

CONFERENCE ON FRONTIERS OF SCIENCE AND ENGINEERING • QATAR •

MEDIA MAKONIDA SUN'IY INTELLEKT TEXNOLOGIYALARINING POTENTIAL XAVFLARI VA ULARNI BARTARAF ETISH USULLARI.

Babadjanov Salohiddin Sobitdjonovich

*O'z Res Harbiy xavfsizlik va mudofaa universiteti Qurolli Kuchlar Bosh shtabi
fakulteti, Istiqbolli harbiy texnologiyalar kafedrasi dotsenti
QK Xizmatchisi p.f.f.d., v.b., professor*

Annotatsiya Ushbu tezisda zamonaviy media makonida sun'iy intellekt (SI) texnologiyalarining jadal rivojlanishi natijasida yuzaga kelayotgan axborot xavfsizligi tahdidlari tahlil qilingan. Xususan, deepfake, avtomatlashtirilgan dezinformatsiya va shaxsga qaratilgan ijtimoiy muhandislik xurujlari kabi potensial xavflar ko'rib chiqiladi. Shuningdek, soxta kontentni aniqlash (AI-detection) tizimlari, raqamli tarmoqlar va kriptografik verifikatsiya kabi proaktiv himoya mexanizmlari hamda axborot muhitini himoya qilish bo'yicha kompleks yondashuvlar taklif etilgan.

Kalit so'zlar: Sun'iy intellekt, media makon, axborot xavfsizligi, deepfake, dezinformatsiya, AI-detection, raqamli tarmoq.

Аннотация: В данных тезисах анализируются угрозы информационной безопасности, возникающие в результате стремительного развития технологий искусственного интеллекта (ИИ) в современном медиапространстве. В частности, рассматриваются такие потенциальные риски, как дипфейки, автоматизированная дезинформация и целенаправленные атаки с использованием социальной инженерии. Также предлагаются проактивные механизмы защиты, включая системы обнаружения поддельного контента (AI-detection), цифровые водяные знаки и криптографическую верификацию, наряду с комплексными подходами к защите информационной среды.

Ключевые слова: искусственный интеллект, медиапространство, информационная безопасность, дипфейк, дезинформация, AI-detection, цифровые водяные знаки.

Abstract: This thesis analyzes the information security threats arising from the rapid development of artificial intelligence (AI) technologies in the modern media space. In particular, it examines potential risks such as deepfakes, automated disinformation, and targeted social engineering attacks. Furthermore, proactive defense mechanisms are

CONFERENCE ON FRONTIERS OF SCIENCE AND ENGINEERING • QATAR •

proposed, including fake content detection systems (AI-detection), digital watermarking, and cryptographic verification, along with comprehensive approaches to protecting the information environment.

Keywords: artificial intelligence, media space, information security, deepfake, disinformation, AI-detection, digital watermarks.

Zamonaviy axborot jamiyatida sun'iy intellekt (SI) texnologiyalari media makonini tubdan o'zgartirib, kontent yaratish, uni tarqatish va tahlil qilish jarayonlarini avtomatlashtirishda mislsiz imkoniyatlar yaratmoqda. Biroq, bu jadal taraqqiyot o'z-o'zidan yangi, murakkab va yashirin axborot xavfsizligi tahdidlarini ham keltirib chiqarmoqda. An'anaviy kiberxavfsizlik tizimlari SI yordamida shakllantirilgan va yuqori darajada ishonchli ko'rinuvchi soxta axborot xurujlariga qarshi doim ham yetarli darajada samarali ishlay olmaydi. Ushbu tezisning asosiy maqsadi — media makonida generativ SI texnologiyalarining qo'llanilishi natijasida yuzaga kelayotgan potentsial xavflarni tizimli tahlil qilish va axborot xavfsizligini ta'minlashning proaktiv usullarini ko'rib chiqish lozim.

Sun'iy intellekt vositasida yuzaga keladigan asosiy xavflar **Deepfake**¹ va vizual manipulyatsiya: Generativ raqobatbardosh tarmoqlar (GAN) kabi zamonaviy neyrotarmoqlar yordamida yaratiladigan "Deepfake" texnologiyasi audio va video kontentni o'ta aniqlik bilan qalbakilashtirish imkonini bermiqda. Bu texnologiya jamoatchilik fikrini sun'iy ravishda manipulyatsiya qilish, shaxslarning obro'siga putur yetkazish va hatto ijtimoiy-siyosiy beqarorlikni keltirib chiqarish kabi jiddiy xavflarni o'z ichiga oladi.

Avtomatlashtirilgan dezinformatsiya va feyk xabarlar: Katta til modellari (**LLM**)² orqali inson yozgan matndan uslubiy va mantiqiy jihatdan farq qilmaydigan feyk xabarlarni qisqa vaqt ichida ommaviy ravishda generatsiya qilish mumkin. Botlar tarmog'i orqali tarqatiladigan bu turdagi xurujlar ijtimoiy tarmoqlarda sun'iy vahima

¹ Deepfake — bu sun'iy intellekt yordamida yaratilgan soxta (lekin juda real ko'rinadigan) video, audio yoki rasm hisoblanadi. Unda odam go'yoki aytmagan gapni aytgandek yoki qilmagan ishni qilgandek ko'rsatiladi.

² LLM — bu Large Language Model (katta til modeli) degani. Bu sun'iy intellektning bir turi bo'lib, u matnni tushunish, yozish va odamga o'xshab muloqot qilish uchun yaratilgan.

CONFERENCE ON FRONTIERS OF SCIENCE AND ENGINEERING QATAR

uyg'otishga va algoritmik aks-sado (**echo chamber**³) effektini kuchaytirishga xizmat qiladi.

Ijtimoiy muhandislikdagi yangi bosqich: SI algoritmlari foydalanuvchilarning tarmoqdagi raqamli izlarini (**digital footprint**⁴) chuqur tahlil qilib, ularning psixologik portretini shakllantira oladi. Bu esa o'ta aniq va shaxsga qaratilgan (**spear-phishing**⁵) axborot hujumlarini avtomatlashtirilgan tarzda amalga oshirishga zamin yaratadi.

Texnik va algoritmik himoya choralari: Soxta kontentni aniqlash tizimlari (**AI-detection**⁶): Sun'iy intellekt tomonidan yaratilgan tahdidlarga qarshi kurashish uchun xuddi shunday kuchli va ixtisoslashgan SI modellarini qo'llash zarur. Xususan, videolardagi g'ayritabiiy pikseller almashinuvini, yorug'lik tushishidagi anomaliyalarni yoki matnlardagi sintaktik va semantik g'alizliklarni aniqlovchi "**neyron tarmoq**" (**inglizcha: neural network**)⁷ detektorlarini media platformalar arxitekturasiga integratsiya qilish muhim ahamiyat kasb etadi.

Raqamli moybo'yoqlar va **kriptografik**⁸ (**cryptographic**) verifikatsiya: Media fayllarning haqiqiyligini tasdiqlash uchun generatsiya jarayonida ko'zga ko'rinmas raqamli moybo'yoqlar (**invisible watermarking**) joylashtirish amaliyoti tobora kengaymoqda. Shuningdek, kontentning dastlabki manbasi va o'zgarish tarixini o'zgarimas tarzda saqlovchi kriptografik hamda **blockchain**⁹ usullarini qo'llash orqali ishonchli raqamli axborot muhitini yaratish mumkin.

Xulosa Media makonida sun'iy intellekt xavflari kundan-kunga murakkablashib borayotgan bir sharoitda, himoya strategiyasi faqatgina reaktiv (hujumdan keyingi) emas, balki proaktiv (hujumning oldini oluvchi) mexanizmlarga asoslanishi talab etiladi. Axborot xavfsizligini ta'minlash ilg'or algoritmik yechimlar, qat'iy kiberxavfsizlik

³ Echo chamber — bu odam faqat o'z fikriga o'xshash qarashlarni eshitadigan va qarama-qarshi fikrlar deyarli uchramaydigan muhitdir. Ya'ni, siz nima ishonsangiz, atrofingiz (ijtimoiy tarmoqlar, do'stlar, yangiliklar) aynan shu fikrni "qaytarib" beradi.

⁴ Digital footprint — bu siz internetda qoldiradigan barcha izlar, ya'ni onlayn faoliyatingiz natijasida yig'iladigan ma'lumotlar to'plamidir.

⁵ Spear-phishing — bu aniq bir odam yoki tashkilotni nishonga olgan, juda puxta o'ylangan firibgarlik turi. Bu oddiy phishingdan farqli ravishda, umumiy emas, balki sizga moslab tayyorlanadi.

⁶ AI-detection — bu matn, rasm, video yoki audio sun'iy intellekt tomonidan yaratilganmi yoki yo'qmi, shuni aniqlash jarayonidir.

⁷ Neyron tarmoq — bu inson miyasidan ilhomlangan sun'iy intellekt modeli bo'lib, u ma'lumotlarni o'rganish va tahlil qilish uchun ishlatiladi.

⁸ Kriptografik — bu ma'lumotlarni himoya qilish va yashirish bilan bog'liq tushuncha. U Cryptography (kriptografiya) sohasiga tegishli.

⁹ Blockchain — bu ma'lumotlarni xavfsiz va o'zgartirib bo'lmaydigan tarzda saqlaydigan raqamli tizim (texnologiya). U Distributed Ledger Technology ga asoslangan — ya'ni ma'lumotlar bitta joyda emas, balki ko'plab kompyuterlarda saqlanadi.

CONFERENCE ON FRONTIERS OF SCIENCE AND ENGINEERING • QATAR •

protokollari hamda ommaviy media savodxonlikni oshirishni o'z ichiga olgan kompleks yondashuvni taqozo etadi. Milliy va global media platformalar darajasida kontentni verifikatsiya qilishning ishonchli, avtomatlashtirilgan tizimlarini joriy etish kechiktirib bo'lmaz zaruratdir.

Internet manbalari

- 1) IBM Security. (2023). AI and Automation in Cybersecurity. <https://www.ibm.com/security>
- 2) MIT Technology Review. (2024). How AI-generated disinformation is changing the modern media landscape. <https://www.technologyreview.com/>
- 3) ENISA (European Union Agency for Cybersecurity). (2022). Artificial Intelligence Threat Landscape. (AI vositasida yuzaga keladigan axborot xurujlari bo'yicha rasmiy Yevropa hisoboti).

CONFERENCE ON FRONTIERS OF SCIENCE AND ENGINEERING • QATAR •

Bu yerda Deepfake, LLM, Echo chamber, Digital footprint, Spear-phishing, AI-detection, Neyron tarmoq, Kriptografiya va Blockchain kabi barcha asosiy tushunchalar vizual tarzda va ularning ta'riflari bilan birga ko'rsatilgan

Yuqorida ko'rsatilgan atamalar asosida tayyorlangan infografika:

