EURASIAN JOURNAL OF ACADEMIC RESEARCH
Innovative Academy Research Support Center
UIF = 8.1 | SJIF = 5.685 www.in-academy.uz

SUN'IY INTELEKT VA CHUQUR O'RGANISHLAR ASOSIDA KIBERXAVFSIZLIKNI YAXSHILASH

Raxmonjonov G'iyosjon Ibroxim o'g'li
Toshkent davlat yuridik universiteti, xalqaro huquq va qiyosiy huquqshunoslik fakulteti, 3-bosqich talabasi
https://doi.org/10.5281/zenodo.7317536

ARTICLE INFO

Received: 03rd November 2022
Accepted: 12th November 2022
Online: 14th November 2022

KEY WORDS

kiberxavfsizlik, sun'iy intellekt, xavfsizlik metodlari, chuqur o'rganish.

ABSTRACT

So'nggi paytlarda sun'iy intellekt imkoniyatlaridan kiberxavfsizlik yo'nalishi bo'yicha keng doirada foydalanishga urinishlar bo'lmoqda. Shuning uchun, ushbu maqolada foydalanuvchilarning tarmoqqa kirish autentifikatsiyasi, tarmoqdagi vaziyatdan xabardorlik, xavfli xatti-harakatlar monitoringgi va noodatiy harakatlarni aniqlash sohalaridagi sun'iy intellektdan foydalanish bo'yicha mavjud bo'lgan bir qator so'nggi adabiyotlar va yutuqlar tahlil qilinadi. Mazkur maqolada, shuningdek, sun'iy intellekt va kiberxavfsizlikning o'zaro aloqador jihatlari ham batafsil tahlil etiladi.

Jamiyatning o'zaro bog'liqligi va hayotimizning texnologiyalarga tayanganligi darajasi ortgani sari, xavfsizlik masalalari va strategiyalarining ahamiyati ortib bormoqda. Tizimlarimizni va jamiyatimizni himoya qilish maqsadida yanada samarali kiberxavfsizlik usullarini ishlab chiqish masalasi kun tartibidagi eng aktual masalalardan biri bo'lib qolmoqda.

Kiberxavfsizlik tushunchasi tizimlarning har xil hujumlar va tahdidlardan himoyalanganligini, foydalanuvchilarni xavfsiz hamda samarali xizmatlar bilan ta'minlash holatini anglatuvchi tushuncha bo'lib, bu borada turli metodlar va choralarni qo'llashni taqozo etadi. Shu sababli, ushbu maqolada keltirilgan kiberxavfsizlik tushunchasi tashqi va tizim ichidagi tahdidlarni (ba'zi tadqiqotlarda **tarmoq xavfsizligi** deb nomlanadi) o'z ichiga oladi. Ushbu tahdidlar tizimlarning muntazam ishlashiga jiddiy ta'sir qiladi, shuning uchun kiberxavfsizlikning maqsadi tahdidlarni iloji boricha to'xtatib qolish va jinoiy harakatlar sodir etilishidan oldin aniqlash, hodisadan keyin tiklanish talablarini o'z vaqtida va samarali qondirishdir.

Turli tashkilotlarning kibertahdidlardan himoyalanishda sun'iy intellektdan foydalanishning ahamiyatini chuqur anglay boshlaganligi sababli so'nggi yillarda kiberxavfsizlikni ta'minlash bilan bog'liq masalalarda sun'iy intellektga asoslangan yechimlarni ishlab chiqishga urinishlar sezilarli darajada ko'paydi. Mazkur o'zgarishga kompyuter texnologiyalari sohasidagi rivojlanishlar ham muhim faktorlardan biri sifatida xizmat qilmoqda. Masalan, Stanford universitetining "AI Index 2019 Report" hisobotiga ko'ra, bulutli infrastrukturada katta hajmli rasmlarni klassifikatsiya qilish tizimini yaratishga sarflanadigan vaqt 2017-yil

noyabrdagi taxminan uch soatlik ko‘rsatkichdan 2019-yil iyuldagi 88 soniyalik natijagacha kamaygan.

Tadqiqotlar shuni ko‘rsatmoqdaki, kiberxavfsizlik sohasidagi sun’iy intellekt bozori 2016-yildagi 1 milliard dollarni tashkil etgan bo‘lsa, bu ko‘rsatkich 2025-yilga borib 34,8 milliard AQSH dollarigacha o‘sishi kutilmoqda.

Kiberxavfsizlik sohasida sun’iy intellektning ahamiyati bu darajada ortib borayotgani, boshqa sohalarda bo‘lgani singari, o‘zining ijobiy va salbiy taraflariga ega hamda mazkur jarayonlarni huquqiy jihatdan tartibga solish o‘ta murakkab bo‘lib, kuchli ilmiy va amaliy ko‘nikmalarni talab etadi.

Kiberxavfsizlik vazifalari hujumlarning oldini olish, hujumlarni aniqlash, tekshiruvlar o‘tkazish, tahdidlarni tasniflash va tahlil qilish, kiberxavfsizlik tizimlarini o‘qitish va modellashtirishdan iborat.

Hujumlarning oldini olish (profilaktika) – bu dasturiy ta’minotdagi zaifliklar sonini kamaytirishga qaratilgan harakatlar. Masalan, Kouliaridis va boshqalar tomonidan amalga oshirilgan ilmiy izlanishda bunga odatiy misol mavjud bo‘lib, unda Android tizimlarida ilovalarning zararliligini aniqlaydigan avtomatik o‘rganish tizimlari tavsiflangan[1]. Bunda ilovalarning xususiyatlari (xarakteristikalari) yig‘iladi hamda tasniflagichlar (klassifikatorlar) datasetlarda ilovalar bo‘yicha o‘qitiladi.

Hatto Random Forest yetakchi bo‘lgan ishlatiladigan tasniflash usullari bo‘yicha statistik ma’lumotlar ham mavjud.

Microsoft kompaniyasi sun’iy intellekt va kiberxavfsizlik mavzularini hamda ularning o‘zaro aloqasini quyidagi tazrda klassifikatsiya qilishni taklif qildi[2]:

– sun’iy intellekt yordamida kiberxavfsizlikni yaxshilash (kiberxavfsizlikda sun’iy intellektdan foydalanish),

– sun’iy intellekt yordamidagi kiberhujumlar (kiberhujumlarni kuchaytirish uchun sun’iy intellektdan foydalanish),

– sun’iy intellekt tizimlarining kiberxavfsizligi (sun’iy intellekt tizimlariga hujumlar),

– zararli axborot operatsiyalarida sun’iy intellektdan foydalanish (sun’iy intellektdan foydalangan holda soxta narsalar (feyklar)).

Biz ushbu klassifikatsiyaga murojaat qilamiz va maqolaning asosiy qismi ushbu bo‘linishga muvofiq tuzilgan.

Sun’iy intellekt yordamida kiberxavfsizlikni yaxshilash

2021 yilda AV-Test instituti 125 milliondan ortiq yangi zararli dasturlarni aniqlagan[3]. Zararli dasturlarning yangi variantlarini aniqlash uchun mashinaviy o‘rganish metodlarining avvalgi shablonlarni umumlashtirish qobiliyati kengaytirilgan himoya tizimini yaratish uchun asosiy vositalardan biri hisoblanadi.

Shuni ta’kidlash lozimki, Google Scholar platformasida “ML for malware detection”

[1] Kouliaridis, Vasileios, and Georgios Kambourakis. "A comprehensive survey on machine learning techniques for android malware detection." Information 12.5 (2021): 185.

[2] Applications for artificial intelligence in Department of Defense cyber missions https://blogs.microsoft.com/on-the-issues/2022/05/03/artificialintelligence-department-of-defense-cyber-missions/

[3] AV-Test Institute https://www.av-test.org/en/statistics/malware/

so‘rovi kiritilganda, bu mavzu bo‘yicha 20,000 dan ortiq ilmiy maqolalar mavjudligini ko‘rish mumkin[4].

Ushbu sohada chuqur o‘rganish metodi ham faol qo‘llanilmoqda. Yuan va boshqalarning 2014-yilda chop etilgan mazkur ilmiy izlanishida Xitoyning asosiy texnologiyalar uchun davlat granti asosida yaratilgan tizimi tasvirlangan[5]. Zararli dasturlarni aniqlash uchun chuqur o‘rganish modellarining qiziqarli qiyosiy tahlili Vinayakumar va boshqalarning 2019-yilgi ilmiy ishida o‘z ifodasini topgan[6]. Bunday ishlarning barchasi amaliy ahamiyatga ega, masalan, Microsoft 365 Defender ham chuqur o‘rganishdan foydalanadi[7].

Qayd etib o‘tishimiz lozimki, “dasturlar” tushunchasi bu yerda faqat kod sifatida tushunilmasligi kerak. Masalan, Tajaddodianfar va boshqalarning 2020-yildagi ilmiy izlanishida fishing URL manzillarini aniqlash uchun chuqur o‘rganish modeli tasvirlangan[8]. Va bu shunga o‘xshash ko‘plab ilmiy ishlardan faqat bitta misol, xolos. Umuman olganda, fishing hujumlari juda xilma-xil bo‘lib, ularni aniqlash uchun mashinani o‘rganishdan foydalanish 2008 yilda tasvirlangan[9].

Hujumlarni aniqlash shubhali xatti-harakatlarni tezda sezishni va ular sodir bo‘lganda to‘g‘ridan-to‘g‘ri ogohlantirishni o‘z ichiga oladi. Maqsad hujumlarga tezda javob berish, shu jumladan hujum ko‘lamini aniqlash, hujumchilar uchun kirishlarni yopish va tajovuzkorlar foydalanishi mumkin bo‘lgan zaifliklarni (“orqa eshiklar” (back doors”) va shu kabilar) yo‘q qilishdir.

Umuman olganda, noma’lum hujum shablonlarini qidirish ko‘p sonli noto‘g‘ri pozitivlarga (“false positive”) olib kelishi mumkin[10]. Adabiyotlarda ta’kidlanishicha, shubhali faoliyatlarni aniqlashning asosiy muammosi aniq xavfsizlik ogohlantirishlarini topish va noto‘g‘ri pozitivlar soni orqali yetarli qamrovni ta’minlash o‘rtasida to‘g‘ri muvozanatni ta’minlashdir.

Hujum haqida ogohlantirish uchun mashinaviy o‘rganishdan foydalanish bilan bog‘liq quyidagi yo‘nalishlarni ajratishimiz mumkin[11]:

(1) potensial hujumlar haqida ogohlantirishlarni birinchi o‘ringa qo‘yish[12],

[4] ML for malware detection https://scholar.google.com/scholar?q=ml+for+malware+detection&hl=en&as _sdt=0,5

[5] Yuan, Zhenlong, et al. "Droid-sec: deep learning in android malware detection." Proceedings of the 2014 ACM conference on SIGCOMM. 2014.

[6] Vinayakumar, R., et al. "Robust intelligent malware detection using deep learning." IEEE Access 7 (2019): 46717-46738

[7] Using fuzzy hashing and deep learning to counter malware detection evasion techniques https://www.microsoft.com/security/blog/2021/07/27/combing-through-thefuzz-using-fuzzy-hashing-and-deep-learning-to-counter-malware-detectionevasion-techniques/

[8] Tajaddodianfar, Farid, Jack W. Stokes, and Arun Gururajan. "Texception: a character/word-level deep learning model for phishing URL detection." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.

[9] Basnet, Ram, Srinivas Mukkamala, and Andrew H. Sung. "Detection of phishing attacks: A machine learning approach." Soft computing applications in industry. Springer, Berlin, Heidelberg, 2008. 373-383.

[10] Shenfield, Alex, David Day, and Aladdin Ayesh. "Intelligent intrusion detection systems using artificial neural networks." Ict Express 4.2 (2018): 95-99.

[11] Applications for artificial intelligence in Department of Defense cyber missions https://blogs.microsoft.com/on-the-issues/2022/05/03/artificialintelligence-department-of-defense-cyber-missions/

[12] Mishra, Preeti, et al. "A detailed investigation and analysis of using machine learning techniques for intrusion detection." IEEE Communications Surveys & Tutorials 21.1 (2018): 686-728.

(2) uzoq muddat davom etadigan katta va uzoqroq xakerlik kampaniyalarining bir qismi bo‘lgan ko‘p bosqichli xakerlik urinishlarini aniqlash[13],

(3) kompyuterda ham, tarmoqda ham zararli dasturlarning izlarini aniqlash[14],

(4) muayyan tashkilot orqali amalga oshiriladigan zararli dasturiy ta’minot oqimini aniqlash. “Living off The Land” (LotL) deb nomlanuvchi mazkur kiberhujumlarda tajovuzkorlar hujum harakatlarini amalga oshirish uchun tashkilotdagi legal dasturlardan foydalananishadi[15].

(5) hujum tarqalishining oldini olish uchun tezkor javob zarur bo‘lganda avtomatlashtirilgan hujumlarni yumshatish yondashuvlarini aniqlash. Masalan, avtomatlashtirilgan tizim, agar to‘lov dasturining harakatlari bilan bog‘liq bo‘lgan ogohlantirishlar ketma-ketligi aniqlansa, tarmoq ulanishini o‘chirib qo‘yishi va qurilmani bloklashi mumkin[16].

Tekshirish va tuzatish (hujumlardan keyingi qayta tiklanish) – bu xavfsizlik buzilishidan keyin qo‘llaniladigan usullar bo‘lib, mijozlarga xavfsizlik buzilishlari, shu jumladan buzilish darajasi, ta’sirlangan qurilmalar va ma’lumotlar ro‘yxati, hujumning tarqalishi va hodisa sabablari to‘g‘risida yaxlit tushuncha berish uchun mo‘ljallangan bo‘lib, bu yetarlicha yangi soha hisoblanadi. Misollar sifatida asarlarni nomlash mumkin[17]. Boshqa tomondan, ko‘p sonli hujumlar ular haqida katta hajmli ma’lumot to‘plashga olib keldi, shuning uchun tadqiqot qilinishi mumkin bo‘lgan yetarlicha materiallar mavjud. Jumladan, ushbu mavzu bo‘yicha DARPA’ning hujumlarni atributlash bo‘yicha qiziqarli taqdimoti bunga yaqqol misol bo‘la oladi[18].

Sun’iy intellekt yutuqlaridan tahdidlarni yuqori darajada tahlil qilishda ham qo‘llaniladi. Masalan, turli kampaniyalar (buzib kirishlar va boshqalar) o‘rtasidagi o‘xshashlikni aniqlashga yordam beradigan hujumlar to‘g‘risidagi ma’lumotlarni tahlil qilish uchun freymvorklar taqdim etilgan[19].

Sun’iy intellekt yordamidagi kiberhujumlar

Sun’iy intellekt vositasida sodir etilayotgan kiberhujumlar borasida hujumkor (tajovuzkor) sun’iy intellekt iborasini ishlatishimiz mumkin.

Bir guruh olimlar sun’iy intellekt yordamida sodir etiladigan kiberhujumlarning quyidagicha tarmoqlarga bo‘ladilar[20].

1. Bashorat qilish (prognozlash) – ilgari kuzatilgan ma’lumotlar asosida ba’zi prognozlarni amalga oshirish. Mashinaviy o‘rganish hujumining misoli – harakatga (tebranishga) asoslangan smartfondagi

[13] O’sha joyda.

[14] Alsaheel, Abdulellah, et al. "{ATLAS}: A sequence-based learning approach for attack investigation." 30th USENIX Security Symposium (USENIX Security 21). 2021.

[15] Ongun, Talha, et al. "Living-Off-The-Land Command Detection Using Active Learning." 24th International Symposium on Research in Attacks, Intrusions and Defenses. 2021.

[16] Kok, S., et al. "Ransomware, threat and detection techniques: A review." Int. J. Comput. Sci. Netw. Secur 19.2 (2019): 136.

[17] Noor, Umara, et al. "A machine learning framework for investigating data breaches based on semantic analysis of adversary’s attack patterns in threat intelligence repositories." Future Generation Computer Systems 95 (2019): 467-487.

[18] Enhanced Attribution https://www.enisa.europa.eu/events/cti-euevent/cti-eu-event-presentations/enhanced-attribution/

[19] Gao, Peng, et al. "Enabling efficient cyber threat hunting with cyber threat intelligence." 2021 IEEE 37th International Conference on Data Engineering (ICDE). IEEE, 2021.

[20] Yamin, Muhammad Mudassar, et al. "Weaponized AI for cyber attacks." Journal of Information Security and Applications 57 (2021): 102722.

tugmachalarni aniqlash bo‘lib[21], yuqoridagi boshqa misollar ijtimoiy tarmoq foydalanuvchilari uchun sezgir ma’lumotlarni bashorat qilish (hujum uchun zaif havolani qidirish)[22], dasturiy ta’minotning zaif tomonlarini qidirish va boshqalarda namoyon bo‘ladi[23].

2. Generatsiya – sun’iy intellekt yordamida kontent yaratish. Ushbu generatsiyaning tajovuzkor maqsadlariga media ma’lumotlarini soxtalashtirish[24], parollarni tanlash[25], trafikni o‘zgartirish[26] kabilarni misol qilib keltirishimiz mumkin. Oxirgisi (ingliz tilidagi adabiyotlarda – “traffic-space attacks”) aslida trafikni tahlil qilish (bosqinlarni aniqlash) uchun ishlatiladigan mashinaviy o‘rganish tizimiga nisbatan qarama-qarshi hujumdir. Hujumning maqsadi haqiqiy bosqinni yashirishdan iborat.

Dipfeyklar (“deepfakes”) – ushbu kategoriyadagi tajovuzkor sun’iy intellektning yana bir turi hisoblanadi. Dipfeykni qaysidir ma’noda mediafayl deb baholashimiz mumkin bo‘lib, ular chuqur o‘rganish metodi yordamida yaratiladi. Texnologiya fishing hujumini amalga oshirayotganda jabrlanuvchining ovozi yoki yuziga taqlid qilib, o‘zini jabrlanuvchi sifatida ko‘rsatish uchun ishlatilishi mumkin[27].

3. Analiz – bu, ma’lumotlar yoki modeldan foydali ma’lumotlarni olish yoki tahlil qilish jarayoni, ya’ni hujumga uchragan ML modelini o‘rganish, masalan, tasnifga ta’sir qiluvchi haqiqiy omillarni aniqlashdir. Bu tushuntirish yondashuvlaridan (LIME, SHAPLEY va boshqalar) foydalanishni anglatadi. Hujum qilingan modelning ishlashini tushunish samarali hujumlarni yaratish yoki bosqinlarni yashirish uchun juda muhimdir. Agar hujumga uchragan model mavjud bo‘lmasa, unda bunday tajribalar uning soyali nusxasida o‘tkazilishi mumkin.

4. Qidiruv – bu belgilangan mezonlar bo‘yicha hujum qilish uchun ma’lumot yoki obyektlarni topish vazifasi.

Bir nechta buzib kirilgan kameralardagi tasvirlardan odamni qidirish (identifikatsiya qilish)[28,29], ijtimoiy

[21] Abdul Rehman Javed, Mirza Omer Beg, Muhammad Asim, Thar Baker, and Ali Hilal Al-Bayatti. 2020. AlphaLogger: Detecting motion-based side-channel attack using smartphone keystrokes. Journal of Ambient Intelligence and Humanized Computing (2020), 1–14.

[22] Y. Abid, Abdessamad Imine, and Michaël Rusinowitch. 2018. Sensitive Attribute Prediction for Social Networks Users. In EDBT/ICDT Workshop.

[23] Serguei A. Mokhov, Joey Paquet, and Mourad Debbabi. 2014. The Use of NLP Techniques in Static Code Analysis to Detect Weaknesses and Vulnerabilities. In Advances in Artificial Intelligence, Marina Sokolova and Peter van Beek (Eds.). Springer International Publishing, Cham, 326–332.

[24] Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. 2019. CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning. In 28th USENIX Security Symposium (USENIX Security 19). USENIX Association, Santa Clara, CA, 461–47.

[25] Vernit Garg and Laxmi Ahuja. 2019. Password Guessing Using Deep Learning. In 2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC). IEEE, 38–40.

[26] Dongqi Han, Zhiliang Wang, Ying Zhong, Wenqi Chen, Jiahai Yang, Shuqiang Lu, Xingang Shi, and Xia Yin. 2020. Practical traffic-space adversarial attacks on learning-based nidss. arXiv preprint arXiv:2005.07519 (2020).

[27] Yisroel Mirsky and Wenke Lee. 2021. The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR) 54, 1 (2021), 1–41.

[28] Rahman, Tanzila, Mrigank Rochan, and Yang Wang. "Video-based person re-identification using refined attention networks." 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2019.

[29] X. Zhu, X. Jing, X. You, X. Zhang, and T. Zhang. 2018. Video-Based Person Re-Identification by Simultaneously Learning Intra-Video and InterVideo

tarmoqlardagi postlarni semantik tahlil qilish orqali ehtimoliy insayderlarni qidirish[30], ochiq manbalardan ma'lumotlarni yig'ishda hujjatlarni izohlash (referat, umumlashtirish) (OSINT – ochiq razvedka)[31] kabilarni misol sifatida keltirish mumkin.

5. Qaror qabul qilish – bu strategik rejani ishlab chiqish yoki operatsiyani (hujumni) muvofiqlashtirish vazifalari. Sun'y intellekt bo'yicha misollar – avtonom bot tarmog'ini boshqarish uchun to'dadan foydalanish[32] va optimal tarmoq hujumlarini rejalashtirish[33].

Tadqiqotchilarning bir guruhi hujumlarni avtomatlashtirishni mashinaviy o'rganish metodidan foydalanmay turib ham amalga oshirish mumkinligini ta'kidlaydi[34], ammo kuchaytirishni o'rganish ("reinforcement learning") hujumlarni amalga oshirishda asosiy vosita bo'lish uchun barcha imkoniyatlarga ega. Microsoft o'z hisobotida kiberhujumlarda sun'iy intellektdan foydalanish tajribali ishtirokchilar bilan boshlanishini kutmoqda[35], ammo hamkorlik darajasini oshirish va ishlatilgan vositalarni tijoratlashtirish orqali jarayon tezda kengroq ekotizimga tarqaladi. Xususan, hujumchilarning vositalari MITRE atlasida tasvirlanganidek, mudofaani chetlab o'tishning umumiy asosiy taktikalarini o'z ichiga oladi[36]. Hujumkor sun'iy intellektning eng muvaffaqiyatli ishlatiladigan avtomatlashtirish tizimlaridan biri bu ijtimoiy tarmoqlardagi botlar hisoblanadi[37]. Hujum harakatlarini avtomatlashtirishning yana bir misoli kuchaytirishni o'rganishdan foydalangan holda avtomatlashtirilgan penetratsion testlardir (penetration test)[38].

Mashinaviy o'rganish biometrik autentifikatsiya tizimlariga hujum qilish uchun ishlatiladi: soxta ovoz va shunga o'xshash narsalar[39].

Yuqorida biz mashinaviy o'rganish orqali fishing hujumlarini aniqlash haqida gaplashdik. Ammo mashinaviy o'rganish fishing hujumlarini yaratishda ham qo'llaniladi[40]. Bundan maqsad himoya

Distance Metrics. IEEE Transactions on Image Processing 27, 11 (2018), 5683–5695.

[30] Gavai, Gaurang, et al. "Detecting insider threat from enterprise social and online activity data." Proceedings of the 7th ACM CCS international workshop on managing insider security threats. 2015.

[31] Zhou, Qingyu, et al. "Neural document summarization by jointly learning to score and select sentences." arXiv preprint arXiv:1807.02305 (2018).

[32] Aniello Castiglione, Roberto De Prisco, Alfredo De Santis, Ugo Fiore, and Francesco Palmieri. 2014. A botnet-based command and control approach relying on swarm intelligence. Journal of Network and Computer Applications 38 (2014), 22–33.

[33] John A. Bland, Mikel D. Petty, Tymaine S. Whitaker, Katia P. Maxwell, and Walter Alan Cantrell. 2020. Machine Learning Cyberattack and Defense Strategies. Computers & Security 92 (2020), 101738.

[34] B. Buchanan, J. Bansemer, D. Cary, et al., Automating Cyber Attacks: Hype and Reality, Center for Security and Emerging Technology, November 2020. https://cset.georgetown.edu/wp-content/uploads/CSET-AutomatingCyber-Attacks.pdf

[35] How cyberattacks are changing according to new Microsoft Digital Defense Report https://www.microsoft.com/security/blog/2021/10/11/howcyberattacks-are-changing-according-to-new-microsoft-digital-defense-report/

[36] Virtualization/Sandbox Evasion, Technique T1497 – Enterprise | MITRE ATT&CK https://attack.mitre.org/techniques/T1497/

[37] Himelein-Wachowiak, McKenzie, et al. "Bots and misinformation spread on social media: Implications for COVID-19." Journal of Medical Internet Research 23.5 (2021): e26933.

[38] Deep Exploit https://github.com/13o-bbrbbq/machine_learning_security/tree/master/DeepExploit

[39] Biggio, Battista, et al. "Adversarial biometric recognition: A review on biometric system security from the adversarial machine-learning perspective." IEEE Signal Processing Magazine 32.5 (2015): 31-41.

[40] AlEroud, Ahmed, and George Karabatis. "Bypassing detection of URLbased phishing attacks using generative adversarial deep neural networks."

tizimlarini chetlab o‘tish, yanada jozibali tarkib yaratish va foydalanuvchilarni zararli havolani bosishga, tizimga dasturiy ta’minotni o‘rnatishga va boshqalarga undashdir.

Hujumkor harakatlarga misollar qatoriga parollarni moslashtirish[41], dasturlarning manba kodini chalkashtirib yuborish[42], trafikni maskalash[43], bot tarmog‘ini boshqarish[44] kiradi.

Microsoft tomonidan tashkil etilgan alohida hisobot (“workshop”) hujumkor sun’iy intellektga bag‘ishlangan[45]. Sun’iy intellektdan foydalangan holda hujumlar, shuningdek, sun’iy intellekt bo‘yicha Milliy xavfsizlik komissiyasining (NSCAI) juda batafsil hisobotida ham ko‘rib chiqiladi[46].

Sun’iy intellekt tizimlariga hujumlar

Mazkur tushuncha kompyuter xavfsizligi uchun yetarlicha yangi maydon hisoblanadi. Hujumlar sun’iy intellekt tizimlarining o‘ziga qaratilishi mumkin (aslida mashinaviy o‘rganish tizimlari). Amalga oshirilgan har qanday mashinaviy o‘rganish tizimi avvalo dasturga ega. Ammo muammo shundaki, bunday ilovalar uchun xavfsizlikni tahlil qilishning an’anaviy usullari qo‘llanilmaydi, chunki ushbu dasturlarning xavfsizligi bilan bog‘liq muammolarni an’anaviy usullar bilan hal qilib bo‘lmaydi. Albatta, dasturning buzilgan ishlash muhiti muammolarga olib keladi.

Ammo bu asosiy muammo emas. Mashinani o‘rganish tizimlari ma’lumotlarga bog‘liq. Taqdim etilgan o‘quv ma’lumotlari asosida tizim ba’zi umumlashmalarni ishlab chiqadi, keyinchalik ular haqiqiy (sinov) ma’lumotlarni qayta ishlashda qo‘llaniladi. Shunday qilib, mashinani o‘rganish konveyerining turli bosqichlarida ma’lumotlarning modifikatsiyasi va bunday tizimlar umuman ishlamasligi yoki aksincha, hujumchiga kerakli natijalarni berishi mumkinligiga olib keladi. Bunday holda, maxsus o‘zgartirilgan ma’lumotlar, umuman aytganda, “toza” ma’lumotlar bilan bir xil bo‘ladi, ularni ajratib bo‘lmaydi.

Bundan tashqari, mashg‘ulotlar har doim ma’lum bir o‘quv-ma’lumotlar to‘plamida amalga oshirilganligi sababli, umumiy yig‘indi butunlay noma’lum bo‘lib qoladi.

Shuningdek, ish bosqichidagi ma’lumotlarning “o‘zgarishi” hech qanday zararli harakatlarsiz sodir bo‘lishi mumkin (va ko‘pincha sodir bo‘ladi ham). Faqat ma’lumotlarning o‘zi shunday tartibga solinganligi sababli, bunday holda, hujumlar aniq ma’lumotlarning maxsus o‘zgarishi yoki tizim noto‘g‘ri ishlaydigan (umuman ishlamaydigan) ma’lumotlarning maxsus almashinuvi deb ataladi. Umuman

---

Proceedings of the Sixth International Workshop on Security and Privacy Analytics. 2020.

[41] B. Hitaj, P. Gasti, G. Ateniese, F. Perez-Cruz, PassGAN: A Deep Learning Approach for Password Guessing, NeurIPS 2018 Workshop on Security in Machine Learning (SecML’18), December 2018.

[42] S. Datta, DeepObfusCode: Source Code Obfuscation through Sequence-to-Sequence Networks In: Arai, K. (eds) Intelligent Computing. Lecture Notes in Networks and Systems, vol 284. Springer, Cham.

[43] J. Li, L. Zhou, H. Li, L. Yan and H. Zhu, “Dynamic Traffic Feature Camouflaging via Generative Adversarial Networks,” 2019 IEEE Conference on Communications and Network Security (CNS), 2019, pp. 268-276

[44] Castiglione, Aniello, et al. "A botnet-based command and control approach relying on swarm intelligence." Journal of Network and Computer Applications 38 (2014): 22-33.

[45] Implications of Artificial Intelligence for Cybersecurity: A Workshop, National Academy of Sciences, 2019. https://www.nationalacademies.org/our-work/implications-of-artificialintelligence-for-cybersecurity-a-workshop.

[46] National Security Commission on Artificial Intelligence report https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

olganda, bu – mashinaviy o‘rganish tizimlarining barqarorligi muammosi. Hozirda bu masalaga katta e’tibor qaratilmoqda, chunki bu muhim dasturlarda (avionika, yadro xavfsizligi va boshqalar) sun’iy intellekt tizimlaridan foydalanishga to‘sqinlik qiladigan asosiy faktor hisoblanadi[47].

Mashinaviy o‘rganish tizimlariga qilingan hujumlarning yana bir nomi – qarama-qarshi misollar[48]. Shunday qilib, tizimlarga dushmanlik ta’siri an’anaviy zaifliklar shaklida, shuningdek, yangi kategoriya orqali amalga oshirilishi mumkin: qarama-qarshi misollar.

An’anaviy zaifliklarga misol sifatida, masalan, Tensorflow dasturiy ta’minot paketidagi zaifliklar to‘g‘risidagi hisobotni ko‘rsatish mumkin[49], bu tabiiy ravishda uni ishlatadigan sun’iy intellekt tizimlarida zaifliklar mavjudligini anglatadi.

Sun’iy intellekt dasturiy infratuzilmasiga qilingan hujumlar bi qator ilmiy izlanishlar doirasida o‘rganilgan. Nyu-York universiteti tadqiqotchilari shuni aniqladilarki, aksariyat sun’iy intellekt dasturlari yuklab olingan sun’iy intellekt modellarining yaxlitligini tekshirmaydi, an’anaviy dasturiy ta’minot bilan qabul qilingan amaliyotdan farqli o‘laroq, bu yerda to‘liq fayllar/kutubxonalarni kriptografik tekshirish o‘n yildan ortiq vaqtdan beri standart amaliyot bo‘lib kelgan[50]. Ommaviy ma’lumotlar to‘plamlari belgilashda xatolarga yo‘l qo‘yishi mumkin, bu tabiiy ravishda ular yordamida o‘qitilgan tizimlarning ishlashiga ta’sir qiladi[51].

Qarama-qarshi misollar odatda hujum harakatlarini qo‘llash nuqtasi (mashinani o‘rganish konveyerining bosqichi) va hujumchining tizim haqidagi bilimlari (oq quti, qora quti) bo‘yicha tasniflanadi.

Zararli axborot operatsiyalarida sun’iy intellektdan foydalanish

Mashinaviy o‘rganish va kompyuter grafikasidagi yutuqlar davlat va nodavlat subyektlarning sintetik media va chuqur feyklar deb nomlangan yuqori sifatli audiovizual kontentni ishlab chiqarish va tarqatish imkoniyatlarini kengaytirdi. “Deepfake” yaratish uchun sun’iy intellekt texnologiyalari endi haqiqiy odamlar, sahnalar va voqealardan farq qilmaydigan kontentni yaratishi mumkin. Bunday kontent haqiqatan ham milliy xavfsizlikka tahdid solishi mumkin.

Turli xil signallarni, shu jumladan yuqori sifatli audiovizual tasvirlarni sintez qilish uchun generativ sun’iy intellekt usullarini kuchaytirish kiberxavfsizlik uchun muhimdir. Personalizatsiyada sun’iy intellektdan chuqur feyklar yaratish uchun foydalanish ijtimoiy muhandislik operatsiyalari samaradorligini oshirishi mumkin (dastur o‘zini haqiqiy shaxs sifatida ko‘rsatadi) va masalan, oxirgi foydalanuvchilarni tajovuzkorlarga tizimlar

[47] Namiot, Dmitry, Eugene Ilyushin, and Ivan Chizhov. "Ongoing academic and industrial projects dedicated to robust machine learning." International Journal of Open Information Technologies 9.10 (2021): 35-46. (in Russian)

[48] Ilyushin, Eugene, Dmitry Namiot, and Ivan Chizhov. "Attacks on machine learning systems-common problems and methods." International Journal of Open Information Technologies 10.3 (2022): 17-22.

[49] Tensorflow : Vulnerability Statistics https://www.cvedetails.com/product/53738/Google-Tensorflow.html

[50] Gu, Tianyu, Brendan Dolan-Gavitt, and Siddharth Garg. "Badnets: Identifying vulnerabilities in the machine learning model supply chain." arXiv preprint arXiv:1708.06733 (2017).

[51] Northcutt, Curtis G., Anish Athalye, and Jonas Mueller. "Pervasive label errors in test sets destabilize machine learning benchmarks." arXiv preprint arXiv:2103.14749 (2021).

va ma’lumotlarga kirish huquqini berishga ishontirishi mumkin[52].

Kengroq miqyosda sun’iy intellekt texnikasi va sintetik muhitlarning ishlab chiqaruvchi kuchi mudofaa va milliy xavfsizlikka muhim ta’sir ko‘rsatadi. Ushbu usullar raqiblar tomonidan dunyo rahbarlari va qo‘mondonlarining ishonchli bayonotlarini yaratish, ishonchli soxta bayroq operatsiyalarini soxtalashtirish va soxta yangiliklar yaratish uchun ishlatilishi mumkin[53].

Jorjiya Tech universiteti tomonidan olib borilgan tadqiqotlar shuni ko‘rsatadiki, sintetik ommaviy axborot vositalarining tarqalishi yana bir bezovta qiluvchi ta’sirga ega edi: zararli subyektlar chuqur feyklar davrida ishonchni yo‘qotish bilan birga keladigan rad etishning yangi shakllaridan foydalanib, haqiqiy voqealarni “soxta” deb atashdi. Video va fotosurat dalillari, masalan, vahshiyliklar tasvirlari “soxta” deb nomlanadi. “Yolg‘onchi dividend” deb nomlanuvchi sintetik ommaviy axborot vositalarining tarqalishi odamlarni haqiqiy ommaviy axborot vositalarini “soxta” deb atashga undaydi va ularning xatti-harakatlari uchun ishonchli rad etishni keltirib chiqaradi[54].

Microsoft yuqorida nomi keltirib o‘tilgan taqdimotiga ko‘ra, sintetik ommaviy axborot vositalari va ularning qo‘llanilish sohalari vaqt o‘tishi bilan tobora takomillashib borishini kutish mumkin, shu jumladan dunyodagi haqiqiy voqealar bilan chuqur feyklarning majburiy almashinuvi va real vaqtda chuqur feyklarning sintezi kuzatiladi. Haqiqiy vaqtda ishlab chiqarishlar yordamida tabiiy bosh holati, yuz ifodalari va bayonotlarga ega bo‘lgan ishonchli, interaktiv yolg‘onchilarni (masalan, telekonferensiyalarda paydo bo‘ladigan va inson tomonidan boshqariladigan) yaratish mumkin. Qayd etib o‘tish kerakki, kelajakda biz audio va vizual kanallar orqali real vaqtda ishonchli suhbatlarda avtonom tarzda ishtirok eta oladigan sun’iy ravishda yaratilgan odamlar muammosiga duch kelishimiz mumkin. Tabiiyki, bunday sharoitda chuqur feyklarni aniqlash juda dolzarb vazifaga aylanadi.

DARPA Semantic Forensics (SemaFor) dasturi fikrimizga yaqqol misol bo‘la oladi. SemaFor dasturi ommaviy axborot vositalarini tahlil qilish uchun innovatsion semantik texnologiyalarni ishlab chiqishga qaratilgan. Ushbu texnologiyalar multimodal media aktivlari yaratilganligini yoki manipulyatsiya qilinganligini aniqlaydigan semantik aniqlash algoritmlarini o‘z ichiga oladi. Atribut algoritmlari multimodal ommaviy axborot vositalari ma’lum bir tashkilot yoki shaxsdan kelib chiqishi haqida xulosa beradi.

Xarakterlash algoritmlari multimodal ommaviy axborot vositalari zararli maqsadlarda yaratilganmi yoki boshqarilganmi, degan savolga javob beradi. Ushbu SemaFor texnologiyalari dushmanning dezinformatsiya kampaniyalarini aniqlash, cheklash va tushunishga yordam beradi.

Sun’iy intellekt va kiberxavfsizlikning kelajakdagi istiqboli

World Economic Forum (Jahon iqtisodiy forumi) tomonidan e’lon qilingan The 2019

[52] Fedushko, Solomia. "Artificial Intelligence Technologies Using in Social Engineering Attacks." (2020).

[53] Smith, Hannah, and Katherine Mansted. "Weaponised deep fakes." (2020).

[54] The Liar’s Dividend: The Impact of Deepfakes and Fake News on Politician Support and Trust in Media https://gvu.gatech.edu/research/projects/liars-dividend-impact-deepfakes-andfake-news-politician-support-and-trust-media

Global Risks Report hisobotida kiberhujumlarni eng global va xavfli beshta xavfdan biri sifatida ta'riflagan. Masalan, 2018-yilning 1-yarmida umumiy hisobda 3.3 milliard kiberhujumlar amalga oshirilgan bo'lib, bu 2017-yildagi barcha kiberhujumlar sonidan 70%ga ortiqroqdir. Mazkur turdagi hujumlar, shuningdek, juda tezlik bilan sodir etilmoqda. Microsoft o'rganishlarining ko'rsatishicha, 2018-yildagi hujumlarning 60 foizi bir soatdan kamroq vaqt davomida amalga oshirilgan bo'lib, "malware"ning yangi shakllariga tayangan.

Sun'iy intellekt bu holatlarni 3 xil yo'l bilan kamaytirishi hamda kiberxavfsizlik sohasida sarflanayotgan inson kapitali va xarajatlarni tejashga yordam berishi mumkin. *Birinchidan,* sun'iy intellekt tizimning chidamliligini oshiradi, ya'ni turli tahdidlar va hujumlar sharoitida ham tizim o'zini o'zi tekshirish va nazorat qilish orqali kutilgan darajada ishlashda davom etadi. *Ikkinchidan,* sun'iy intellekt tizimning javob berish qobiliyatini yaxshilaydi, ya'ni tizim kiberhujumlarni avtomatik tarzda bartaraf etishni boshlaydi, erishilgan natijalar asosida kelajakdagi qarshilik ko'rsatish strategiyalarini belgilaydi va har bir holatdan so'ng yanada samarali tarzda kontroperatsiyalarni amalga oshirishni boshlaydi. *Uchinchidan,* sun'iy intellekt tizimni yanada kuchli qiladi, ya'ni tizimning hujumlarni erta aniqlashiga va mazkur holatlar bilan bog'liq ma'lumotlarni o'zida saqlab qolishiga yordam beradi.

Yuqoridagi uchta faktorning natijasi o'laroq, kiberxavfsizlik sohasida sun'iy intellektdan foydalanish taktik va strategik foydalarini namoyon etadi. Taktik jihatdan, sun'iy intellekt tizimning xavfsizligini oshiradi va uning tahdidlarga nisbatan ojizligini kamaytiradi. Strategik jihatdan, sun'iy intellekt kibermaydonda sodir etiladigan jinoyatlar dinamikasini kamaytirishga yordam beradi. Kiberxavfsizlik sohasidagi eng katta muammolardan biri ancha paytgacha kiberhujumlarni sodir etgan shaxslarni aniqlash bilan bog'liq bo'lib kelayotgan edi. Sun'iy intellekt tuzilmalari kiberhujumlarga uni sodir etgan hujumchilarning identifikatsiyasidan qat'iy nazar, hatto ular anonim bo'lgan taqdirda ham, qarshi javob qaytarish imkoniyatiga ega.

References:

1. Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. 2019. CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning. In 28th USENIX Security Symposium (USENIX Security 19). USENIX Association, Santa Clara, CA, 461–47.

2. Vernit Garg and Laxmi Ahuja. 2019. Password Guessing Using Deep Learning. In 2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC). IEEE, 38–40.

3. Д.Е. Намиот, Е.А. Ильюшин, И.В. Чижов. Искусственный интеллект и кибербезопасность. International Journal of Open Information Technologies ISSN: 2307-8162 vol. 10, no. 9, 2022.

4. Applications for artificial intelligence in Department of Defense cyber missions https://blogs.microsoft.com/on-the-issues/2022/05/03/artificialintelligence-department-of-defense-cyber-missions/

5. Fedushko, Solomia. “Artificial Intelligence Technologies Using in Social Engineering Attacks.” (2020).
6. Smith, Hannah, and Katherine Mansted. “Weaponised deep fakes.” (2020).
7. The Liar’s Dividend: The Impact of Deepfakes and Fake News on Politician Support and Trust in Media https://gvu.gatech.edu/research/projects/liars-dividend-impact-deepfakes-andfake-news-politician-support-and-trust-media