UO‘K 81

SUN’IY INTELLEKT VA MASHINA O‘RGANISHNING LINGVISTIK EKSPERTOLOGIYADAGI INNOVATSION YO‘NALISHLARI

Abduraxmonova Vazira Umurzoq qizi,
Respublika sud ekspertiza markazi sud eksperti

Annotatsiya. *Mazkur maqolada lingvistik ekspertologiyada sun’iy intellekt va mashina o‘rganish texnologiyalaridan foydalanishning zamonaviy yo‘nalishlari tahlil qilinadi. Raqamli kommunikatsiya muhitining kengayishi natijasida ijtimoiy tarmoq postlari, messenjer yozishmalari, audio-video transkriptlar, multimodal kontent va metadata kabi yangi turdagi lingvistik obyektlarni ekspertiza qilish zarurati ortib borayotganligi asoslanadi. Tadqiqotda sun’iy intellekt va mashina o‘rganish texnologiyalarining katta hajmdagi matnlarni tezkor tahlil qilish, kontekstual va semantik xususiyatlarni aniqlash, baholovchi-emotsional mazmunni klassifikatsiya qilish hamda ekspert xulosalarining tekshiriluvchanligini oshirishdagi imkoniyatlari yoritiladi. Shuningdek, lingvistik ekspertologiyada reproduktivlik, verifikatsiya va izohlanuvchanlik tamoyillarining ahamiyati ochib beriladi. Maqolada “human-in-the-loop” modeli asosida sun’iy intellekt vositalarini ekspert faoliyatiga integratsiya qilish, milliy va ekspertologik korpuslarni shakllantirish, ularni o‘zbek tilining lingvistik xususiyatlari, lotin-kirill yozuvi, transliteratsiya, sheva, internet slengi va noformal yozuv shakllarini hisobga olgan holda takomillashtirish masalalariga alohida e’tibor qaratiladi. Tadqiqot natijalari sun’iy intellekt va mashina o‘rganish texnologiyalaridan foydalanish lingvistik ekspertizaning aniqligi, tezkorligi, obyektivligi va metodik shaffofligini oshirishga xizmat qilishini ko‘rsatadi. Shu bilan birga, mazkur texnologiyalar ekspertni to‘liq almashtirmasligi, balki ekspert nazorati va yakuniy lingvistik-huquqiy interpretatsiya bilan uyg‘un holda qo‘llanishi zarurligi asoslanadi.*

Kalit so‘zlar: *lingvistik ekspertologiya, sud-lingvistik ekspertiza, sun’iy intellekt, mashina o‘rganish, tabiiy tilni qayta ishlash, ekspert tizimlari, lingvistik korpus, izohlanuvchanlik, verifikatsiya, reproduktivlik, avtomatlashtirilgan tahlil, human-in-the-loop.*

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ИННОВАЦИОННЫЕ НАПРАВЛЕНИЯ МАШИННОГО ОБУЧЕНИЯ В ЛИНГВИСТИЧЕСКОЙ ЭКСПЕРТОЛОГИИ

Аннотация. *В статье анализируются современные направления применения технологий искусственного интеллекта и машинного обучения в лингвистической экспертологии. Обосновывается необходимость исследования новых видов лингвистических объектов, таких как публикации в социальных сетях, переписка в мессенджерах, аудио- и видеотранскрипты, мультимодальный контент и метаданные, в условиях стремительного расширения цифровой коммуникационной среды. Рассматриваются возможности технологий искусственного интеллекта и машинного обучения в оперативном анализе больших объёмов текстовой информации, выявлении контекстуальных и семантических особенностей, классификации оценочной и эмоциональной лексики, а также повышении проверяемости экспертных выводов. Особое внимание уделяется принципам воспроизводимости, верификации и объяснимости результатов автоматизированного анализа. Рассматривается модель «human-in-the-loop», предусматривающая интеграцию технологий искусственного интеллекта в экспертную деятельность при сохранении контроля со стороны специалиста. Отдельно раскрывается значение создания национальных и специализированных экспертологических корпусов с учётом особенностей узбекского языка, параллельного использования латиницы и кириллицы, транслитерации, диалектных элементов, интернет-сленга и неформальных форм письменной речи. Делается вывод о том, что применение искусственного интеллекта и машинного обучения способствует повышению точности, оперативности, объективности и методологической прозрачности лингвистической экспертизы. Вместе с тем подчёркивается необходимость использования данных технологий в сочетании с профессиональной экспертизой и окончательной лингвистической и правовой интерпретацией.*

Ключевые слова: *лингвистическая экспертология, судебно-лингвистическая экспертиза, искусственный интеллект, машинное обучение, обработка естественного языка, экспертные системы, лингвистический корпус, объяснимость, верификация, воспроизводимость, автоматизированный анализ, human-in-the-loop.*

INNOVATIVE DIRECTIONS OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN LINGUISTIC EXPERTISE

Abstract. *This article analyzes current directions in the application of artificial intelligence and machine learning technologies in forensic linguistic expertology. The expansion of the digital communication environment has resulted in the emergence of new types of linguistic objects requiring expert analysis, including social media posts, messenger correspondence, audio and video transcripts, multimodal content, and metadata. The study examines the potential of artificial intelligence and machine learning technologies for the rapid analysis of large volumes of textual data, identification of contextual and semantic features, classification of evaluative and emotional content, and improvement of the verifiability of expert conclusions. Particular attention is paid to the principles of reproducibility, verification, and explainability of automated analytical results. The article also considers the "human-in-the-loop" model, which integrates artificial intelligence technologies into expert practice while maintaining professional human control over the analytical and interpretative process. Special emphasis is placed on the development of national and specialized forensic linguistic corpora that take into account the specific features of the Uzbek language, including the parallel use of Latin and Cyrillic scripts, transliteration, dialectal elements, Internet slang, and informal written forms. The study concludes that the application of artificial intelligence and machine learning can contribute to improving the accuracy, efficiency, objectivity, and methodological transparency of forensic linguistic examination. At the same time, these technologies should not completely replace the human expert but should be used in combination with professional linguistic expertise and final legal and linguistic interpretation.*

Keywords: *forensic linguistic expertology, forensic linguistics, artificial intelligence, machine learning, natural language processing, expert systems, linguistic corpus, explainability, verification, reproducibility, automated analysis, human-in-the-loop.*

Kirish. Bugungi kunda raqamli texnologiyalarning jadal rivojlanishi va kommunikatsiya vositalarining kengayishi natijasida insonlarning yozma va og‘zaki muloqot shakllari sezilarli darajada o‘zgarib bormoqda. Ijtimoiy tarmoqlar, messenjerlar, elektron axborot resurslari va boshqa raqamli platformalarda yaratilayotgan katta hajmdagi lingvistik ma’lumotlar sud-lingvistik ekspertiza uchun yangi turdagi tadqiqot obyektlarini yuzaga keltirmoqda. Jumladan, ijtimoiy tarmoq postlari, messenjer yozishmalari, audio-video materiallarning transkriptlari, multimodal kontent, metadata hamda platformaga xos lingvistik izlar ekspert tahlilining tarkibiy qismiga aylanmoqda.

An’anaviy lingvistik tahlil usullari katta hajmdagi raqamli ma’lumotlarni qisqa vaqt ichida qayta ishlash, ularning takrorlanuvchi lingvistik xususiyatlarini aniqlash va statistik jihatdan baholash imkoniyatlari nuqtayi nazaridan muayyan cheklovlarga ega. Shu sababli lingvistik ekspertologiyada kompyuter texnologiyalari, xususan, sun’iy intellekt (SI) va mashina o‘rganishi (MO‘) asosidagi vositalardan foydalanish zamonaviy ilmiy-amaliy yo‘nalishlardan biriga aylanmoqda. Bunday texnologiyalar katta hajmdagi matnlarni avtomatik qayta ishlash, semantik va kontekstual bog‘lanishlarni aniqlash, lingvistik birliklarni tasniflash hamda ekspert faoliyatini optimallashtirish imkonini beradi. Lingvistik ekspertologiyada sun’iy intellekt texnologiyalarining ahamiyati faqat tahlil jarayonini avtomatlashtirish bilan chegaralanmaydi. Zamonaviy ekspertiza uchun olingan natijaning ilmiy asoslanganligi, tekshiriluvchanligi va qayta ishlab ko‘rilishi ham muhim metodologik talab hisoblanadi. Shu nuqtayi nazardan, reproduktivlik, verifikatsiya va izohlanuvchanlik tamoyillari sun’iy intellekt asosidagi ekspert tizimlarining ishonchliligini ta’minlovchi muhim mezonlar sifatida namoyon bo‘ladi. Bir xil kirish ma’lumotlari va parametrlar asosida qayta tahlil qilinganda natijaning takrorlanishi, model natijalarining nazorat ma’lumotlari asosida tekshirilishi hamda model qarorlarining lingvistik mezonlar orqali tushuntirilishi ekspert xulosasining daliliy qiymatini oshiradi. Shuningdek, sun’iy intellekt texnologiyalaridan lingvistik ekspertologiyada foydalanishda inson ekspertining o‘rnini saqlab qolish muhim ahamiyatga ega. “Human-in-the-loop” modeli doirasida sun’iy intellekt katta hajmdagi ma’lumotlarni qayta ishlash, lingvistik birliklarni aniqlash va dastlabki analitik natijalarni shakllantirishda yordamchi vosita sifatida qo‘llanadi, yakuniy lingvistik interpretatsiya va huquqiy baholash esa ekspert tomonidan amalga oshiriladi. Bunday yondashuv avtomatlashtirilgan tahlilning tezkorligi bilan professional ekspertizaning metodologik imkoniyatlarini uyg‘unlashtirishga xizmat qiladi.

Mazkur masalaning yana bir muhim jihati lingvistik ekspertologiya uchun maxsus korpus va ma’lumotlar bazalarini yaratish bilan bog‘liq. O‘zbek tilining morfologik xususiyatlari, lotin va kirill

yozuvining parallel qo‘llanishi, transliteratsiya, sheva elementlari, internet slengi, qisqartmalar va noformal yozuv shakllarining keng tarqalganligi umumiy til korpuslaridan tashqari, ekspertiza vazifalariga yo‘naltirilgan maxsus annotatsiyalangan korpuslarni shakllantirish zaruratini yuzaga keltiradi. Shu bois lingvistik ekspertologiyani raqamlashtirish jarayonida sun’iy intellekt va mashina o‘rganish texnologiyalarini milliy til resurslari, metodik standartlar hamda lingvist, ekspert va yuristlarning institutsional hamkorligi bilan uyg‘unlashtirish dolzarb ilmiy-amaliy vazifa hisoblanadi. Bunda asosiy maqsad ekspertni texnologiya bilan almashtirish emas, balki ekspert faoliyatining aniqligi, tezkorligi, obyektivligi va tekshiriluvchanligini oshiruvchi raqamli metodologiyani shakllantirishdan iborat.

Mavzuga doir adabiyotlar tahlili. Lingvistik ekspertologiyada sun’iy intellekt va mashina o‘rganish texnologiyalaridan foydalanish masalasi raqamli matnlarni avtomatik tahlil qilish, ekspert xulosalarining tekshiriluvchanligi, izohlanuvchanligi hamda lingvistik korpuslarni shakllantirish masalalari bilan bog‘liq holda o‘rganilmoqda.

A.V. Zubov sud lingvistikasida raqamli matnlarni tahlil qilish jarayonida SI va MO‘ texnologiyalari ekspert tahlilining samaradorligi hamda aniqligini oshirishini ta’kidlaydi. M.T. Ribeiro, S. Singh va C. Guestrinlarning ishlarida katta hajmdagi matn va audio korpuslarini tezkor qayta ishlash, statistik va neyron modellardan foydalanish, natijalarni izohlash, reproduktivlik va verifikatsiya masalalariga e’tibor qaratiladi.

D.A. Nemirov raqamli dalillar va sud ekspertizasi metodologiyasi uyg‘unligini muhim omil sifatida ko‘rsatad . A. Joulin va hammualliflari esa matnlarni avtomatik klassifikatsiya qilishda leksik, kontekstual va sintaktik belgilarni birgalikda hisobga olish imkoniyatlarini yoritadi.

O‘zbek tilida lingvistik ekspertologiyani raqamlashtirish nuqtayi nazaridan A.A. Usmonovning qarashlari muhimdir. U sud-lingvistik ekspertiza uchun yaratiladigan korpuslarda morfologik belgilash, lemmatizatsiya, POS-tagging, lotin-kirill transliteratsiyasi, aralash yozuv, sheva va ijtimoiy dialekt elementlari hamda ekspertologik annotatsiyalangan ma’lumotlar bo‘lishi zarurligini ta’kidlaydi.

Adabiyotlar tahlili shuni ko‘rsatadiki, lingvistik ekspertologiyada SI va MO‘dan foydalanish faqat texnik avtomatlashtirish masalasi emas. U lingvistik korpuslarning sifati, metodik standartlar, ekspert nazorati, natijalarni izohlash va verifikatsiya qilish kabi masalalarni ham qamrab oladi.

Tadqiqot metodologiyasi. Tadqiqotda lingvistik ekspertologiyada sun’iy intellekt va mashina o‘rganish texnologiyalarining qo‘llanish imkoniyatlarini aniqlash uchun nazariy-tahliliy yondashuvdan foydalanildi. Tadqiqotning asosiy metodlarini nazariy tahlil, qiyosiy tahlil, korpus tahlili, kontekstual-semantik tahlil va ekspertologik yondashuv tashkil etdi.

Nazariy tahlil orqali lingvistik ekspertologiya, sun’iy intellekt, mashina o‘rganish, tabiiy tilni qayta ishlash va lingvistik korpuslarga oid ilmiy qarashlar o‘rganildi. Qiyosiy tahlil asosida an’anaviy lingvistik ekspertiza usullari bilan avtomatlashtirilgan tahlil imkoniyatlari qiyoslandi. Korpus tahlili lingvistik birliklarning real kommunikativ muhitda qo‘llanishini aniqlash va ularning kontekstual-semantik xususiyatlarini o‘rganish imkoniyatlarini baholashga yo‘naltirildi. Ekspertologik yondashuv esa avtomatlashtirilgan natijalarni mutaxassis tomonidan tekshirish va yakuniy lingvistik-huquqiy interpretatsiya bilan uyg‘unlashtirish zaruratini asoslashga xizmat qildi.

Tadqiqotda “human-in-the-loop” modeli alohida metodik tamoyil sifatida qaraldi. Unga ko‘ra, SI katta hajmdagi ma’lumotlarni qayta ishlash, lingvistik birliklarni aniqlash va dastlabki analitik natijalarni shakllantirishda yordamchi vosita bo‘ladi, yakuniy lingvistik interpretatsiya va huquqiy baholash esa ekspert tomonidan amalga oshiriladi.

Olingan natijalar va ularning tahlili. Lingvistik ekspertologiyada raqamli kommunikatsiya muhitining keskin kengayishi ekspert oldiga yangi turdagi obyektlarni — ijtimoiy tarmoq postlari, messenjer yozishmalari, audio-video transkriptlar, ko‘pkanalli (multimodal) kontent, metadata va platformaga xos izlarni — tahlil qilish vazifasini qo‘ydi. Hozirgi kunda ommaviy axborot vositalari, internet jahon ijtimoiy tarmoqlarning jad’al rivojlanishi lingvistik ekspertizaga bo‘lgan talabni oshirmoqda. Bunday holatlarda an’anaviy tahlil usullari yetarlicha o‘zini oqlamaydi. Raqamli texnologiyalar shiddat bilan rivojlangan zamonda tadqiqotlar sun’iy intelekt va kompyuter texnologiyasi asosida olib borilganda tadqiqotlar ishonchli, xolis va tezkor tarzda amalga oshiriladi. Mazkur holat lingvistik ekspertologiyada kompyuter texnologiyalari, xususan sun’iy intellekt va mashina o‘rganishiga tayangan innovatsion yechimlarni joriy etishni ilmiy-amaliy zaruratga aylantirmoqda. Bu borada A.V. Zubov sud lingvistikasida raqamli matnlarni tahlil qilish jarayonida SI va MO‘ texnologiyalari ekspert tahlilining samaradorligi hamda aniqligini sezilarli darajada oshirishini alohida ta’kidlaydi.

Marco Tulio Ribeiro, Sameer Singh va Carlos Guestrinning fikricha: SI va MO‘ning lingvistik ekspertologiyadagi o‘rni, avvalo, ikki yo‘nalishda ko‘rinadi:

1. Katta hajmdagi matn/audio korpuslarini tezkor tahlil qilish, shuningdek, xulosalarning tekshiriluvchanligini oshirish uchun statistik va neyron modellardan foydalanish;

2. Ekspert xulosasining metodik “shaffofligi”ni ta’minlaydigan, natijalarni izohlash (explainability) va xatoliklarni boshqarish (uncertainty management) mexanizmlarini ishlab chiqish.

Mazkur yondashuv lingvistik ekspertologiyada sun’iy intellekt texnologiyalarining faqat avtomatlashtirish vositasi emas, balki ekspert xulosalarining ilmiy asoslanganligi va tekshiriluvchanligini ta’minlovchi metodologik mexanizm sifatida ham muhim ekanini ko‘rsatadi. Ayniqsa, ushbu tamoyil ekspert tomonidan chiqarilgan xulosaning qanday lingvistik belgilar va statistik omillarga tayangan holda shakllanganini izohlash imkonini beradi. Bu esa sud-huquqiy ekspertizada xulosaning daliliy kuchini oshirish bilan birga, uning qayta tekshirilishiga ham xizmat qiladi. D.A.Nemirovning ta’kidlashicha, bu ikki yo‘nalishning uyg‘unlashuvi lingvistik ekspertiza sifatini oshirishi bilan birga, sud jarayonida dalilning ishonchliligi, metodning tekshiruvdan o‘tishi va qayta qo‘llanishi kabi mezonlarni ham mustahkamlaydi. Bizningcha, Nemirov tomonidan qayd etilgan mazkur integrativ yondashuv lingvistik ekspertologiyada “inson eksperti + algoritmik tahlil” modelining shakllanishiga olib keladi. Bunda SI va MO texnologiyalari ekspertning o‘rnini to‘liq egallamaydi, balki murakkab til birliklarini aniqlash, statistik tendensiyalarni ko‘rsatish va xulosaning verifikatsiyasini ta’minlash orqali ekspert faoliyatining ilmiy aniqligini kuchaytiradi. Shu sababli izohlanuvchanlik va tekshiriluvchanlik tamoyillari kelajak lingvistik ekspertizasining asosiy metodologik mezonlaridan biri sifatida qaralishi mumkin. Masalan, Armand Joulin va hammualliflari ta’kidlashicha “baholovchi-emotsional leksika” tushunchasi an’anaviy tahlilda ekspertning talqiniga suyanishi mumkin matnlarni avtomatik klassifikatsiya qilishda kontekstual belgilar, leksik birliklar hamda sintaktik bog‘lanishlarning birgalikdagi roli muhim ekanini ta’kidlaydi. Tadqiqotchilar fikricha, mashinaviy o‘qitish modellari til birliklarini alohida emas, balki ularning qo‘llanish muhiti bilan birga tahlil qilishi sababli baholovchi yoki emotsional mazmunni aniqlash imkoniyati kengayadi.Ushbu fikr shuni anglatadiki, model “so‘zning o‘zi”ni emas, uning kontekstda qanday ishlatilishini ham hisobga oladi; ekspert esa natijani tekshirib, lingvistik izoh bilan mustahkamlaydi. Mazkur yondashuv lingvistik ekspertologiyada kontekst omilining ustuvor ahamiyatga ega ekanini ko‘rsatadi. Chunki bir xil leksik birlik turli kommunikativ vaziyatlarda neytral, kinoyaviy, haqoratomuz yoki emotsional ma’no kasb etishi mumkin. Shu sababli MO‘ modellari tomonidan kontekstual tahlilning qo‘llanilishi baholovchi birliklarni aniqlash aniqligini oshiradi hamda ekspertning subyektiv talqin ehtimolini kamaytiradi

M.T. Ribeiro, S. Singh va C. Guestrinlarning ta’kidlashicha, lingvistik ekspertologiyada SI/MO metodlarini qo‘llashda uchta metodik prinsip muhim:

(a) reproduktivlik — bir xil kirish ma’lumoti bir xil parametrlar bilan qayta ishlansa, bir xil natija chiqishi;

(b) verifikatsiya — natijaning nazorat to‘plamlarida tekshirilishi;

(c) izohlanish — model qarori asoslari lingvistik mezonlar bilan ifodalanishi.

Tadqiqotchilar fikriga ko‘ra, bir xil parametrlar asosida qayta ishlangan ma’lumot bir xil natija berishi, model natijalari nazorat to‘plamlarida tekshirilishi hamda qaror qabul qilish jarayoni lingvistik mezonlar orqali izohlanishi zarur Ushbu metodik prinsiplar lingvistik ekspertologiyada sun’iy intellekt va mashinaviy o‘qitish texnologiyalaridan foydalanishda ilmiy ishonchlilikni ta’minlaydi. Ayniqsa, reproduktivlik prinsipi ekspert xulosalarining qayta tekshiriluvchanligini kuchaytirsa, verifikatsiya modeli amaliy sinovdan o‘tkazish orqali natijaning barqarorligini baholash imkonini beradi. Izohlanish talabi esa avtomatlashtirilgan tizimlarning “yopiq mexanizm” sifatida emas, balki lingvistik mezonlarga tayanuvchi tahliliy vosita sifatida qo‘llanishiga xizmat qiladi. Shu sababli zamonaviy lingvistik ekspertizada SI/MO modellari faqat texnik aniqlik bilan emas, balki metodik shaffoflik va huquqiy asoslanganlik bilan ham baholanishi lozim.

Til modellari va sun’iy intelektning rivojlanishi fanga yangi imkoniyatlar bilan bir qatorda turli xatarlarni ham olib keldi. Bizningcha, mashinaviy o‘qitish va sun’iy intellekt texnologiyalarining ilm-fanga kirib kelishi katta hajmdagi ma’lumotlarni tezkor qayta ishlash, murakkab bog‘lanishlarni aniqlash hamda tadqiqot natijalarining aniqligini oshirish imkonini bermoqda. Ayniqsa, inson uchun uzoq vaqt talab qiladigan statistik va korpus tahlillari avtomatlashtirilishi ilmiy tadqiqot samaradorligini sezilarli darajada kuchaytiradi. Shu bilan birga, mazkur texnologiyalar bilan bog‘liq ayrim xavf va cheklovlar ham mavjud. Jumladan, algoritmik xatoliklar, ma’lumotlar bazasidagi noto‘g‘ri yoki biryoqlama namunalar natijasida modelning noobyektiv qaror chiqarishi ehtimoli saqlanib qoladi. Bundan tashqari, SI tomonidan yaratilgan matnlarning inson nutqiga yaqinlashuvi mualliflikni aniqlash, akademik halollik va axborot ishonchliligi bilan bog‘liq yangi muammolarni yuzaga keltirmoqda.

Ribeiro va hammualliflari izohlanuvchi NLP konsepsiyasi doirasida generativ modellarni ekspert faoliyatiga integratsiya qilish masalasini tahlil qilib, quyidagi ishlarni bajarishi mumkinligini ta’kidlaydi:

(a) katta hajmdagi yozishmalarni tematik klasterlash va qisqacha mazmunini chiqarish;

(b) termin va frazeologizmlarni kontekst bo‘yicha izohlash;

(c) savol-javob rejimida dastlabki gipotezalarni shakllantirish;

(d) hujjatlarni standartlashtirilgan shaklda rasmiylashtirishga ko‘maklashish.

Ularning amaliy samaradorligi inson nazorati bilan uyg‘unlashgandagina yuqori bo‘lishini ta’kidlaydi. Mualliflarning fikricha, LLM(katta til model)lar katta hajmdagi matnlarni tizimlashtirish, semantik bog‘lanishlarni aniqlash va dastlabki analitik tavsiyalarni shakllantirishda samarali vosita bo‘lsa-da, model natijalari ekspert tomonidan qo‘shimcha verifikatsiyadan o‘tkazilishi zarur. Bizningcha, “human-in-the-loop” modeli lingvistik ekspertologiya uchun eng maqbul yondashuvlardan biri hisoblanadi. Chunki bunda SI tizimlari tahlil jarayonini tezlashtiradi, ekspert esa yakuniy interpretatsiya va huquqiy baholash ustidan nazoratni saqlab qoladi. Natijada avtomatlashtirish va insoniy ekspertiza o‘rtasida muvozanat yuzaga keladi. Tadqiqotlar shuni ko‘rsatib turibdiki SI/MO samaradorligi ko‘p jihatdan ma’lumot sifatiga bog‘liq.

A.V. Zubovning ta’kidlashicha, lingvistik ekspertologiya uchun yaratiladigan korpuslar oddiy til korpuslaridan farqli ravishda huquqiy ahamiyatga ega janrlarni (ariza, shikoyat, tahdid, reklama matni, ijtimoiy tarmoq posti va kommentlar), shuningdek normativ va no-normativ yozuv shakllarini (slang, xato imlo, transliteratsiya, aralash yozuv) ham qamrab olishi lozim. Tadqiqotda janr va platforma omillarini hisobga olish zarurligi alohida qayd etiladi, chunki ijtimoiy tarmoq tili bilan rasmiy hujjat tili o‘rtasidagi tafovut avtomatik modellarning xatolik profiliga bevosita ta’sir ko‘rsatadi.

Mazkur yondashuv lingvistik ekspertologiyada korpus tanlash masalasi faqat texnik jarayon emas, balki metodologik ishonchlilik omili ekanini ko‘rsatadi. Chunki ekspert tahlil qilayotgan matnning kommunikativ muhiti, janri va yozuv shakli til birliklarining semantik-pragmatik xususiyatlarini o‘zgartirishi mumkin. Masalan, ijtimoiy tarmoqlarda uchraydigan qisqartmalar, transliteratsiya, emotiv birliklar yoki slang unsurlari rasmiy hujjat tilidan keskin farq qiladi. Shu sababli korpus tarkibida turli janr va platformalarga oid materiallarning mavjud bo‘lishi sun’iy intellekt modellari hamda avtomatik tahlil tizimlarining aniqligi va obyektivligini oshirishga xizmat qiladi. Aks holda model faqat “standart” til namunalariga moslashib qolib, real kommunikativ muhitdagi matnlarni noto‘g‘ri talqin qilishi ehtimoli ortadi.

A.A. Usmonov o‘zbek tilida sud-lingvistik ekspertiza uchun yaratiladigan korpuslar umumiy lingvistik korpuslardan funksional jihatdan farq qilishini ta’kidlaydi. Tadqiqotchining fikricha, bunday korpuslarda:

(a) morfologik boylikni to‘liq aks ettiruvchi belgilangan korpuslar (*lemmatizatsiya, POS-tagging, morfologik tahlil*);

(b) lotin-kiril transliteratsiyasi va aralash yozuvni normallashtirish;

(c) sheva va ijtimoiy dialekt elementlarini belgilash;

(d) ekspertologik yorliqlar bilan anotatsiyanlangan dateslar.

Olim umumiy maqsadli korpuslar ekspertiza vazifalarini to‘liq qamrab ololmasligini, shu sababli sud-lingvistik tahlil uchun “vazifa yo‘naltirilgan” annotatsiyalangan datasetlar zarurligini qayd etadi. Ushbu yondashuvdan kelib chiqib aytish mumkinki, o‘zbek tilida lingvistik ekspertologiyani raqamlashtirish jarayonida korpusning hajmidan ko‘ra uning strukturaviy va funksional mosligi muhimroq hisoblanadi. Chunki ekspertiza jarayonida matnning nafaqat grammatik qurilishi, balki kommunikativ maqsadi, emotsional yo‘nalishi va pragmatik ta’siri ham tahlil qilinadi. Ayniqsa, o‘zbek internet segmentida lotin va kirill yozuvining parallel qo‘llanishi, sheva unsurlari, transliteratsiya hamda noformal yozuv shakllari keng uchrashi sababli standart korpuslar real nutq muhitini to‘liq aks ettira olmaydi. Shu bois ekspertologik annotatsiyaga ega maxsus korpuslar yaratish sun’iy intellekt asosidagi tahlil tizimlarining aniqligi, moslashuvchanligi va huquqiy ishonchliligini oshirishga xizmat qiladi.

Korpuslarni O‘zbekiston sharoitida joriy etishning tashkiliy, metodik va etik masalalari xususida A.A. Usmonovning qayd etishicha o‘zbek tilida sun’iy intellekt asosidagi lingvist ekspert tizimlarini rivojlantirish uchun uchta yo‘nalishga taynagan holda amalga oshirilishi lozim:

1) milliy til resurslari (annotatsiyalangan korpuslar, lug‘atlar, termin bazalari);

2) metodik standartlar (vazifa tavsifi, ma’lumot tayyorlash protokoli, modelni test qilish tartibi);

3) kadrlar va institutsional hamkorlik (*ekspertlar, lingvistlar, yuristlar*).

Tadqiqotda ayniqsa, “metodika+texnologiya” integratsiyasi muhimligi ekspert tixzimlarining samaradorligini belgilovchi muhim omil sifatida baholanadi, aks holda, model natijasi ekspert xulosasiga to‘g‘ri singdirilmaydi. Mazkur yondashuvdan kelib chiqib aytish mumkinki, lingvistik ekspert tizimlarining samaradorligi faqat algoritm aniqligi bilan emas, balki milliy til resurslari sifati, metodik standartlarning mavjudligi hamda turli soha mutaxassislarining hamkorligi bilan belgilanadi. Shu jihatdan, sun’iy intelekt

asosidagi ekspert tizimlarini o‘zbek tiliga moslashtirishda kompleks yondashuv muhim metodologik omil hisoblanadi.

Muhokama. Mazkur yondashuvdan kelib chiqib aytish mumkinki, lingvistik ekspert tizimlarining samaradorligi faqat algoritm aniqligi bilan emas, balki milliy til resurslari sifati, metodik standartlarning mavjudligi hamda turli soha mutaxassislarining hamkorligi bilan belgilanadi. Shu jihatdan, SI asosidagi ekspert tizimlarini o‘zbek tiliga moslashtirishda kompleks yondashuv muhim metodologik omil hisoblanadi. Ushbu bobda lingvistik ekspertologiyada matn tahlili, til korpuslari va sun’iy intellekt texnologiyalarining o‘rni keng yoritilgan. Asosiy xulosa shundan iboratki, zamonaviy lingvistik ekspertiza an’anaviy intuitiv tahlildan raqamli va avtomatlashtirilgan metodlarga o‘tmoqda.

Matn tahlili dasturlari, korpus texnologiyalari, ma’lumotlar bazalari hamda SI/MO (sun’iy intellekt va mashina o‘rganish) modellari ekspert xulosalarining aniqligi, tezkorligi va tekshiriluvchanligini oshirishga xizmat qiladi. Bobda OCR, morfologik va sintaktik tahlil, korpus va konkordans tizimlari, stilometriya, plagiatni aniqlash hamda semantik-pragmatik tahlil vositalarining ekspertologik ahamiyati asoslab berilgan. Ayniqsa, korpus texnologiyalari orqali til birliklarining real qo‘llanish kontekstini aniqlash, bahsli birliklarning pragmatik ma’nosini statistik va empirik dalillar bilan tasdiqlash mumkinligi ko‘rsatib o‘tilgan.

Shu bilan birga, ma’lumotlar bazalari ekspert faoliyatining audit izi, versiyalash va qayta tekshiriluvchanlik tamoyillarini ta’minlovchi metodologik asos sifatida baholangan. Sun’iy intellekt va mashina o‘rganish texnologiyalarining lingvistik ekspertologiyaga kirib kelishi esa katta hajmdagi matnlarni avtomatik tahlil qilish, semantik bog‘lanishlarni aniqlash va ekspert ishini optimallashtirish imkonini bermoqda.

Biroq, SI tizimlari inson ekspertini to‘liq almashtirmasligi, balki “human-in-the-loop” modeli asosida ekspert nazorati bilan ishlashi kerakligi ta’kidlanadi. Chunki izohlanuvchanlik, verifikatsiya va reproduktivlik prinsiplari sud-lingvistik ekspertizada hal qiluvchi mezon hisoblanadi. Shuningdek, o‘zbek tilida korpus resurslari va ekspert tizimlarini rivojlantirish zarurligi, ayniqsa kirill-lotin yozuvi, transliteratsiya, internet slengi va multimodal birliklarni qamrab oluvchi milliy korpuslarni yaratish dolzarb vazifa ekani qayd etilgan. Umuman olganda, bob lingvistik ekspertologiyaning kelajagi raqamli texnologiyalar, korpus lingvistikasi va sun’iy intellekt bilan chambarchas bog‘liq ekanini ilmiy jihatdan asoslab beradi.

Xulosa. Tadqiqot natijalari sun’iy intellekt va mashina o‘rganish texnologiyalari lingvistik ekspertologiyaning zamonaviy rivojlanishidagi muhim innovatsion yo‘nalishlardan biri ekanligini ko‘rsatdi. Ushbu texnologiyalar katta hajmdagi lingvistik ma’lumotlarni tezkor qayta ishlash, matnlarning semantik va kontekstual xususiyatlarini aniqlash, lingvistik birliklarni avtomatik klassifikatsiya qilish hamda ekspert xulosalarining tekshiriluvchanligini oshirish imkonini beradi.

Lingvistik ekspertologiyada SI va MO‘ texnologiyalaridan foydalanishning metodik jihatdan muhim shartlari reproduktivlik, verifikatsiya va izohlanuvchanlik tamoyillarini ta’minlashdan iborat. Ushbu tamoyillar avtomatlashtirilgan tahlil natijalarining ilmiy asoslanganligi, qayta tekshirilishi va ekspert xulosasida tushunarli tarzda ifodalanishiga xizmat qiladi.

O‘zbek tilida sun’iy intellekt asosidagi lingvistik ekspert tizimlarini rivojlantirish uchun maxsus ekspertologik korpuslarni yaratish alohida ahamiyatga ega. Bunday korpuslar o‘zbek tilining morfologik xususiyatlari bilan bir qatorda lotin-kirill yozuvi, transliteratsiya, sheva elementlari, internet slengi, qisqartmalar, aralash va noformal yozuv shakllarini ham qamrab olishi kerak.

Sun’iy intellekt va mashina o‘rganish texnologiyalari lingvistik ekspertni to‘liq almashtiruvchi vosita sifatida emas, balki ekspert faoliyatini qo‘llab-quvvatlovchi texnologik vosita sifatida qo‘llanishi lozim. “Human-in-the-loop” modeli doirasida avtomatlashtirilgan tahlilning tezkorligi inson ekspertining kasbiy tajribasi va lingvistik-huquqiy interpretatsiyasi bilan uyg‘unlashtiriladi.

Umuman olganda, lingvistik ekspertologiyaning kelajakdagi rivoji raqamli texnologiyalar, korpus lingvistikasi, sun’iy intellekt va mashina o‘rganish texnologiyalarining o‘zaro integratsiyasi bilan chambarchas bog‘liq. Bunday integratsiya ekspertiza jarayonining aniqligi, tezkorligi, obyektivligi va metodik shaffofligini oshirishga xizmat qiladi.

ADABIYOTLAR:

1. Zubov, A.V. Sud lingvistikasida raqamli matnlarni tahlil qilish: zamonaviy yondashuvlar / A.V. Zubov. — Moskva: Infra-M, 2021. 4-6.b.

2. Ribeiro, M.T. Explainable NLP: a survey of methods and evaluation / M.T. Ribeiro, S. Singh, C. Guestrin // Journal of Artificial Intelligence Research. — 2021. — № 71. 51-55.b.

3. Nemirov, D.A. Raqamli dalillar va sud ekspertizasi metodologiyasi / — Moskva: Yurayt, 2023. — 109-111.b.

4. Joulin, A. Bag of Tricks for Efficient Text Classification / A. Joulin, E. Grave, P. Bojanowski, T. Mikolov // Proceedings of EACL. — 2017. — № 2. — B. 427–431.

5. Usmonov, A.A. O‘zbek tilida NLP va sud-lingvistik ekspertiza uchun korpuslar: metodik tavsiyalar / A.A. Usmonov // O‘zbekiston filologiyasi. — 2024. — № 2. — 83-86.b.

6. Biber, D. Corpus Linguistics: Investigating Language Structure and Use / D. Biber, S. Conrad. — London: Routledge, 2021. — 356 p.

7. Casey, E. (2011). Raqamli dalillar va kompyuter jinoyati: Sud tibbiyoti, kompyuterlar va internet (3-nashr). Academic Press.

8. Coulthard, M., & Johnson, A. (Tahrirchilar). (2010). Sud tilshunosligi bo‘yicha Routledge qo‘llanmasi. Routledge.

9. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Til tushunish uchun ikki tomonlama transformerlarni oldindan o‘qitish. NAACL-HLT 2019, 4171–4186.

10.Egbert, J. Issues in Corpus Representativeness and Balance / J. Egbert, D. Biber // International Journal of Corpus Linguistics. — 2020. — № 25(1). — B. 1–25.