

SUN'IY INTELLEKT VOSITALARI TALABA FAOLIYATIDA KELITIRIB CHIQARADIGAN AXBOROT TAHDIDLARINING PEDAGOGIK TASNIFI

Sh.M.Mustafojev¹ 

Mas'ul muallif: Mustafojev
Shahzod Mamanazar o'g'li, Buxoro
Xalqaro Universiteti mustaqil
izlanuvchisi.

e-mail: shahzod@buxdpi.uz

Received: 04 Oktabr 2026

Accepted: 09 Oktabr 2026

Published: 14 Oktabr 2026

Copyright: © 2026 by the
author(s). Licensee Buxoro xalqaro
universiteti ilmiy axborotnomasi.
This article is an open access
article distributed under the terms
and conditions of the Creative
Commons Attribution (CC-BY-4.0)
license.

Iqtibos uchun: Mustafojev, S. M.
SUN'IY INTELLEKT VOSITALARI
TALABA FAOLIYATIDA KELITIRIB
CHIQARADIGAN AXBOROT
TAHDIDLARINING PEDAGOGIK
TASNIFI / S. M. Mustafojev //
Buxoro xalqaro universiteti ilmiy
axborotnomasi. — 2026. — T. 10,
№ 1. — B. 457–462.

1. Buxoro Xalqaro Universiteti mustaqil izlanuvchisi

Annotatsiya. Maqolada oliy ta'lim talabalari generativ sun'iy intellekt vositalari bilan kundalik ishlaganda duch keladigan axborot tahdidlarini pedagogik maqsadda tasniflash masalasi ko'rib chiqilgan. Uchta oliy ta'lim muassasasida 586 talaba o'rtasida so'rov o'tkazildi. Respondentlarning 71,3 foizi shaxsiy ma'lumotlarini sun'iy intellektga kiritgan, 64,7 foizi olingan natijani tekshirmay ishlatgan, 43,6 foizi esa rasmiy xizmat o'rniga uchinchi shaxslar yaratgan Telegram-botlardan foydalanadi. OWASP, NIST va boshqa mavjud tasniflar muhandislar uchun tuzilgan, shu bois ularni to'g'ridan-to'g'ri o'quv mazmuniga aylantirib bo'lmaydi. Muallif tahdidlarni talabaniy sun'iy intellekt bilan muloqoti bosqichlariga qarab to'rt guruhga ajratadi: kiruvchi ma'lumotlar, chiquvchi natija, kirish va hisob qaydnomasi hamda kognitiv-xulqiy tahdidlar. Har bir guruh uchun shakllantirilishi kerak bo'lgan himoya ko'nikmasi belgilangan. Tasnifdan o'quv mazmunini tanlash va keyslar tuzishda qanday foydalanish mumkinligi misollarda ko'rsatilgan.

Kalit so'zlar: sun'iy intellekt, axborot tahdidlari, pedagogik tasnif, gallyutsinatsiya, prompt-inyeksiya, deepfake, shaxsga doir ma'lumotlar, talabalar, xavfsiz xulq-atvor.

ПЕДАГОГИЧЕСКАЯ КЛАССИФИКАЦИЯ ИНФОРМАЦИОННЫХ УГРОЗ, ВОЗНИКАЮЩИХ В ДЕЯТЕЛЬНОСТИ СТУДЕНТА ПРИ ИСПОЛЬЗОВАНИИ СРЕДСТВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. В статье рассматривается задача педагогической классификации информационных угроз, возникающих при повседневном взаимодействии студентов вузов с генеративными средствами искусственного интеллекта. Приводятся результаты опроса 586 студентов трёх вузов: 71,3% вводили в ИИ персональные данные, 64,7% использовали ответ без проверки, 43,6% обращаются к ИИ через сторонние Telegram-боты. Показано, что существующие классификации (OWASP, NIST и др.) построены с инженерной точки зрения и не могут быть напрямую переведены в учебное содержание. Предложено делить угрозы по этапам акта взаимодействия студента с ИИ на четыре группы – входные данные, выходной результат, доступ и учётная запись, когнитивно-поведенческие угрозы – с указанием формируемого защитного умения для каждой группы. На примерах показано применение классификации при отборе содержания и проектировании кейсов.

Ключевые слова: искусственный интеллект, информационные угрозы, педагогическая классификация, галлюцинация, промпт-инъекция, дипфейк, персональные данные, студенты, безопасное поведение.

PEDAGOGICAL CLASSIFICATION OF INFORMATION THREATS ARISING FROM THE USE OF ARTIFICIAL INTELLIGENCE TOOLS IN STUDENT ACTIVITY

Abstract. The paper addresses the task of classifying, for pedagogical purposes, the information threats that arise in students' everyday interaction with generative AI tools. Results of a survey of 586 students from three universities are reported: 71.3% had entered personal data into AI tools, 64.7% had used AI output without verification, and 43.6% access AI through third-party Telegram bots rather than official services. It is shown that existing classifications (OWASP, NIST and others) are built from an engineering standpoint and cannot be translated directly into curriculum content. The author proposes grouping threats by the stages of a student's interaction act with AI into four categories – input data, output result, access and account, cognitive-behavioural threats – and specifies the protective skill to be developed for each. The use of the classification in content selection and case design is illustrated with examples.

Keywords: artificial intelligence, information threats, pedagogical classification, hallucination, prompt injection, deepfake, personal data, students, safe behaviour.

Kirish. Axborot xavfsizligi sohasida tahdidlarni tasniflash xavflarni aniqlash, ularning xususiyatlarini tizimlashtirish va tegishli himoya choralarini ishlab chiqishning muhim metodologik vositalaridan biri hisoblanadi. Biroq tahdidlar tasnifining mazmuni va undan foydalanish maqsadi muayyan faoliyat sohasi hamda foydalanuvchining

ehtiyotlariga qarab farqlanadi. Xususan, muhandislik yondashuvda tahdidlar tasnifi, asosan, axborot tizimlarini loyihalash, zaifliklarni aniqlash va texnik himoya mexanizmlarini ishlab chiqishga xizmat qiladi. Pedagogik nuqtayi nazardan esa bunday tasnifning asosiy vazifasi ta'lim oluvchilarda muayyan xavflarni anglash, ularni baholash, oldini olish va xavfsiz faoliyat yuritish uchun zarur bo'lgan bilim hamda ko'nikmalarni shakllantirishdan iborat. Shu sababli, sun'iy intellekt xavfsizligi bo'yicha muhandislik va siyosiy-huquqiy yondashuvlar doirasida ishlab chiqilgan tahdidlar tasniflarini oliy ta'lim amaliyotiga o'zgartirishsiz tatbiq etish yetarli natija bermasligi mumkin. Ularni talabalarining faoliyat xususiyatlari, raqamli muhitdagi kundalik ehtiyotlari va ta'lim maqsadlariga muvofiq ravishda pedagogik jihatdan qayta ko'rib chiqish zarur.

Mazkur muammo "Sun'iy intellekt muhitida axborot xavfsizligi" maxsus kursining mazmunini shakllantirish jarayonida amaliy jihatdan namoyon bo'ldi. Dastlab kurs mavzularini tanlashda OWASP tashkiloti tomonidan katta til modellari uchun ishlab chiqilgan "Top 10" tahdidlar ro'yxati asosiy manbalardan biri sifatida qabul qilindi. Biroq mazkur ro'yxatdagi ayr

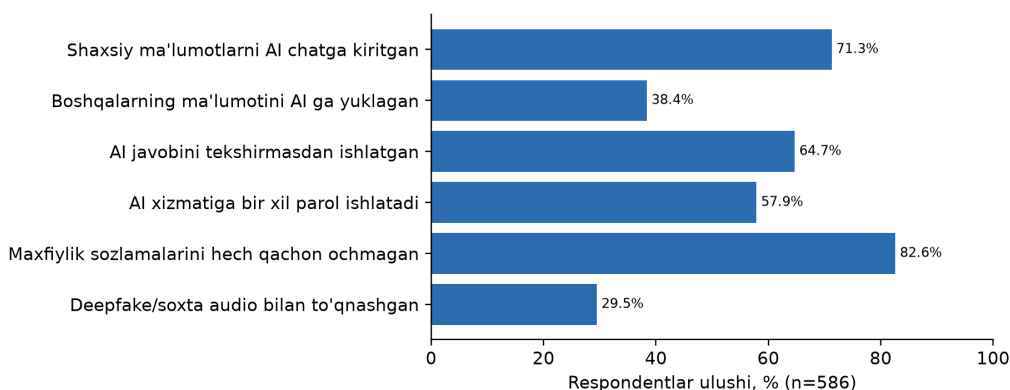
Tadqiqot ikki o'zaro bog'liq bosqichda amalga oshirildi. Birinchi bosqichda sun'iy intellekt tizimlari bilan bog'liq tahdidlarni tasniflashga doir mavjud xalqaro yondashuvlar mazmuniy jihatdan tahlil qilindi. Jumladan, OWASP tashkilotining katta til modellari uchun ishlab chiqilgan LLM Top 10 tahdidlar ro'yxati [1], NIST tomonidan taqdim etilgan sun'iy intellekt risklarini boshqarish doiraviy hujjati [2], L. Weidinger va hammualliflarining til modellari faoliyati natijasida yuzaga kelishi mumkin bo'lgan zararlar taksonomiyasi [3] hamda ENISA tashkilotining tahdidlar landshafti [4] o'rganildi. Mazkur manbalar asosida sun'iy intellektdan foydalanish jarayonida yuzaga keladigan xavflarning asosiy turlari, ularning xususiyatlari va tasniflash mezonlari qiyosiy tahlil qilindi. Shuningdek, 2023–2025-yillar davomida sun'iy intellekt texnologiyalari bilan bog'liq holda qayd etilgan real insidentlar to'g'risidagi ochiq manbalarda e'lon qilingan ma'lumotlar to'planib, tahlil qilindi. Ushbu bosqichning asosiy vazifasi mavjud nazariy va amaliy tasniflarning mazmunini aniqlash hamda ularni talabalarining raqamli faoliyatida uchraydigan xavflar bilan qiyoslash uchun zarur asosni shakllantirishdan iborat bo'ldi.

Tadqiqotning ikkinchi bosqichida 2023–2024-o'quv yilida uchta oliy ta'lim muassasasida tahsil olayotgan talabalar o'rtasida anonim so'rov o'tkazildi. So'rovda "Informatika o'qitish metodikasi", "Axborot tizimlari va texnologiyalari", "Dasturiy injiniring" hamda "Pedagogika va psixologiya" ta'lim yo'nalishlarining 2–3-kurslarida tahsil olayotgan jami 586 nafar talaba ishtirok etdi. So'rovnoma 28 ta savoldan iborat bo'lib, unda talabalarining sun'iy intellekt vositalaridan foydalanish xususiyatlari, axborot xavfsizligi nuqtayi nazaridan xavfli deb baholanishi mumkin bo'lgan xatti-harakatlari hamda mavjud tahdidlar to'g'risidagi xabardorlik darajasini aniqlashga e'tibor qaratildi. Ochiq turdagi savollar respondentlarga o'z tajribalarida uchragan vaziyatlarni mustaqil ravishda bayon etish imkonini berdi. Bundan tashqari, tadqiqotda 64 nafar o'qituvchi ham ishtirok etdi va ular bilan sun'iy intellektdan foydalanish jarayonida kuzatiladigan xavflar, talabalarining axborot xavfsizligiga oid xatti-harakatlari hamda ta'lim jarayonida uchraydigan muammolarni aniqlash maqsadida o'z tuzilmali suhbatlar o'tkazildi. So'rovnoma va suhbatlar natijalari talabalar faoliyatida uchraydigan tahdidlarni pedagogik nuqtayi nazardan tahlil qilish uchun empirik ma'lumotlar bazasini shakllantirdi.

So'rov natijalari talabalarining sun'iy intellekt vositalaridan foydalanish darajasi va shakllari turlicha ekanligini ko'rsatdi. Xususan, respondentlarning 94,2 foizi kamida bitta matnli chat-yordamchidan haftasiga bir martadan ko'proq foydalanishini bildirgan. Tarjima va matnlarni tahrirlash xizmatlaridan foydalanish ko'rsatkichi 61,4 foizni, tasvir yaratish vositalaridan foydalanish 47,8 foizni, dasturiy kod yozishga ko'maklashuvchi yordamchilardan foydalanish esa 38,1 foizni tashkil etgan. Kod yozish vositalaridan, asosan, texnik yo'nalishlarda tahsil olayotgan talabalar foydalanishi qayd etilgan. Shuningdek, respondentlarning 26,3 foizi ovoqli ma'lumotlarni matnga aylantiruvchi ilovalardan foydalanishini bildirgan. Ushbu natijalar generativ sun'iy intellekt vositalari talabalarining nafaqat matn yaratish va tahrirlash, balki dasturlash, tasvir bilan ishlash hamda nutqni qayta ishlashga oid faoliyatlariga ham kirib borganini ko'rsatadi.

Tadqiqotda alohida e'tibor qaratilgan jihatlardan biri sun'iy intellekt xizmatlaridan foydalanish kanallari bilan bog'liq bo'ldi. So'rov ma'lumotlariga ko'ra, respondentlarning 43,6 foizi sun'iy intellekt vositalaridan rasmiy veb-saytlar yoki mobil ilovalar orqali emas, balki uchinchi shaxslar tomonidan yaratilgan Telegram-botlar vositasida foydalanishini bildirgan. Mazkur ko'rsatkich talabalarining axborot xavfsizligi bilan bog'liq xatarlarini baholashda faqat sun'iy intellekt tizimining o'ziga emas, balki ushbu tizimga kirish uchun foydalaniladigan vositachilarga ham e'tibor qaratish zarurligini ko'rsatadi. Xususan, norasmiy botlar orqali yuborilgan ma'lumotlarning kim tomonidan qayta ishlanishi, qayerda saqlanishi va qanday maqsadlarda foydalanilishi har doim ham foydalanuvchiga ma'lum bo'lmisligi mumkin. Natijada, shaxsiy yoki akademik ma'lumotlarning oshkor etilishi, ularning uchinchi shaxslar tomonidan qayta ishlanishi hamda hisob ma'lumotlarining xavfsizligiga doir qo'shimcha tahdidlar yuzaga kelishi ehtimoli mavjud. Shu bois, talabalarining sun'iy intellektdan foydalanish amaliyotini o'rganishda foydalanilayotgan vositalarning funksional imkoniyatlari bilan bir qatorda, ularning rasmiyligi, ishonchliligi va ma'lumotlarni qayta ishlash shartlarini anglash darajasini ham hisobga olish lozim.

Xavfli xulq-atvor ko'rsatkichlari. 1-rasmda keltirilgan.



1-rasm. Talabalarining sun'iy intellekt vositalari bilan muloqotidagi xavfli xulq-atvor ko'rsatkichlari (n=586)

Ochiq savollarga berilgan javoblar bu raqamlar ortida nima turganini tushunishga yordam beradi. "Bu shunchaki dastur, u ma'lumotni qayerga yuborishini o'ylamaganman" (2-kurs). "Google ham hamma narsani biladi, buning nima farqi bor" (3-kurs). "O'qituvchi tekshirmaydi, sun'iy intellekt yozganini o'zim yozgandek topshiraman" (3-kurs). Uchala javobning umumiy jihati shundaki, talaba xavfni inkor etmaydi. U shunchaki xavfni ko'rmaydi, chunki buni unga hech kim ko'rsatmagan.

OWASP ro'yxatiga prompt-inyeksiya, xavfsiz bo'lmagan chiqish ma'lumotlarini qayta ishlash, o'qitish ma'lumotlarini zaharlash, xizmatni rad etish, ta'minot zanjiri zaifliklari, maxfiy ma'lumotlarni oshkor qilish, xavfsiz bo'lmagan plaginlar, haddan ortiq vakolat, ortiqcha ishonch va model o'g'irlanishi kiritilgan [1]. Ulardan to'rttasi (zaharlash, xizmatni rad etish, ta'minot zanjiri, model o'g'irlanishi) talaba faoliyatiga umuman taalluqli emas. Qolganlari esa "nima buziladi" degan nuqtai nazardan yozilgan va "talaba nima qilishi kerak" degan savolga javob bermaydi. NIST hujjatida risklar ishonchlilik xususiyatlari bo'yicha guruhlangan. Tashkilot rahbari uchun bu qulay, lekin talabaga mavhum tuyuladi. L.Weidinger taksonomiyasi esa jamiyat miqyosidagi zararlarga qaratilgan. Bu manbalarining barchasi foydali, biroq o'quv mazmunini tuzish uchun boshqacha asos kerak.

Biz tahdidlarni talabaning sun'iy intellekt vositasi bilan bitta muloqoti qanday bosqichlardan o'tishiga qarab tasniflashni taklif qilamiz. Soddalashtirib aytganda, muloqot to'rt bosqichdan iborat: talaba vositaga kiradi, ma'lumot kiritadi, natija oladi, so'ng natijani baholab, undan foydalanadi. Har bir bosqichning o'z tahdidlari bor (1-jadval).

1-jadval

Sun'iy intellekt tahdidlarining pedagogik tasnifi

Tahdid guruhi	Tipik tahdidlar	Talaba faoliyatidagi misol	Shakllantiriladigan himoya ko'nikmasi
I. Kiruvchi ma'lumotlar (prompt bosqichi)	Shaxsiy ma'lumotlarni oshkor qilish; uchinchi shaxslar ma'lumotlarini uzatish; maxfiy hujjatlarni yuklash; ma'lumotning o'qitishga ishlatilishi; noma'lum vositachi orqali uzatish; transchegaraviy uzatish	Guruh ro'yxatini "jadval qilib ber" deb yuklash; amaliyot o'tagan maktab o'quvchilari baholarini tahlil qildirish	Ma'lumotni tasniflash (ochiq / ichki / maxfiy / shaxsiy); anonimlashtirish; "yuklashdan oldin uch savol" qoidasi
II. Chiquvchi natija (javob bosqichi)	Gallyutsinatsiya (soxta fakt, manba, iqtibos); zaif yoki zararli kod; mualliflik huquqi buzilishi; xolislik buzilishi; bilvosita prompt-inyeksiya orqali zararli havola	Kurs ishida mavjud bo'lmagan maqolaga havola; sun'iy intellekt yozgan kodda SQL-inyeksiya zaifligi	"Ikki manba qoidasi"; manbani birlamchi manbadan tekshirish; kodni xavfsizlik nuqtai nazaridan ko'rib chiqish; mualliflikni ko'rsatish
III. Kirish va hisob qaydnomasi	Soxta ilova va kengaytmalar; parolni qayta ishlatish; API kalitlarni oshkor qilish; sessiya o'g'irlanishi; ortiqcha OAuth ruxsatlari; "bepul Premium" fishingi	"ChatGPT for Chrome" nomli soxta kengaytma; Telegram-botga universitet pochta paroli bilan kirish	Rasmiy manbani aniqlash; noyob parol va parol menejeri; ikki bosqichli autentifikatsiya; ruxsatlarni minimallashtirish
IV. Kognitiv-xulqiy	Avtomatlashtirishga ortiqcha ishonish; tanqidiy fikrlash pasayishi; deepfake va ovoz klonlash; shaxsiylashtirilgan fishing; qaramlik; "sun'iy intellekt aytdi" argumenti	Dekanat nomidan kelgan mukammal xatga ishonish; "onam"ning klonlangan ovozi bilan pul so'rovi	Shubhali belgilarni tanish; "ikkinchi kanal orqali tasdiqlash"; sun'iy intellektni yordamchi, hakam emas deb qabul qilish; refleksiv kundalik

Taklif etilgan pedagogik tasnif doirasida sun'iy intellekt vositalaridan foydalanish jarayonida yuzaga keladigan tahdidlarning har bir guruhini alohida ko'rib chiqish va ularning talabalarining axborot xavfsizligi kompetensiyasini shakllantirish bilan bog'liq jihatlarni aniqlash muhim ahamiyatga ega.

Birinci guruhga shaxsga doir ma'lumotlarning maxfiyligi va ularni qonunga muvofiq qayta ishlash bilan bog'liq tahdidlar kiritiladi. Mazkur guruhning o'ziga xos jihati shundaki, u nafaqat axborot xavfsizligi, balki shaxsning huquqlari va

qonuniy manfaatlarini himoya qilish bilan ham bevosita bog'liq. O'zbekiston Respublikasining «Shaxsga doir ma'lumotlar to'g'risida»gi qonunida shaxsga doir ma'lumotlarni qayta ishlash va ularni transchegaraviy uzatish bilan bog'liq munosabatlar huquqiy jihatdan tartibga solingan [5]. Shu nuqtayi nazardan, talabani guruhdoshlariga tegishli ism-familiyalari, telefon raqamlari yoki baholarni tashqi sun'iy intellekt tizimiga yuklashi ma'lumotlarni qayta ishlashning qonuniyligi, tegishli rozilikning mavjudligi va axborotni himoya qilish talablari bilan bog'liq masalalarni yuzaga keltirishi mumkin. Ayniqsa, mazkur ma'lumotlar xorijiy serverlarda qayta ishlanganda, foydalanuvchi ma'lumotlarni saqlash va ulardan foydalanish shartlarini yetarlicha bilmasligi qo'shimcha xavflarni keltirib chiqaradi. 2023-yilda yirik korporatsiya muhandislari tomonidan maxfiy dasturiy kodning chat-botga kiritilishi va buning ortidan kompaniyada bunday vositalardan foydalanishga cheklovlar joriy etilishi ushbu turdagi xavflarning amaliy ahamiyatini ko'rsatuvchi misollardan biridir [6]. Mazkur tadqiqot doirasida o'tkazilgan so'rov natijalariga ko'ra, respondentlarning 38,4 foizi boshqa shaxslarga tegishli ma'lumotlarni sun'iy intellekt vositalariga yuklaganini bildirgan. Bu respondentlarning aksariyati bunday harakatlarning huquqiy jihatlari bilan bog'liq talablarni bilmaganini ham qayd etgan. Ushbu natijalar talabalarda shaxsga doir ma'lumotlarni himoya qilish, ma'lumotlarni tashqi tizimlarga uzatishdan oldin xavflarni baholash va tegishli huquqiy talablarni hisobga olish ko'nikmalarini shakllantirish zarurligini ko'rsatadi.

Ikkinchi guruhga sun'iy intellekt tomonidan yaratilgan axborotning ishonchliligi, natijalarni tanqidiy baholash hamda modelga taqdim etilgan ma'lumotlar orqali yuzaga keladigan manipulyativ ta'sirlar bilan bog'liq tahdidlar kiritiladi. Ushbu guruh oliy ta'lim uchun alohida ahamiyatga ega, chunki talabalar sun'iy intellekt tomonidan yaratilgan matnlar, dalillar va ilmiy manbalardan o'quv hamda tadqiqot faoliyatida foydalanadilar. Z. Ji va hammualiflarining tadqiqotlarida katta til modellarida kuzatiladigan gallyutsinatsiya hodisasi, ya'ni haqiqatga mos kelmaydigan, biroq ishonchli ko'rinadigan axborotning yaratilishi tahlil qilingan [7]. Mazkur hodisa sun'iy intellekt natijalarini tekshirmasdan qabul qilish ta'lim jarayonida xato ma'lumotlarning tarqalishiga sabab bo'lishi mumkinligini ko'rsatadi. Bunga 2023-yilda AQShda sud jarayoni davomida advokatning chat-bot tomonidan yaratilgan, amalda mavjud bo'lmagan oltita sud pretsedentiga havola qilishi bilan bog'liq holat misol bo'la oladi [8].

Tadqiqot doirasidagi mashg'ulotlarda talabalar tomonidan sun'iy intellektdan olingan 2140 ta faktik mazmundagi javob tekshirildi. Natijada javoblarning 31,4 foizi tasdiqlanmadi, ilmiy manbalarga berilgan havolalar orasida esa tasdiqlanmaganlari 36,2 foizni tashkil etdi. Ushbu natijalar sun'iy intellekt yaratgan axborotni mustaqil manbalar orqali tekshirish, ilmiy havolalarning mavjudligi va ishonchliligini aniqlash ko'nikmalarini rivojlantirish zarurligini ko'rsatadi. Ikkinchi guruh doirasidagi yana bir muhim tahdid bilvosita prompt-inyeksiya bilan bog'liq. K. Greshake va hammualiflari tadqiqotlarida tashqi hujjatlar yoki veb-sahifalar tarkibiga yashirilgan ko'rsatmalar modelning ishlash jarayoniga ta'sir ko'rsatishi mumkinligi namoyish etilgan [9]. Masalan, talaba xulosa tayyorlash maqsadida sun'iy intellekt tizimiga veb-sahifani taqdim etganda, ushbu sahifaga joylashtirilgan zararli ko'rsatmalar modelning kutilmagan yoki xavfli natijalar yaratishiga olib kelishi mumkin. Shu sababli, talabalarda nafaqat yaratilgan javoblarni tekshirish, balki sun'iy intellektga taqdim etilayotgan tashqi manbalarni ham tanqidiy baholash ko'nikmalarini shakllantirish lozim.

Uchinchi guruh an'anaviy kiberxavfsizlik tahdidlarini qamrab oladi, biroq ularning namoyon bo'lish shakllari sun'iy intellekt xizmatlari va ulardan foydalanish kanallari bilan bog'liq ravishda o'zgaradi. Ushbu guruhga soxta ilovalar, zararli brauzer kengaytmalari, fishing havolalari, hisob ma'lumotlarini noqonuniy yig'ish hamda norasmiy vositachilar orqali amalga oshiriladigan firibgarlik holatlari kiradi. 2023-yilda soxta ChatGPT brauzer kengaytmalari orqali foydalanuvchilarning sessiya ma'lumotlarini o'g'irlash bilan bog'liq holatlar qayd etilgan [10]. Bundan tashqari, bepul premium xizmatlarni va'da qiluvchi havolalar va Telegram-botlar orqali parollarni qo'lga kiritishga qaratilgan urinishlar ham foydalanuvchilar uchun xavf tug'diradi. Tadqiqot natijalariga ko'ra, respondentlarning 57,9 foizi sun'iy intellekt xizmatlarida boshqa xizmatlar uchun ham foydalanadigan parollarini qo'llashini bildirgan. Bu holat hisob ma'lumotlarini himoya qilish, turli xizmatlarda alohida parollardan foydalanish, ikki bosqichli autentifikatsiyani qo'llash va xizmatlarning haqiqiyligini tekshirishga oid amaliy ko'nikmalarni shakllantirish zarurligini ko'rsatadi.

To'rtinchi guruhga sun'iy intellekt natijalariga ortiqcha ishonish, avtomatlashtirilgan qarorlarga tanqidsiz yondashish, manipulyativ axborot ta'siriga berilish va deepfake texnologiyalari orqali amalga oshiriladigan ijtimoiy muhandislik tahdidlari kiritiladi. Ushbu guruhning o'ziga xosligi shundaki, mazkur xavflarni bartaraf etish faqat texnik himoya vositalarini qo'llash bilan cheklanmaydi. U talabalarning axborotni baholash odatlari, mustaqil fikrlashi va qaror qabul qilish jarayonlarini rivojlantirishni ham talab qiladi. R. Parasuraman va D. Manzingning avtomatlashtirilgan tizimlarga ishonch bilan bog'liq tadqiqotlarida avtomatlashtirishga ortiqcha tayanish insonning nazorat qilish va tekshirish faoliyatiga ta'sir ko'rsatishi mumkinligi yoritilgan [11]. Sun'iy intellekt yordamchilarining ravon, izchil va ishonchli ohangda javob berishi foydalanuvchida ushbu natijalarni yetarlicha tekshirmasdan qabul qilish moyilligini kuchaytirishi mumkin. Shu bilan birga, deepfake texnologiyalari orqali ovoz va videoning soxtalashtirilishi ijtimoiy muhandislik tahdidlarining yangi shakllarini yuzaga keltirmoqda [12]. 2024-yilda Gonkongda moliya xodimining videokonferensiyada rahbarlari va hamkasblarining sun'iy yaratilgan tasvirlariga ishongan holda qariyb 25 million AQSh dollari miqdoridagi mablag'ni o'tkazgani haqidagi holat ushbu xavfning amaliy oqibatlarini ko'rsatadi [13]. Tadqiqot so'rovida respondentlarning 29,5 foizi soxta ovoz yoki videoga duch kelganini, 12,8 foizi esa bunday materiallar ta'sirida aldanib qolganini bildirgan. Ushbu ma'lumotlar talabalarda sun'iy intellekt natijalariga nisbatan tanqidiy munosabat, axborot manbalarini mustaqil tekshirish, shubhali vaziyatlarda qo'shimcha tasdiq olish va manipulyatsiyani aniqlash ko'nikmalarini rivojlantirish muhimligini asoslaydi.

Mazkur to'rt guruhli tasnifning muhim xususiyatlaridan biri shundaki, real hayotdagi bitta hodisa bir vaqtning o'zida bir nechta tahdid guruhiga tegishli bo'lishi mumkin. Masalan, talaba noma'lum Telegram-bot orqali kurs ishini va

unga ilova qilingan korxona hisobotini sun'iy intellekt tizimiga yuklagan vaziyatni ko'rib chiqamiz. Bunda, birinchi navbatda, hujjat tarkibidagi shaxsiy yoki xizmatga oid ma'lumotlarning oshkor etilishi birinchi guruh tahdidlariga taalluqli bo'ladi. Noma'lum vositachi orqali xizmatdan foydalanish va hisob ma'lumotlarining xavfsizligi bilan bog'liq xavflar uchinchi guruhga kiradi. Agar model tashqi manbada joylashtirilgan zararli ko'rsatmalar ta'sirida noto'g'ri natija shakllantirsa, ikkinchi guruh tahdidlari ham yuzaga keladi. Talabaning olingan xulosani tekshirmasdan akademik faoliyatida qo'llashi esa to'rtinchi guruh bilan bog'liq muammoni namoyon etadi. Demak, tahdidlarni o'zaro aloqasiz hodisalar sifatida o'rganishdan ko'ra, ularning birgalikda namoyon bo'lishini hisobga olgan holda tahlil qilish pedagogik jihatdan maqsadga muvofiqdir.

Shu asosda o'quv mazmunini tanlashda tahdidlarni alohida nazariy mavzular ko'rinishida bayon etish bilan cheklanmasdan, ularni yaxlit vaziyatlar va amaliy keyslar orqali o'rgatish taklif etiladi. Bunday yondashuv talabalarga xavfning mohiyatini anglash, uning yuzaga kelish shartlarini aniqlash va tegishli himoya choralari tanlash imkonini beradi. Tasnifning pedagogik afzalligi har bir tahdid guruhini talabaning muayyan xatti-harakati va ushbu harakat bilan bog'liq shakllantirilishi lozim bo'lgan ko'nikmalar bilan bog'lash imkoniyatida namoyon bo'ladi. Mazkur tamoyil asosida o'quv mazmunini matritsa usulida shakllantirish mumkin. Bunda matritsaning gorizontol o'qida tahdid guruhlari, vertikal o'qida esa axborot xavfsizligi kompetensiyasining motivatsion, kognitiv, faoliyatga oid va refleksiv komponentlari joylashtiriladi. Ularning kesishish nuqtalari muayyan mavzu doirasida shakllantirilishi zarur bo'lgan bilim, ko'nikma va xulq-atvor ko'rsatkichlarini belgilashga xizmat qiladi.

Taklif etilgan tasnif asosida 24 ta keynsni o'z ichiga olgan o'quv materiallari banki shakllantirildi. Keyslar uchta manba asosida ishlab chiqildi: ochiq axborot manbalarida yoritilgan real insidentlar, so'rov ishtirokchilari tomonidan bayon etilgan va shaxsiy ma'lumotlardan xoli tarzda umumlashtirilgan vaziyatlar hamda maxsus modellashtirilgan holatlar. Har bir keys to'rtta asosiy qismdan tashkil topdi: muammoli vaziyat tavsifi, tahliliy savollar, kutiladigan xavfsiz harakatlar hamda tegishli tahdid guruhi va kompetensiya ko'rsatkichlari bilan bog'liqlik. Bunday tuzilma o'quv topshiriqlarini ishlab chiqish, ularning mazmunini muayyan ta'lim natijalari bilan muvofiqlashtirish va talabalar faoliyatini baholash imkonini beradi.

Masalan, amaliyot hisoboti ustida ishlayotgan talaba yettinchi sinf o'quvchilarining baholari aks etgan jadvalni oladi. Guruhdoshi unga ushbu jadvalni ChatGPT kabi vositaga yuklash va tahlilni avtomatik bajarishni tavsiya qiladi. Mazkur vaziyatda talaba, avvalo, voyaga yetmaganlarga tegishli shaxsiy ma'lumotlarni himoya qilish bilan bog'liq birinchi guruh tahdidini aniqlashi lozim. Agar tahlil norasmiy vositachi orqali amalga oshirilishi rejalashtirilgan bo'lsa, uchinchi guruhga mansub xavflarni ham hisobga olishi kerak. Kutiladigan xavfsiz harakatlar qatoriga ma'lumotlarni uzatishning huquqiy asoslarini tekshirish, ism-familiyalarni shaxsni bevosita aniqlash imkonini bermaydigan shartli belgilar bilan almashtirish va zarur bo'lmagan ma'lumotlarni olib tashlash kiradi. Muqobil yechim sifatida tahlilni mahalliy elektron jadval vositalari yordamida bajarish mumkin. Agar ma'lumotlar allaqachon tashqi tizimga yuklangan bo'lsa, talaba xizmatning mavjud funksiyalaridan foydalanib suhbat yoki yuklangan ma'lumotlarni o'chirishi, zarur hollarda xizmat ko'rsatuvchiga ma'lumotlarni o'chirish haqida murojaat qilishi hamda vaziyat yuzasidan mas'ul rahbar yoki tegishli shaxsni xabardor qilishi lozim. Bunda ma'lumotlarni o'chirish imkoniyati va tartibi foydalanilgan xizmatning siyosati hamda amaldagi huquqiy talablarga bog'liq ekanligi ham e'tiborga olinadi.

Taklif etilgan tasnif yangi texnologiyalar va tahdid shakllarining paydo bo'lishini hisobga olish imkonini beruvchi ochiq tizim sifatida qaraladi. Xususan, 2024–2026-yillarda rivojlanib borayotgan agentli sun'iy intellekt tizimlari foydalanuvchi nomidan muayyan amallarni mustaqil bajarishi, xotira funksiyalari esa foydalanuvchi haqidagi ma'lumotlarni uzoq muddat davomida saqlashi mumkin. Ovozli rejimlardan foydalanish jarayonida esa atrofdagi suhbatlarning qayta ishlanishi bilan bog'liq maxfiylik masalalari yuzaga kelishi ehtimoli mavjud. Ushbu holatlar tegishli xususiyatlariga qarab taklif etilgan to'rt guruh doirasida tahlil qilinishi mumkin. Bunda tasnifning asosiy vazifasi talabadan barcha tahdidlar ro'yxatini yod olishni talab qilish emas, balki unga yangi vositalar bilan ishlashda mustaqil ravishda xavflarni aniqlash va baholash imkonini beradigan fikrlash usulini shakllantirishdan iborat. Shu maqsadda talaba o'z faoliyatini to'rtta asosiy savol asosida tahlil qilishi mumkin: tizimga qanday ma'lumot kiritilmoqda, undan qanday natija olinmoqda, xizmatga qanday vosita orqali kirilmoqda va ushbu jarayon talaba hamda boshqa shaxslarning manfaatlariga qanday ta'sir ko'rsatishi mumkin? Mazkur savollar axborot xavfsizligi kompetensiyasining amaliy mazmunini ochib berishga xizmat qiladi.

Xulosa. Tadqiqot natijalari talabalarining generativ sun'iy intellektdan foydalanish jarayonida xavfli xatti-harakatlarga yo'l qo'yishi ko'pincha xavflarni ataylab inkor etishi bilan emas, balki ularni yetarlicha anglamasligi va o'z harakatlarining ehtimoliy oqibatlarini oldindan baholay olmasligi bilan bog'liq bo'lishi mumkinligini ko'rsatadi. Muhandislik maqsadlarida ishlab chiqilgan mavjud tahdidlar tasniflari xavflarning texnik mohiyatini tushuntirishda muhim ahamiyatga ega bo'lsa-da, ularni bevosita pedagogik maqsadlarga moslashtirmasdan qo'llash talabalarining kundalik faoliyatiga xos muammolarni to'liq qamrab olmaydi. Taklif etilgan to'rt guruhli pedagogik tasnif tahdidlarni talabalarining real xatti-harakatlari, mazkur harakatlar bilan bog'liq xavflar va shakllantirilishi zarur bo'lgan kompetensiyalar bilan tizimli ravishda bog'lash imkonini beradi. Shu jihatdan, tasnif o'quv mazmunini tanlash, amaliy keyslarni loyihalash hamda diagnostik topshiriqlarni ishlab chiqish uchun metodologik asos vazifasini bajaradi. Tasnifning ta'lim jarayonidagi samaradorligi tajriba-sinov ishlari doirasida tekshirilgan bo'lib, ushbu bosqichda olingan natijalar alohida ilmiy maqolada yoritilishi rejalashtirilgan.

- [1] OWASP Top 10 for Large Language Model Applications. Version 1.1. – OWASP Foundation, 2023. – 33 p.
- [2] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). – Gaithersburg: NIST, 2023. – 48 p.
- [3] Weidinger L., Mellor J., Rauh M. et al. Ethical and social risks of harm from Language Models. – arXiv:2112.04359, 2021. – 64 p.
- [4] ENISA Threat Landscape 2024. – Athens: European Union Agency for Cybersecurity, 2024. – 138 p.
- [5] O'zbekiston Respublikasining "Shaxsga doir ma'lumotlar to'g'risida"gi Qonuni. 2019-yil 2-iyul, O'RQ-547-son.
- [6] Gurman M. Samsung bans staff's AI use after spotting ChatGPT data leak // Bloomberg. – 02.05.2023.
- [7] Ji Z., Lee N., Frieske R. et al. Survey of Hallucination in Natural Language Generation // ACM Computing Surveys. – 2023. – Vol. 55, № 12. – P. 1–38.
- [8] Weiser B. Here's what happens when your lawyer uses ChatGPT // The New York Times. – 27.05.2023.
- [9] Greshake K., Abdelnabi S., Mishra S., Endres C., Holz T., Fritz M. Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection // Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security. – New York: ACM, 2023. – P. 79–90.
- [10] Tal N. "FakeGPT": New variant of fake-ChatGPT Chrome extension stealing Facebook ad accounts // Guardio Labs. – 09.03.2023.
- [11] Parasuraman R., Manzey D.H. Complacency and Bias in Human Use of Automation: An Attentional Integration // Human Factors. – 2010. – Vol. 52, № 3. – P. 381–410.
- [12] Westerlund M. The Emergence of Deepfake Technology: A Review // Technology Innovation Management Review. – 2019. – Vol. 9, № 11. – P. 39–52.
- [13] Chen H., Magramo K. Finance worker pays out \$25 million after video call with deepfake "chief financial officer" // CNN. – 04.02.2024.
- [14] Kasneci E., Sessler K., Küchemann S. et al. ChatGPT for good? On opportunities and challenges of large language models for education // Learning and Individual Differences. – 2023. – Vol. 103. – Article 102274.
- [15] Vuorikari R., Kluzer S., Punie Y. DigComp 2.2: The Digital Competence Framework for Citizens. – Luxembourg: Publications Office of the European Union, 2022. – 133 p.