Data Augmentation For Q&A Dataset In Healthcare

Report

Alexandra Matveeva
Faculty of Digital Transformation
ITMO University
Saint Petersburg, Birzhevaya, 4
alexandra.matveeva.s@gmail.com

Darya Murova
Faculty of Digital Transformation
ITMO University
Saint Petersburg, Birzhevaya, 4

Alexander Semiletov
Faculty of Digital Transformation
ITMO University
Saint Petersburg, Birzhevaya, 4

Igor Chernov
Faculty of Digital Transformation
ITMO University
Saint Petersburg, Birzhevaya, 4

January 9, 2024

Abstract

The increasing significance of Question and Answering (Q&A) systems in healthcare necessitates the development of robust datasets. However, the availability of labeled data in healthcare domains is often limited, hindering the performance of machine learning models. This paper explores the application of data augmentation techniques to enrich Q&A datasets in healthcare, with a focus on various augmentation methods for text datasets and the utilization of Large Language Models (LLMs) for augmentation. The experiment involved fine-tuning the T5 model for a specific Question-Answer task and served as a methodological study to determine the feasibility and effectiveness of different approaches to data augmentation. The experiments revealed that the most robust strategy involved the use of extractive generalization techniques to obtain context and abstract generalization techniques for answers. It is important to note that in this particular scenario, the evaluation metrics, specifically BLEU and ROUGE-L, showed lower values compared to the fine-tuning process applied directly to the original dataset.

Keywords Data augmentation · Large Language Models · Question and Answering · Data generation · T5 · Fine-Tuning Large Language Models

1 Introduction

In the changing world of healthcare, incorporating advanced technologies like Question and Answering (Q&A) systems is crucial to make things more accessible and efficient. But to make these systems better, we need good datasets. The problem is, there's not enough labeled data available for healthcare, and that makes it harder to develop and improve these systems in Russian in particular. Addressing this constraint, data augmentation techniques emerge as a promising avenue to enrich existing datasets and enhance the performance of machine learning models tailored for Q&A tasks in healthcare.

Data augmentation of text refers to the process of creating variations in a given text dataset to increase its diversity and size. This technique is commonly employed in natural language processing (NLP) and machine learning to enhance model performance, especially when dealing with limited labeled data. The goal is to expose the model to a broader range of examples, improving its generalization and robustness.

Standard Data Augmentation Approaches A comprehensive overview of traditional augmentation methods, including synonym replacement, back translation, and paraphrasing. The paper [1] showcases how straightforward data

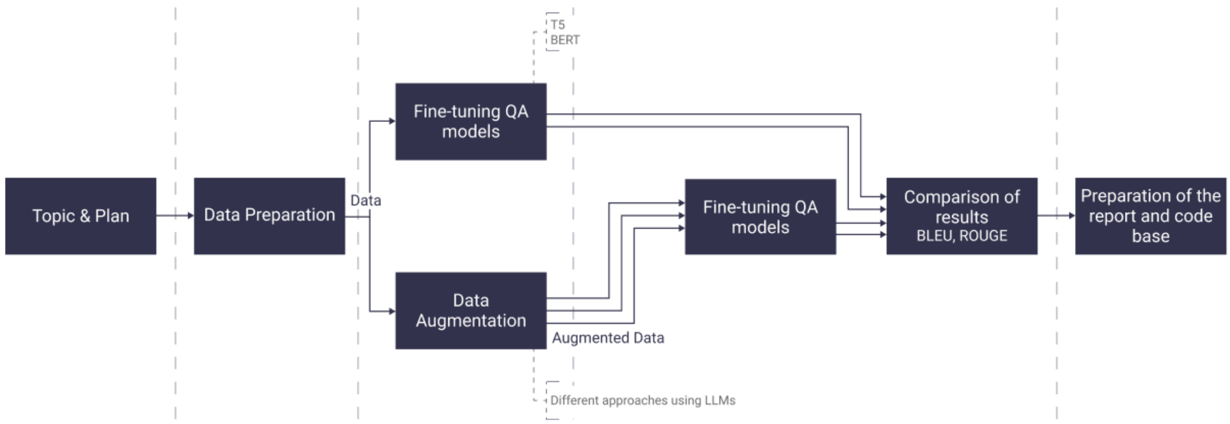

Figure 1: Project execution sequence

augmentation techniques, such as synonym replacement, random insertion, random swap, and random deletion, can enhance performance in text classification tasks using knowledge bases like WordNet [2].

Data Augmentation Using Large Language Models Language models that are pre-trained, like BERT, have shown substantial improvements across various NLP tasks. In the study [3] authors explore various transformer-based pre-trained models, including auto-regressive models (GPT-2), auto-encoder models (BERT), and seq2seq models (BART), specifically focusing on conditional data augmentation. The paper demonstrates that a straightforward yet effective approach involves adding class labels to the beginning of text sequences to condition the pre-trained models for data augmentation.

2 Research Part

The project aims to address the shortage of labeled datasets in Russian for health-related question and answer systems. While there are many datasets available in English, the lack of equivalent resources in Russian is a serious problem. The main goal is to create a robust Russian-language Q&A dataset for healthcare using data extension techniques and fine-tuning of the T5 model. On Figure 1 demonstrated the pipeline of the project:

- Data Gathering: Utilize web scraping techniques to gather data from healthcare-related websites in the format Question-Context-Answer triplets.
- Data Preparation: Preprocess the gathered data, addressing any inconsistencies or noise.
- Data Augmentation: Apply Large language Models for data augmentation. Various approaches have been used in this study:
  - Abstractive answer summarization
  - Context chunking and question-answer pairs generation
  - Abstractive context summarization and question-answer pairs generation
  - Extractive summarization of the context and abstractive summarizarion of the answer
- Fine-tuning the T5 and BERT models: Fine-tune models on the different datasets:
  - Base Russian dataset
  - Base Russian dataset concatenated with augmented answers
  - Base Russian dataset + augmented answers concatenated with generated Q-A pairs
  - Extractive summarized context and abstractive summarized answer (for Russian dataset)
  - Abstractive summarized context and generated Q-A pairs (for English dataset)
  - Base Russian dataset concatenated with extractive summarized context and abstractive summarized answer
  - Base English dataset concatenated with abstractive summarized context and generated Q-A pairs

2.1 Data Gathering

In the pursuit of advancing our capabilities in the field of medical Q&A, we have undertaken a comprehensive review of various datasets curated for this purpose. The success of any Q&A system is heavily dependent on the quality and relevance of the data it is trained on. To ensure that our system is optimally tailored to address the nuances of medical inquiries, we have scrutinized several datasets, evaluating each against a set of predefined criteria.

The table 1 summarizes our findings, providing essential details such as the dataset name, associated link, reasons why a particular dataset may not be suitable for our specific requirements, and the final decision on whether to include it in our training dataset. This meticulous review process is crucial in guiding our selection of a dataset that aligns with our objectives and ensures the robustness of our medical Q&A system.

Table 1: Review of the existing datasets

Dataset Name	Why not suitable for us	Use?
Corpus for Evidence Based Medicine Summarization (Mollá, 2010)	Some information can be outdated, as the dataset was updated only in 2014. That could potentially lead to noise in training data	No
CLEF QA4MRE Alzheimer's task (Peñas et al, 2012)	Might contain updated data	No
BioASK datasets (2012-2020)	Closed dataset	No
TREC LiveQA-Med (Ben Abacha et al, 2017)	Contains question-driven summaries of answers to consumer health questions, which could lead to noise	No
MEDIQA-2019 datasets on NLI, RQE, and QA (Ben Abacha et al., 2019)	Does not contain context	No
MEDIQA-AnS dataset of question-driven summaries of answers (Savery et al., 2020) [4]	Use summarizations in answers	No
MedQuaD Collection of 47k QA pairs (Ben Abacha and Demner-Fushman, 2019)	-	Yes
Medication QA Collection (Ben Abacha et al., 2019)	Very small sample	No
Consumer Health Question Summarization (Ben Abacha and Demner-Fushman, 2019)	Have only context and summary answer	No
emrQA: QA on Electronic Medical Records (Pampari et al., 2018). Scripts to generate emrQA from i2b2 data	Scripts for generation of data, not the data itself	No
EPIC-QA dataset on COVID-19 (Goodwin et al., 2020)	Focuses only on Covid-19	No
BiQA Corpus (Lamurias et al., 2020) [5]	Scripts for generation of data, not the data itself. Also does not provide the trustable contexts	No
HealthQA Dataset (Zhu et al., 2019) [6]	Dataset with private access	No

In conclusion, we have successfully identified a dataset that aligns with our project's requirements, providing a solid foundation for the development of our medical Q&A system. The chosen dataset not only meets the criteria for relevance and diversity in medical queries but also addresses potential challenges that may arise during training.

2.2 Data Preparation

The medQUAD dataset comprises the following information for each instance:

- Question
- Answer
- Link to the Website Page

2.2.1 Issues Encountered

During the data preparation process, several challenges and issues were identified:

1. Data Removal: A significant portion of the data had to be removed, possibly due to irrelevance, inconsistency, or unavailability of associated web pages.

2. Expired Website Links: Many website pages linked in the dataset no longer exist, causing difficulties in retrieving context and validating answers.
3. Lack of Answer Start Token: The absence of a designated answer start token posed challenges in identifying where the answer begins within the context.
4. Texts in English: While the dataset is intended for medical question answering, the texts were found to be in English, potentially deviating from the expected language.

2.2.2 Data Preparation Steps

To address the challenges and ensure the dataset's usability, the following steps were undertaken:

1. Formulation of Dataset Structure: The dataset structure was defined, encompassing the necessary fields for questions, answers, and website page links.
2. Website Context Parsing: Context for each instance was parsed from the associated website to provide comprehensive information for training and evaluation.
3. Answer Start Token Identification: Within each context, the token marking the start of the answer was identified. This step aimed to establish clear boundaries for extracting answers.
4. Translation to Russian: All text fields, initially in English, were translated to Russian to align with the desired language for the medical question answering task.

2.3 LLM-based data augmentation

In the pursuit of enhancing the diversity and informativeness of our dataset for question answering, we employed a two-step data augmentation approach. The first step involved generating summaries for existing answers, while the second step included context chunking followed by question and answer generation. The models utilized for these tasks were:

- Question Generation Model: mrm8488/t5-base-finetuned-question-generation-ap
- Answer Generation Model: valhalla/t5-base-qa-qg-hl

2.3.1 Results

Summaries for Answers The summarization process resulted in 885 augmented answers. This step aimed to condense and capture the essential information within the given answers.

Context Chunking, Question, and Answer Generation

- Number of Generated Items: 5198
- Context Chunking: The context was divided into chunks of 512 length, allowing us to capture information comprehensively. This process aimed to enrich the variety of questions on every topic.
- Observations:
  - Question Variety: The generated questions demonstrated a diverse range, capturing nuanced aspects of the context.
  - Answer Quality: Most generated answers were correct and informative, contributing positively to the dataset.
  - Simplicity of Some Answers: While the majority of answers were satisfactory, a few were overly simple, such as advising to "call the doctor."
  - Challenges: For 40 observations, answers were not successfully generated, indicating potential limitations or complexities in the context.

The data augmentation process using Language Model Models has proven effective in enhancing our dataset for medical question answering. The generated summaries and diversified questions contribute to a more comprehensive training set. However, attention should be given to the simplicity of some answers and the occasional failure to generate answers for certain contexts. These insights will guide further refinement of the augmentation process to ensure the dataset's quality and diversity, ultimately improving the robustness of our question answering system.

2.4 Methodology for English Data Context Summarization and Q-A Pairs Generation

In the pursuit of augmenting our dataset for improved diversity and informativeness, we employed a multi-step approach involving context summarization and the generation of question-answer pairs. The key components of our methodology include:

1. Abstractive Summarization:
   - Model Used: t5-base
   - Objective: Utilizing the T5 transformer model for abstractive summarization, we condensed the context information, aiming to retain essential details while reducing redundancy.
2. Question Generation:
   - Model Used: t5-base-finetuned-question-generation-ap
   - Objective: Generating diverse and contextually relevant questions based on the summarized context. This step contributes to a more varied and comprehensive dataset.
3. Question Answering:
   - Model Used: t5-base
   - Objective: Employing the T5 model for question answering to generate corresponding answers for the questions generated in the previous step. This adds a dynamic Q-A pair to the dataset.
4. Translation:
   - Tool Used: Google Cloud Translation API
   - Objective: Translating the generated Q-A pairs to enhance language diversity and ensure that the augmented dataset includes multilingual variations.

2.4.1 Results

The application of the outlined methodology yielded the following results:

- Context Summarization: The abstractive summarization using the t5-base model effectively condensed the context while maintaining essential information.
- Question Generation: The t5-base-finetuned-question-generation-ap model successfully generated a diverse set of questions that were contextually aligned.
- Question Answering: Utilizing the t5-base model for question answering produced corresponding answers, creating enriched Q-A pairs for each instance.
- Translation The Google Cloud Translation API was employed to translate the generated Q-A pairs, promoting language diversity within the augmented dataset.

The multi-step methodology, encompassing context summarization, question generation, question answering, and translation, has proven effective in augmenting our dataset for English data. The resulting Q-A pairs, enriched with varied questions and contextually relevant answers, contribute to a more robust and diverse training set. The inclusion of translation further enhances language diversity, ensuring the dataset's applicability in multilingual contexts. These augmentations are crucial for training models that exhibit versatility and effectiveness in handling a wide range of queries in the medical domain.

2.5 Methodology for Russian Data Context and Answer Summarization

Also we decided to perform another approach of data augmentation directly with russian dataset. Such approach include steps: extractive summarization of the context and absrtactive summarization of the answer. The key components of this methodology include:

1. Extractive Summarization:
   - Model Used: TextRank
   - Objective: Using the TextRank framework for extractive summarization, we identified meaningful sentences that best capture the essence of the context.
2. Abstractive Summarization:
   - Model Used: rut5-baze-gazeta

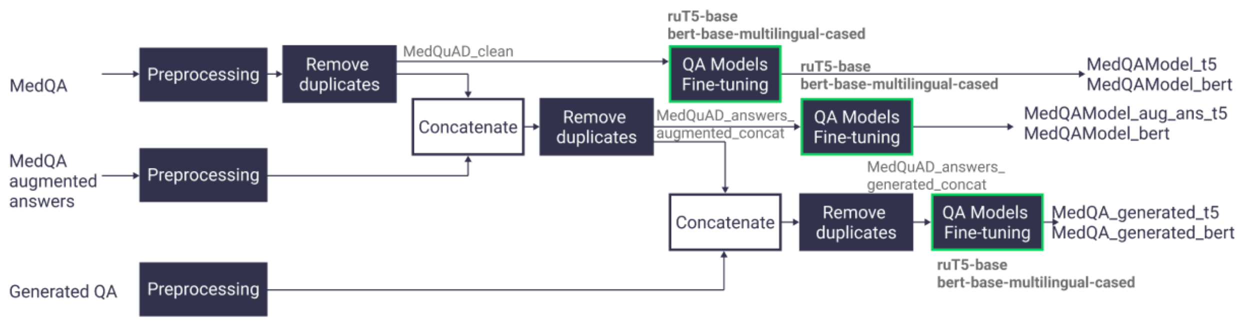

Figure 2: Stages of the first part fine-tuning

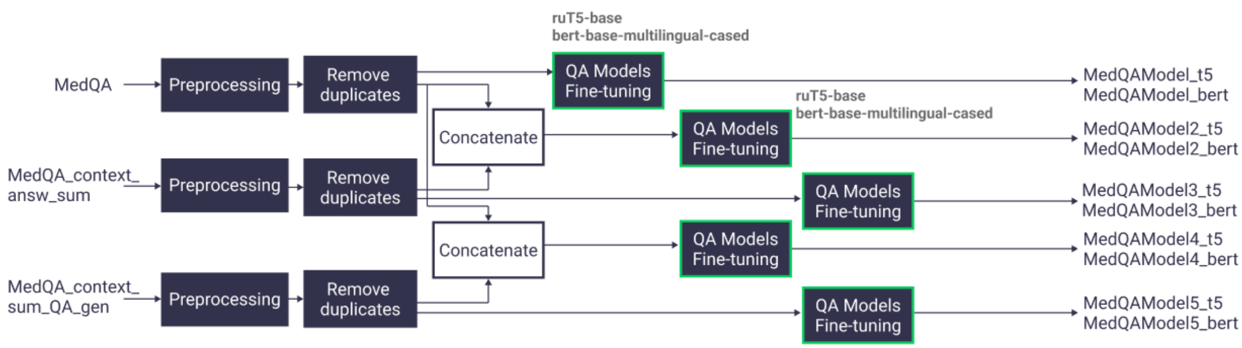

Figure 3: Stages of the second part fine-tuning

- Objective: using rut5-base-gazeta model, it was possible to reduce overly large responses at the the main point of the answer was retained.

2.6 Model fine-tuning for Q&A

The fine-tuning phase of the research aims to assess the robustness and effectiveness of the resulting datasets in enhancing the performance of Question and Answering systems. Two prominent models, ruT5-base and bert-base-multilingual-cased, are employed to evaluate the augmented datasets. A number of evaluation metrics such as ROUGE-L and BLUE were used to assess the results. As a first step of the fine-tuning part three datasets were used, the figure 2 demonstrates main parts of this research stage. Data preprocessing stage includes:

- removal all samples with insignificant context (description of the page)
- removal extra words from all columns (e.g. 'описание', 'резюме' etc.)

Following the completion of data preprocessing, an iterative concatenation of datasets was conducted. A model was fine-tuned on the base dataset, followed by the concatenation of one of the augmented datasets with base dataset. Subsequently, duplicate instances were entirely removed, and another model underwent pre-training on the resulting dataset. This process was iteratively reiterated, employing the dataset derived from the generation of question-answer pairs.

Subsequently, additional experiments were initiated, specifically involving diverse approaches to context summarization and the formation of response-question pairs. The algorithmic process closely resembles the previously described one, as can be seen on figure 3, with the notable distinction that no iterative concatenation is applied. Instead, the models are individually trained on augmented data and subsequently concatenated independently with the baseline. In essence, this approach diverges from the creation of a singular extensive dataset, as observed in the initial set of experiments. The rationale behind adopting this strategy is to meticulously examine the nuanced impact of distinct augmentation types on the efficacy of pre-training Question and Answering models.

3 Results Analysis

After the fine-tuning process, it was found that the most favorable outcomes were achieved when combining the baseline dataset with the augmented method involving extractive summarization of contexts and abstractive summarization of responses for the Russian-language dataset. Additionally, noteworthy metrics were attained by the model fine-tuned on the base dataset concatenated with augmented responses. However, it's important to note that the second result shows a marginal difference of 0.01 for BLEU and 0.02 for ROUGE-L. This minor change is attributed to the relatively small increase in data, and it falls short of establishing this augmentation variant as robust. When examining the overall table 2 of results, it is evident that the model tuned on the original dataset exhibits the most favorable quality scores. These results can be explained by the discrepancies in the distribution of lengths in the question-answer pairs, especially with respect to generation methods. In addition, it is evident that the model fine-tuned on the underlying data shows more accurate responses in the inference phase (as seen in the table 3).

Table 2: Fine-tuning results

	ruT5-base (Fine-Tuned)			bert-base-multilingual-cased		
	BLEU	ROUGE-L (precision)	Epochs	BLEU	ROUGE-L (precision)	Epochs
Base Dataset	**0.921**	**0.942**	100	0.121	0.398	5
Base Dataset + Augmented Answers	0.912	0.923	156	0.059	0.102	5
Base dataset + Augmented Answers + Generated Q-A Pairs	0.834	0.890	42	0.036	0.084	5
Extractive Summarised Context and Abstractive Summarised Answers (RUS)	0.769	0.808	81	-	-	-
Abstractive Summarised Context and Generated Q-A Pairs (ENG)	0.878	0.8823	100	0.002	0.010	5
Base Dataset + Extractive Summarised Context and Abstractive Summarised Answers (RUS)	**0.914**	**0.926**	79	0.002	0.20	5
Base Dataset + Extractive Summarised Context and Generated Q-A Pairs (ENG)	0.801	0.865	81	-	-	-

Table 3: Results on inference mode of the models

	ruT5-base
Base Dataset	Вопрос: Что такое холецистит? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: холецистит является воспалительным процессом, который возникает в желчном пузыре. от этой патологии сегодня страдает около 20% жителей всего мира. Вопрос: Нужно ли обращаться к врачу, если есть симптомы холецистита? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: позвоните своему врачу, если: у вас есть симптомы холецистита.
Base Dataset + Augmented Answers	Вопрос: Нужно ли обращаться к врачу, если есть симптомы холецистита? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: позвоните своему врачу, если: у вас есть признаки холецистита.
Base dataset + Augmented Answers + Generated Q-A Pairs	Вопрос: Что такое холецистит? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: холецистит Вопрос: Нужно ли обращаться к врачу, если есть симптомы холецистита? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: позвоните своему врачу

Continued on the next page

Continued on the next page

	ruT5-base
Extractive Summarised Context and Abstractive Summarised Answers (RUS)	Вопрос: Что такое холецистит? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: холецистит это воспалительный процесс, который возникает в желчном пузыре. это может быть связано с тяжелым заболеванием, которые могут появиться у людей старше 50 лет. Вопрос: Нужно ли обращаться к врачу, если есть симптомы холецистита? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: если есть признаки холецистита, позвоните своему врачу, если есть признаки. если у вас нет признаков и симптомов заболеваниюю если у вас нет признаков и симптомов заболевания.
Extractive Summarised Context and Generated Q-A Pairs (ENG)	Вопрос: Что такое холецистит? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: воспаление возникает в желчном пузыре Вопрос: Нужно ли обращаться к врачу, если есть симптомы холецистита? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: воспаление возникает в желчном пузыре

Continued on the next page

Continued on the next page

	ruT5-base
Base Dataset + Extractive Summarised Context and Abstractive Summarised Answers (RUS)	Вопрос: Что такое холецистит? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: холецистит - это воспалительный процесс, который возникает в желчном пузыре. от этой патологии сегодня страдает около 20 жителей всего мира. причем наиболее подвержены заболеванию женщины старше Вопрос: Нужно ли обращаться к врачу, если есть симптомы холецистита? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: если у вас есть симптомы холецистита, позвоните своему врачу, если у вас нет симптомов. если у вас нет симптомов заболевания, позвоните врачу и позоните ему, когда у вас нет симптомов заболевания.
Base Dataset + Extractive Summarised Context and Generated Q-A Pairs (ENG)	Вопрос: Что такое холецистит? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: воспаление является воспалительным процессом, который возникает в желчном пузыре Вопрос: Нужно ли обращаться к врачу, если есть симптомы холецистита? Контекст: Холецистит является воспалительным процессом, который возникает в желчном пузыре. От этой патологии сегодня страдает около 20% жителей всего мира. Причем наиболее подвержены заболеванию женщины старше 50 лет. Воспаление выявляют у людей и другого возраста. Причем для детей и подростков характерна бескаменная форма. Зачастую болезнь диагностируется в развитых странах. Обусловлено это особым образом жизни и пищевым поведением. Ответ: позвоните врачу, если есть признаки холецистита

4 Conclusion

This article demonstrates the effectiveness of using large language models such as T5 to efficiently increase the amount of data. One of the identified factors contributing to the creation of inaccurate question-answer pairs is limited knowledge of the T5 model in the field of healthcare. To solve this problem in the context of increasing the volume of data by generating question-answer pairs, it is assumed that models with significantly large amounts of pre-training data, such as Llama-2 and Mistral, could offer more promising solutions. As for fine-tuning models on augmented data, the following advantages and disadvantages can be distinguished:

- Advantages: The model converges faster
- Disadvantages: Control the length of the generated pairs during augmentation, since if, for example, the answers in augmented datasets are shorter, then the model on the inference stage will most likely not be able to generate a more accurate answer

References

[1] Wei, J., Zou, K. (2019). Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint* arXiv:1901.11196.

[2] Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. MIT press.

[3] Kumar, V., Choudhary, A., Cho, E. (2020). Data augmentation using pre-trained transformer models. arXiv preprint arXiv:2003.02245.

[4] Savery, M., Abacha, A. B., Gayen, S., & Demner-Fushman, D. (2020). Question-driven summarization of answers to consumer health questions. *Scientific Data*, 7(1), 322.

[5] Lamurias, A., Sousa, D., & Couto, F. M. (2020). Generating biomedical question answering corpora from Q&A forums. *IEEE Access*, 8, 161042-161051.

[6] Zhu, M., Ahuja, A., Wei, W., & Reddy, C. K. (2019, May). A hierarchical attention retrieval model for healthcare question answering. In The World Wide Web Conference (pp. 2472-2482).
Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.