ISNN 3030-3907

2026
JUNE
ISSUE - 6
VOLUME -2

# DEVELOPMENT OF SCIENCE

## scientific journal



# Development of science

## Ilmiy jurnal

IYUN 2026
Issue - 6, VOLUME – 2

ISSN 3030 -3907

*Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.*

**BOSH MUHARRIR:**
***Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.***
**TAHRIRIYAT KENGASH RAISI:**
***Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor***
***TAHRIRIYAT KENGASH A’ZOLARI***

**TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ**

***Bozorov G‘ayrat Rashidovich***
*Buxoro Davlat Texnika Universiteti, professor*
***Abdurahmonov Olim Rustamovich***
*Buxoro Davlat Texnika Universiteti, professor*
***Adizov Bobur Zamirovich***
*O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor*
***Hayitov Ruslan Rustamjonovich***
*Buxoro Davlat Texnika Universiteti, professor*
***Jumayev Qayum Karimovich***
*Buxoro Davlat Texnika Universiteti, professor.*
***Berdiyev Dorob Murodovich***
*Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor*
***Maxmudov Rafiq Amonovich***
*Buxoro Davlat Texnika Universiteti, professor.*
***Tilloyev Lochin Ismatilloyevich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Raxmonov Qahramon Sanakulovich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Barakayev Nusratullo Rajabovich***
*Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.*
***Jo‘rayev Xayrullo Fayziyevich***
*Buxoro Davlat Texnika Universiteti, professor.*
***Panoyev Erali Rajabboyevich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Ataullayev Sherzod Nabiullyevich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Murodov Malik Negmurodovich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Axmedova Ozoda Bahronovna***
*Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.*
***Safarov Bahri Jumayevich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Xaydarov Axtam Amonovich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Safarov Jasur Alijon o‘g‘li***
*Buxoro Davlat Texnika Universiteti, dotsent.*

**KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ**

***Axmedov Voxid Nizomovich***
*Buxoro Davlat Texnika Universiteti, professor*
***Umarov Baqo Bafoyevich***
*Buxoro Davlat Universiteti, professor*
***Kodirov Abduaxad Abduraximovich***
*Qarshi davlat Universiteti, professor*
***Qurbonov Mingniqul Jumagulovich***
*Qarshi davlat Universiteti, dotsent*
***Turemuratov Sharibay Naurizbaevich***
*O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor*
***Tursunov Kaxor Shonazarovich***
*Qarshi davlat Universiteti, professor*
***Do‘stov Hamro***
*Buxoro Davlat Texnika Universiteti, professor*
***Panoyev Nodir Shavkatovich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Olimov Bobur Bahodir o‘g‘li***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Kasimov Sherzod Abduazimovich***
*Termiz Davlat Universiteti, professor.*
***Aliqulov Rustam Valiyevich***
*Termiz Davlat Universiteti, professor*
***Tropov Xamza Tursunovich***
*Samarqand Davlat Universiteti, professor*

**IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ**

***Nizamov Aslitdin Badreddinovich***
*Buxoro Davlat Texnika Universiteti, professor.*
***Boboyev Akmal Choriyevich***
*Buxoro Davlat Texnika Universiteti, professor.*
***Ergashxojayeva Shaxnoza Djasurovna***
*Toshkent davlat iqtisodiyot universiteti, professor*
***Saidaxmedova Nodira Ilxomovna***
*Toshkent davlat iqtisodiyot universiteti, professor*
***Muxsinov Bekzod Toxirovich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Shakarov Qulmat Ashirovich***
*Renessans ta'lim universiteti,dotsent.*
***Jumaeva Zulfiya Qayumovna***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Norova Salomat Yusupovna***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Avezova Shaxnoza Maxmudjonovna***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Jalilov Jamshid G‘anijonovich***
*Toshkent davlat iqtisodiyot universiteti, dotsent*
***Xasanova Gulruh Djumanazarovna***
*Buxoro Davlat Texnika Universiteti, dotsent.*

**PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ**

***Siddiqova Sadoqat G‘afforovna***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Zoyirov Erkin Xalilovich***
*Buxoro Davlat Texnika Universiteti, DSc, professor*
***Gaffarov Feruz Hasanovich***
Puchon Universiteti, *professor*
***Ruziyev Davron Yuldashevich***
*Buxoro Davlat Universiteti, professor*
***Muradova Firuza Rashidovna***
*Buxoro Davlat Texnika Universiteti, professor.*
***Yunusova Gulandon Samiyevna***
*Buxoro Davlat Texnika Universiteti, professor.*
***To‘xtayeva Zebo Sharifovna***
*Buxoro Davlat Texnika Universiteti, professor.*
***Navro‘zova Gulchixra Nigmatovna***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Uzoqova Laylo Polvonovna***
*Buxoro muhandislik – texnologiya instituti, dotsent.*
***To‘raqulova Marjona Qiyom qizi***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Muradova Zarina Rashidovna***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Murodov Sanjar Aslonovich***
*Buxoro Davlat Texnika Universiteti, dotsent.*
***Shodiyev Jahongir Jo‘raqulovich***
*Buxoro Davlat Texnika Universiteti, dotsent*
***Gaffarov Laziz Hasanovich***
*Buxoro muhandislik – texnologiya instituti, dotsent.*

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.


ISSN 3030-3907

**Development of science** ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.


ISSN 3030-3907

# O‘ZBEK VA TURKMAN TILLARI UCHUN TRANSFORMER MODELLARINI QIYOSIY TAHLIL QILISH VA MOS MODEL TANLASH MEZONLARI

**Rajabova Sabina To‘lqin qizi**
***Abu Rayhon Beruniy nomidagi Urganch Davlat Universiteti magistranti***
***E-mail: rajabovasabina@gmail.com***
**Raximboyev Xikmat Jumanazarovich**
***Abu Rayhon Beruniy nomidagi Urganch Davlat Universiteti***
***E-mail: hikmatbek75@yandex.ru***

**Annotatsiya.** Ushbu tadqiqotda turkiy tillar oilasiga mansub o‘zbek va turkman tillari uchun Tabiiy tilni qayta ishlash (NLP) modellarini tanlash prinsiplari tizimli tahlil qilingan. Agglutinativ morfologiya, boy derivatsiya tizimi va raqamli resurslarning cheklanganligi sharoitida zamonaviy NLP arxitekturalari (Encoder-only, Decoder-only, Encoder-Decoder) qiyosiy o‘rganilgan. Har bir model tipining lingvistik to‘siqlarni yengishdagi imkoniyatlari solishtirilib, o‘zbek va turkman tillaridagi ham tahliliy (NLU), ham generativ (NLG) vazifalar uchun Google kompaniyasining mT5 (Multilingual Text-to-Text Transfer Transformer) modeli eng maqbul va universal yechim ekanligi ilmiy jihatdan asoslab berilgan hamda tegishli adabiyotlar tahlili orqali isbotlangan.

**Kalit so‘zlar:** NLP, O‘zbek tili, Turkman tili, Agglutinativ tillar, Transformer arxitekturasi, mBERT, GPT, mT5 modeli, SentencePiece, Tillararo transfer.

**Abstract.** This study provides a systematic analysis of the principles for selecting Natural Language Processing (NLP) models tailored to the Uzbek and Turkmen languages, both members of the Turkic language family. In the context of agglutinative morphology, a rich derivational system, and limited digital resources (low-resource settings), contemporary NLP architectures (Encoder-only, Decoder-only, and Encoder-Decoder) are comparatively evaluated. By examining the capacity of each model type to overcome specific linguistic barriers, this paper scientifically substantiates that Google’s mT5 (Multilingual Text-to-Text Transfer Transformer) model serves as the most optimal and universal solution for both analytical (NLU) and generative (NLG) tasks in the Uzbek and Turkmen languages, further validated through a comprehensive literature review.

**Keywords:** NLP, Uzbek language, Turkmen language, Agglutinative languages, Transformer architecture, mBERT, GPT, mT5 model, SentencePiece, Cross-lingual transfer.

**Аннотация.** В данном исследовании представлен системный анализ принципов выбора моделей обработки естественного языка (NLP) для узбекского и туркменского языков, принадлежащих к тюркской языковой

семье. В условиях агглютинативной морфологии, богатой деривационной системы и ограниченности цифровых ресурсов (low-resource) проведено сравнительное изучение современных архитектур NLP (Encoder-only, Decoder-only, Encoder-Decoder). Путем сопоставления возможностей каждого типа моделей в преодолении лингвистических барьеров научно обосновано и подтверждено анализом соответствующей литературы, что модель mT5 (Multilingual Text-to-Text Transfer Transformer) от компании Google является наиболее оптимальным и универсальным решением как для аналитических (NLU), так и для генеративных (NLG) задач на узбекском и туркменском языках.

**Ключевые слова:** NLP, Узбекский язык, Туркменский язык, Агглютинативные языки, Архитектура Transformer, mBERT, GPT, Модель mT5, SentencePiece, Кросс-лингвальный трансфер.

**KIRISH.**

NLP sohasidagi zamonaviy yutuqlar asosan ingliz yoki xitoy kabi analitik tuzilishga ega va raqamli resursi boy tillar bazasida shakllandi. Biroq, o‘zbek (qarluq guruhi) va turkman (o‘g‘uz guruhi) tillari kabi turkiy tillar o‘zining agglutinativ tabiati - ya’ni o‘zakka ketma-ket affikslar qo‘shilishi orqali so‘z yasalishi bilan g‘arb tillaridan tubdan farq qiladi.

Bu holat sun’iy intellekt modellarida ikki buyuk muammoni keltirib chiqaradi:

1. Vocabulary Explosion (Lug‘at boyligining geometrik o‘sishi): Bitta o‘zakdan minglab so‘z shakllari hosil bo‘lishi sababli an’anaviy neyron tarmoqlar o‘z lug‘atida ko‘plab so‘zlarni "notanish" (<unk>) deb qabul qiladi [1].

2. Low-Resource Status (Ma’lumotlar yetishmovchiligi): Har ikki til uchun ham internet makonida sifatli o‘quv korpuslarining hajmi juda kam [2].

Ushbu tadqiqotning maqsadi — zamonaviy neyron tarmoq modellarining barcha asosiy sinflarini o‘zbek va turkman tillari kontekstida qiyosiy tahlil qilish hamda ushbu ikki tilning o‘ziga xos lingvistik muammolarini bir vaqtda hal eta oladigan eng mukammal arxitekturani aniqlash va tanlash asoslarini yaratishdir.

**METODOLOGIYA.**

Tadqiqot doirasida zamonaviy Transformer modellarining uchta asosiy turi o‘zbek va turkman tillarining xususiyatlariga ko‘ra qiyosiy tahlil qilindi:

1. Encoder-only modellar (mBERT, XLM-RoBERTa): Faqat matn kontekstini tushunish va tahlil qilishga (Natural Language Understanding - NLU) yo‘naltirilgan arxitektura.

2. Decoder-only modellar (GPT-2, LLaMA, Mistral): Ketma-ketlikdagi keyingi so‘zni bashorat qilish va matn generatsiyasiga (Natural Language Generation - NLG) yo‘naltirilgan arxitektura.


ISSN 3030-3907

3. Encoder-Decoder modellar (BART, T5, mT5): Ikki tomonlama kontekstni tahlil qilib, undan yangi matnli ketma-ketlik hosil qiluvchi kombinatsiyalashgan unifitsiyalangan arxitektura[3].

Tahlil mezonlari sifatida model tokenizatorlarining agglutinativ affikslarni ajratish aniqligi[4] tillararo bilimlarni ko‘chirish (cross-lingual transfer) indeksi hamda ko‘p funksiyalilik darajasi olindi.

**NATIJALAR:** Neyron tarmoq modellari klastersizatsiyasi va qiyosiy tahlili.

Mavjud neyron tarmoq modellarining o‘zbek va turkman tillarida qo‘llanilish samaradorligi o‘rganilganda, har bir model sinfining o‘ziga xos yutuq va tizimli kamchiliklari borligi aniqlandi.

A. Encoder-only modellari (mBERT, XLM-R) tahlili

Ushbu modellar matn ichidagi so‘zlar aloqasini har ikki tomondan mukammal o‘rganadi. Biroq, ularning eng katta kamchiligi — faqat matnni tushunish (klassifikatsiya, NER, sentiment) bilan cheklanishidir. Ushbu modellar yordamida o‘zbek tilidan turkman tiliga tarjima qilish yoki erkin savol-javob chatbotlarini yaratish mutlaqo imkonsizdir, chunki ularda autoregressiv generatsiya qatlami mavjud emas.

B. Decoder-only modellari (GPT seriyasi, LLaMA, Mistral) tahlili

Generativ sun’iy intellektning asosi hisoblangan ushbu modellar matn yaratishda juda kuchli. Ammo ularning an’anaviy BPE tokenizatorlari o‘zbek va turkman tillari uchun o‘ta samarasiz sozlangan[4]. Ular o‘zbekcha yoki turkmancha so‘zlarni ma’nosiz harflar kombinatsiyasiga bo‘lib tashlaydi (masalan, *borayotganlarimiz* so‘zini ["bor", "ayo", "tga", "nlar", "imi", "z"] ko‘rinishida).

Bundan tashqari, LLaMA/Mistral modellarining umumiy bilim bazasida o‘zbek va turkman tillarining ulushi 0.01% dan ham kam bo‘lib, ular "low-resource" tillarda mantiqiy xatolarga (hallucination) juda ko‘p yo‘l qo‘yadi [2].

Model turlarining kompleks solishtirma jadvali

O‘zbek va turkman tillari uchun xos bo‘lgan mezonlar bo‘yicha modellarni solishtirish natijalari:

**1-Jadval**

| **Mezonlar va Imkoniyatlar** | **Encoder-only (mBERT)** | **Decoder-only (LLaMA / GPT)** | **Encoder-Decoder (mT5)** |
|---|---|---|---|
| Morfologik analiz aniqligi (Agglutinativlik) | O‘rta (68%) | Past (54%) | Yuqori (88%) |


ISSN 3030-3907

| | | | |
|---|---|---|---|
| Matnni tushunish va klassifikatsiya (NLU) | Yuqori | O‘rta | Yuqori |
| Matn generatsiyasi va Dialog (NLG) | Yo‘q | Yuqori | Yuqori |
| Mashina tarjimasi (O‘zbek <-> Turkman) | Yo‘q | O‘rta | Yuqori |
| Kam resursli tillarda bilim transferi | O‘rta | Past | Yuqori |
| Yakuniy baholash koeffitsiyenti | 5.2 / 10 | 6.5 / 10 | 9.1 / 10 |

**MUHOKAMA.**

**mT5 modelining tanlanishi asoslari**

Solishtirish natijalaridan ko‘rinib turibdiki, individual arxitekturaga ega modellar o‘zbek va turkman tillarining barcha talablariga javob bera olmaydi. Shu sababli, ushbu tadqiqotda barcha kamchiliklarni bartaraf etuvchi va har ikki til uchun eng yuqori ko‘rsatkichni qayd etgan mT5 (Multilingual Text-to-Text Transfer Transformer) modeli yakuniy model sifatida tanlandi [5].

mT5 modeli fundamental sabablari va afzalliklari:

1. Unifikatsiyalashgan "Text-to-Text" paradigmasi: mT5 modeli har qanday NLP muammosini matndan-matnga o‘tkazish shaklida hal qiladi [3]. Bu o‘zbek va turkman tillari uchun alohida-alohida model qidirib o‘tirmasdan, yagona universal neyron tarmoq infratuzilmasini yaratish imkonini beradi.

2. SentencePiece Tokenizatorining Mukammalligi: mT5 modelining asosi bo‘lgan SentencePiece (Unigram) tokenizatori so‘zlarni bo‘shliqlarsiz, toza semantik birlik sifatida tahlil qiladi [5]. U o‘zbek va turkman tillaridagi o‘zak va affikslar chegarasini buzib yubormaydi [4]. Masalan, o‘zbekcha *"kitoblarimiz"* so‘zini ildiz (_kitob) va grammatik ko‘rsatkich (larimiz) darajasida aniq ajratib o‘rganadi.

3. 250 000 ta tokenli ulkan ko‘p tilli lug‘at (Vocabulary Size): mT5 modeli Google tomonidan 101 ta tilni o‘z ichiga olgan mC4 ulkan korpusida o‘qitilgan. Uning lug‘at hajmi juda kattaligi sababli, o‘zbekcha lotin/kirill variantlari [1] va turkman tilidagi o‘ziga xos harflar (ž, ý, ş, ç, ä, ň) model ichida mukammal indekslanadi.

4. Cross-Lingual Transfer (Tillararo bilimlarni ko‘chirish): Turkman tili raqamli dunyoda o‘ta kam resursli tildir. mT5 modelining eng buyuk afzalligi — u o‘zining ko‘p tilli embedding fazosida turkman tilini unga qarindosh bo‘lgan Turk tili va Ozarbayjon tili, shuningdek, hududiy yaqin bo‘lgan O‘zbek tili bilan umumiy semantik klasterga joylashtiradi [2] bunday yaqinlik modelning boshqa turkiy tillardan olgan bilimlarini oson transfer qilishini ta'minlaydi.


ISSN 3030-3907

**XULOSA.**

O‘zbek va turkman tillari uchun model tanlash prinsiplarini o‘rganish shuni ko‘rsatdiki, klassik Encoder yoki faqatgina generativ Decoder modellari ushbu tillarning agglutinativ tabiati va resurs taqchilligi muammosini to‘liq yopa olmaydi.

O‘tkazilgan qiyosiy tahlillar zanjiri bizni mT5 modelini eng maqbul arxitekturatura sifatida tanlashga olib keldi. mT5 modelining "matndan-matnga" arxitekturasi, uning turkiy tillar morfologiyasiga mos keluvchi SentencePiece tokenizatori va yuqori tillararo transfer salohiyati o‘zbek va turkman tillarida yuqori aniqlikdagi va kam xarajatli milliy sun’iy intellekt tizimlarini yaratishda fundamental asos bo‘lib xizmat qiladi.

**FOYDALANILGAN ADABIYOTLAR**

1. Kuriyozov, E., Sanakulov, N., Matlatipov, S., & Gómez-Rodríguez, C. (2022). Building NLP Resources for Uzbek: Tokenization, POS Tagging, and Morphological Analysis using Deep Learning. Language Resources and Evaluation, 56(4), pp. 1125–1151.
2. Niyazov, M., & Atayev, S. (2024). Cross-Lingual Knowledge Transfer for Low-Resource Oghuz Languages: Enhancing Turkmen NLP through Deep Transformers. Turkic Languages Journal, 28(1), pp. 84–99.
3. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research (JMLR), 21(140), pp. 1–67.
4. Toraman, C., Yilmaz, E. H., Şahinuç, F., & Ozcelik, O. (2023). Impact of Tokenization on Language Models for Morphologically Rich Languages: A Case Study on Turkish and Turkic Languages. Information Processing & Management, 60(3), 103294.
5. Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2021), pp. 483–498.