

## PRE-TRAINED MODELLAR YORDAMIDA QISQA MATNLARNI TASNIFLASH ALGORITMLARINI TADQIQ QILISH: QIYOSIY TAHLIL VA HYBRID YONDASHUV

**Raxmatullayeva Dilbarxon Otabek qizi**

*Vatan universiteti Kommunikatsiya va axborot texnologiyalari kafedra 70610101-  
Kompyuter tizimlari va ularning dasturiy ta'minoti mutaxassisligi 2-kurs magistranti*

### **Annotatsiya**

*Mazkur maqolada qisqa matnlarni tasniflashda zamonaviy oldindan o'qitilgan (pre-trained) til modellaridan foydalanish samaradorligi tadqiq qilingan. Qisqa matnlar — ijtimoiy tarmoq xabarlari, mahsulot sharhlari, foydalanuvchi so'rovlari va SMS bildirishnomalar — leksik kambag'allik, kontekst yetishmasligi va xususiyatlar siyrakligi (feature sparsity) muammolari bilan ajralib turadi. Tadqiqotda BERT, RoBERTa, DistilBERT, mBERT (multilingual BERT) va XLM-RoBERTa modellari beshta jamoatchilik datasetlarida (AG News, TREC, SST-2, Twitter Sentiment140 va o'zbek tilidagi mahalliy korpus) qiyosiy baholangan. Asosiy metodologik hissa sifatida BERT/XLM-R asoslangan kontekstual embeddinglarni Convolutional Neural Network (CNN) lokal xususiyatlari va Bidirectional GRU (BiGRU) ketma-ket bog'liqliklari bilan birlashtiruvchi diqqat (attention) mexanizmiga ega yangi hybrid arxitektura — STC-Hybrid (Short Text Classification Hybrid) modeli taklif etilgan. Eksperimental natijalar shuni ko'rsatdiki, taklif etilgan hybrid model standart fine-tuning yondashuviga nisbatan o'rtacha F1-ko'rsatkichni 2,4–4,1 foiz punktga oshirgan, ayniqsa kam resursli o'zbek tilida ( $F1 = 0,887$ ) sezilarli yutuqqa erishilgan. Olingan natijalar hybrid arxitekturaning leksik siyraklik muammosini hal etishda samaradorligini tasdiqlaydi va kam resursli tillar uchun tasniflash tizimlarini ishlab chiqishda amaliy ahamiyatga ega.*

**Kalit so'zlar:** *qisqa matn tasnifi, pre-trained modellar, BERT, RoBERTa, XLM-RoBERTa, mBERT, DistilBERT, hybrid arxitektura, transfer learning, tabiiy tilni qayta ishlash, o'zbek tili, kam resursli tillar.*

### **1. Kirish**

So'nggi o'n yillikda raqamli aloqa hajmining keskin oshishi va ijtimoiy tarmoqlarning rivojlanishi natijasida tabiiy tilni qayta ishlash (Natural Language Processing — NLP) sohasida qisqa matnlar bilan ishlash dolzarb yo'nalishga aylandi. Twitter, Telegram, Instagram kabi platformalarda har kuni milliardlab qisqa matnli xabarlar generatsiya qilinadi.

Ushbu matnlar mijozlar fikrlarini tahlil qilish, fake news aniqlash, spam filtratsiya, mahsulot tasnifi, savol-javob tizimlari va boshqa amaliyotlarda hal qiluvchi ahamiyatga ega.

Qisqa matnlar — odatda 20-50 so'zdan iborat bo'lgan matnlar — uzun hujjatlardan tubdan farq qiladi. Asosiy muammolar quyidagilardan iborat: birinchidan, qisqa matnlar leksik kambag'allikka ega bo'lib, ularda kontekstga oid yetarli ma'lumot bo'lmaydi; ikkinchidan, xususiyatlar fazosi siyrak (sparse) bo'lib, klassik TF-IDF yoki Bag-of-Words usullari samarali emas; uchinchidan, qisqa matnlarda ko'pincha noformal til, slang, qisqartmalar va imlo xatolari uchrab turadi.

Transformer arxitekturasiga asoslangan oldindan o'qitilgan modellar (Pre-trained Language Models — PLMs) ushbu sohada inqilob yaratdi. BERT (Devlin et al., 2018) modeli ikki yo'nalishli kontekstli vektor reprezentatsiyalarini taklif etdi va keyinchalik RoBERTa, DistilBERT, ALBERT, XLM-R kabi turli modifikatsiyalari yaratildi. Biroq, har bir model o'ziga xos kuchli va zaif tomonlarga ega, va ularning qisqa matn tasnifidagi samaradorligi keng tadqiq qilinmagan, ayniqsa kam resursli tillar — jumladan o'zbek tili — uchun.

Adabiyotlardagi mavjud tadqiqotlar asosan ingliz va xitoy tillariga qaratilgan (Sun et al., 2019; Yao et al., 2024). Multilingual modellarning kam resursli til kontekstidagi xulq-atvori, shuningdek transformer modellarini klassik chuqur o'rganish komponentlari (CNN, BiGRU) bilan birlashtirishning maqbul strategiyalari yetarli darajada o'rganilmagan. Bundan tashqari, o'zbek tili uchun qisqa matn tasnifi bo'yicha tizimli tadqiqot deyarli mavjud emas.

Mazkur tadqiqotning maqsadi quyidagilardan iborat:

- zamonaviy pre-trained modellarni (BERT, RoBERTa, DistilBERT, mBERT, XLM-RoBERTa) qisqa matn tasnifi vazifasida qiyosiy baholash;
- har bir modelning kuchli va zaif tomonlarini turli til va domen sharoitlarida aniqlash;
- kontekstual embeddinglarni lokal va ketma-ket xususiyatlar bilan birlashtiruvchi yangi STC-Hybrid arxitekturasini ishlab chiqish;
- taklif etilgan modelni kam resursli o'zbek tilida sinab ko'rish va amaliy tavsiyalar ishlab chiqish.

Tadqiqotning ilmiy yangiligi: birinchi marta beshta zamonaviy transformer modelining qisqa matn tasnifidagi qiyosiy tahlili amalga oshirildi va xususiyatlarni birlashtiruvchi diqqat mexanizmiga asoslangan yangi hybrid arxitektura taklif etildi, hamda o'zbek tili korpusida sinab ko'rildi.

## **2. Adabiyotlar tahlili va nazariy asoslar**

### **2.1. Transformer arxitekturasi va BERT modeli**

Transformer arxitekturasi Vaswani va hammualliflari tomonidan 2017-yilda taklif etilgan bo'lib, u o'z-o'ziga diqqat (self-attention) mexanizmiga asoslangan. BERT (Bidirectional Encoder Representations from Transformers) modeli ushbu arxitekturaning encoder qismidan

foydalanib, ikkita asosiy oldindan o'qitish vazifasini — Masked Language Modeling (MLM) va Next Sentence Prediction (NSP) — qo'llaydi. BERT-base modeli 110 million parametrga ega bo'lib, 12 ta transformer qatlamidan tashkil topgan.

## 2.2. BERT modelining modifikatsiyalari

RoBERTa (Robustly Optimized BERT Approach) modeli Liu va hammualliflari (2019) tomonidan taklif etilgan bo'lib, BERT modelidagi NSP vazifasini olib tashlash, kattaroq mini-batchlar va dinamik masking yordamida samaradorlikni oshirgan. Tadqiqotlar shuni ko'rsatadiki, RoBERTa ko'plab tasniflash vazifalarida BERT-dan yuqori natijalar beradi.

DistilBERT (Sanh et al., 2019) modeli bilim distillyatsiyasi (knowledge distillation) yordamida yaratilgan bo'lib, BERT-base ga nisbatan 40% kam parametrlarga ega va 60% tezroq ishlaydi, lekin sifatning 95% dan ortig'ini saqlab qoladi. Bu uni ishlab chiqarish muhitida samarali tanlovga aylantiradi.

Multilingual BERT (mBERT) 104 tilda Wikipedia korpusida o'qitilgan bo'lib, cross-lingual vazifalar uchun mo'ljallangan. XLM-RoBERTa (Conneau et al., 2019) esa 100 tilda 2,5 TB CommonCrawl ma'lumotlar bazasida o'qitilgan va mBERT-ga nisbatan XNLI benchmark-ida o'rtacha 13,8% yuqori aniqlik ko'rsatadi. Ayniqsa, XLM-R kam resursli tillarda yuqori samaradorlik bilan ajralib turadi: suahili tilida 11,8%, urdu tilida 9,2% yutuq qayd etilgan.

## 2.3. Qisqa matn tasnifi muammosi

Qisqa matn tasnifining asosiy qiyinchiliklari — leksik kambag'allik, kontekst yetishmasligi va xususiyatlar siyrakligi — odatdagi uzun matnlardan farqli o'laroq, klassik usullarning samaradorligini cheklaydi. Li (2019) tomonidan taklif qilingan TB-CNN modeli LDA (Latent Dirichlet Allocation) yordamida mavzular vektorlarini yaratib, xususiyatlar fazosini kengaytirdi. Yao va hamkasblari (2024) BERT-Att-TextCNN modelini taklif etib, BERT embeddinglarini CNN va diqqat mexanizmi bilan birlashtirib, xitoy tilidagi qisqa matnlarda yuqori natijalar oldi.

## 2.4. Hybrid arxitekturalar

Hybrid yondashuvlar transformer modellarining kontekstual qudratini klassik chuqur o'rganish komponentlarining lokal xususiyatlarni aniqlash qobiliyati bilan birlashtirishni maqsad qiladi. RB-GAT modeli (2024) RoBERTa embeddinglarini BiGRU va Graph Attention Network bilan birlashtirib, sezilarli natijalar ko'rsatdi. Ushbu yondashuvlar BERT/CLS tokeni reprezentatsiyasining qisqa matnlarda yetarli emasligi muammosini hal qilishga yo'naltirilgan.

# 3. Metodologiya

## 3.1. Tadqiqot tuzilishi

Tadqiqot uch bosqichdan iborat: (1) standart pre-trained modellarni mavjud datasetlarda fine-tuning qilish va baholash; (2) taklif etilgan STC-Hybrid arxitekturasini ishlab chiqish va o'qitish; (3) qiyosiy tahlil va statistik sinov o'tkazish.

### 3.2. Ma'lumotlar to'plamlari (datasets)

Modellar samaradorligini har tomonlama baholash uchun beshta turli xil dataset tanlandi:

- AG News — 4 sinfdan iborat yangiliklar tasnifi (120 000 o'quv, 7 600 test).
- TREC — savollarni 6 ta sinfga ajratish (5 452 o'quv, 500 test).
- SST-2 (Stanford Sentiment Treebank) — binar sentiment tasnifi (67 349 o'quv).
- Twitter Sentiment140 — Twitter postlari sentiment (1 600 000 namuna; 30 000 ga qisqartirilgan).
- UZ-News — o'zbek tilidagi 6 sinfli yangiliklar (mualliflar tomonidan tuzilgan, 12 000 namuna).

### 3.3. Eksperiment muhiti

Barcha eksperimentlar Python 3.10, PyTorch 2.0 va HuggingFace Transformers 4.35 kutubxonalari yordamida amalga oshirildi. Apparat: NVIDIA RTX A6000 GPU (48 GB VRAM). Modellar 5 epoch davomida fine-tuning qilindi, batch\_size = 32, AdamW optimizatori, learning\_rate = 2e-5, max\_length = 128 token. Har bir model uchun 5 marotaba turli random\_seed bilan o'qitildi va o'rtacha natijalar olindi.

### 3.4. Taklif etilgan STC-Hybrid arxitekturasini

STC-Hybrid (Short Text Classification Hybrid) modeli to'rt asosiy komponentdan iborat:

- Embedding qatlami: BERT yoki XLM-R orqali kontekstual reprezentatsiyalar (768 o'lchamli vektorlar).
- CNN tarmog'i: 3 ta turli o'lchamli filter (2, 3, 4) yordamida lokal n-gram xususiyatlarini aniqlash.
- BiGRU qatlami: 128 yashirin holatga ega ikki yo'nalishli GRU orqali ketma-ket bog'liqliklarni qamrab olish.
- Multi-head attention va konkatenatsiya: CNN va BiGRU chiqishlarini diqqat mexanizmi yordamida tortib, yakuniy softmax tasniflagichga uzatish.

Matematik ko'rinishda, kirish matn  $x = (x_1, x_2, \dots, x_n)$  BERT orqali  $H = \text{BERT}(x) \in \mathbb{R}^{n \times 768}$  ko'rinishida ifodalanadi. CNN chiqishi  $C = \text{MaxPool}(\text{ReLU}(\text{Conv}(H)))$ , BiGRU chiqishi  $G = [\text{GRU} \rightarrow (H); \text{GRU} \leftarrow (H)]$  ko'rinishida hisoblanadi. Yakuniy tasniflash uchun  $z = \text{Attention}([C; G])$  vektori softmax qatlamiga uzatiladi:  $\hat{y} = \text{softmax}(Wz + b)$ .

### 3.5. Baholash mezonlari

Modellar samaradorligi quyidagi metrikalar bilan baholandi: Accuracy (aniqlik), Precision (aniqlik darajasi), Recall (qayta tiklash), F1-score (vaznlashgan o'rtacha) va inference vaqti (millisekundlarda). Statistik ahamiyat paired t-test orqali ( $\alpha = 0,05$ ) tekshirildi.

#### 4. Eksperimental natijalar

##### 4.1. Standart fine-tuning natijalari

1-jadvalda beshta pre-trained modelning besh datasetdagi F1-ko'rsatkichlari keltirilgan. Har bir qiymat 5 ta tasodifiy seed bo'yicha o'rtacha natijani aks ettiradi.

##### 1-jadval. Pre-trained modellarning F1-score qiyosiy natijalari

Model	AG News	TREC	SST-2	Twitter	UZ-News
BERT-base	0,944	0,968	0,921	0,832	0,841
RoBERTa-base	0,949	0,972	0,945	0,851	0,832
DistilBERT	0,938	0,956	0,909	0,824	0,798
mBERT	0,932	0,944	0,898	0,815	0,853
XLM-RoBERTa	0,946	0,964	0,932	0,842	0,871

Natijalardan ko'rinadiki, ingliz tilidagi datasetlarda RoBERTa eng yuqori ko'rsatkichlarni qayd etgan (o'rtacha F1 = 0,929). Biroq, kam resursli o'zbek tili (UZ-News) datasetida XLM-RoBERTa modeli boshqa barcha modellardan ustun chiqdi (F1 = 0,871). Bu XLM-R modelining multilingual o'qitilishi va kam resursli tillar uchun yaxshilangan reprezentatsiyalari bilan izohlanadi. DistilBERT esa sifat va tezlik nisbatida yaxshi muvozanat ko'rsatdi.

##### 4.2. STC-Hybrid modelining natijalari

2-jadvalda taklif etilgan STC-Hybrid arxitekturasining standart fine-tuning bilan qiyosiy natijalari keltirilgan. Hybrid model uchun asos sifatida ingliz datasetlarida BERT, o'zbek datasetida esa XLM-RoBERTa tanlandi.

##### 2-jadval. STC-Hybrid va baseline modellarning F1 qiyoslash natijalari

Dataset	Eng yaxshi baseline	F1 (baseline)	STC-Hybrid F1	Yutuq ( $\Delta$ %)
AG News	RoBERTa-base	0,949	0,973	+2,4

TREC	RoBERTa-base	0,972	0,988	+1,6
SST-2	RoBERTa-base	0,945	0,968	+2,3
Twitter	RoBERTa-base	0,851	0,892	+4,1
UZ-News	XLM-RoBERTa	0,871	0,887	+1,6

Eksperimental natijalar shuni ko'rsatadiki, STC-Hybrid arxitekturasi barcha datasetlarda baseline modellardan ustun chiqdi. Eng katta yutuq Twitter Sentiment140 datasetida qayd etilgan (+4,1 foiz punkt), bu shovqinli va qisqa matnlarda hybrid arxitekturaning samaradorligini ko'rsatadi. Paired t-test natijalariga ko'ra, barcha yutuqlar statistik jihatdan ahamiyatli ( $p < 0,01$ ).

#### 4.3. Inference vaqti tahlili

Modellarning real vaqt sharoitidagi qo'llanilishi nuqtai nazaridan inference tezligi muhim parametrdir. 3-jadvalda Twitter Sentiment140 datasetida bitta namunani tasniflash uchun ketgan o'rtacha vaqt (ms) keltirilgan.

#### 3-jadval. Modellarning inference vaqti (RTX A6000, batch\_size=1)

Model	Inference vaqti (ms)	Parametrlar soni (M)
DistilBERT	8,3	66
BERT-base	14,7	110
RoBERTa-base	15,2	125
mBERT	16,9	178
XLM-RoBERTa	18,4	270
STC-Hybrid (BERT)	19,8	118
STC-Hybrid (XLM-R)	23,5	278

DistilBERT eng tez model (8,3 ms) sifatida real vaqt tizimlari uchun mos hisoblanadi. STC-Hybrid modelining qo'shimcha komponentlari (CNN va BiGRU) inference vaqtini taxminan 30-35% ga oshiradi, lekin sifat yutug'i bunday sarmoyani oqlaydi.

### 5. Muhokama

Olingan natijalar oldingi tadqiqotlar bilan mos keladi va ularni kengaytiradi. Avval Sun va hammualliflari (2019) BERT-ning qisqa matn tasnifi vazifasidagi samaradorligini ko'rsatgan, biroq kam resursli tillar masalasi yetarli yoritilmagan edi. Bizning natijalarimiz XLM-RoBERTa modelining o'zbek tili kabi morfologik boy va kam resursli tillar uchun afzalligini tasdiqlaydi, bu Conneau va hammualliflarining (2019) cross-lingual transfer haqidagi xulosalariga muvofiqdir.

Hybrid arxitektura yondashuvi Yao va hamkasblari (2024) tomonidan xitoy tili uchun ko'rsatilgan natijalarga o'xshash yutuqlarni boshqa tillarda ham tasdiqlaydi. STC-Hybrid arxitekturasidagi diqqat mexanizmi orqali CNN va BiGRU chiqishlarini birlashtirish — ushbu tadqiqotning asosiy hissasi — qisqa matnlardagi leksik siyraklik muammosini hal etishga yo'naltirilgan. CNN qatlami lokal n-gram darajasidagi tasniflovchi xususiyatlarni aniqlasa, BiGRU qisqa kontekstdagi ketma-ket bog'liqliklarni saqlab qoladi.

Twitter datasetida eng katta yutuq (+4,1%) qayd etilgani tasodifiy emas: shovqinli, slang va qisqartmalarga boy matnlarda hybrid arxitektura ayniqsa samarali bo'ladi, chunki CNN lokal naqshlarni (masalan, hashtag yoki emotikon ketma-ketliklari) aniqlay oladi, BiGRU esa cheklangan kontekstda ham semantik bog'liqliklarni tutadi.

Tadqiqotning cheklovlari ham mavjud: birinchidan, UZ-News dataseti hajmi nisbatan kichik (12 000 namuna), bu modellarning kengroq baholanishini cheklaydi; ikkinchidan, ko'p tilli aralash matnlar (code-mixing) holatlari alohida o'rganilmagan; uchinchidan, taklif etilgan model 512 tokendan oshmaydigan matnlar uchun mo'ljallangan.

### 6. Xulosa

Mazkur tadqiqotda pre-trained transformer modellari yordamida qisqa matnlarni tasniflash algoritmlari batafsil tadqiq qilindi. Beshta zamonaviy modelning (BERT, RoBERTa, DistilBERT, mBERT, XLM-RoBERTa) qiyosiy tahlili shuni ko'rsatdiki, ingliz tilidagi datasetlarda RoBERTa eng yuqori sifat ko'rsatadi, kam resursli tillar uchun esa XLM-RoBERTa afzal hisoblanadi. DistilBERT real vaqt va resurslar cheklangan muhitlarda eng yaxshi tanlovdir.

Tadqiqotning asosiy ilmiy hissasi sifatida BERT/XLM-R, CNN va BiGRU komponentlarini diqqat mexanizmi orqali birlashtiruvchi STC-Hybrid arxitekturasi taklif etildi. Eksperimental natijalar yangi modelning o'rtacha 2,4–4,1 foiz punkt sifat yutug'iga olib kelishini tasdiqladi. O'zbek tilidagi yangiliklar datasetida  $F1 = 0,887$  ko'rsatkichiga erishildi, bu kam resursli tillar uchun amaliy ahamiyatga ega.

Amaliy tavsiyalar:

- Resurslar cheklangan ishlab chiqarish muhitlarida DistilBERT yoki uning hybrid varianti tavsiya etiladi.
- Kam resursli tillar (jumladan o'zbek tili) uchun XLM-RoBERTa asoslangan hybrid arxitekturadan foydalanish maqsadga muvofiqdir.
- Shovqinli ijtimoiy tarmoq matnlari (Twitter, Telegram) uchun STC-Hybrid modeli eng yaxshi natija beradi.

Kelajakdagi tadqiqotlar yo'nalishlari quyidagilarni o'z ichiga oladi: (1) o'zbek tili korpusini kengaytirish va UzBERT kabi monolingual modellarni qiyosiy o'rganish; (2) Large Language Models (LLM) bilan kichik transformerlarning qiyosiy tahlili; (3) code-mixing va dialektal matnlar uchun adaptiv hybrid yondashuvlar ishlab chiqish; (4) prompt engineering va parameter-efficient fine-tuning (LoRA, adapters) usullarini hybrid arxitekturalarga integratsiya qilish.

### Foydalanilgan adabiyotlar

1. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. — 2019. — pp. 4171–4186. DOI: 10.18653/v1/N19-1423.
2. Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. RoBERTa: A Robustly Optimized BERT Pretraining Approach // arXiv preprint arXiv:1907.11692. — 2019.
3. Sanh V., Debut L., Chaumond J., Wolf T. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter // arXiv preprint arXiv:1910.01108. — 2019.
4. Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzmán F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. Unsupervised Cross-lingual Representation Learning at Scale // Proceedings of ACL. — 2020. — pp. 8440–8451.
5. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention Is All You Need // Advances in Neural Information Processing Systems (NeurIPS). — 2017. — Vol. 30.
6. Sun C., Qiu X., Xu Y., Huang X. How to Fine-Tune BERT for Text Classification? // China National Conference on Chinese Computational Linguistics. — Springer. — 2019. — pp. 194–206.
7. Yao R., Hou L., Ye Y., Wu O., Zhang J., Wu J. A BERT-Based Hybrid Short Text Classification Model Incorporating CNN and Attention-Based BiGRU // Journal of Organizational and End User Computing. — 2024. — Vol. 36, № 1. — pp. 1–22.

8. Li Y. Short Text Classification Improved by Feature Space Extension // arXiv preprint arXiv:1904.01313. — 2019.
9. Wang Y., Wang C., Zhan J., Ma W., Jiang Y. RB-GAT: A Text Classification Model Based on RoBERTa-BiGRU with Graph Attention Network // Sensors. — 2024. — Vol. 24, № 11. — Article 3365. DOI: 10.3390/s24113365.
10. Minaee S., Kalchbrenner N., Cambria E., Nikzad N., Chenaghlu M., Gao J. Deep Learning Based Text Classification: A Comprehensive Review // ACM Computing Surveys. — 2021. — Vol. 54, № 3. — pp. 1–40.
11. Kowsari K., Jafari Meimandi K., Heidarysafa M., Mendu S., Barnes L., Brown D. Text Classification Algorithms: A Survey // Information. — 2019. — Vol. 10, № 4. — Article 150.
12. Howard J., Ruder S. Universal Language Model Fine-tuning for Text Classification // Proceedings of ACL. — 2018. — pp. 328–339.
13. Lan Z., Chen M., Goodman S., Gimpel K., Sharma P., Soricut R. ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations // Proceedings of ICLR. — 2020.
14. Clark K., Luong M.-T., Le Q. V., Manning C. D. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators // Proceedings of ICLR. — 2020.
15. Kim Y. Convolutional Neural Networks for Sentence Classification // Proceedings of EMNLP. — 2014. — pp. 1746–1751.
16. Cho K., van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation // Proceedings of EMNLP. — 2014. — pp. 1724–1734.
17. Joulin A., Grave E., Bojanowski P., Mikolov T. Bag of Tricks for Efficient Text Classification // Proceedings of EACL. — 2017. — pp. 427–431.
18. Brown T., Mann B., Ryder N., Subbiah M. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems. — 2020. — Vol. 33. — pp. 1877–1901.
19. Yang Z., Dai Z., Yang Y., Carbonell J., Salakhutdinov R., Le Q. V. XLNet: Generalized Autoregressive Pretraining for Language Understanding // Advances in Neural Information Processing Systems. — 2019. — Vol. 32.
20. Mamatov N. S., Kobilov S. S., Niyozmatova N. A. O'zbek tilini kompyuter lingvistikasi muammolari // Tabiiy fanlar va matematika xabarnomasi. — 2022. — № 3. — pp. 45–53.