

XAVFSIZ SUN'YI INTELLEKT ARXITEKTURALARI: GENERATIV AI TIZIMLARIDA ISHONCHLILIK VA KIBERXAVFSIZLIKNI TA'MINLASH MODELLARI

Djumabayeva Nodira Yuldashevna

Toshkent amaliy fanlar universiteti, Tarmoqlar iqtisodiyoti va buxgalteriyasi kafedrası katta o'qituvchisi

E-mail: yuldashevna1310@gmail.com

ANNOTATSIYA:

Ushbu maqolada generativ sun'iy intellekt tizimlarida xavfsizlik va ishonchlilikni ta'minlash masalalari tahlil qilinadi. Zamonaviy AI arxitekturalarining rivojlanishi bilan birga, ularning kiberxavfsizlikka ta'siri, ma'lumotlar maxfiyligi va noto'g'ri natijalar berish xavfi dolzarb muammoga aylanmoqda. Maqolada xavfsiz sun'iy intellekt arxitekturalarini yaratishning asosiy yondashuvlari, jumladan, modelni himoyalash, adversarial hujumlarga qarshi choralar va tizim ishonchliligini oshirish usullari ko'rib chiqiladi. Shuningdek, generativ AI tizimlarida etik me'yorlar va nazorat mexanizmlarining ahamiyati ham yoritiladi. Tadqiqot natijalari AI tizimlarini xavfsiz va barqaror ishlashini ta'minlashda muhim ilmiy-amaliy ahamiyatga ega.

Kalit so'zlar: sun'iy intellekt, generativ AI, kiberxavfsizlik, ishonchlilik, xavfsiz arxitektura, ma'lumotlar maxfiyligi, adversarial hujumlar, model himoyasi, etik AI, xavfsizlik modellari

ABSTRACT:

This article analyzes the issues of ensuring security and reliability in generative artificial intelligence systems. With the development of modern AI architectures, their impact on cybersecurity, data confidentiality, and the risk of erroneous results are becoming an urgent problem. The article considers the main approaches to creating secure artificial intelligence architectures, including model protection, measures against adversarial attacks, and methods for increasing system reliability. It also highlights the importance of ethical standards and control mechanisms in generative AI systems. The research results have important scientific and practical significance in ensuring the safe and stable operation of AI systems.

Keywords: artificial intelligence, generative AI, cybersecurity, reliability, secure architecture, data confidentiality, adversarial attacks, model protection, ethical AI, security models

АННОТАЦИЯ:

В данной статье анализируются вопросы обеспечения безопасности и надежности в системах генеративного искусственного интеллекта. С развитием современных архитектур ИИ их влияние на кибербезопасность, конфиденциальность данных и риск ошибочных результатов становятся актуальной проблемой. В статье рассматриваются основные подходы к созданию безопасных архитектур искусственного интеллекта, включая защиту моделей, меры против атак со стороны злоумышленников и методы повышения надежности системы. Также подчеркивается важность этических стандартов и механизмов контроля в системах генеративного ИИ. Результаты исследования имеют важное научное и практическое значение для обеспечения безопасной и стабильной работы систем ИИ.

Ключевые слова: искусственный интеллект, генеративный ИИ, кибербезопасность, надежность, безопасная архитектура, конфиденциальность данных, атаки со стороны злоумышленников, защита моделей, этический ИИ, модели безопасности

So'nggi yillarda sun'iy intellekt texnologiyalari, ayniqsa generativ AI tizimlari juda tez rivojlanib bormoqda. Bunday tizimlar matn, rasm, audio va boshqa turdagi kontentlarni yaratishda keng qo'llanilmoqda. Natijada ular ta'lim, tibbiyot, biznes va boshqa ko'plab sohalarda muhim ahamiyat kasb etmoqda. Shu bilan birga, generativ sun'iy intellektidan foydalanish jarayonida xavfsizlik va ishonchlilik bilan bog'liq muammolar ham yuzaga kelmoqda. Ayniqsa, ma'lumotlar maxfiyligi, noto'g'ri yoki zararli kontent yaratish ehtimoli, hamda turli kiberhujumlarga nisbatan zaifliklar ushbu tizimlarning eng dolzarb muammolaridan biri hisoblanadi. Shu sababli, xavfsiz sun'iy intellekt arxitekturalarini ishlab chiqish va ularning barqaror ishlashini ta'minlash bugungi kunda muhim vazifalardan biri bo'lib qolmoqda. Mazkur maqolada generativ AI tizimlarida ishonchlilik va kiberxavfsizlikni ta'minlashga qaratilgan yondashuvlar tahlil qilinadi. Shuningdek, zamonaviy xavfsizlik modellari va ularni amaliyotda qo'llash imkoniyatlari ko'rib chiqiladi.

Generativ sun'iy intellekt tizimlari zamonaviy axborot texnologiyalarining eng muhim yo'nalishlaridan biri sifatida shakllanib bormoqda. Ular katta hajmdagi ma'lumotlar asosida yangi kontent yaratish qobiliyatiga ega bo'lib, turli sohalarda samaradorlikni oshirishga xizmat qiladi. Biroq bunday tizimlarning keng qo'llanilishi bilan birga, ularning xavfsizligi va ishonchliligini ta'minlash masalasi ham alohida e'tibor talab etadi. Avvalo, generativ AI tizimlarining ishlash prinsipi murakkab neyron tarmoqlarga asoslangan bo'lib, ular o'qitish jarayonida katta hajmdagi ma'lumotlarni qayta ishlaydi. Shu jarayonda ma'lumotlar sifati va ularning himoyalanganligi muhim rol o'ynaydi. Agar model noto'g'ri yoki zararli ma'lumotlar asosida o'qitilsa, u holda tizim noto'g'ri xulosalar chiqarishi yoki xavfli kontent yaratishi mumkin. Shu sababli, ma'lumotlarni oldindan tozalash, filtratsiya qilish va verifikatsiya qilish jarayonlari muhim bosqich hisoblanadi. Bundan tashqari, generativ AI tizimlari turli xil kiberhujumlarga ham duch kelishi mumkin. Jumladan, adversarial hujumlar orqali modelning ishlashiga ataylab ta'sir ko'rsatilib, noto'g'ri natijalar olish mumkin. Bunday hujumlarda kirish ma'lumotlariga kichik o'zgartirishlar kiritish orqali modelni chalg'itish maqsad qilinadi. Shu nuqtai nazardan, modelni bunday tahdidlardan himoya qilish uchun maxsus himoya mexanizmlarini ishlab chiqish zarur. Masalan, robust o'qitish (robust training), adversarial trening va modelni sinovdan o'tkazish usullari keng qo'llaniladi. Yana bir muhim jihat – bu tizimlarning ishonchliligini ta'minlashdir. Generativ AI natijalarining to'g'riligi va izchilligi foydalanuvchi ishonchini shakllantiradi. Shu bois model natijalarini tekshirish, monitoring qilish va xatoliklarni aniqlash mexanizmlarini joriy etish zarur. Ayrim hollarda "human-in-the-loop" yondashuvi, ya'ni inson nazoratini tizimga qo'shish orqali natijalar sifatini oshirish mumkin. Shuningdek, sun'iy intellekt tizimlarida ma'lumotlar maxfiyligini ta'minlash ham muhim masalalardan biridir. Foydalanuvchi ma'lumotlarining oshkor bo'lishi katta xavf tug'dirishi mumkin. Shu sababli, differensial maxfiylik (differential privacy), shifrlash texnologiyalari va xavfsiz hisoblash usullaridan foydalanish tavsiya etiladi. Bu yondashuvlar ma'lumotlarni himoyalash bilan birga, modelning samaradorligini ham saqlab qolishga yordam beradi. Bundan tashqari, generativ AI tizimlarida etik me'yorlarga amal qilish masalasi ham muhim hisoblanadi. Tizim tomonidan yaratilayotgan kontent jamiyatga salbiy ta'sir ko'rsatmasligi kerak. Shu sababli, zararli yoki noto'g'ri axborotni aniqlash va cheklash mexanizmlarini ishlab chiqish zarur. Bu esa AI tizimlarini boshqarishda mas'uliyatli yondashuvni talab qiladi. Generativ sun'iy intellekt tizimlarida xavfsizlik va ishonchlilikni ta'minlash kompleks yondashuvni talab etadi. Bu jarayonda texnik, tashkiliy va etik choralar uyg'un holda qo'llanilishi lozim. Faqat shundagina AI tizimlaridan samarali va xavfsiz foydalanish imkoniyati yaratiladi.

Mazkur maqolada generativ sun'iy intellekt tizimlarida xavfsizlik va ishonchlilikni ta'minlash masalalari atroflicha tahlil qilindi. Tadqiqot natijalari shuni ko'rsatadiki, bunday

tizimlarning samarali ishlashi nafaqat ularning funksional imkoniyatlariga, balki ularning himoyalanganlik darajasiga ham bevosita bog'liq. Generativ AI tizimlarida yuzaga kelishi mumkin bo'lgan asosiy xavflar – noto'g'ri ma'lumotlar asosida ishlash, adversarial hujumlar, ma'lumotlar maxfiylikning buzilishi va etik me'yorlarga rioya qilinmasligi bilan bog'liq. Shu sababli, ushbu muammolarni bartaraf etish uchun kompleks yondashuv zarur ekanligi aniqlandi. Xususan, ma'lumotlarni sifatli tayyorlash, modelni mustahkam o'qitish, xavfsizlik mexanizmlarini joriy etish va doimiy monitoring olib borish muhim ahamiyatga ega.

Bundan tashqari, inson nazorati va etik tamoyillarga amal qilish generativ AI tizimlarining ishonchliligini oshirishda muhim omil bo'lib xizmat qiladi. Kelajakda sun'iy intellekt texnologiyalarining yanada rivojlanishi bilan birga, ularning xavfsizligini ta'minlash bo'yicha yangi yondashuvlar va metodlar ishlab chiqilishi zarur. Umuman olganda, generativ sun'iy intellekt tizimlarini xavfsiz va ishonchli tarzda rivojlantirish nafaqat texnologik, balki ijtimoiy va axloqiy jihatdan ham muhim vazifa hisoblanadi.

Foydalanilgan adabiyotlar

1. Abduqodirov A.A. «Axborot texnologiyalari va tizimlari». – Toshkent: O'qituvchi, 2023.
2. Karimov M.M. «Axborot xavfsizligi asoslari». – Toshkent: Fan va texnologiya, 2021.
3. Tursunov X.T. «Sun'iy intellekt asoslari». – Toshkent: Universitet nashriyoti, 2021.
4. Raxmonov D.A. «Dasturiy injiniring va tizimlar xavfsizligi». – Toshkent: Innovatsiya nashriyoti, 2022.
5. Yo'ldoshev B.S. «Kiberxavfsizlik asoslari». – Toshkent: Fan, 2021.
6. Axmedov Q.R. «Ma'lumotlar bazasi va ularni boshqarish tizimlari». – Toshkent: Oliy ta'lim nashriyoti, 2018.
7. Ismoilov N.N. «Kompyuter tarmoqlari va xavfsizlik». – Toshkent: Tafakkur, 2020.
8. Jo'rayev S.M. «Zamonaviy axborot texnologiyalari». – Toshkent: Sharq, 2022.