SCIENCE HUB REVIEW
2026
Volume 1
Issue 5
www.sciencehubreview.uz
ILMIY ELEKTRON JURNAL

SUN’IY INTELLEKT TIZIMLARIDA O‘ZBEK TILINI QO‘LLASHNING NAZARIY VA AMALIY MUAMMOLARI

Soatova Sitora Yunus qizi

Osiyo texnologiyalar universiteti, magistrant

soatovasitora3@gmail.com

Annotatsiya

Ushbu maqolada zamonaviy sun’iy intellekt (SI) va tabiiy tilni qayta ishlash (NLP) texnologiyalarida o‘zbek tilining tutgan o‘rni, kam resursli (low-resource) til sifatida duch kelayotgan lingvistik, texnik va grafik-imloviy muammolari chuqur tahlil qilinadi. Shuningdek, agglyutinativ morfologiya, raqamli korpuslar tanqisligi va alifbo xilma-xilligi bilan bog‘liq to‘siqlarni bartaraf etishning ilmiy-amaliy yechimlari hamda istiqbolli yo‘nalishlari taklif etiladi.

Kalit so‘zlar

sun’iy intellekt (SI), tabiiy tilni qayta ishlash (NLP), yirik til modellari (LLM), agglyutinativ til, til korpusi, tokenizatsiya, imlo standartlari.

Kirish

Jadal rivojlanib borayotgan raqamli asrda sun’iy intellekt (SI) texnologiyalari jamiyatning barcha jabhalariga chuqur kirib bormoqda. Tabiiy tilni qayta ishlash (NLP – Natural Language Processing) va yirik til modellari (LLM – Large Language Models) sohasi bugungi kunda sun’iy intellektning eng tez rivojlanayotgan yo‘nalishlaridan biridir. Dunyodagi ko‘plab asosiy tillar uchun bu borada ulkan muvaffaqiyatlarga erishilgan bo‘lsa-da, o‘zbek tili kabi kam resursli (low-resource) tillar uchun SI muhitida qator fundamental va amaliy muammolar saqlanib qolmoqda.

Ushbu maqolada o‘zbek tilining sun’iy intellekt tizimlariga integratsiya qilinishidagi asosiy muammolar, ularning kelib chiqish sabablari va ularni hal etish yo‘llari ilmiy jihatdan tahlil qilinadi.

1. O‘zbek tilining lingvistik xususiyatlari va ularning SI uchun murakkabligi

O‘zbek tili turkiy tillar oilasiga mansub bo‘lib, o‘zining boy morfologik va sintaktik tuzilishiga ega. Bu xususiyatlar an’anaviy neyron tarmoqlar va zamonaviy til modellari uchun bir qator o‘ziga xos qiyinchiliklarni tug‘diradi:

Agglyutinativ (qo‘shimchali) tuzilish: O‘zbek tilida so‘zlar o‘zakka qo‘shimchalarning zanjirsimon ketma-ketlikda qo‘shilishi orqali yasaladi. Bitta fe’l yoki ot o‘zagi asosida o‘nlab, hatto yuzlab grammatik shakllarni hosil qilish mumkin (masalan: «ishla-t-il-ma-gan-lik-lar-i-dan»). Bu esa lug‘at hajmining (vocabulary size) haddan tashqari kengayib ketishiga va tokenizatsiya (tokenization — matnni kichik bo‘laklarga ajratish) jarayonida so‘z ma’nosining parchalanib ketishiga olib keladi.

Erkin so‘z tartibi (SOV – Subject-Object-Verb): Ingliz tilidagi qat’iy SVO tartibidan farqli o‘laroq, o‘zbek tilida gap bo‘laklarining o‘rni nisbatan erkin bo‘lib, mantiqiy urg‘u va uslubiy kontekstga qarab o‘zgarishi mumkin. Bu sintaktik tahlilni (syntactic parsing) va mashinali tarjima algoritmlarini murakkablashtiradi.

Ko‘p ma’nolilik (polisemiya) va omonimiya: Shakldosh so‘zlarning (masalan, «ot», «yoz», «ko‘k», «qo‘y») matn kontekstiga qarab mutlaqo boshqa grammatik turkum va ma’no kasb etishi modellar tomonidan noto‘g‘ri talqin qilinish ehtimolini oshiradi.

2. Asosiy ilmiy va texnik muammolar

2.1. Korpuslar va ma’lumotlar bazasining yetishmasligi (Low-Resource Challenge)

Hozirgi kunda zamonaviy til modellarini o‘qitish uchun milliardlab sifatli va tahrirlangan matnli ma’lumotlar (tokenlar) talab etiladi. O‘zbek tilidagi internet segmenti, raqamli kutubxonalar va ochiq ma’lumotlar bazalari ingliz, rus yoki xitoy tillari bilan solishtirganda juda kichik ulushni (global veb-kontentning 0,1 foizidan kamrog‘ini) tashkil etadi. Mavjud matnlarning katta qismi ham zamonaviy imlo qoidalariga to‘liq mos kelmaydi yoki past sifatli mashinaviy tarjimalardan iborat bo‘lib, bu neyron tarmoqlarning «gallyutsinatsiya» va grammatik xatolarga yo‘l qo‘yishiga sabab bo‘ladi.

2.2. Grafik va imlo xilma-xilligi muammosi

O‘zbek tili tarixiy va ijtimoiy-siyosiy omillar tufayli bir nechta yozuv tizimlarida faoliyat yuritib kelmoqda, bu esa ma’lumotlarni yagona raqamli standartga keltirishda jiddiy to‘siq yaratadi:

Kirill alifbosi: Hozirgi kunda ham ommaviy axborot vositalari, adabiy manbalar va kundalik muloqotda keng qo‘llaniladi.

Lotin alifbosi: Rasmiy hujjatlar, ta’lim tizimi va davlat boshqaruvida asosiy yozuv sifatida ishlatiladi.

Arab yozuvi: Tarixiy qo‘lyozma manbalar va xorijdagi o‘zbek diasporasi vakillari tomonidan qo‘llaniladi.

Ayni paytda lotin yozuviga asoslangan o‘zbek alifbosida ham yagona texnik standartning to‘liq qaror topmaganligi (O', G' harflari o‘rnida apostrofning turli Unicode belgilari — ', ‘, ’, ‘ ishlatilishi, shuningdek «sh», «ch» digraflarining alohida harf sifatida tanilmasligi) tokenizatorlar uchun bitta so‘zni o‘nlab turli ko‘rinishlarda qabul qilishga olib kelmoqda.

2.3. Semantik tahlil va milliy kontekstni anglashdagi cheklovlar

Aksariyat global til modellari o‘zbek tilini boshqa vositachi tillar (asosan rus yoki ingliz tili) orqali o‘rganadi. Bu holat o‘zbek tiliga xos bo‘lgan milliy-madaniy kontekstlar, frazeologik iboralar, xalq maqollari va hududiy shevalarni noto‘g‘ri tushunishga yoki ma’noni yo‘qotgan holda so‘zma-so‘z tarjima qilib yuborishga olib keladi.

1-jadval. O‘zbek tilining SI tizimlaridagi asosiy muammolari va yechim yo‘nalishlari

Muammo yo‘nalishi	Asosiy sabab va oqibatlari	Taklif etilayotgan yechim
Morfologik murakkablik	Agglyutinativ tuzilish tufayli so‘z shakllarining ko‘pligi va tokenizatsiya samarasizligi.	O‘zbek tili morfologiyasiga moslashtirilgan maxsus subword tokenizatorlar yaratish.
Ma’lumotlar tanqisligi	Ochiq va tahrirlangan milliy matn korpuslari hajmining kamligi.	Davlat va ilmiy markazlar ishtirokida yirik Milliy til korpusini shakllantirish.
Grafik va imlo farqlari	Kirill-lotin parallelizmi va tutuq belgilarining Unicode xilma-xilligi.	Matnlarni avtomatik normallashtiruvchi va konvertatsiya qiluvchi dasturiy modullar joriy etish.

Muammo yo‘nalishi	Asosiy sabab va oqibatlari	Taklif etilayotgan yechim
Semantik va sheva farqlari	Vositachi tillar orqali o‘qitish oqibatida milliy kontekst va shevalarning yo‘qolishi.	Mahalliy LLM modellarini to‘g‘ridan-to‘g‘ri o‘zbekcha ma’lumotlarda fine-tuning qilish.

3. Muammolarni hal etish yo‘llari va istiqbollar

O‘zbek tilini sun’iy intellekt tizimlarida to‘laqonli joriy etish va raqamli mustaqillikni ta’minlash uchun quyidagi tizimli chora-tadbirlarni amalga oshirish zarur:

Katta hajmli o‘zbek tilidagi korpuslarni yaratish: Davlat va ilmiy muassasalar ko‘magida ilmiy maqolalar, badiiy adabiyotlar, qonun hujjatlari, ensiklopediyalar va ishonchli ommaviy axborot vositalari materiallarini o‘z ichiga olgan yirik, toza va morfologik teglangan ochiq matnlar bazasini shakllantirish.

Mahalliy yirik til modellarini (LLM) rivojlantirish: O‘zbek tilining agglyutinativ xususiyatlarini chuqur tushunadigan maxsus tokenizatorlar (BPE/Unigram asosidagi morfologik moslashuvchan algoritmlar) va o‘zbek tiliga ixtisoslashgan neyron tarmoq arxitekturalarini yaratish.

Imlo konvertatsiyasi va normallashtirishni avtomatlashtirish: Kirill va lotin yozuvlari o‘rtasidagi hamda Unicode belgilaridagi imlo farqlarini avtomatik ravishda yagona adabiy standartga keltiruvchi yuqori aniqlikdagi dasturiy modullarni SI modellarining dastlabki ishlov berish (pre-processing) bosqichiga integratsiya qilish.

Nutqni tanish (ASR) va sintezlash (TTS) tizimlarini kengaytirish: O‘zbek tilining turli hududiy shevalari va talaffuz xususiyatlarini o‘z ichiga olgan audio ma’lumotlar bazalarini boyitish orqali ovozli yordamchilar sifatini oshirish.

Xulosa

O‘zbek tilining sun’iy intellekt tizimlaridagi muammolari asosan sifatli raqamli ma’lumotlar yetishmasligi, tilning murakkab agglyutinativ tuzilishi va grafik-imlo xilma-xilligi bilan bog‘liq. Ushbu muammolarni bartaraf etish nafaqat axborot texnologiyalari sohasidagi ilmiy yutuq, balki o‘zbek tilining global raqamli makonda munosib o‘rin egallashini ta’minlashning strategik sharti hisoblanadi. Lingvistlar va sun’iy intellekt mutaxassislarining hamkorlikdagi tadqiqotlarini kuchaytirish hamda ochiq manbali milliy loyihalarni qo‘llab-quvvatlash bu borada keskin ijobiy burilish yasaydi.

Foydalanilgan adabiyotlar

1. Jurafsky, D., & Martin, J. H. (2024). Speech and Language Processing (3rd ed.). Stanford University.

2. Mengliev, B. R. (2020). O‘zbek tilining kompyuter lingvistikasi va milliy korpus asoslari. Toshkent: Fan va texnologiya.

3. Mansurov, B., & Mansurov, A. (2021). UzBERT: Pretraining a BERT Model for Uzbek. arXiv preprint arXiv:2108.09814.

4. O‘zbekiston Respublikasi Prezidentining «Sun’iy intellekt texnologiyalarini jadal joriy etish uchun shart-sharoitlar yaratish chora-tadbirlari to‘g‘risida»gi qarori.