UniPublish Journal of Science
Vol. 1 No. 2 (2026) · b. 110-117

ILMIY MAQOLA

Sun’iy intellekt va tabiiy tilni qayta ishlash (NLP): rivojlanish bosqichlari, imkoniyatlar va muammolar

Ismatullayev Maxsudxon Nodirxon o‘g‘li[1]

[1]O‘zbekiston davlat jahon tillari universiteti, Ingliz filologiyasi fakulteti, 4-kurs talabasi

MAQOLA HAQIDA

Qabul qilindi
17-noyabr 2026

Nashr etildi
17-noyabr 2026

ISSN
ariza berilgan

Kalit so‘zlar: sun’iy intellekt, tabiiy tilni qayta ishlash, mashinaviy o ʻ qitish, transformer

Onlayn o‘qish
unipublish.org

Annotatsiya. Ushbu maqolada sun’iy intellekt (SI) va tabiiy tilni qayta ishlash (Natural Language Processing, NLP) sohasining rivojlanish bosqichlari, asosiy vazifalari va amaliy qo‘llanilishi tahlil qilinadi. NLP qoidaga asoslangan tizimlardan statistik usullarga, so‘ngra neyron tarmoqlar va transformer arxitekturasiga asoslangan yirik til modellariga (LLM) qadar bosib o‘tgan yo‘li ko‘rib chiqiladi. Shuningdek, til ta’limi va tilshunoslikdagi imkoniyatlar, o‘zbek tili uchun NLP rivojlantirishdagi muammolar hamda sohaning axloqiy masalalari yoritiladi.

Iqtibos uchun: Ismatullayev, M. (2026). Sun’iy intellekt va tabiiy tilni qayta ishlash (NLP): rivojlanish bosqichlari, imkoniyatlar va muammolar. *UniPublish Journal of Science*, 1(2), 110-117. https://unipublish.org/articles/suniy-intellekt-va-tabiiy-tilni-qayta-ishlash-nlp-rivojlanish-bosqichl

Maqolaning to‘liq matni va adabiyotlar ro‘yxati keyingi sahifalarda.

© 2026 Mualliflar · CC BY 4.0

Sun'iy intellekt va tabiiy tilni qayta ishlash (NLP): rivojlanish bosqichlari, imkoniyatlar va muammolar

Nov 17, 2026 · @Makhsudkhon

Annotatsiya

Ushbu maqolada sun’iy intellekt (SI) va tabiiy tilni qayta ishlash (Natural Language Processing, NLP) sohasining rivojlanish bosqichlari, asosiy vazifalari va amaliy qo‘llanilishi tahlil qilinadi. NLP qoidaga asoslangan tizimlardan statistik usullarga, so‘ngra neyron tarmoqlar va transformer arxitekturasiga asoslangan yirik til modellariga (LLM) qadar bosib o‘tgan yo‘li ko‘rib chiqiladi. Shuningdek, til ta’limi va tilshunoslikdagi imkoniyatlar, o‘zbek tili uchun NLP rivojlantirishdagi muammolar hamda sohaning axloqiy masalalari yoritiladi.

Kalit so‘zlar: sun’iy intellekt, tabiiy tilni qayta ishlash, mashinaviy o‘qitish, transformer, yirik til modellari, korpus lingvistikasi, o‘zbek tili.

Abstract

This article analyses the development stages, core tasks and practical applications of artificial intelligence (AI) and natural language processing (NLP). It traces the field’s path from rule-based systems through statistical methods to neural networks and large language models (LLMs) built on the transformer architecture. The article also discusses opportunities in language teaching and linguistics, the challenges of developing NLP for the Uzbek language, and the ethical issues of the field.

Keywords: artificial intelligence, natural language processing, machine learning, transformer, large language models, corpus linguistics, Uzbek language.

Kirish

Til insoniyatning asosiy aloqa vositasi va bilimlarni saqlash shaklidir. Shu sababli kompyuterga inson tilini tushunish va yaratishni o‘rgatish sun’iy intellektning eng muhim maqsadlaridan biri bo‘lib kelgan. Bugungi kunda tarjima dasturlari, ovozli yordamchilar,

chat-botlar va matn tahrirlovchi vositalar kundalik hayotimizning bir qismiga aylandi. Ularning barchasi NLP texnologiyalariga tayanadi.

So‘nggi yillarda yirik til modellarining paydo bo‘lishi bu sohani tub o‘zgarishlarga olib keldi. Ta’lim, tarjima, jurnalistika, huquq va tibbiyot kabi tilga bog‘liq ko‘plab kasblarda ish usullari o‘zgarmoqda. Bu esa filologlar va til o‘qituvchilari uchun NLP asoslarini bilishni zarur qilib qo‘ymoqda.

Maqolaning maqsadi — SI va NLP sohasining rivojlanish yo‘li, asosiy vazifalari va qo‘llanilishini tizimli yoritish hamda o‘zbek tili uchun mavjud muammolarni aniqlash.

Vazifalar:

1. NLP tushunchasi va uning SI tarkibidagi o‘rnini aniqlash.
2. NLP rivojlanishining asosiy bosqichlarini tavsiflash.
3. Asosiy vazifalar va qo‘llanilish sohalarini, ayniqsa til ta’limidagi imkoniyatlarni ko‘rsatish.
4. O‘zbek tili uchun NLP rivojlantirishdagi muammolarni tahlil qilish.
5. Sohaning axloqiy masalalarini ko‘rib chiqish.

NLP tushunchasi va uning sun’iy intellektdagi o‘rni

Tabiiy tilni qayta ishlash — kompyuterlarga inson tilidagi matn va nutqni tahlil qilish, tushunish va yaratishni o‘rgatuvchi fan sohasi. U tilshunoslik, informatika va sun’iy intellekt tutashgan joyda shakllangan. D. Jurafsky va J. Martin NLP’ni hisoblash lingvistikasi bilan chambarchas bog‘liq amaliy yo‘nalish sifatida ta’riflaydi [5].

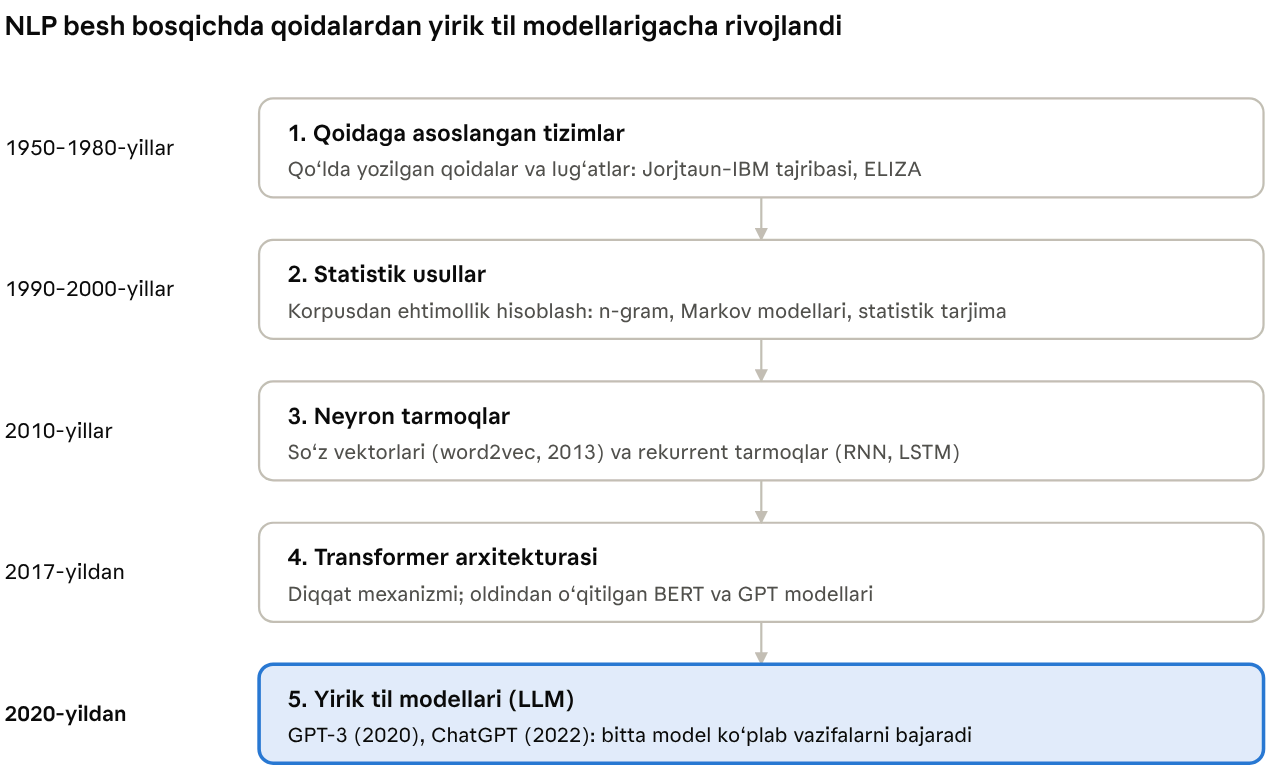
NLP odatda ikki katta yo‘nalishga bo‘linadi:

- **Tabiiy tilni tushunish (NLU)** — matn ma’nosini, niyatni, kayfiyatni va undagi ma’lumotlarni aniqlash.
- **Tabiiy tilni yaratish (NLG)** — berilgan ma’lumot yoki so‘rov asosida ravon va mazmunli matn hosil qilish.

Sun’iy intellekt tarkibida NLP mashinaviy o‘qitish (machine learning) va chuqur o‘qitish (deep learning) usullariga tayanadi. Til modellari katta hajmdagi matnlardan qonuniyatlarni o‘rganadi: qaysi so‘z qaysi kontekstda kelishi, gap qanday tuzilishi va ma’no qanday ifodalanishi. Shu ma’noda zamonaviy NLP tilshunoslikning an’anaviy kategoriyalarini (morfologiya, sintaksis, semantika, pragmatika) statistik va hisoblash vositalari orqali modellashtiradi.

NLP rivojlanish bosqichlari

NLP tarixi taxminan 75 yilni qamrab oladi va besh asosiy bosqichga bo‘linadi. Har bir bosqichda tilni modellashtirish usuli tubdan o‘zgargan.



1-rasm. NLP rivojlanish bosqichlari · 5 bosqich

Har bir keyingi bosqich oldingisining cheklovlarini yengib o‘tgan: qoidalar o‘rnini statistika, statistika o‘rnini esa o‘z-o‘zidan o‘rganuvchi neyron modellar egallagan.

1. Ilk g‘oyalar va qoidaga asoslangan tizimlar (1950–1980-yillar). A. Tyuring 1950-yilda mashina fikrlay oladimi, degan savolni qo‘yib, uni til orqali sinash g‘oyasini taklif qildi [1]. 1954-yilda Jorjtaun universiteti va IBM tajribasida rus tilidagi gaplar ingliz tiliga avtomatik tarjima qilindi [10]. 1966-yilda J. Veyzenbaum yaratgan ELIZA dasturi oddiy shablonlar yordamida psixoterapevt bilan suhbatga taqlid qildi [2]. Bu davr tizimlari tilshunoslar qo‘lda yozgan grammatik qoidalar va lug‘atlarga tayangan. Ular tor sohalarda ishlagan, ammo tilning ko‘p ma’noliligi va istisnolari oldida ojiz bo‘lgan.

2. Statistik davr (1990–2000-yillar). Raqamli matn korpuslari ko‘payishi bilan qoidalar o‘rnini ehtimollik modellari egalladi. N-gram modellari, yashirin Markov modellari va statistik mashinaviy tarjima shu davrda keng tarqaldi. Tizim endi qoidani yodlamay, misollardan ehtimollikni hisoblab o‘rgandi.

3. Neyron tarmoqlar davri (2010-yillar). 2013-yilda T. Mikolov va hamkasblari taklif qilgan word2vec usuli so‘zlarni ma’nosiga ko‘ra vektor ko‘rinishida ifodalash imkonini berdi [3]. Rekurrent neyron tarmoqlar (RNN, LSTM) gap ketma-ketligini qayta ishlashda statistik usullardan ustun keldi.

4. Transformer davri (2017-yildan). A. Vasvani va hamkasblarining “Attention Is All You Need” maqolasi transformer arxitekturasini taqdim etdi [4]. Diqqat (attention) mexanizmi modelga gapdagi barcha so‘zlarning o‘zaro bog‘lanishini bir vaqtda hisobga olish imkonini berdi. Uning asosida BERT [6] va GPT kabi oldindan o‘qitilgan modellar yaratildi.

5. Yirik til modellari davri (2020-yildan). 175 milliard parametrli GPT-3 modeli maxsus o‘qitishsiz, faqat bir nechta misol asosida yangi vazifalarni bajara olishini ko‘rsatdi [7]. 2022-yil noyabrida ChatGPT’ning ommaga taqdim etilishi bilan LLM’lar kundalik vositaga aylandi. Bugungi modellar matn yozish, tarjima, dasturlash va tahlil kabi vazifalarni bitta tizimda bajaradi.

Asosiy vazifalar va qo‘llanilish sohalari

NLP bir qator bazaviy vazifalardan iborat bo‘lib, amaliy tizimlar ularni birlashtirib quriladi.

Vazifa	Mazmuni	Amaliy misol
Tokenizatsiya va morfologik tahlil	Matnni so‘z va qo‘shimchalarga ajratish	Imlo tekshirgichlar
Gap bo‘laklarini aniqlash (POS tagging)	Har bir so‘zning turkumini belgilash	Grammatik tahlil vositalari
Nomlangan obyektlarni aniqlash (NER)	Shaxs, joy, tashkilot nomlarini topish	Hujjatlardan ma’lumot ajratish
Kayfiyat tahlili	Matnning ijobiy yoki salbiy ohangini aniqlash	Mijozlar fikrini o‘rganish
Mashinaviy tarjima	Matnni bir tildan boshqasiga o‘girish	Onlayn tarjimonlar
Nutqni tanish va sintez	Ovozni matnga va matnni ovozga aylantirish	Ovozli yordamchilar
Matnni umumlashtirish	Uzun matnning qisqa mazmunini yaratish	Yangiliklar dayjesti

Vazifa	Mazmuni	Amaliy misol
Savol-javob va dialog	So‘rovga mos javob berish	Chat-botlar

Til ta’limidagi qo‘llanilishi. Filologiya va til o‘qitish uchun NLP alohida imkoniyatlar ochadi:

- **Yozma ishlarni avtomatik baholash** — insho va esselarni grammatika, leksika va mantiqiy tuzilish bo‘yicha tahlil qilib, o‘quvchiga darhol izoh berish.
- **Og‘zaki nutqni baholash** — nutqni tanish texnologiyasi orqali talaffuz va ravonlikni tekshirish.
- **Suhbat amaliyoti** — chat-botlar bilan istalgan vaqtda chet tilida muloqot qilish.
- **Moslashtirilgan mashqlar** — o‘quvchi darajasiga (masalan, CEFR A2 yoki B2) mos matn va topshiriqlar yaratish.
- **O‘qituvchi uchun yordam** — dars rejasi, test savollari va baholash mezonlarini tayyorlashni tezlashtirish.

Tilshunoslikdagi qo‘llanilishi. Korpus lingvistikasi NLP vositalari yordamida millionlab so‘zdan iborat matnlarni tahlil qiladi. Bu so‘z chastotasi, kollokatsiyalar, til o‘zgarishlari va uslubiy xususiyatlarni empirik asosda o‘rganish imkonini beradi. Shuningdek, avtomatik tarjima natijalarini qiyosiy tahlil qilish tarjimashunoslikning yangi yo‘nalishiga aylangan.

O‘zbek tili uchun NLP: xususiyatlar va muammolar

O‘zbek tili NLP sohasida “kam resursli til” (low-resource language) hisoblanadi: unga oid raqamli korpuslar, belgilangan ma’lumotlar to‘plamlari va tayyor vositalar ingliz yoki rus tiliga nisbatan ancha kam. Buning bir necha sababi bor.

1. **Agglyutinativ morfologiya.** O‘zbek tilida bitta o‘zakka ko‘plab qo‘shimchalar qo‘shiladi (masalan, *kitob-lar-imiz-dan*). Natijada bitta so‘zning juda ko‘p shakli paydo bo‘ladi, bu esa lug‘atga asoslangan modellar uchun qiyinchilik tug‘diradi.
2. **Ikki yozuvning parallel qo‘llanilishi.** Rasmiy lotin yozuvi bilan bir qatorda kirill yozuvi ham keng ishlatiladi. Bundan tashqari, lotin yozuvida “o‘” va “g‘” harflari turli belgilar bilan (‘, ’, ` yoki ′) yoziladi. Bu matnlarni bir xil ko‘rinishga keltirishni talab qiladi.
3. **Belgilangan ma’lumotlar tanqisligi.** Gap bo‘laklari, nomlangan obyektlar yoki kayfiyat bo‘yicha belgilangan katta korpuslar kam. Ularni yaratish tilshunos mutaxassislarning ko‘p mehnatini talab qiladi.
4. **Sheva va so‘zlashuv uslubi.** Ijtimoiy tarmoqlardagi matnlarda adabiy me’yordan chetlanish, rus va ingliz so‘zlarining aralashuvi ko‘p uchraydi.

Shunga qaramay, so‘nggi yillarda sezilarli siljishlar bor. Masalan, B. Mansurov va A. Mansurov o‘zbek tili uchun oldindan o‘qitilgan UzBERT modelini yaratib, uning ko‘p tilli modellardan ayrim vazifalarda ustunligini ko‘rsatdi [8]. Zamonaviy ko‘p tilli LLM’lar ham o‘zbek tilida matn tushunish va yaratishni tobora yaxshiroq bajarmoqda. Biroq ularning sifati hali yuqori resursli tillar darajasiga yetmagan.

Bu yo‘nalishda filologlarning o‘rni katta: sifatli korpuslar yaratish, ma’lumotlarni belgilash, imlo me’yorlarini standartlashtirish va model natijalarini lingvistik baholash aynan til mutaxassislarining ishidir.

Muammolar va axloqiy masalalar

NLP texnologiyalarining tez rivojlanishi bir qator jiddiy muammolarni ham yuzaga chiqardi.

- **“Gallyutsinatsiya”.** Til modellari ishonchli ohangda, lekin noto‘g‘ri ma’lumot yoki mavjud bo‘lmagan manbalarni keltirishi mumkin. Shu sababli ilmiy va ta’limiy ishlarda ularning javoblarini tekshirish shart.
- **Tarafkashlik (bias).** Modellar o‘qitilgan matnlardagi stereotiplarni o‘zlashtirib, ularni takrorlashi mumkin. E. Bender va hamkasblari juda katta modellarning bunday xavflari haqida ogohlantirgan [9].
- **Akademik halollik.** Talabalar yozma ishlarni to‘liq SI yordamida tayyorlashi baholash tizimiga tahdid soladi. Buning yechimi taqiq emas, balki SI’dan foydalanish qoidalarini aniq belgilash va jarayonga yo‘naltirilgan baholashdir.
- **Maxfiylik.** Foydalanuvchi kiritgan shaxsiy ma’lumotlar va matnlar himoyasi muhim masala bo‘lib qolmoqda.
- **Til tengsizligi.** Yirik modellar asosan ingliz tilidagi ma’lumotlarda o‘qitiladi. Natijada kam resursli tillar so‘zlashuvchilari texnologiyadan teng foydalana olmaydi.
- **Resurs sarfi.** Yirik modellarni o‘qitish katta hisoblash quvvati va energiya talab qiladi.

Xulosa

NLP qoidaga asoslangan tajribalardan bugungi yirik til modellarigacha bo‘lgan yo‘lni taxminan 75 yilda bosib o‘tdi. Har bir bosqichda tilni modellashtirish usuli o‘zgardi: qo‘lda yozilgan qoidalar statistikaga, statistika neyron tarmoqlarga, neyron tarmoqlar esa transformer asosidagi umumiy modellarga o‘rin bo‘shatdi.

Til ta’limi va tilshunoslik uchun NLP katta imkoniyatlar ochadi: yozma va og‘zaki nutqni avtomatik baholash, moslashtirilgan mashqlar va korpus tadqiqotlari shular jumlasidandir. Shu bilan birga, gallyutsinatsiya, tarafkashlik, akademik halollik va maxfiylik kabi muammolar texnologiyadan ongli va tanqidiy foydalanishni talab qiladi.

O‘zbek tili uchun asosiy vazifa — sifatli korpuslar va belgilangan ma’lumotlar to‘plamlarini yaratish hamda imlo me’yorlarini raqamli muhitda standartlashtirishdir. Bu ishda filologlar va dasturchilarning hamkorligi hal qiluvchi ahamiyatga ega.

Foydalanilgan adabiyotlar

1. Turing A. M. Computing Machinery and Intelligence // Mind. — 1950. — Vol. 59, No. 236. — P. 433–460.
2. Weizenbaum J. ELIZA — A Computer Program for the Study of Natural Language Communication between Man and Machine // Communications of the ACM. — 1966. — Vol. 9, No. 1. — P. 36–45.
3. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space // arXiv:1301.3781. — 2013.
4. Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need // Advances in Neural Information Processing Systems 30 (NeurIPS). — 2017.
5. Jurafsky D., Martin J. H. Speech and Language Processing. 3rd ed. draft. — Stanford University, 2024.
6. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. — 2019. — P. 4171–4186.
7. Brown T. B., Mann B., Ryder N. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems 33 (NeurIPS). — 2020.
8. Mansurov B., Mansurov A. UzBERT: pretraining a BERT model for Uzbek // arXiv:2108.09814. — 2021.
9. Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of FAccT ’21. — 2021. — P. 610–623.
10. Hutchins W. J. The Georgetown-IBM Experiment Demonstrated in January 1954 // Machine Translation: From Real Users to Research (AMTA 2004). — Springer, 2004. — P. 102–114.