

SUN'IY INTELLEKT ASOSIDAGI TIZIMLARDA AXBOROT
XAVFSIZLIGI MUAMMOLARI

Arziqulov Husan Normurod o'g'li

Navoiy davlat universiteti

Raqamli texnologiyalar kafedrası dotsenti

Annotatsiya: Ushbu ilmiy tezisdagi sun'iy intellekt asosidagi tizimlarda axborot xavfsizligini ta'minlash bilan bog'liq dolzarb muammolar tizimli ravishda tahlil qilingan. Tadqiqot davomida sun'iy intellekt modellarida uchraydigan ma'lumotlar sizib chiqishi, adversarial hujumlar, data poisoning, model inversion hamda modelga qarshi boshqa hujum turlari ilmiy nuqtai nazardan ko'rib chiqilgan.

Kalit so'zlar: Sun'iy intellekt, axborot xavfsizligi, kiberxavfsizlik, adversarial hujumlar, data poisoning, model inversion, ma'lumotlar maxfiyligi, federativ o'rganish, differential privacy, explainable AI.

Kirish: So'nggi yillarda sun'iy intellekt (SI) texnologiyalarining jadal rivojlanishi axborot tizimlarining samaradorligini sezilarli darajada oshirdi. Tibbiyot, moliya, ta'lim, sanoat va davlat boshqaruvi kabi muhim sohalarda SI asosidagi tizimlar qaror qabul qilish jarayonlarini avtomatlashtirish va optimallashtirishda keng qo'llanilmoqda. Biroq, mazkur texnologiyalarning murakkab arxitekturasini va katta hajmdagi ma'lumotlarga tayanishi axborot xavfsizligi bilan bog'liq yangi tahdidlar va muammolarni yuzaga keltirmoqda. Shu sababli SI tizimlarida axborot xavfsizligini ta'minlash masalasi zamonaviy ilmiy tadqiqotlarning ustuvor yo'nalishlaridan biri hisoblanadi.

Tadqiqotning dolzarbligi: So'nggi yillarda SI va axborot havfsizlik bo'yicha ko'plab ilmiy tadqiqotlar olib borilmoqda. Tadqiqotlar shuni

ko'rsatadiki, mashinaviy o'rganish (Machine YOSH OLIMLAR ILMIY-AMALIY KONFERENSIYASI in-academy.uz/index.php/yo 116 Learning – ML) va chuqur o'rganish (Deep Learning – DL) axborot havfsizlik tizimlarining samaradorligini oshirishda muhim omillar hisoblanadi. SI asosida ishlaydigan tahdidlarni aniqlash tizimlari an'anaviy signatura (imzo) asosida ishlaydigan antivirus dasturlariga qaraganda samaraliroq bo'lib, noma'lum tahdidlarni ham aniqlay oladi (Smith et al., 2023). Kiberhujumlarning murakkablashib borishi sababli, klassik xavfsizlik devorlarini (firewall) mustahkamlash uchun avtomatlashtirilgan SI algoritmlaridan foydalanish tavsiya etiladi (Brown & Miller, 2022). Anomaliyalarni aniqlashda generativ adversarial tarmoqlar (GANs) va kuchaytiruvchi o'rganish (Reinforcement Learning) usullari yaxshi natija bermoqda (Garcia et al., 2024). Shuningdek, SI texnologiyalari asosida ishlaydigan axborot havfsizlik tizimlarining kamchiliklari ham mavjud. Masalan, ba'zi tadqiqotlar sun'iy intellekt asosida ishlovchi tizimlarning noto'g'ri ijobiy natijalar (False Positives) ko'rsatish ehtimoli yuqoriligini ta'kidlamogda. Bu esa tizimlarning takomillashtirilishini talab etadi[1]. An'anaviy axborot tizimlaridan farqli ravishda, sun'iy intellekt modellari nafaqat dasturiy kodga, balki o'qitish ma'lumotlariga, model parametrlariga va qaror chiqarish mexanizmlariga ham tayanadi. Ushbu komponentlarning har biri hujumlar uchun potensial nishon bo'lishi mumkin. Masalan, noto'g'ri yoki ataylab buzilgan ma'lumotlar asosida o'qitilgan model noto'g'ri qarorlar qabul qilishi ehtimoli yuqori. Bu esa ayniqsa kritik infratuzilmalar va xavfsizlikka bevosita ta'sir qiluvchi sohalarda jiddiy xavf tug'diradi. Shuning uchun SI asosidagi tizimlarda axborot xavfsizligi muammolarini chuqur tahlil qilish va ilmiy asoslangan himoya mexanizmlarini ishlab chiqish dolzarb ahamiyat kasb etadi.

Asosiy qism: Sun'iy intellekt tizimlarida uchraydigan axborot xavfsizligi muammolarini bir necha asosiy toifalarga ajratish mumkin.

Birinchi muammo - ma'lumotlar sizib chiqishi hisoblanadi. SI modellarini o'qitishda ko'pincha shaxsiy yoki maxfiy ma'lumotlar ishlatiladi. Agar ushbu ma'lumotlar yetarli darajada himoyalansa, tashqi hujumchilar tomonidan noqonuniy ravishda qo'lga kiritilishi mumkin. Masalan, tibbiy diagnostika tizimlarida bemorlarning shaxsiy ma'lumotlari saqlanadigan ma'lumotlar bazasiga ruxsatsiz kirish holatlari kuzatilgan. SI tizimlarida ma'lumotlar sizib chiqishining asosiy manbalari va ularning ulushi tasvirlangan.

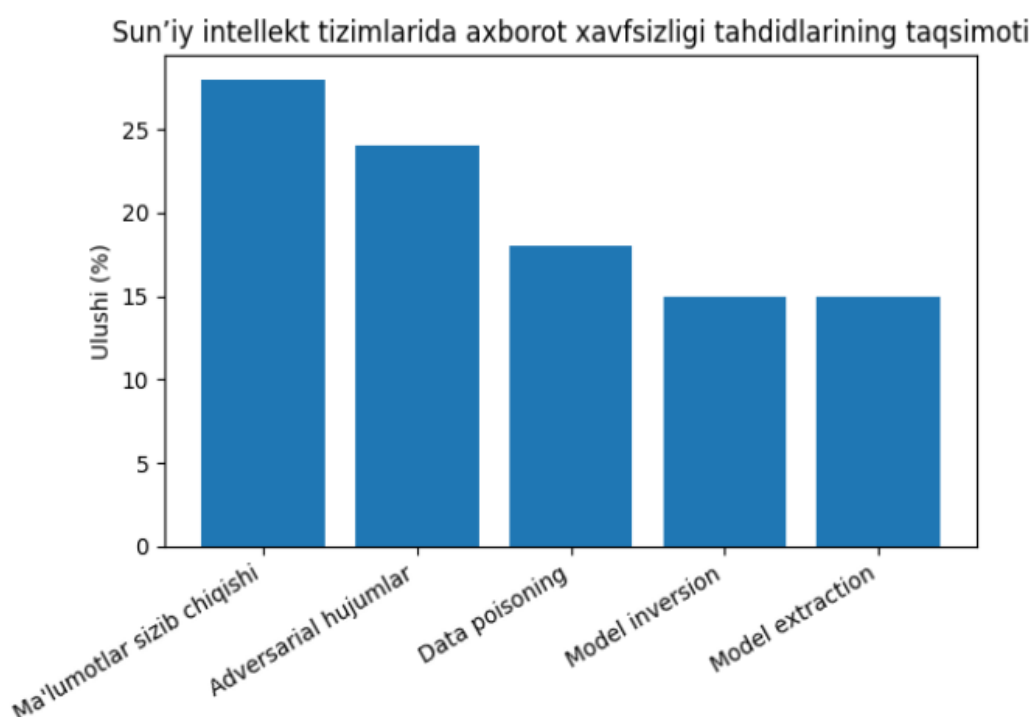
Ikkinchi muammo - modelga qarshi hujumlar (model attacks) bo'lib, ular modelning ichki tuzilmasini yoki xatti-harakatini o'rganishga qaratilgan. Jumladan, model inversion hujumlari orqali hujumchi model chiqishlaridan foydalanib, o'qitish jarayonida ishlatilgan asl ma'lumotlarni qisman tiklashi mumkin. Nazariy misol sifatida, yuzni aniqlash modeli orqali foydalanuvchilarning yuz tasvirlarini qayta tiklashga urinish holatlarini keltirish mumkin.

Uchinchi muammo - adversarial hujumlar bo'lib, bunda kirish ma'lumotlariga juda kichik, ammo maqsadli o'zgartirishlar kiritish orqali modelni noto'g'ri qaror qabul qilishga majbur etiladi. Masalan, avtonom transport vositalarida yo'l belgilariga deyarli sezilmas belgilar qo'shish orqali tizimni noto'g'ri harakatga undash mumkin. adversarial hujumlarning klassifikatsiyasi va ularning SI modellariga ta'siri ko'rsatilgan.

To'rtinchi muammo - data poisoning, ya'ni o'qitish ma'lumotlarini ataylab buzish orqali model sifatini pasaytirishdir. Bu holatda hujumchi o'qitish to'plamiga noto'g'ri yoki zararli ma'lumotlarni kiritadi. Natijada model noto'g'ri xulosalar chiqaradi. Masalan, spam filtrlash tizimlariga

noto'g'ri belgilangan elektron xabarlar qo'shilishi filtrning ishonchliligini kamaytiradi.

Sun'iy intellekt asosidagi tizimlarda axborot xavfsizligiga tahdid soluvchi muammolar turli ko'rinishlarda namoyon bo'lib, ularning uchrash chastotasi va ta'sir darajasi bir xil emas.



Rasmda keltirilgan ma'lumotlardan ko'rinadiki, ma'lumotlar sizib chiqishi va adversarial hujumlar eng keng tarqalgan tahdidlar hisoblanadi. Bu holat SI tizimlarida ma'lumotlar maxfiyligini himoyalash hamda modelning tashqi ta'sirlarga bardoshliligini oshirish zarurligini ko'rsatadi.

Taklif etilayotgan yechimlar: Yuqoridagi muammolarni bartaraf etish uchun zamonaviy ilmiy yondashuvlar taklif etilmoqda. Ulardan biri — kriptografik himoya usullari bo'lib, ma'lumotlarni shifrlash va xavfsiz hisoblash mexanizmlarini qo'llash orqali ma'lumotlar maxfiyligini ta'minlaydi. Xususan, homomorf shifrlash texnologiyalari shifrlangan ma'lumotlar ustida to'g'ridan-to'g'ri hisoblash imkonini beradi.



Federativ o'rganish yondashuvi esa ma'lumotlarni markazlashtirmasdan, ularni foydalanuvchi qurilmalarida saqlagan holda modelni o'qitish imkonini yaratadi. Bu usul ma'lumotlar sizib chiqish xavfini sezilarli darajada kamaytiradi. 3-rasmda federativ o'rganish arxitekturasini va an'anaviy markazlashtirilgan o'rganish modeli taqqoslangan.

Shuningdek, differential privacy usuli model chiqishlariga statistik shovqin qo'shish orqali individual ma'lumotlarning aniqlanish ehtimolini pasaytiradi. Explainable AI (XAI) yondashuvlari esa model qarorlarining izohlanishini ta'minlab, yashirin va zararli xatti-harakatlarni aniqlashga yordam beradi.

Xulosa qilib aytganda, sun'iy intellekt asosidagi tizimlarda axborot xavfsizligi muammolari ko'p qatlamli va murakkab xususiyatga ega. Ma'lumotlar sizib chiqishi, adversarial hujumlar, data poisoning va modelga qarshi hujumlar SI tizimlarining ishonchliligi va barqarorligiga jiddiy tahdid soladi. Tadqiqot natijalari shuni ko'rsatadiki, ushbu muammolarni hal etish faqatgina an'anaviy xavfsizlik choralarini emas, balki kriptografiya, federativ o'rganish, differential privacy va explainable AI kabi zamonaviy ilmiy yondashuvlarni kompleks qo'llashni talab etadi. Kelgusida SI tizimlarining xavfsizligini ta'minlash bo'yicha tadqiqotlarni chuqurlashtirish ularning jamiyatda keng va ishonchli joriy etilishiga xizmat qiladi.

Adabiyotlar ro'yxati.

1. Toshniyozov Farrux O'ktamovich TATU Samarqand fililali assistenti Sun'iy intellekt asosida axborot xavfsizlikni ta'minlash
2. Kurakin, A., Goodfellow, I., & Bengio, S. (2017). Adversarial Examples in the Physical World. International Conference on Learning Representations (ICLR).

3. Xu, W., Evans, D., & Qi, Y. (2018). Feature Squeezing: Detecting Adversarial Examples in Deep Neural Networks. Proceedings of the 2018 Network and Distributed System Security Symposium (NDSS).
— Adversarial hujumlarni aniqlash bo'yicha yondashuvlar.