O‘ZBEK TILINI RAQAMLASHTIRISHDA SUN’IY INTELLEKT TEXNOLOGIYALARI: FILOLOGIK MUAMMOLAR VA AMALIY YECHIMLAR

Norqulova Gulhayo Avloberdiyevna
O‘zbekiston Jahon tillari Universiteti

Annotatsiya. *Maqolada o‘zbek tilining raqamli muhitdagi faoliyatini kengaytirish masalasi sun’iy intellekt va zamonaviy kompyuter lingvistikasi imkoniyatlari bilan bog‘liq holda ko‘rib chiqiladi. Tadqiqotning asosiy e’tibori elektron til resurslarini shakllantirish, o‘zbek tili korpusining ilmiy qiymati, avtomatik matn tahlili, imlo nazorati, mashina tarjimasi hamda nutq texnologiyalariga qaratilgan. O‘zbek tilining grammatik tuzilishi, so‘z shakllarining xilma-xilligi va raqamli matnlarning notekisligi AI tizimlari uchun qanday lingvistik vazifalarni yuzaga keltirishi tahlil qilinadi. Shuningdek, filolog, dasturchi va ma’lumotlar mutaxassisining hamkorligiga asoslangan amaliy model taklif etiladi. Tadqiqot natijasida o‘zbek tiliga mo‘ljallangan sun’iy intellekt vositalarining ishonchliligi faqat texnik algoritmga emas, balki sifatli va belgilangan lingvistik ma’lumotlar bazasiga ham bog‘liqligi asoslanadi.*

Kalit so‘zlar : *o‘zbek tili, sun’iy intellekt, raqamli filologiya, kompyuter lingvistikasi, til korpusi, tabiiy tilni qayta ishlash, mashina tarjimasi, avtomatik tahrir, nutq texnologiyalari, lingvistik resurs.*

Аннотация. *В статье анализируются возможности использования технологий искусственного интеллекта в процессе цифровизации узбекского языка, а также связанные с этим филологические проблемы. Рассматриваются вопросы формирования корпусов узбекского языка, автоматической обработки текстов, машинного перевода, автоматического редактирования и развития речевых технологий. Особое внимание уделяется качеству цифровых лингвистических ресурсов, обогащению языковых баз данных и практическим решениям, направленным на расширение применения узбекского языка в современных информационных технологиях.*

Ключевые слова: *узбекский язык, искусственный интеллект, цифровизация, цифровая филология, языковой корпус, машинный перевод, автоматическое редактирование, речевые технологии.*

Abstract*. This article analyzes the possibilities of using artificial intelligence technologies in the digitalization of the Uzbek language and the philological problems associated with this process. It examines the development of Uzbek language corpora, automatic text processing, machine translation, automatic editing, and speech technologies.*

Particular attention is paid to the quality of digital linguistic resources, the enrichment of language databases, and practical solutions aimed at expanding the use of the Uzbek language in modern information technologies.

Keywords: *Uzbek language, artificial intelligence, digitalization, digital philology, language corpus, machine translation, automatic editing, speech technologies.*

KIRISH

Raqamli texnologiyalar taraqqiyoti tilning mavjud bo‘lish shakllarini ham o‘zgartirmoqda. Ilgari til materiallari asosan kitob, qo‘lyozma yoki bosma nashr ko‘rinishida saqlangan bo‘lsa, bugun katta hajmdagi matnlar elektron bazalarda jamlanmoqda. Natijada tilshunoslikning tadqiqot obyekti kengayib, matnlarni avtomatik qayta ishlash, qidirish va statistik tahlil qilish imkoniyatlari paydo bo‘ldi.

Bu holat filologiyani axborot texnologiyalari bilan yaqinlashtirib, raqamli filologiya deb ataluvchi fanlararo yo‘nalishning rivojiga zamin yaratmoqda.

O‘zbek tili uchun mazkur jarayon alohida ahamiyatga ega. 2025-yil 21oktabrdagi rasmiy ma’lumotda o‘zbek tilidagi adabiyot, maqola, hisobot va ilmiy ishlarning muayyan qismi hali yetarli darajada raqamlashtirilmagani, mavjud elektron materiallar esa turli manbalarda tarqoq saqlanayotgani qayd etilgan. Shu munosabat bilan “O‘zbek tili korpusi”ni yaratish vazifasi ilgari surilgan [1]. 2025-yil 22oktabrdagi PF–189-son Farmonda esa 2026-yil yakuniga qadar kamida 15 milliard parametrni o‘z ichiga oladigan “O‘zbek tili korpusi”ni yaratish va uni kamida beshta ochiq katta til modeliga joriy etish vazifasi belgilangan [2].

Mazkur maqolaning maqsadi — o‘zbek tilini raqamlashtirishda sun’iy intellektning imkoniyatlarini filologik nuqtayi nazardan ko‘rsatish, asosiy

muammolarni tasniflash va ularni kamaytirishga xizmat qiladigan amaliy takliflarni ishlab chiqishdir.

TADQIQOT METODOLOGIYASI

Tadqiqotda tavsifiy, qiyosiy-tahliliy va tizimli yondashuvlardan foydalanildi.

Mavzuga doir rasmiy hujjatlar va ilmiy qarashlar o‘rganilib, ularning o‘zbek tili uchun raqamli resurslar yaratish bilan bog‘liq jihatlari umumlashtirildi. Lingvistik muammolarni belgilashda o‘zbek tilining grammatik va leksik xususiyatlari nazariy jihatdan hisobga olindi. Maqolada keltirilgan davlat siyosatiga oid ma’lumotlar rasmiy manbalarga tayangan holda berildi.

3. RAQAMLI FILOLOGIYANING O‘ZBEK TILI UCHUN AHAMIYATI

Raqamli filologiya filologik bilimlarni elektron ma’lumotlar, dasturiy vositalar va hisoblash metodlari bilan birlashtiradi. Uning asosiy afzalligi katta hajmdagi materialni tizimli ravishda kuzatish imkoniyatidir. Masalan, ma’lum bir so‘zning turli janrlardagi chastotasi, yangi terminlarning ommalashuvi yoki muayyan grammatik shaklning qo‘llanishi korpus asosida aniqlanishi mumkin.

O‘zbek tili uchun elektron korpus yaratish faqat matnlarni internetdan yig‘ish bilan tugamaydi. Korpusga kiritilgan materialning manbasi, janri, yaratilgan vaqti, muallifi va boshqa tavsiflari qayd etilishi, matn esa texnik jihatdan tozalanishi kerak.

Bundan tashqari, so‘zlarning grammatik shakllarini avtomatik aniqlash uchun morfologik belgilash tizimi zarur. Bunday resurslar tilshunoslik tadqiqotlari bilan bir qatorda ta’limiy dasturlar, qidiruv tizimlari va AI modellarini ishlab chiqishda ham xizmat qiladi.

4. O‘ZBEK TILI VA SUN’IY INTELLEKT: LINGVISTIK VAZIFALAR

Tabiiy tilni qayta ishlash texnologiyalari matn va nutqni kompyuter yordamida tushunish hamda qayta ishlashga yo‘naltirilgan. Biroq har bir tilning o‘ziga xos tuzilishi model oldiga turli talablarni qo‘yadi. O‘zbek tilida grammatik ma’nolarning qo‘shimchalar orqali ifodalanishi so‘zning ko‘plab shakllarini yuzaga keltiradi.

Shuning uchun AI tizimi so‘zning faqat lug‘aviy ko‘rinishini emas, uning grammatik tarkibi va gapdagi vazifasini ham hisobga olishi kerak.

Yana bir muhim masala — kontekst. Bir xil birlik turli gaplarda turlicha ma’no anglatishi mumkin. Demak, avtomatik tarjima, savol-javob yoki matn tasnifi tizimlarining sifati faqat lug‘at hajmiga emas, kontekstni qayta ishlash qobiliyatiga

ham bog‘liq. Bu jarayonda filologik tahlil natijalari AI modelini o‘qitish va tekshirish uchun zarur bo‘lgan mezonlarni shakllantirishga yordam beradi.

“O‘ZBEK TILI KORPUSI” VA MA’LUMOT SIFATI

Sun’iy intellekt tizimining natijasi ko‘p jihatdan unga berilgan ma’lumotlarning sifatiga bog‘liq. Raqamlashtirilgan matnlarda kodlash xatolari, takroriy materiallar, imloviy nomuvofiqliklar, skanerlangan hujjatlardagi OCR xatolari va turli yozuv shakllari uchrashi mumkin. Shu sababli korpus yaratishning texnik bosqichidan oldin

lingvistik saralash va tozalash jarayoni yo‘lga qo‘yilishi lozim.

2025-yil 10-iyuldagi 425-son qaror bilan 2025–2026-yillarda AI sohasidagi ustuvor loyihalar belgilangan va “O‘zbek tili korpusi” uchun ma’lumotlar to‘plamini shakllantirish bo‘yicha alohida vazifa nazarda tutilgan [3]. Bu hujjat korpusni yaratishning amaliy jarayonida ma’lumot yig‘ish, texnik hujjatlarni tayyorlash va ochiq modellar asosida milliy modelni sinovdan o‘tkazish kabi bosqichlarni ko‘rsatadi.

Korpusning ilmiy qiymatini oshirish uchun unda badiiy adabiyot, ilmiy uslub, ommaviy axborot vositalari, rasmiy hujjatlar, ta’limiy matnlar va zamonaviy internet diskursining turli namunalarini muvozanatli tarzda jamlash maqsadga muvofiq.

Bunda mualliflik huquqi va ma’lumotlardan foydalanish shartlari ham alohida hisobga olinishi kerak.

AVTOMATIK TAHRIR VA MASHINA TARJIMASI

O‘zbek tilidagi avtomatik tahrir vositalari foydalanuvchining imlo va ayrim grammatik xatolarini aniqlash, ehtimoliy to‘g‘ri variantni taklif qilish hamda matn sifatini oshirishga xizmat qilishi mumkin. Bunday tizim uchun zamonaviy imlo lug‘ati, so‘z shakllari ro‘yxati, grammatik qoidalar va real matnlardan olingan xatoto‘g‘ri juftliklar kerak bo‘ladi. Shu sababli avtomatik tahrirni yaratish filologik lug‘atshunoslik bilan bevosita bog‘liq.

Mashina tarjimasida ham o‘zbek tilining morfologiyasi va gap qurilishi muhim rol o‘ynaydi. So‘zma-so‘z moslik har doim ham tabiiy tarjimani ta’minlamaydi. Iboralar, ko‘p ma’noli birliklar, milliy-madaniy realiyalar va uslubiy farqlar kontekstual tahlilni talab qiladi. Shuning uchun avtomatik tarjima natijasi, ayniqsa ilmiy va rasmiy matnlarda, mutaxassis tomonidan tekshirilishi kerak.

NUTQ TEXNOLOGIYALARI VA AMALIY QO‘LLANISH

Nutqni matnga aylantirish va matnni ovozli shaklda o‘qish texnologiyalari o‘zbek tilining raqamli xizmatlardagi ishtirokini kengaytirishi mumkin. Bunday vositalar ta’lim platformalari, elektron xizmatlar, qidiruv tizimlari va turli axborot

xizmatlarida qo‘llanishi ehtimolga ega. Ularni takomillashtirish uchun turli talaffuz, nutq tezligi va akustik sharoitlarni qamrab oladigan ovozli ma’lumotlar bazasi kerak.

Bu yo‘nalishda fonetika va dialektologiya bo‘yicha an’anaviy tadqiqotlar ham amaliy ahamiyat kasb etadi. Chunki turli hududiy talaffuz xususiyatlari hisobga olinmasa, nutqni tanish tizimining ayrim foydalanuvchilarda aniqligi pasayishi mumkin. Demak, raqamli filologiya mavjud filologik bilimlarni texnologik mahsulotga aylantirish imkonini beradi.

MUAMMOLAR VA AMALIY YECHIMLAR

O‘zbek tilini AI texnologiyalariga moslashtirishda bir necha asosiy muammo ajralib turadi. Birinchisi, raqamli matnlarning tarqoqligi va formatlar xilma-xilligi.

Ikkinchisi, lingvistik jihatdan belgilangan ochiq ma’lumotlar hajmining yetarli darajada keng emasligi. Uchinchisi, terminologik va imloviy variantlarni yagona mezon asosida muvofiqlashtirish zarurati. To‘rtinchisi, avtomatik tizim natijalarini baholash uchun o‘zbek tiliga mos mezonlar va sinov to‘plamlarini ishlab chiqish masalasi.

Ushbu muammolarni kamaytirish uchun quyidagi amaliy model taklif etiladi: avvalo, raqamlashtiriladigan manbalar yagona texnik standartga keltiriladi; keyin filologlar tomonidan matnlar saralanib, lingvistik belgilar bilan ta’minlanadi; undan so‘ng ma’lumotlar AI modellarini o‘qitish va sinash uchun alohida to‘plamlarga ajratiladi. Yakuniy bosqichda model natijalari avtomatik ko‘rsatkichlar bilan birga mutaxassis bahosi orqali tekshiriladi.

Bunda filolog mutaxassisning roli kamaymaydi, aksincha o‘zgaradi. U ma’lumotlarni tanlash, til me’yorlarini belgilash, terminlarni izohlash, xatolarni tasniflash va model natijasini lingvistik nuqtayi nazardan baholashda ishtirok etadi.

Shu sababli oliy ta’limda kompyuter lingvistikasi, korpus lingvistikasi va AI asoslarini filologik fanlar bilan integratsiyalash foydali bo‘lishi mumkin.

XULOSA

O‘zbek tilini raqamlashtirish bugungi filologiyaning faqat texnik emas, balki ilmiy-metodologik masalalaridan biridir. Sun’iy intellektning matn va nutq bilan ishlash imkoniyatlari o‘zbek tiliga oid katta hajmdagi materialni qayta ishlash, yangi lug‘atlar va korpuslar yaratish, avtomatik tahrir va tarjima vositalarini takomillashtirish uchun sharoit yaratadi.

Tahlil shuni ko‘rsatadiki, texnologik imkoniyatning o‘zi yetarli emas. Sifatli AI vositasi uchun tartibli elektron korpus, aniq lingvistik belgilash, ishonchli lug‘atlar va mutaxassis ekspertizasi zarur. O‘zbekistonda “O‘zbek tili korpusi”ni yaratish va

AI asosidagi loyihalarni amalga oshirish bo‘yicha 2025–2026-yillarda belgilangan vazifalar mazkur yo‘nalishning amaliy ahamiyatini ko‘rsatadi [1–3].

Shu bois o‘zbek tili kelajakdagi raqamli muhitda to‘laqonli ishlashi uchun filologiya va axborot texnologiyalari o‘rtasidagi hamkorlikni kuchaytirish, ochiq lingvistik resurslarni ko‘paytirish va AI natijalarini doimiy filologik nazoratdan o‘tkazish muhimdir.

FOYDALANILGAN ADABIYOTLAR

6. O‘zbekiston Respublikasi Raqamli texnologiyalar vazirligi. “Bugun o‘zbek tiliga davlat tili maqomi berilganiga 36 yil bo‘ldi”. 21.10.2025.

7. O‘zbekiston Respublikasi Raqamli texnologiyalar vazirligi. “Sun’iy intellekt yangi bosqichda: Prezident Farmoni qabul qilindi”. 27.10.2025.

8. O‘zbekiston Respublikasi Vazirlar Mahkamasining 2025-yil 10-iyuldagi

425-son qarori. “2025–2026-yillarda sun’iy intellekt texnologiyalari sohasida ustuvor loyihalarni amalga oshirish chora-tadbirlari to‘g‘risida”.

9. Jurafsky, D., Martin, J. H. Speech and Language Processing. 3rd ed. Draft.

10. Manning, C. D., Raghavan, P., Schütze, H. Introduction to Information Retrieval. Cambridge University Press, 2008.