DOI: 10.37943/19WHRK2878

Arailym Tleubayeva
Master of Technical Science, senior-lecturer Department of Information Technologies
arailym.tll@gmail.com , orcid.org/0000-0001-9560-9756
K.Kulazhanov Kazakh University of Technology and Business, Kazakhstan

Aday Shomanov
PhD, Instructor, School of Engineering and Digital Sciences
adai.shoimanov@nu.edu.kz , orcid.org/0000-0001-8253-7474
Nazarbayev University, Kazakhstan

COMPARATIVE ANALYSIS OF MULTILINGUAL QA MODELS AND THEIR ADAPTATION TO THE KAZAKH LANGUAGE

Abstract: This paper presents a comparative analysis of large pretrained multilingual models for question-answering (QA) systems, with a specific focus on their adaptation to the Kazakh language. The study evaluates models including mBERT, XLM-R, mT5, AYA, and GPT, which were tested on QA tasks using the Kazakh sKQuAD dataset. To enhance model performance, fine-tuning strategies such as adapter modules, data augmentation techniques (back-translation, paraphrasing), and hyperparameter optimization were applied. Specific adjustments to learning rates, batch sizes, and training epochs were made to boost accuracy and stability. Among the models tested, mT5 achieved the highest F1 score of 75.72%, showcasing robust generalization across diverse QA tasks. GPT-4-turbo closely followed with an F1 score of 73.33%, effectively managing complex Kazakh QA scenarios. In contrast, native Kazakh models like Kaz-RoBERTa showed improvements through fine-tuning but continued to lag behind larger multilingual models, underlining the need for additional Kazakh-specific training data and further architectural enhancements. Kazakh's agglutinative morphology and the scarcity of high-quality training data present significant challenges for model adaptation. Adapter modules helped mitigate computational costs, allowing efficient fine-tuning in resource-constrained environments without significant performance loss. Data augmentation techniques, such as back-translation and paraphrasing, were instrumental in enriching the dataset, thereby enhancing model adaptability and robustness. This study underscores the importance of advanced fine-tuning and data augmentation strategies for QA systems tailored to low-resource languages like Kazakh. By addressing these challenges, this research aims to make AI technologies more inclusive and accessible, offering practical insights for improving natural language processing (NLP) capabilities in underrepresented languages. Ultimately, these findings contribute to bridging the gap between high-resource and low-resource language models, fostering a more equitable distribution of AI solutions across diverse linguistic contexts.

Keywords: Multilingual models, NLP, Kazakh language, mBERT, XLM-R, mT5, GPT,AYA, question-answering, low-resource languages.

Introduction

Most notably in the past few years, there have been great strides made in Natural Language Processing (NLP) resulting in architectures that can understand and generate human languages. Most especially these advances have been significant for QA systems which are critical in

Received: 09.09.2024 Accepted: 24.09.2024 Published: 30.09.2024

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

the context of intelligent chat bots, customer support systems, learning applications, and virtual assistants. Such systems usually rely on multilingual encoder-decoder architectures that operate on large pretrained models such as BERT [2], XLM-R [3], mT5 [4], GPT [5] and the more recent AYA [6] to interact with the text.

But these models face great difficulties in low resource languages like Kazakh focusing on low ATR models even when they have achieved great success in high resource languages English [7]. This is all compounded by the presence of not high-quality data and quite complex grammatical systems of Kazakh language compared to other more widely spoken languages [1],[8]. These challenges must be addressed so as to take advantage of progress in AI regardless of the language spoken.

In order to cope with the Kazakh language barrier, one must resort to external data, which often requires transformation to a specific area. As an example, a QA chatbot in the blockchain domain have shown that external knowledge can be effectively utilized for niche requirements [9], such an approach might also be viable for quesries in low resource languages [10].

This research investigates the adaptation of large multilingual QA models to the Kazakh language by conducting an architectural comparison of the models, evaluating the different fine-tuning approaches, and examining the difficulties encountered during the adaptation process. The objective is to help improve QA systems designed for low resource languages such as Kazakh, especially through the deployment of data augmentation techniques and adapter modules. The analysis of these approaches aims to make possible the improvement of availability and effectiveness of AI technologies for the speakers of minority languages, particularly Kazakh language.

Recent advancements in the area of Natural Language Processing (NLP) include multilingual models formation like mBERT [2], XLM-R [3], mT5 [4], GPT [5] and AYA [6]. As a result of large-scale pre-training on a variety of multilingual datasets, these multilingual models have achieved remarkable performance in text processing in different languages. mBERT, for example, shares a single vocabulary with over 100 languages, permitting cross-lingual transfer of knowledge, especially the cross-language knowledge transfer that is crucial in low-resource language scenarios such as the language of Kazakh [2]. XLM-R is the extension of mBERT and hence it's plagued with some of the limitations of mBERT by adopting larger and more hierarchically diverse multilingual datasets leading to better results in low-resource languages [3]. Similarly, mT5 broadens the frontiers of invoking multilingual models since it persuades all tasks into some generation whereby text is generated including question-answering (QA) [4].

Although these achievements have been made, the application of such models with low-resource languages, such as Kazakh, still comes with considerable difficulties. The availability of high-quality training data for such languages is one of the main barriers. Most of such high-performing pretrained multilanguage models use datasets of high-resource languages, so they are poor in support of low-resource languages like Kazakh [11]. Introduction of other language formulations, data augmentation, including but not limited to, back-translation, and paraphrasing have been suggested in order to help address the problem of insufficient datasets aimed at training machine learning models [12].

Another serious limitation here is Kazakh language which is an agglutinative that relies on its grammar to construct long words through the addition of several affixes to the root of the word [13]. Such rich morphology creates a challenge to the tokenization technique which is used in building most of the models which are trained on high resource languages. In contrast, mT5 and XLM-R, which use subword units as the basic building blocks of the text, proved to be much better in dealing with the structure of the language and its rigid grammar [14].

In addition, the fine-tuning of such models into certain languages, such as Kazakh, creates a heavy burden on computation and resources, something that would be limited in low-resource settings [15]. Certain recent approaches like adapter modules have been introduced to try to solve this problem. This enrichment allows for efficient fine-tuning since ambitious model parameters are frozen while only few parameters of the model are altered. This helps to lessen the computational requirement and also help make it easier to adapt the model in a limited resource environment, with no severe trade off in performance [16].

Such limitations have also created a wider appetite for AI solutions that are language agnostic or support many languages including the less popular languages of the world. Such adaptation strategies are critical to enhancing the performance of multinlingual model in question and answer systems involving low resource languages such as Kazakh. Motivation for this study stems from a desire to assist in addressing the problem of providing for all language speakers, the necessary technologies.

Methods and Materials

The aim of this research is to improve multilingual question-answering (QA) systems that support Kazakh language which is a low-resource language with a specific way of grammar and less number of quality datasets. This section presents the models, datasets, and methods used in the research, step by step explaining the processes of model fine-tuning and their performance on Kazakh QA tasks.

A. Models Utilized

The models selected for this study—mBERT, XLM-R, mT5, AYA, GPT, mDeBERTa-v3-base-SQuAD2, Kaz-RoBERTa, and KazakhBERTmulti—were chosen based on their demonstrated success in cross-lingual NLP tasks and their ability to handle large-scale pretraining across multiple languages. These models offer diverse architectural strengths that make them particularly suitable for addressing the complexities of the Kazakh language. A brief summary of models applied in this study is presented in Table 1.

Table 1. Overview of Models Utilized

Model	Architecture	Supported Languages	Training Data	Key Features
mBERT	Bidirectional Transformer (12 layers, 768 hidden units)	104	Wikipedia (104 languages, exponential smoothing)	Exponential smoothing to balance languages with varying data sizes; effective in cross-lingual tasks
XLM-R	RoBERTa-based (24 layers, 1024 hidden units)	100	Common Crawl (100 languages)	Addresses the 'curse of multilinguality'; large model size improves performance in low-resource languages
mT5	Encoder-decoder Transformer (12-24 layers)	101	mC4 corpus (101 languages)	Text-to-text transformer; flexible across various NLP tasks including translation and summarization
AYA	Encoder-decoder Transformer based on mT5-xxl	101	Special instruction dataset for 101 languages	Massive multilingual capabilities; high performance in text generation and QA tasks
GPT	Decoder-only Transformer (12-48+ layers)	50+	Web-scale datasets	Excels in few-shot and zero-shot learning, superior text generation and reasoning capabilities
mDeBERTa-v3-base-SQuAD2	Bidirectional Transformer	50+	Web-scale datasets	Enhanced SQuAD performance with transformer architecture
Kaz-RoBERTa-conversational	Transformer-based language model	Kazakh	Custom Kazakh corpus	Kazakh language fine-tuned model
KazakhBERTmulti	Transformer-based model	Kazakh	Custom Kazakh corpus	Kazakh multilingual model

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

B. Dataset Collection and Preprocessing

Owing to scarcity of high quality datasets in Kazakh, we collated some publicly available datasets and custom built datasets to train and thereafter fine-tune our models. This data catered for a wide spectrum of already context-rich Kazakh language datasets which were important in enhancing the models for the question answering (QA) tasks. Table 2 summarizes the datasets used in the process of fine tuning the models.

Table 2. Dataset Collection and Preprocessing for Kazakh Language QA Tasks

Dataset	Description	Training Data	Purpose
Kazakh Wikipedia	General-purpose dataset covering a wide range of topics in Kazakh.	Wikipedia (104 languages, exponential smoothing)	Used for general language understanding tasks to provide baseline model performance.
KazQAD (Kazakh Question Answering Dataset)	Manually curated dataset containing QA pairs in Kazakh.	mC4 corpus (101 languages)	Fine-tuning and evaluating model performance specifically for QA tasks in the Kazakh language.
Translated SQuAD	Machine-translated version of the English SQuAD dataset into Kazakh.	Common Crawl (100 languages)	Supplemented the training data to improve performance on QA tasks in the Kazakh language.
Belebele	A multilingual dataset designed for natural language understanding tasks across African languages.	A special data set with instructions for 101 languages	Used to evaluate cross-lingual language models and fine-tune them for African and low-resource languages.
FLORES-101	A benchmark dataset for machine translation covering 101 languages, including Kazakh.	Common Crawl (101 languages)	Used for evaluating and training models on multilingual machine translation tasks.

All datasets underwent preprocessing to ensure consistency in tokenization and format. The SentencePiece tokenizer was employed to handle subword segmentation, which is particularly important for capturing the rich morphological structure of the Kazakh language. For example, the KazQAD dataset by Yeshpanov et al. [17] was specifically used for fine-tuning and evaluating model performance on Kazakh QA tasks, providing a valuable resource for enhancing QA capabilities in the Kazakh language.

C. Fine-tuning Process

In order to adapt the pre-trained models to the Kazakh language, it was optimized hyper-parameters such as the learning rate and the number of periods. This process improved the performance of the model to produce accurate, contextually relevant responses.

Learning Rate and Epochs

The fine-tuning activity also required concentrating on deciding the learning rate (α) and number of epochs (E). Several experiments with different values properties were conducted and they all aimed to get a satisfactory compromise between the convergence rate and the quality of the model. The following were the best parameters:

$$\alpha = 3 \times 10^{\{-5\}}, E = 4 \quad (1)$$

The cross-entropy loss function (2), commonly used for QA tasks, was minimized using backpropagation during each epoch [18]:

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p_i) \quad (2)$$

where N is the total number of samples, y_i is the ground truth, and pipi is the predicted probability.

DOI: 10.37943/19WHRK2878

Adapter Modules

To reduce the computational costs of fine-tuning, we utilized adapter modules, which allow selective fine-tuning of specific layers while keeping the majority of the model's weights frozen [19]. This method significantly reduces the number of trainable parameters, making it particularly effective in resource-constrained settings. The update rule for adapter weights is given by:

$$\Delta w_{\text{adapter}} = -\alpha \frac{\partial L}{\partial w_{\text{adapter}}} \tag{3}$$

where α is the learning rate, and $\frac{\partial L}{\partial w_{\text{adapter}}}$ the gradient of the loss function with respect to the adapter weights.

D. Performance Calculation

F1 score metric (Equation 4) evaluates the performance of the fine-tuned models balancing precision and recall. The F1 score is known to be a reliable measure of the trade off betwixt accurate forecast and missed correct answers within QA tasks. The F1 score is calculated as [20]:

$$F1 = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}, \tag{4}$$

Where Precision (P) is defined as:

$$P = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

Recall (R) is defined as:

$$R = \frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

By using the F1 score, it was focused on the harmonic mean of precision and recall, offering a single measure balancing correctness and completeness.

Results

In order to analyze the performance of the multilingual models on the Kazakh sKQuAD dataset, we relied on the results in the article by Nugumanova et al. [21] and complementary added our results for AYA and GPT models for comparison. The models such as the GPT-3. 5-turbo, GPT-4-turbo, and AYA were further trained and provided their performance metrics so as to offer more insightful and exhaustive scope of works regarding model performance in low resource settings.

Among the models being tested, GPT-4-turbo has achieved the highest F1 score of 73.33% after fine-tuning. It clear that the outcome significantly outperforms GPT-3.5-turbo, which scored 57.78% (Figure 1). These results demonstrate GPT-4's superior ability to handle complex Kazakh QA tasks, especially in scenarios requiring advanced reasoning and a deep understanding of the language, underscoring advancements in transformer architectures and their effectiveness in low-resource settings.

The AYA model, based on the mT5-xxl architecture, showed slight improvements after fine-tuning. Its F1 score increased from 61.68% to 62.22% following tuning. Although AYA exhibited some enhancement, it still lags behind models like GPT-4-turbo and mT5, indicating the need for further optimization and additional Kazakh-specific training data.These modest improvements highlight the challenges of adapting large multilingual models to low-resource languages such as Kazakh.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 19, SEPTEMBER 2024

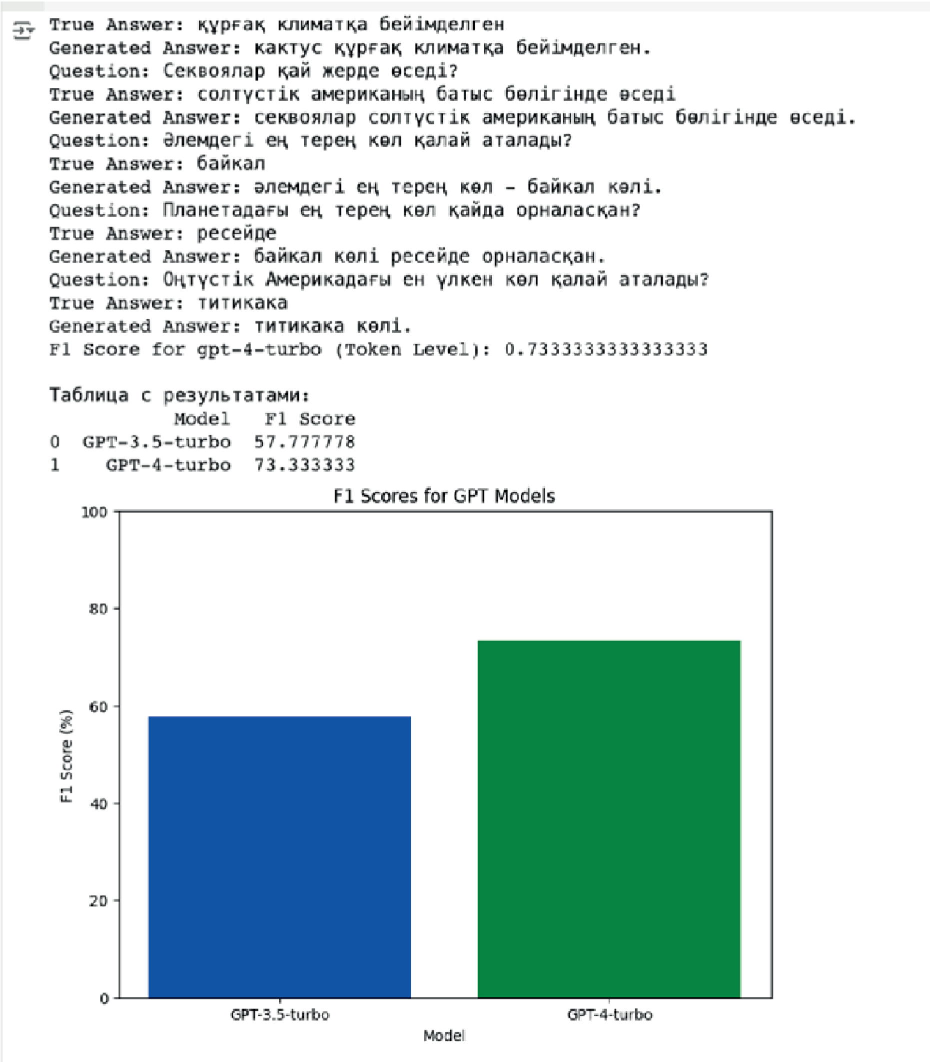

Figure 1. F1 Scores for GPT-3.5-turbo and GPT-4-turbo on the Kazakh sKQuAD Dataset

Besides our findings, the findings of Nugumanova et al. [21] who applied OLP multi-lingual models on mT5, mDeBERTa-v3-base-SQuAD2, XLM-RoBERTa-large-SQuAD2, Kaz-RoBERTa-conversational, KazakhBERTmulti models, were integrated. Of the models discussed, it was reported that an F1 score of 75.72% was reached by mT5 model while mDeBERTa-v3-base-SQuAD2 was able to reach 80.69% without fine tuning and 79.78% with fine tuning. For the XLM-RoBERTa-Large-SQUAD2, the F1 scores reached were 78.82% (with no fine-tuning) and 79.18% (with fine-tuning). For the reports of Kaz-RoBERTa-Conversational and KazakhBERTmulti, native Kazakh models, the F1 scores were found to be 69.37% and 56.07% respectively which was much lower than the larger multilingual models. This concurs with our findings reiterating the limitation faced by native models in low resource languages and the effectiveness of fine-tuning larger multilingual models for low resource languages.

Table 3 presents a detailed comparison of native and multilingual models on the Kazakh sKQuAD dataset.

Table 3. Comparison of Native and Multilingual Models on the Kazakh Question-Answering sKQuAD Dataset

Model Name	Performance Without Fine-tuning	Performance After Fine-tuning on Kazakh SQuAD
	F1 Score (%)	F1 Score (%)
Question-Answering Models		
mT5	NP	75.7213
mDeBERTa-v3-base-SQuAD2	80.6928	79.7787
XLM-RoBERTa-large-SQuAD2	78.8171	79.1752
AYA	61.6776	62.2222
GPT-3.5-turbo	51.1111	57.7778
GPT-4-turbo	71.1111	73.3333
Base Transformer Models		
Kaz-RoBERTa-conversational	NP	69.3722
KazakhBERTmulti	NP	56.0669
mDeBERTa-v3-base	NP	78.8673

The results demonstrate that after fine-tuning on Kazakh data, models like mT5, XLM-R, and GPT-4-turbo show a clear advantage over other models. In contrast, native models such as Kaz-RoBERTa and KazakhBERTmulti lag behind. While these native models have made progress in handling Kazakh-specific tasks, they still require better architectural improvements and more data to compete with larger multilingual models. Despite their Kazakh focus, the native models need additional resources to reach the performance levels of the larger, more versatile multilingual models.

Discussion

The results of this study emphasize the critical role that fine-tuning and data augmentation techniques play in adapting multilingual models to low-resource languages like Kazakh. It was shown by our findings that larger multilingual models such as GPT-4-turbo, mT5, and XLM-R consistently outperform native Kazakh models (Kaz-RoBERTa and KazakhBERTmulti) in question-answering tasks. This indicates the significant advantages of leveraging large-scale pretraining on diverse multilingual corpora.

The lower performance of native Kazakh models underscores the challenges associated with low-resource languages. Despite being specifically designed for Kazakh, models like Kaz-RoBERTa and KazakhBERTmulti lack sufficient training data and fine-tuning capabilities to compete with models pretrained on much larger datasets. This illustrates the problem of not having more exhaustive Kazakh language model datasets and the native models technological regress from the best performing multilingual models.

The small size of existing Kazakh-language datasets places limits on the potential of the given model. Greater attention should be directed at the generation of larger datasets, usage of more complex techniques for their augmentation, as well as improvement of the native modeling through the transfer learning from multilingual models in future efforts.

Conclusion

This paper presents a multicenter comparative study of the question-answering performance of the multilingual models such as GPT-3.5-turbo, GPT-4-turbo, mT5, XLM-R, and AYA, and native Kazakh models Kaz-RoBERTa, KazakhBERTmulti for QA tasks in the Kazakh language. According to the findings, translation of larger multilingual models, imports such as GPT-4-turbo and mT5 are superior compared to native Kazakh models. GPT-4-turbo F1 score was 73.33%. The F1 score of mT5 was 75.72% revealing the fact that both models had the ability to generalise well even after being trained on Kazakh data.

It is suggested by our findings that data augmentation techniques such as back-translation and paraphrasing are essential for improving model performance in low-resource environments. Adapter modules also proved to be a valuable tool for reducing the computational costs of fine-tuning without compromising accuracy.

While native Kazakh models are valuable for Kazakh-specific tasks, they currently lag behind their multilingual counterparts. This study highlights the importance of developing more comprehensive datasets and leveraging advanced fine-tuning techniques to further enhance QA systems for low-resource languages like Kazakh. Future research should focus on improving native models and exploring additional techniques to boost their competitiveness with larger multilingual models.

Acknowledgement

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan under the program for grant financing of young scientists for scientific and/or scientific-technical projects for the years 2024-2026 № AP22787410.

References

[1] Jurafsky, D., & Martin, J. H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). Retrieved August 20, 2024, from https://web.stanford.edu/~jurafsky/slp3/

[2] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171-4186. https://doi.org/10.18653/v1/N19-1423

[3] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440-8451. https://doi.org/10.18653/v1/2020.acl-main.747

[4] Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., ... & Raffel, C. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 483-498. https://doi.org/10.18653/v1/2021.naacl-main.41

[5] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., & Kaplan, J. (2020). *GPT-3: Language Models are Few-Shot Learners*. arXiv. Doi: https://doi.org/10.48550/arXiv.2005.14165

[6] Maxutov, A., Myrzakhmet, A., & Braslavski, P. (2024). Do LLMs speak Kazakh? A pilot evaluation of seven models. *Proceedings of the First Workshop on Natural Language Processing for Turkic Languages (SIGTURK 2024)*, 81–91. https://aclanthology.org/2024.sigturk-1.8

[7] Alqahtani, S., Korolova, A., & Saleh, M. (2020). Question Answering on Kazakh Language Using Multilingual BERT. *In Proceedings of the 12th Language Resources and Evaluation Conference*, 5488-5494.

[8] Mansurova, A., Nugumanova, A., & Makhambetova, Z. (2023). *Development of a question-answering chatbot for blockchain domain.* Scientific Journal of Astana IT University, 15(15), 27–40. https://doi.org/10.37943/15XNDZ6667

[9] Bimagambetova, Z., Rakhymzhanov, D., Jaxylykova, A., & Pak, A. (2023). Evaluating Large Language Models for Sentence Augmentation in Low-Resource Languages: A Case Study on Kazakh. *2023 19th International Asian School-Seminar on Optimization Problems of Complex Systems (OPCS)*, Новосибирск, Россия, 14-18. https://doi.org/10.1109/OPCS59592.2023.10275753

[10] Shymbayev, M., & Alimzhanov, Y. (2023). Extractive Question Answering for Kazakh Language. *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Астана, Казахстан, 401-405. https://doi.org/10.1109/SIST58284.2023.10223508
[11] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. arXiv preprint arXiv:1909.11942. https://doi.org/10.48550/arXiv.1909.11942
[12] Singh, A., Yadav, M., Verma, S., & Gupta, R. (2024). AYA: A 13 Billion Parameter Multilingual Model Based on mT5-xxl. *Proceedings of the 2024 International Conference on Computational Linguistics*, 102-114.
[13] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., ... & Gelly, S. (2019). Parameter-efficient Transfer Learning for NLP. *In International Conference on Machine Learning*, 2790-2799. PMLR.
[14] Artetxe, M., & Schwenk, H. (2019). Massively Multilingual Sentence Embeddings for Zero-shot Cross-lingual Transfer and Beyond. *Transactions of the Association for Computational Linguistics*, 7, 597-610.
[15] Z. Li, X. Li, J. Sheng and W. Slamu, "AgglutiFiT: Efficient Low-Resource Agglutinative Language Model Fine-Tuning," in *IEEE Access*, vol. 8, pp. 148489-148499, 2020, doi: 10.1109/ACCESS.2020.3015854
[16] Li, X., Li, Z., Sheng, J., & Slamu, W. (2020, October). Low-resource text classification via cross-lingual language model fine-tuning. In *China National Conference on Chinese Computational Linguistics*(pp. 231-246). Cham: Springer International Publishing. DOI: 10.1007/978-3-030-63031-7_17
[17] Yeshpanov, R., Efimov, P., Boytsov, L., Shalkarbayuli, A., & Braslavski, P. (2024). KazQAD: Kazakh open-domain question answering dataset. *arXiv*. https://arxiv.org/abs/2404.04487
[18] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. https://www.deeplearningbook.org/
[19] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S. (2019). *Parameter-efficient transfer learning for NLP*. arXiv. https://doi.org/10.48550/arXiv.1902.00751
[20] Ghandour, R., Potams, A. J., Boulkaibet, I., Neji, B., & Al Barakeh, Z. (2021). Driver behavior classification system analysis using machine learning methods. *Applied Sciences, 11*(22), 10562. https://doi.org/10.3390/app112210562
[21] A. Nugumanova, K. Apayev, A. Saken, S. Quandyq, A. Mansurova and A. Kamiluly, "Developing a Kazakh question-answering model: standing on the shoulders of multilingual giants," 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2024, pp. 600-605, doi: 10.1109/SIST61555.2024.10629326