga aylanishmoqda. Masalan, **Kaspersky Adaptive Anomaly Detection** tizimi o‘zida AI modelni mujassamlashtirgan bo‘lib, tizimdagi g‘ayritabiiy xatti-harakatlarni real vaqt rejimida aniqlay oladi.

AI asosidagi IDS/IPS tizimlarining asosiy afzalligi — **aniq, moslashuvchan va doimiy yangilanadigan himoya mexanizmi**dir. Ular statik hujum signaturalariga tayanmasdan, yangi turdagi hujumlarni ham aniqlashga qodir. Xususan, **Recurrent Neural Networks (RNN)** va **Decision Tree-based ensemble modellari** bu borada samaradorligi isbotlangan.

AI orqali foydalanuvchi ma’lumotlarini himoya qilish imkoniyatlari

AI nafaqat hujumlarni aniqlash, balki ma’lumotlarni **proaktiv tarzda himoya qilish**da ham qo‘llaniladi. Misol uchun, **access control (kirish huquqlari)** ni tahlil qilishda AI foydalanuvchi odatlarini o‘rganib, g‘ayrioddiy harakatlarni to‘xtatadi. Shuningdek, **homomorfik shifrlash** va **federated learning** kabi AI bilan bog‘liq texnikalar orqali shaxsiy ma’lumotlar tahlil qilinadi, lekin markaziy serverda saqlanmaydi — bu esa xavfsizlikni kuchaytiradi.

Deepfake texnologiyalari va ularning xavfi

Deepfake — bu **Generative Adversarial Networks (GANs)** asosida ishlovchi texnologiya bo‘lib, u mavjud video, audio yoki rasmga realistik soxta elementlarni

qo‘shishga imkon beradi. Masalan, kimningdir yuzini boshqa bir video orqali “tiktirib”, u gapirmagan so‘zlarni aytayotgandek ko‘rsatish mumkin.

So‘nggi yillarda deepfake texnologiyasi yordamida siyosiy arboblar, aktyorlar, hatto oddiy odamlar nomidan noto‘g‘ri kontentlar tarqatilgan. 2022-yilda Ukraina prezidenti Zelenskiyning “taslim bo‘lishga chaqirgan” soxta videosi butun dunyo bo‘ylab jiddiy xavotir uyg‘otgan edi (BBC, 2022).

Jamiyat va siyosatga ta’siri

Deepfake'lar nafaqat shaxsiy obro‘, balki **ijtimoiy ishonch, davlat barqarorligi va axborot xavfsizligi** uchun ham tahdid tug‘diradi. Ular orqali siyosiy provokatsiyalar, yolg‘on guvohliklar, moliyaviy firibgarliklar amalga oshirilmoqda. Amerika Qo‘shma Shtatlarida deepfake’lar orqali moliyaviy firibgarlik holatlari 2023-yilda 65% ga oshgan (FTC Report, 2023).

Aniqlash algoritmlari: CNN, Autoencoder, Frequency Analysis

Deepfake’larni aniqlashda **Convolutional Neural Networks (CNN)**, **Autoencoder**, hamda **Frequency Domain Analysis** asosida ishlaydigan algoritmlar qo‘llaniladi. Ulardan ba'zilari:

- *XceptionNet:* Yuzning mikro ifodalarini tahlil qilib, video manipulyatsiyasini aniqlaydi.
- *MesoNet:* Yuqori darajada siqilgan videolarda deepfake izlarini aniqlashga moslashgan.
- *FFT-based Detection:* Surat yoki videodagi past chastotali o‘zgarishlarni ko‘rib chiqadi.

Ilmiy ishlanmalar: DeepFake Detection Challenge. 2020-yilda Facebook, Microsoft va Amazon hamkorligida o‘tkazilgan **DeepFake Detection Challenge (DFDC)** loyihasi bu sohadagi ilmiy yondashuvlar rivojiga turtki bo‘ldi. Loyiha natijalariga ko‘ra, **EfficientNet**, **ResNet50**, **Hybrid CNN-LSTM** modellari eng samarali aniqlash mexanizmlari sifatida ajralib chiqdi (Dolhansky et al., 2020).

Differential Privacy: Shaxsiy ma’lumotlarni himoyalash

An’anaviy himoya va Differential Privacy farqi. An’anaviy ma’lumotlar himoyasi usullari, masalan, shifrlash yoki anonymization, ma’lumotlar bazasidan shaxsiy identifikatorlarni olib tashlash orqali ishlaydi. Biroq bu usullar ko‘pincha rekonstruktsiya hujumlari oldida ojiz qoladi. Bunga qarshi eng zamonaviy yondashuvlardan biri — **Differential Privacy (DP)** hisoblanadi.

DP har bir foydalanuvchining hissa qo‘shgan ma’lumotlarini himoya qiladi, hatto butun modelga kirgan taqdirda ham, *biror foydalanuvchining ishtiroki aniqlanmasligini* kafolatlaydi (Dwork, 2006).

DP algoritmlari (Laplace, Gaussian noise)

Differential Privacy algoritmlarida **matematik shovqin** (noise) kiritiladi:

• **Laplace mechanism**: javoblarga tasodifiy shovqin qo‘shib, maxfiylikni saqlaydi.

• **Gaussian mechanism**: ko‘proq statistik barqarorlik kerak bo‘lganda qo‘llaniladi.

Ushbu algoritmlar real taqsimotga yaqin natijalar beradi, lekin foydalanuvchining haqiqiy ma’lumotini yashirib turadi.

Google, Apple’ning real loyihalari

DP hozirda **Google Chrome**, **Apple iOS** tizimlarida joriy etilgan. Masalan, Apple foydalanuvchi klaviatura yozuvlari orqali qaysi emoji'lar mashhur ekanini aniqlaydi, lekin **foydalanuvchining individual yozuvlarini serverga yubormaydi** — bu DP mexanizmi orqali amalga oshadi.

Talabalar ma’lumotlari kontekstida qo‘llash

Ta’lim sohasida DP yordamida talabalar reytingi, imtihon natijalari yoki ishtirok statistikasi kabi sezgir ma’lumotlar xavfsiz tahlil qilinadi. Bu ayniqsa **onlayn ta’lim tizimlari** uchun muhim: foydalanuvchi xatti-harakatlari tahlil qilinadi, lekin shaxsiy identifikatsiya yo‘q qilinadi.

XULOSA

Sun’iy intellekt texnologiyalarining jadal rivojlanishi kiberxavfsizlik sohasida katta imkoniyatlar yaratmoqda. Shu bilan birga, bu jarayon ma’lumotlar himoyasi, yolg‘on kontentlar (deepfake) va shaxsiylik masalalarida yangi xavf va muammolarni yuzaga keltirmoqda. Maqolada tahlil qilingan holatlar shuni ko‘rsatadiki:

AI asosidagi tahdidlar – ayniqsa deepfake texnologiyasi – nafaqat shaxsiy hayotga, balki davlat xavfsizligiga ham tahdid solmoqda;

Differential Privacy kabi zamonaviy yondashuvlar esa ma’lumotlar himoyasini ta’minlashda istiqbolli va ilmiy asoslangan vosita hisoblanadi;

AI yordamida hujumlarni aniqlash tizimlari (IDS/IPS) klassik yondashuvlarga qaraganda ancha samarali, ayniqsa o‘zgaruvchan kiberxavflar sharoitida.

Shunga asoslanib, quyidagi tavsiyalarni ilgari surish mumkin:

AI texnologiyalarini qo‘llashda yuridik va axloqiy standartlar ishlab chiqilishi zarur. Deepfake texnologiyasining noqonuniy ishlatilishini oldini olish uchun global qonunchilikka muvofiq algoritmik nazorat tizimlari joriy etilishi lozim.

O‘zbekistonda ham Differential Privacy kabi ilg‘or maxfiylik texnologiyalarini ta’lim va sog‘liqni saqlash sohalariga tadbiq etish maqsadga muvofiq. Ayniqsa talabalar, bemorlar, ijtimoiy aholini qamrab oluvchi axborot tizimlarida.

AI asosidagi kiberxavfsizlik tizimlarini amaliyotda keng joriy qilish kerak. Bu ayniqsa moliya, transport va hukumat portallarida tahdidlarni oldindan aniqlash imkonini beradi.

Akademik doirada deepfake aniqlash bo‘yicha ilmiy izlanishlar rag‘batlantirilishi lozim. O‘zbek tilidagi kontentlar uchun chuqur o‘rganilgan modellar yaratish dolzarb.

FOYDALANILGAN ADABIYOTLAR:

1. IBM Security. (2023). *Cost of a Data Breach Report*. https://www.ibm.com/reports/data-breach
2. Dolhansky, B., et al. (2020). *The Deepfake Detection Challenge Dataset*. arXiv:2006.07397.
3. **McKinsey Global Institute.** (2020). *The Future of Work in Technology*. https://www.mckinsey.com
4. **Dwork, C.** (2006). Differential Privacy. *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*. Springer. https://doi.org/10.1007/11787006_1
5. **Federal Trade Commission (FTC).** (2023). *Consumer Sentinel Network Data Book 2023*. https://www.ftc.gov
6. **BBC News.** (2022). *Ukraine war: Deepfake video of President Zelensky circulates online*. https://www.bbc.com/news/technology-60780142
7. **Kaspersky Lab.** (2023). *Adaptive Anomaly Detection Overview*. https://www.kaspersky.com
8. **Apple Inc.** (2017). *Learning with Privacy at Scale*. Apple Machine Learning Journal, 1(8). https://machinelearning.apple.com/research/learning-with-privacy-at-scale
9. **IBM Security.** (2023). *Cost of a Data Breach Report 2023*. IBM Corporation. https://www.ibm.com/reports/data-breach
10. **Dolhansky, B., Howes, R., Pflaum, B., Baram, N., Roessler, A., & Ferrer, C. C.** (2020). The Deepfake Detection Challenge (DFDC) Dataset. *arXiv preprint arXiv:2006.07397*. https://arxiv.org/abs/2006.07397