

O‘ZBEK TILI UCHUN SUN’IY INTELLEKT YARATISHDAGI HOZIRGI JARAYONLAR TAHLILI

Murotjonov Abduraxmon Abdullatif o‘g‘li

Qo‘qon universiteti magistranti

E-mail: abduraxmonabdullatif@gmail.com

Annotatsiya. Ushbu maqolada o‘zbek tili uchun sun‘iy intellekt texnologiyalarini yaratish jarayonidagi hozirgi tendensiyalar va ilmiy-amaliy ishlanmalar tahlil qilinadi. So‘nggi yillarda tabiiy tilni qayta ishlash (Natural Language Processing – NLP) sohasida jadal rivojlanish kuzatilmoqda va bu jarayon o‘zbek tili kabi kam resursli tillar uchun ham yangi imkoniyatlarni ochmoqda. Tadqiqotda o‘zbek tiliga moslashtirilgan sun‘iy intellekt modellarini yaratishdagi asosiy muammolar, jumladan lingvistik korpuslarning yetishmasligi, morfologik murakkablik va standartlashtirilgan ma‘lumotlar bazalarining kamligi ko‘rib chiqiladi. Shuningdek, mavjud texnologiyalar, jumladan mashinaviy tarjima, nutqni tanish, matnni avtomatik tahlil qilish va chatbot tizimlarini rivojlantirishdagi yutuqlar ham tahlil qilinadi. Tadqiqot natijalari o‘zbek tilida ishlovchi sun‘iy intellekt tizimlarini takomillashtirish uchun ilmiy asos yaratadi hamda kelajakdagi tadqiqot yo‘nalishlarini aniqlashga yordam beradi.

Kalit so‘zlar: sun‘iy intellekt, o‘zbek tili, tabiiy tilni qayta ishlash, NLP, lingvistik korpus, mashinaviy tarjima, til texnologiyalari.

Abstract. This article analyzes the current processes and developments in creating artificial intelligence technologies for the Uzbek language. In recent years, significant progress has been observed in the field of Natural Language Processing (NLP), opening new opportunities for low-resource languages such as Uzbek. The study examines the main challenges in developing AI models adapted to the Uzbek language, including the lack of large linguistic corpora, morphological complexity, and insufficient standardized datasets. In addition, the research reviews current technological achievements in machine translation, speech recognition, automatic text processing, and chatbot systems designed for the Uzbek language. The findings highlight the importance of developing language resources and computational tools to enhance AI systems working with Uzbek. The study also outlines potential directions for future research in the field of Uzbek language technologies.

Keywords: *artificial intelligence, Uzbek language, natural language processing, NLP, linguistic corpus, machine translation, language technologies.*

Аннотация. *В данной статье анализируются современные процессы и тенденции разработки технологий искусственного интеллекта для узбекского языка. В последние годы наблюдается стремительное развитие области обработки естественного языка (Natural Language Processing – NLP), что открывает новые возможности для языков с ограниченными ресурсами, таких как узбекский. В исследовании рассматриваются основные проблемы создания моделей искусственного интеллекта, адаптированных к узбекскому языку, включая недостаток крупных лингвистических корпусов, морфологическую сложность языка и ограниченность стандартизированных наборов данных. Кроме того, анализируются достижения в области машинного перевода, распознавания речи, автоматического анализа текста и разработки чат-ботов для узбекского языка. Результаты исследования подчеркивают необходимость дальнейшего развития языковых ресурсов и технологий для повышения эффективности систем искусственного интеллекта, работающих с узбекским языком.*

Ключевые слова: *искусственный интеллект, узбекский язык, обработка естественного языка, NLP, лингвистический корпус, машинный перевод, языковые технологии.*

Kirish

Sun'iy intellekt (SI) texnologiyalarining jadal rivojlanishi til texnologiyalari va tabiiy tillarni qayta ishlash (NLP) rivojlanishiga sezilarli ta'sir ko'rsatdi. Mashina tarjimasi, nutqni aniqlash, aqlli dialog tizimlari va avtomatik matnni qayta ishlash kabi SIga asoslangan tizimlar hozirda ta'lim, raqamli aloqa va axborotni boshqarishda keng qo'llanilmoqda. Ushbu texnologiyalar katta lingvistik ma'lumotlar to'plamlariga, murakkab hisoblash modellariga va inson tilini tahlil qilish va yaratishga qodir mashinani o'rganish algoritmlariga tayanadi. Ingliz va xitoy kabi keng tarqalgan tillarda sezilarli yutuqlarga erishilgan bo'lsa-da, raqamli resurslari cheklangan ko'plab tillar hali ham SIga asoslangan til texnologiyalarini rivojlantirishda jiddiy qiyinchiliklarga duch kelmoqda. Bunday tillardan biri boy morfologiya va agglyutinativ grammatik tuzilishga ega turkiy til bo'lgan o'zbek tilidir.

O‘zbek tili uchun sun‘iy intellektni rivojlantirish raqamli transformatsiya va O‘zbekistonda axborot texnologiyalarining kengayishi sharoitida tobora muhim ahamiyat kasb etmoqda. So‘nggi yillarda bir qancha o‘zbek tilshunoslari va tadqiqotchilari o‘zbek tiliga oid hisoblash lingvistikasi va til texnologiyalarini o‘rganishga hissa qo‘shdilar. Masalan, Sh. Safarov, N. Mahmudov va A. Abduazizov o‘zbek tilshunosligi, diskurs tahlili va til tuzilishi bo‘yicha muhim tadqiqotlar olib borishdi, bu esa zamonaviy til texnologiyalari uchun muhim nazariy asos yaratadi. Ularning o‘zbek tilining leksik semantikasi, pragmatik xususiyatlari va strukturaviy xususiyatlari bo‘yicha tadqiqotlari sun‘iy intellekt tizimlari uchun zarur bo‘lgan lingvistik resurslarni rivojlantirishga hissa qo‘shadi.

Ushbu hissalariga qaramay, o‘zbek tili uchun sun‘iy intellekt ilovalarini ishlab chiqish yuqori resursli tillarga nisbatan cheklanganligicha qolmoqda. Asosiy muammolardan biri zamonaviy mashina o‘rganish modellarini o‘qitish uchun zarur bo‘lgan katta izohli korpuslar va standartlashtirilgan lingvistik ma‘lumotlar to‘plamlarining yo‘qligidir (Jurafsky & Martin, 2023). Bundan tashqari, o‘zbek tilining agglyutinativ tabiati hisoblashni qayta ishlash uchun, ayniqsa morfologik tahlil, tokenizatsiya va sintaktik tahlil kabi vazifalarda qo‘shimcha murakkablik yaratadi. Shuning uchun o‘zbek tili uchun sun‘iy intellekt texnologiyalarini ishlab chiqishda ishtirok etadigan mavjud jarayonlarni tahlil qilish juda muhimdir. Bunday tahlil tadqiqotchilarga mavjud texnologik bo‘shliqlarni aniqlash, mavjud yutuqlarni baholash va o‘zbek tilini raqamli til texnologiyalarining global ekotizimiga integratsiya qilish bo‘yicha kelajakdagi tadqiqot yo‘nalishlarini belgilash imkonini beradi.

Natijalar

O‘zbek tili uchun sun‘iy intellektning hozirgi rivojlanishini tahlil qilish lingvistik ma‘lumotlar to‘plamlari, raqamli korpuslar va hisoblash vositalarini yaratishda sezilarli yutuqlarni ko‘rsatmoqda. Biroq, miqdoriy ko‘rsatkichlar shuni ko‘rsatadiki, mavjud resurslar ko‘lami yuqori resursli tillar uchun mavjud resurslarga qaraganda ancha kichikroq. Sun‘iy intellektni rivojlantirishning eng muhim tarkibiy qismlaridan biri bu mashinani o‘rganish modellarini o‘qitish uchun ishlatiladigan matnli va nutqiy ma‘lumotlarning mavjudligidir. Ommaviy foydalanish mumkin bo‘lgan lingvistik resurslarni ko‘rib chiqish shuni ko‘rsatadiki, so‘nggi o‘n yillikda o‘zbek tili ma‘lumotlar to‘plamlari asta-sekin o‘sib bordi. Masalan, onlayn ensiklopedik resurslar, raqamli matn to‘plamlari va ochiq ko‘p tili



ma'lumotlar to'plamlari tabiiy tilni qayta ishlash tizimlari uchun o'quv ma'lumotlarining muhim manbalariga aylandi.

Vikipediyaning o'zbekcha nashri hisoblash tadqiqotlari uchun eng katta erkin kirish mumkin bo'lgan matnli resurslardan biriga aylandi. So'nggi statistik ma'lumotlarga ko'ra, unda 140 000 dan ortiq maqolalar mavjud bo'lib, ular til modellarini o'qitish uchun mos bo'lgan millionlab tuzilgan matnli ma'lumotlar tokenlarini taqdim etadi. Bundan tashqari, Common Crawl va OPUS parallel korpuslari kabi ko'p tilli ma'lumotlar to'plamlari mashina tarjimasini tajribalari uchun tez-tez ishlatiladigan o'zbek matnini o'z ichiga oladi. Nutq texnologiyalari ham rivojlanmoqda, ammo mavjud resurslar cheklanganligicha qolmoqda. Avtomatik nutqni aniqlash tadqiqotlari uchun bir nechta eksperimental o'zbek nutq ma'lumotlar to'plamlari yaratildi. Ushbu ma'lumotlar to'plamlarida tadqiqotchilarga nutqni aniqlash tizimlarini o'rgatish imkonini beruvchi transkripsiyalar bilan birlashtirilgan yozib olingan audio namunalari mavjud.

O'zbek tili uchun sun'iy intellekt tadqiqotlarida qo'llanilayotgan asosiy raqamli resurslar

| Resurs / Ma'lumotlar to'plami | Ma'lumot turi | Taxminiy hajmi | Sun'iy intellekt dagi asosiy qo'llanishi |
|---|---------------------------------|--------------------------------|---|
| O'zbekcha Vikipediya | Matn korpusi | ~140 000 ta maqola | Til modellarini o'qitish, matn tahlili |
| Common Crawl (o'zbek tilidagi qismi) | Veb matn korpusi | ~120–150 million token | NLP modellarini o'qitish |
| OPUS parallel korpusi (Uz–En, Uz– Ru) | Parallel matnlar | ~1–2 million gap juftligi | Mashinaviy tarjima |
| O'zbek nutq ma'lumotlar to'plamlari | Audio + transkripsiya | ~80–120 soat nutq | Nutqni avtomatik tanish |
| Elektron o'zbek lug'atlari | Leksik ma'lumotlar bazasi | ~60 000–80 000 so'z birligi | Morfologik tahlil, leksik ishlov berish |

Quyidagi jadvalda o‘zbek tilidagi sun’iy intellekt tadqiqotlarida hozirda qo‘llaniladigan bir nechta asosiy raqamli resurslar umumlashtirilgan. Jadvalda o‘zbek tilidagi lingvistik resurslar soni asta-sekin o‘shib borayotgan bo‘lsa-da, ushbu ma’lumotlar to‘plamlarining hajmi keng tarqalgan tillar uchun ishlatiladigan ma’lumotlar to‘plamlariga nisbatan nisbatan kichikligicha qolmoqda. Masalan, ingliz NLP modellari ko‘pincha milliardlab tokenlarni o‘z ichiga olgan ma’lumotlar to‘plamlarida o‘qitiladi, o‘zbek resurslari esa odatda millionlab ma’lumotlarni o‘z ichiga oladi. Ushbu cheklovlarga qaramay, natijalar tadqiqot faoliyati va resurslarni rivojlantirishda barqaror o‘shishni ko‘rsatadi. Raqamli korpuslar, ochiq ma’lumotlar to‘plamlari va hisoblash vositalarining paydo bo‘lishi o‘zbek tilidagi sun’iy intellekt uchun zarur bo‘lgan texnologik infratuzilma asta-sekin yaratilayotganidan dalolat beradi. Kattaroq izohli korpuslar va nutq ma’lumotlar to‘plamlarini doimiy ravishda rivojlantirish o‘zbek tilini qayta ishlash uchun mo‘ljallangan AI tizimlarining ishlashi va ishonchligini oshirishda muhim rol o‘ynaydi.

Munozara

Ushbu tadqiqot natijalari o‘zbek tili uchun sun’iy intellekt texnologiyalarini rivojlantirishdagi taraqqiyot va mavjud cheklovlarni ta’kidlaydi. So‘nggi yillarda sezilarli yaxshilanishlar amalga oshirilgan bo‘lsa-da, o‘zbek tilini qayta ishlashning umumiy texnologik ekotizimi keng o‘rganilayotgan tillarga nisbatan dastlabki bosqichda qolmoqda. Natijalar shuni ko‘rsatadiki, asosiy qiyinchiliklar keng ko‘lamli lingvistik ma’lumotlar to‘plamlarining cheklanganligi, tilning strukturaviy murakkabligi va o‘zbek tili uchun maxsus ishlab chiqilgan hisoblash vositalarining nisbatan kamligi bilan bog‘liq.

Eng muhim kuzatuvlardan biri til resurslarining mavjudligi va sun’iy intellekt tizimlarining ishlashi o‘rtasidagi kuchli bog‘liqlikdir. Zamonaviy NLP modellari lingvistik naqshlarni samarali o‘rganish uchun juda ko‘p miqdordagi izohli matn va nutq ma’lumotlarini talab qiladi. O‘zbek tili misolida mavjud korpuslar va ma’lumotlar to‘plamlari muhim boshlang‘ich nuqtani ta’minlaydi, ammo ular hali ham ingliz yoki boshqa asosiy tillar uchun ishlab chiqilganlarga o‘xshash keng ko‘lamli til modellarini o‘qitish uchun yetarli emas. Bu holat raqamli korpuslarni kengaytirish, izohli ma’lumotlar to‘plamlarini yaratish va tadqiqotchilar va ishlab chiquvchilar tomonidan foydalanish mumkin bo‘lgan ochiq lingvistik resurslarni rivojlantirish muhimligini tasdiqlaydi. Yana bir muhim masala o‘zbek tilining o‘zining lingvistik tuzilishiga tegishli. Agglutinativ til sifatida o‘zbek tili boy



morfologik xilma-xillikni namoyish etadi, bu yerda bir nechta grammatik ma'nolar qo'shimchalar orqali ifodalanishi mumkin. Bu xususiyat tilning ifodali salohiyatini boyitsa-da, hisoblashni qayta ishlashda ham qiyinchiliklar tug'diradi. Morfologik tahlil, sintaktik tahlil va semantik talqin kabi vazifalar murakkab so'z birikmalarini qayta ishlashga qodir ixtisoslashgan algoritmlarni talab qiladi. Natijada, lingvistik tadqiqotlar hisoblash modellashtirishga asos bo'la oladigan til tuzilishining batafsil tavsiflarini taqdim etish orqali texnologik rivojlanishni qo'llab-quvvatlashda muhim rol o'ynaydi.

Natijalar, shuningdek, bu sohada fanlararo hamkorlik tobora muhim ahamiyat kasb etayotganini ko'rsatadi. O'zbek tili texnologiyalarini samarali rivojlantirish tilshunoslar, kompyuter olimlari va ma'lumotlar muhandislari o'rtasidagi hamkorlikni talab qiladi. Tilshunoslar til tuzilishi haqida nazariy bilimlar bilan o'rtoqlashadilar, hisoblash mutaxassislari esa lingvistik ma'lumotlarni samarali qayta ishlashga qodir algoritmlar va mashinani o'rganish modellarini loyihalashtiradilar. Ushbu hamkorlik yondashuvi o'zbek tili ilovalari uchun ishonchli AI tizimlarini yaratishni sezilarli darajada tezlashtirishi mumkin. Va nihoyat, tadqiqot shuni ko'rsatadiki, o'zbek tilidagi sun'iy intellekt texnologiyalarining kelajakdagi rivojlanishi asosan raqamli til infratuzilmasiga strategik investitsiyalarga bog'liq. Ommaviy foydalanish mumkin bo'lgan ma'lumotlar to'plamlarini kengaytirish, akademik tadqiqotlarni rag'batlantirish va o'zbek tili resurslarini xalqaro NLP platformalariga integratsiya qilish AI vositalarining sifati va ulardan foydalanish imkoniyatini sezilarli darajada yaxshilashi mumkin. Ushbu yo'nalishlardagi sa'y-harakatlarni davom ettirish orqali o'zbek tili til texnologiyalari global manzarasida yanada to'liqroq namoyon bo'lishi mumkin.

Xulosa

Ushbu tadqiqotda mavjud texnologik resurslar, lingvistik xususiyatlar va davom etayotgan tadqiqot ishlarini tahlil qilish orqali o'zbek tili uchun sun'iy intellektni rivojlantirishning hozirgi holati o'rganildi. Tadqiqot natijalari shuni ko'rsatadiki, o'zbek tili uchun sun'iy intellekt texnologiyalari hali ham rivojlanayotgan bo'lsa-da, so'nggi yillarda, ayniqsa raqamli korpuslar, elektron lug'atlar va eksperimental tilni qayta ishlash vositalarini yaratishda sezilarli yutuqlarga erishildi. Ushbu o'zgarishlar tadqiqotchilar va muassasalar o'rtasida tabiiy tilni qayta ishlashning kengroq sohasida o'zbek tili texnologiyalarini



kengaytirishga qiziqish ortib borayotganini ko'rsatadi. Shu bilan birga, tadqiqot o'zbek tili uchun mo'ljallangan sun'iy intellekt tizimlarining samaradorligini cheklab kelayotgan bir qator muammolarni ta'kidlaydi. Eng muhim cheklov zamonaviy mashina o'rganish modellarini o'qitish uchun zarur bo'lgan keng ko'lamli izohli ma'lumotlar to'plamlarining yetishmasligidir. Bundan tashqari, o'zbek tilining agglyutinativ tabiati, ayniqsa morfologik tahlil, sintaktik tahlil va til modellashtirish kabi sohalarda o'ziga xos hisoblash muammolarini keltirib chiqaradi. Ushbu muammolarni hal qilish ham lingvistik tajribani, ham ilg'or hisoblash usullarini talab qiladi.

Tadqiqot shuningdek, o'zbek tili texnologiyalarini rivojlantirishda fanlararo hamkorlikning muhimligini ta'kidlaydi. Tilshunoslar, kompyuter olimlari va texnologiya ishlab chiquvchilari o'rtasidagi hamkorlik aniqroq til modellari va samarali hisoblash vositalarini yaratishga hissa qo'shishi mumkin. Bundan tashqari, ommaga ochiq til resurslarini kengaytirish va ochiq ilmiy tadqiqotlarni rag'batlantirish bu sohada texnologik taraqqiyotni tezlashtirishda muhim rol o'ynaydi. Umuman olganda, o'zbek tili uchun sun'iy intellekt texnologiyalarini ishlab chiqish tilni global raqamli muhitga integratsiyalash yo'lidagi muhim qadamdir. Davom etayotgan tadqiqotlar, resurslarni ishlab chiqish va texnologik innovatsiyalar sun'iy intellekt tizimlarining ish faoliyatini yaxshilashga va kelajakda o'zbek tilining kengroq raqamli tasvirini ta'minlashga hissa qo'shadi.

Adabiyotlar

1. Abduazizov, A. (2010). Tilshunoslik nazariyasi asoslari. Toshkent: Sharq nashriyoti.
2. Bird, S., Klein, E., & Loper, E. (2009). Python yordamida tabiiy tilni qayta ishlash. O'Reilly Media.
3. Jurafsky, D., & Martin, J. H. (2023). Nutq va tilni qayta ishlash (3-nashr). Stenford universiteti.
4. Mahmudov, N. (2012). Til va nutq masalalari. Toshkent: O'zbekiston Milliy Ensiklopediyasi.
5. Safarov, Sh. (2008). Pragmalingvistika. Toshkent: Fan Publishing.

