

Argumentative Text Generation in Economic Domain

Fishcheva I. N.

Vyatka State University
Kirov, Russia
fishchevain@gmail.com

Osadchiy D.

ITMO University
Saint Petersburg, Russia
da.osadchiy@gmail.com

Bochenina K. O.

ITMO University
Saint Petersburg, Russia
kbochenina@itmo.ru

Kotelnikov E. V.

Vyatka State University
Kirov, Russia
kotelnikov.ev@gmail.com

Abstract

The development of large and super-large language models, such as GPT-3, T5, Switch Transformer, ERNIE, etc., has significantly improved the performance of text generation. One of the important research directions in this area is the generation of texts with arguments. The solution of this problem can be used in business meetings, political debates, dialogue systems, for preparation of student essays. One of the main domains for these applications is the economic sphere.

The key problem of the argument text generation for the Russian language is the lack of annotated argumentation corpora. In this paper, we use translated versions of the Argumentative Microtext, Persuasive Essays and UKP Sentential corpora to fine-tune RuBERT model. Further, this model is used to annotate the corpus of economic news by argumentation. Then the annotated corpus is employed to fine-tune the ruGPT-3 model, which generates argument texts. The results show that this approach improves the accuracy of the argument generation by more than 20 percentage points (63.2% vs. 42.5%) compared to the original ruGPT-3 model.

Keywords: Argumentation Mining, Text Generation, ruGPT-3, RuBERT, XGBoost

DOI: 10.28995/2075-7182-2022-21-211-222

Генерация аргументативных текстов экономической тематики

Фищева И. Н.

Вятский государственный университет
Киров, Россия
fishchevain@gmail.com

Осадчий Д.

Университет ИТМО
Санкт-Петербург, Россия
da.osadchiy@gmail.com

Боченина К. О.

Университет ИТМО
Санкт-Петербург, Россия
kbochenina@itmo.ru

Котельников Е. В.

Вятский государственный университет
Киров, Россия
kotelnikov.ev@gmail.com

Аннотация

Разработка больших и сверхбольших языковых моделей, таких как GPT-3, T5, Switch Transformer, ERNIE и др., позволила в последнее время значительно повысить качество генерации текстов. Одним из важных направлений в этой области является порождение текста с аргументами. Решение такой задачи может быть использовано при проведении деловых совещаний, в политических дебатах, в диалоговых системах, при подготовке студенческих эссе. Одной из основных предметных областей в указанных приложениях является экономическая сфера.

Ключевой проблемой при генерации аргументов для русского языка является дефицит корпусов, размеченных по аргументации. В настоящей работе мы используем переводные версии корпусов Argumentative Microtext, Persuasive Essays и UKP Sentential для дообучения модели RuBERT. Далее построенная модель используется для разметки по аргументации собранного корпуса экономических новостей. Затем размеченный корпус применяется для дообучения модели ruGPT-3, которая порождает аргументационные тексты. Результаты показывают, что такой подход позволяет повысить правильность генерации аргументов более чем на 20 процентных пунктов (63.2% vs. 42.5%) по сравнению с исходной моделью ruGPT-3.

Ключевые слова: анализ аргументации, генерация текстов, BERT, GPT

1 Introduction

Automatic text generation has recently made impressive progress with the development of large and super-large pre-trained language models (Han et al., 2021), such as GPT-3 (Brown et al., 2020), T5 (Raffel et al., 2020), Switch Transformer (Fedus et al., 2021), ERNIE (Sun et al., 2021) and others. These models allow tuning to the problem by updating the weights on a small training dataset (fine-tuning) or without updating weights in few-shot learning (with several training examples, usually from 10 to 100), one-shot learning (one training example) and even zero-shot learning (Brown et al., 2020).

One of the important directions in this field is the generation of texts with arguments (Hua et al., 2020; Gretz et al., 2020; Schiller et al., 2021). Arguments in this case are either found using an information retrieval system (Hua et al., 2020), or generated by a pre-trained language model (Gretz et al., 2020; Al Khatib et al., 2021). An *argument* is a set of statements that includes claim and premises (Stede and Schneider, 2018). A *claim* is a statement that expresses a potentially controversial point of view. *Premises* are statements supporting or refuting this claim.

Systems that generate text with arguments related to a given claim can be used during business meetings to promptly generate arguments; in political debate to back up the position of a speaker; in jurisprudence to search and generate arguments for legislative acts and precedents; in dialogue systems for the selection of arguments to convince the interlocutor; in education for analyzing, generating, and evaluating arguments in student essays. One of the most common domains for these applications is economics.

There are several publicly available pre-trained language models for the Russian language, including RuBERT (Kuratov and Arkhipov, 2019), SBERT (SBERT, 2020), and ruGPT-3 (ruGPT-3, 2020). Recently, several papers have appeared on the argument mining in Russian (Fishcheva and Kotelnikov, 2019; Fishcheva et al., 2021; Salomatina et al., 2021; Ilina et al., 2021).

However, to the best of our knowledge, there have been no work devoted to the generation of argumentative Russian texts so far. We aim to close this gap and propose an approach to build a language model for generating an argumentative text in Russian in the economic domain.

The contribution of this work is as follows:

- a new approach to the generation of Russian argumentative texts is proposed. In this approach, the model RuBERT is first trained using argumentative corpora, which is then used to automatically annotate arguments in an economic text corpus. The annotated corpus is used to fine-tune the ruGPT-3 model, which allows generating texts with arguments, particularly, sentences containing premises related to a given claim;
- presenting the results of experiments with the proposed approach;
- sharing the corpus of pairs $\langle \textit{claim}, \textit{sentence} \rangle$ on economic domain. The corpus includes 660 sentences manually annotated into two classes – “premise” / “non-premise”.

2 Previous Work

2.1 Argumentative Text Generation

Argumentative texts generation includes the following research directions.

1. Application of traditional text generation architectures, including content selection/organization and linguistic implementation components (Carenini and Moore, 2006).
2. Generation of an argumentative text from given small argumentative elements (Walton and Gordon, 2012; Reisert et al., 2015; Wachsmuth et al., 2018; El Baff et al., 2019).
3. Replacing target objects in extracted arguments (Bilu and Slonim, 2016).

4. Using predefined argument templates (Bilu et al., 2020).
5. Application of the encoder-decoder architecture (seq2seq) (Hidey and McKeown, 2019), possibly supplemented by an information retrieval component (Hua et al., 2020; Schiller et al., 2021).
6. Using pre-trained language models for text generation, such as GPT-2 (Gretz et al., 2020; Al Khatib et al., 2021).

The papers closest to our research are (Schiller et al., 2021; Gretz et al., 2020; Al Khatib et al., 2021).

Schiller et al. (Schiller et al., 2021) proposed a neural network model Arg-CTRL for controlled generation of texts based on the well-known CTRL model (Keskar et al., 2019). Controlling the text generation is carried out using the so-called *control codes*, including the topic, stance, and aspect of the argument. The retrieval system and the BERT model were used to extract aspects. In our work, instead of the CTRL model, we use the ruGPT-3 model, which does not require learning from scratch for the Russian language, like CTRL. Also, the BERT model is used not to extract aspects, but to select sentences containing premises.

Gretz et al. (Gretz et al., 2020) investigated claim generation. They suggested using the fine-tuning of the neural network model GPT-2 (Radford et al., 2020) based on the corpus of automatically extracted arguments. Moreover, authors added context to the input data for GPT-2 in order to improve the quality of generation. In our work, we generated premises, not claims. Also, instead of GPT-2, which is mainly an English-language model, we use the Russian-language ruGPT-3 model.

Al-Khatib et al. (Al Khatib et al., 2021) used argumentation knowledge graphs to generate argumentative texts. The concepts contained in such graphs are searched on debate portals and Wikipedia. The found fragments of texts are used for fine-tuning GPT-2. In our work, we use the RuBERT model rather than knowledge graphs to generate training data. In addition, instead of GPT-2 we use the Russian-language model ruGPT-3.

2.2 Argumentation Mining in Russian

The area of argumentation mining for the Russian language has recently attracted more and more attention of researchers.

The first annotated corpus for the Russian language (Fishcheva and Kotelnikov, 2019) was created based on the translation of the English language Argumentative Microtext Corpus (ArgMicro) (Peldszus and Stede, 2015; Skeppstedt et al., 2018). It was then expanded with machine translation of the Persuasive Essays Corpus (PersEssays) (Stab and Gurevych, 2014) and a Joint Argument Annotation Scheme was proposed (Fishcheva et al., 2021). By using XGBoost and BERT, the authors were able to improve the results of automatic classification of “for” / “against” premises.

Salomatina et al. (Salomatina et al., 2021) proposed a method for finding an argumentative structure based on using the patterns of argumentation indicators and their role in the thematic structure of the text. This method can be used in the absence of a sufficient amount of annotated data.

Ilina et al. (Ilina et al., 2021) presented a web resource designed to study argumentation in popular science discourse. A distinctive feature of the argument annotation model is the using of argument persuasiveness weighting. The annotation script includes several procedures that allow the annotator to check the quality of text annotation and evaluate the persuasiveness of the argumentation.

In our work, to the best of our knowledge, the problem of generating argumentative texts in Russian is being investigated for the first time. We are expanding the Russian-language corpus from (Fishcheva et al., 2021) by translating the UKP Sentential Argument Mining Corpus (UKP Sentential) (Stab et al., 2018). Based on the extended corpus, we train RuBERT model, which we then use to annotate sentences of the economic news corpus. The annotated corpus is used to fine-tune the ruGPT-3 model to generate texts containing premises for a given claim.

3 Proposed approach

The scheme of the proposed approach for generating argumentative texts is shown in Fig. 1.

The three argumentative annotated corpora – ArgMicro, PersEssays, and UKP Sentential (see Sec-

tion 4) – serve as input data. They are used to fine-tune the pre-trained RuBERT¹ model (Kuratov and Arkhipov, 2019). This model showed the best results (along with XGBoost² model) in the argument classification task (Fishcheva et al., 2021). We also compared the performance of RuBERT with XGBoost on “premise” / “non-premise” classification task at the sentence level. As a result, RuBERT outperformed XGBoost (see Subsection 5.1) and further we used the RuBERT model.

We collected the corpus of economic news from the internet (see Section 4) – it is denoted as “Un-labeled economic corpus” on the scheme. The fine-tuned RuBERT classifies economic corpus sentences so we obtain economic corpus labeled by “premise” / “non-premise” (see details in Subsection 5.1).

At the next step, fine-tuning of the pretrained ruGPT-3 (ruGPT-3, 2020) model is carried out. Firstly, we choose 3,500 sentences (it is about 5% from the whole corpus) with the highest probability scores of classification as “premise” from RuBERT. Then we take pretrained ruGPT-3 model and fine-tune it on selected 3,500 sentences (see details in Subsection 5.2). To evaluate and compare the accuracy of the original model and the fine-tuned one, 20 prompts were used, for each prompt both models generated 20 potential premises. The 800 (400+400) received sentences were annotated manually (see details in Subsection 5.3).

4 Text Corpora

In our study, we used three existing argument-annotated corpora – ArgMicro, PersEssays, and UKP Sentential, as well as a new economic news corpus.

The Argumentative Microtext Corpus (ArgMicro) was proposed in (Peldszus and Stede, 2015; Skeppstedt et al., 2018). The corpus includes 283 texts on various topics (raising the retirement age, health insurance, school uniforms, etc.). Each text contains one claim about a topic and from 2 to 10 ADUs (argumentative discourse units), annotated as premises “for” or “against” this claim. An ADU is a fragment of text that has a single argumentative value (Stede and Schneider, 2018). In the ArgMicro, an ADU can be either an entire sentence or part of a sentence. Further, we consider each individual ADU as a sentence.

The Persuasive Essays Corpus (PersEssays) was introduced in (Stab and Gurevych, 2014). It contains 399 texts on a wide range of topics (school education, immigration, economic policy, etc.). The ADUs are sentences. The annotation of sentences is done according to four types: main claim, claim, premise, and neutral element. When forming the training corpus for the classification “premise” / “non-premise”, the main claims were excluded, just claims and the actual premises were used as “premises”, and neutral elements – as “non-premises”.

The UKP Sentential Argument Mining Corpus (UKP Sentential) was proposed in (Stab et al., 2018). The corpus includes 25,492 sentences, annotated as “for”, “against” or “non-premise” in relation to one of the eight topics (abortion, cloning, the death penalty, etc.). The corpus contains 489 sentences, the annotation of which is different for different topics. Such sentences were excluded.

(Fishcheva and Kotelnikov, 2019) showed that the best result among the Google Translate, Yandex.Translate and Prompt systems was demonstrated by Google Translate during machine translation of the ArgMicro corpus from English into Russian. Therefore, all three English corpora were translated into Russian using Google Translate.

The characteristics of all the corpora with argumentative annotation are shown in Table 1.

For fine-tuning of the ruGPT-3 model in economics, we used the corpus of Russian-language economic news, collected from the website *Banki.ru*³. The original corpus included 7,759 texts for the period from 01.06.2019 to 12.07.2021. The texts were segmented into sentences using the *Stanza*⁴ library (average text size – 9 sentences) and tokenized using the *NLTK*⁵ library. Further, 10% of the longest and shortest sentences, as well as repetitive sentences were removed. The preprocessing resulted in 68,859 sentences with an average length of 15 words.

¹<https://huggingface.co/DeepPavlov/rubert-base-cased>

²<https://xgboost.readthedocs.io>

³<https://www.banki.ru>

⁴<https://stanfordnlp.github.io/stanza>

⁵<https://www.nltk.org>

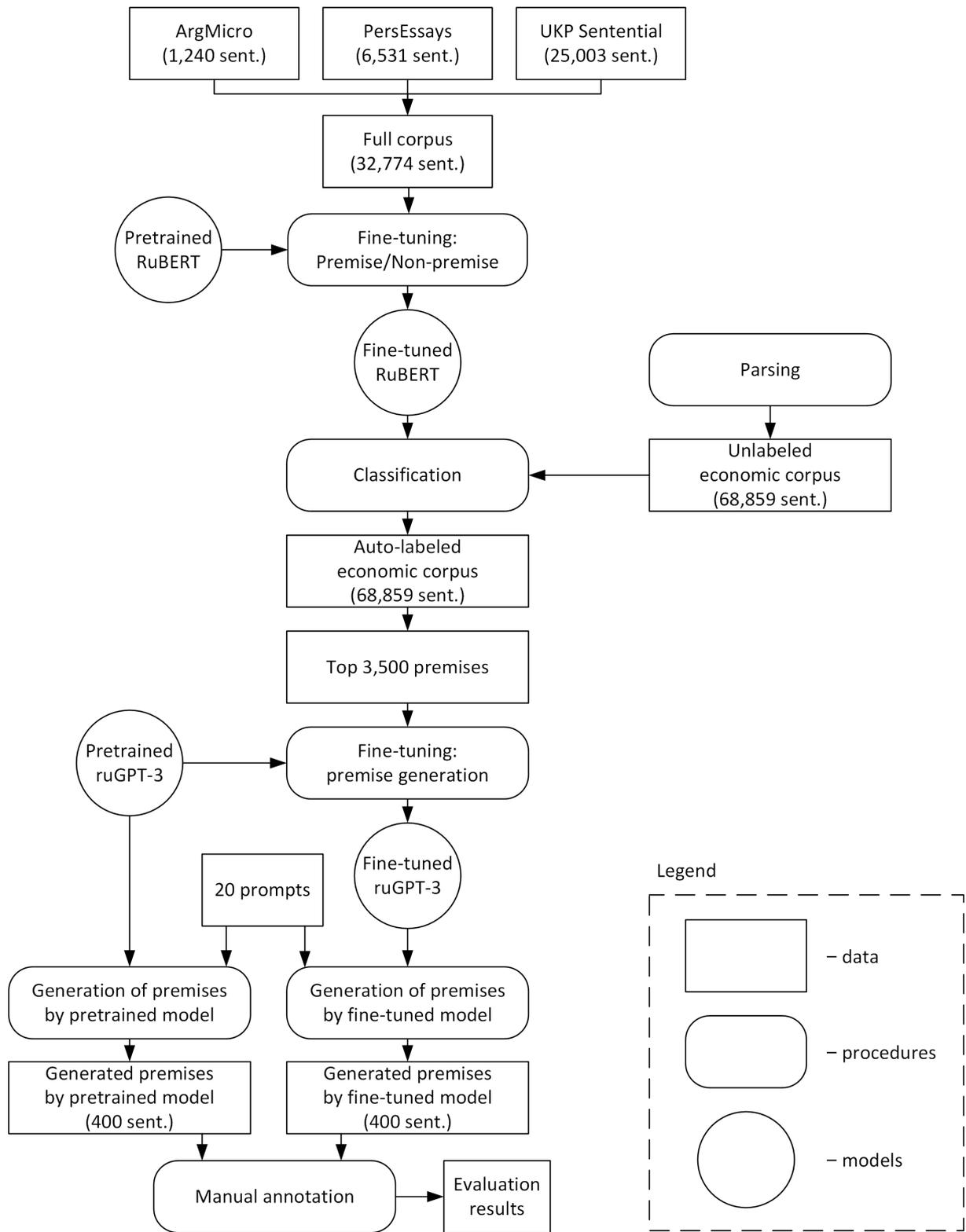


Figure 1: Scheme of the proposed approach.

Table 1: Characteristics of corpora with argumentative annotation.

Corpora	Premise	Non-premise	Total
ArgMicro	1,236	4	1,240
PersEssays	5,302	1,229	6,531
UKP Sentential	11,023	13,980	25,003
Total	17,561	15,213	32,774

5 Result and Discussion

5.1 “Premise” / “Non-premise” Classification

At the first step, we compared the performance of RuBERT and XGBoost, which were the best models in (Fishcheva et al., 2021). To this end, we fine-tuned the pre-trained RuBERT model and we trained the XGBoost classifier on the union of annotated corpora ($\text{ArgMicro} \cup \text{PersEssays} \cup \text{UKP Sentential}$). The problem of classifying sentences into two classes – “premise” and “non-premise” – was set. We used the same 5-fold cross-validation for both models to obtain performance scores. For selection of hyperparameters for XGBoost, 4-fold nested cross-validation was used, for RuBERT – hold-out dataset (validation 20%). The hyperparameters are as follows:

- XGBoost⁶: number of trees = [50, 150, **500**], maximum tree depth = [2, 8, 20, 30];
- RuBERT: number of epochs = [3, **5**], batch size = [4, **8**], learning rate = 10^{-5} .

To train RuBERT, we used class *BertForSequenceClassification* from the *transformers*⁷ library with the following settings: *AdamW* optimizer with default parameters, weight decay = 0.01, context length = 512 tokens.

Following (Fishcheva and Kotelnikov, 2019), we used three types of features for XGBoost:

- lexical features – discursive markers (e.g., “therefore”, “influence”, “firstly”, etc.) and modal words (e.g., “should be”, “can”, “want”, etc.), including negations, 255 features in total;
- punctuation features – comma, colon, semicolon, question and exclamation marks, 5 features in total;
- morphosyntactic features – N-grams based on parts of speech (nouns, pronouns, verbs, adjectives, and adverbs), $N = 2, 3, 4$ (e.g., ADJ+ADJ+NOUN, ADV+VERB), and grammatical features of verbs: time, inclination, person; there are 783 features in total. Morphological analysis was carried out with *Mystem*⁸.

Unlike (Fishcheva and Kotelnikov, 2019), we have not added contextual features of the previous and next sentences, since the UKP Sentential corpus contains only individual sentences.

Cross-validation results for the union corpus ($\text{ArgMicro} \cup \text{PersEssays} \cup \text{UKP Sentential}$) are presented in Table 2.

Table 2: Performance scores for “premise” / “non-premise” classification: macro-averaged F1-score, Precision, Recall and Accuracy (Mean \pm Std Dev)

Model	F1-score	Precision	Recall	Accuracy
XGBoost	0.6800 \pm 0.0066	0.6817 \pm 0.0065	0.6796 \pm 0.0066	0.6832 \pm 0.0064
RuBERT	0.7903\pm0.0051	0.7901\pm0.0051	0.7908\pm0.0050	0.7911\pm0.0051

XGBoost is significantly inferior to RuBERT in this task, in contrast to the results of (Fishcheva et al., 2021), where both models showed comparable performance. This is due to the fact that contextual fea-

⁶Optimal values of the hyperparameters are highlighted in bold – they turned out to be the same on all folds.

⁷https://huggingface.co/docs/transformers/model_doc/bert

⁸<https://yandex.ru/dev/mystem>

tures, which are very important according to (Fishcheva et al., 2021), were not used in training XGBoost in our study.

Thus we decided to take RuBERT as the classification model. After performance assessment, RuBERT with the selected optimal hyperparameters was fine-tuned on the union corpus. The obtained model was used to classify 68,859 sentences of the economic corpus.

5.2 ruGPT-3 Fine Tuning

To build a model for generating argumentative texts, fine-tuning of the pretrained model ruGPT-3 (ruGPT-3, 2020) was held. To this end, 3,500 sentences (approx. 5%) from the economic corpus annotated by the RuBERT model as “premise” with the highest confidence scores were used as training data.

We fine-tuned the ruGPT3Large version (760M parameters) using NVIDIA RTX A6000 video card. The training sentences, which were prefixed with “*потому что*” (“*because*”), were fed individually to the input of the model. We tuned hyperparameters using hold-out strategy: 3,000 sentences were used as training dataset, 500 sentences – as validation dataset. The perplexity was the performance metric.

The hyperparameters were as follows:

- number of epochs = [1..5],
- batch size = [1, 2, 4, 8, 12, 16],
- learning rate = [10^{-6} , 10^{-5} , $5 \cdot 10^{-5}$, 10^{-4}].

The following hyperparameters turned out to be the best: number of epochs 1, batch size 12, learning rate $5 \cdot 10^{-5}$. Perplexity with these hyperparameters is equal to 9.66.

At the next step, the fine-tuned ruGPT-3 model was tested by manual annotation of the generated arguments.

5.3 Manual Annotation of Generated Premises

To test the fine-tuned ruGPT-3 model, we used 20 prompting claims (after each claim, suffix “*потому что*” – “*because*” was added) (Table 3). The original ruGPT3Large model was used as the baseline (zero-shot learning). For each prompt (with suffix “*потому что*”), each model generated 20 sentences – potential premises. The following parameters were used for generation in both models: *Top-K*=50, *Top-p*=0.92. Thus, each model generated 400 sentences.

The overall generated 800 sentences were manually annotated by four annotators. Annotators were provided with *<claim, sentence>* pairs with no information about which model generated the sentence. An argument was considered to be a sentence that could be used to convince an opponent of a given claim. The final decision about the sentence was based on a simple voting of the annotators’ scores – the label was approved if at least 3 from 4 annotators agreed.

As a result, the labels were assigned to 660 sentences (82.5% from 800 sentences) (Table 4): 321 for the fine-tuned model and 339 for the original model. The agreement of the annotators calculated by Krippendorff $\alpha=0.4772$. This level of agreement corresponds to moderate agreement on the Landis & Koch scale (Landis and Koch, 1977).

After manual annotation, the accuracy of the fine-tuned ruGPT-3 model was 63.2% (203 premises out of 321 sentences), and the accuracy of the original model was 42.5% (144 premises out of 339 sentences). Examples of premises generated by both models are shown in Table 5 (English version is presented in Table 6).

5.4 Ethical Considerations

Since the proposed method is not controlled, the model may distort the meaning of the texts used in its fine-tuning. Therefore, the generated arguments may contain incorrect information and mislead users who are not experts. At the same time, large language generation models, including the RuGPT-3 model used in the proposed method, make it possible to generate texts that are indistinguishable from texts written by a person. Given the persuasive nature of the arguments generated by the proposed method, they can be used to deceive or even spread personalized disinformation (Zellers et al., 2019; Schuster et al., 2020). We hope that such technologies will be used only for peaceful purposes to improve and

Table 3: Prompting claims

No.	Prompting claims (Russian)	English translation
1	Банкам следует более широко использовать биометрию	Banks should use biometrics more widely
2	Вместо потребительского кредита лучше взять автокредит	Instead of a consumer loan, it is better to take a car loan
3	В случае высокой инфляции нужно вкладываться в драгоценные металлы	In case of high inflation, you need to invest in precious metals
4	Государственные облигации являются одним из наиболее надежных видов ценных бумаг	Government bonds are one of the most reliable types of securities
5	Деньги нужно вкладывать в акции	Money should be invested in stocks
6	Деньги нужно вкладывать в облигации	Money should be invested in bonds
7	Для сохранения финансов оптимальнее всего использовать депозиты	To save finances, it is best to use deposits
8	Криптовалюты лучше фиатных валют	Cryptocurrencies are better than fiat currencies
9	Лучшая инвестиционная стратегия - негосударственные пенсионные фонды	The best investment strategy - non-state pension funds
10	Лучшей инвестицией является покупка недвижимости	The best investment is the purchase of real estate
11	Наиболее выгодно вкладываться в голубые фишки	The most profitable investment in blue chips
12	Не следует играть на валютном рынке	You should not play in the foreign exchange market
13	Обучение финансовой грамотности зачастую приводит к необоснованной уверенности	Financial literacy training often leads to unwarranted confidence
14	Покупка земельного участка является хорошей инвестиционной стратегией	Buying a plot of land is a good investment strategy
15	При оформлении кредитной карты стоит внимательно отнестись к выбору банка	When applying for a credit card, you should carefully consider the choice of bank
16	Сбережения следует хранить в валюте	Savings should be kept in foreign currency
17	Сбережения следует хранить в долларах	Savings should be kept in dollars
18	Сбережения следует хранить в евро	Savings should be kept in euros
19	Сбережения следует хранить в рублях	Savings should be kept in rubles
20	Храните деньги в той валюте в которой получаете зарплату	Keep money in the currency in which you receive your salary

Table 4: Results of premise generation

Model	Premise	Non-premise	Sum
Original model	144 / 42.5%	195 / 57.5%	339 / 100%
Fine-tuned model	203 / 63.2%	118 / 36.8%	321 / 100%
Overall	347 / 52.6%	313 / 47.4%	660 / 100%

simplify communications in society, and that the openness and independence of the development of such models will not only ensure equal access to them, but also stimulate the development of methods for detecting the use of such models.

Table 5: Examples of text generation (in Russian): claims (suffix “*потому что*” is not shown); generated sentences – premises; annotator scores’ voting results.

Claims	Sentences	Voting
<i>Банкам следует более широко использовать биометрию</i>	Fine-tuned ruGPT-3: <i>она позволяет упростить процедуру взаимодействия с клиентами и упрощает проведение платежа.</i>	4 from 4
	Original ruGPT-3: <i>она может значительно упростить получение и обработку данных о клиентах.</i>	4 from 4
<i>Государственные облигации являются одним из наиболее надежных видов ценных бумаг</i>	Fine-tuned ruGPT-3: <i>они не подвержены обесценению, не утрачивают своей ликвидности, их стоимость остается стабильной.</i>	4 from 4
	Original ruGPT-3: <i>они обеспечены государственной гарантией, которая включает в себя государственный долг.</i>	4 from 4
<i>Обучение финансовой грамотности зачастую приводит к необоснованной уверенности</i>	Fine-tuned ruGPT-3: <i>человек, желающий научиться работать с цифровыми активами, начинает думать, что он сможет легко ими управлять.</i>	3 from 4
	Original ruGPT-3: <i>люди не в состоянии проверить, насколько достоверны и правильны их утверждения.</i>	3 from 4

6 Conclusion

In our study, we proposed an approach for creating a Russian-language model of premise generation for a given claim in the economic domain. First, we compared the RuBERT and XGBoost models on the translated argument corpora ArgMicro, PersEssays, and UKP Sentential. The performance of RuBERT turned out to be better. The fine-tuned RuBERT model was used to annotate sentences from a corpus of economic news. The ruGPT-3 model was fine-tuned on the annotated sentences. The model’s ability to generate argumentative premises was tested using manual annotation. The accuracy of generation of the fine-tuned model was more than 20 percentage points higher than that of the original model (63.2% vs. 42.5%), which confirms the effectiveness of the proposed approach. We made this corpus of 660 manually annotated pairs *<claim, sentence>* publicly available⁹.

In our opinion, the achieved quality of argumentative text generation by the fine-tuned models allows us to speak about the possibility of using such models in practice, for example, for online generation of arguments during business meetings.

As perspective areas of research, we plan to study the input context influence on GPT-type models quality.

Acknowledgements

This work was supported by Russian Science Foundation, project № 22-21-00885, <https://rscf.ru/en/project/22-21-00885/>.

⁹https://github.com/kotelnikov-ev/economic_argument_generation

Table 6: Examples of text generation (English version of Table 5): claims (suffix “because” is not shown); generated sentences – premises; annotator scores’ voting results.

Claims	Sentences	Voting
<i>Banks should use biometrics more widely</i>	<u>Fine-tuned ruGPT-3:</u> <i>it allows you to simplify the procedure of interaction with customers and simplifies the payment process.</i>	4 from 4
	<u>Original ruGPT-3:</u> <i>it can greatly simplify the acquisition and processing of customer data.</i>	4 from 4
<i>Government bonds are one of the most reliable types of securities</i>	<u>Fine-tuned ruGPT-3:</u> <i>they are not subject to depreciation, do not lose their liquidity, their value remains stable.</i>	4 from 4
	<u>Original ruGPT-3:</u> <i>they are backed by a state guarantee, which includes the public debt.</i>	4 from 4
<i>Financial literacy training often leads to unwarranted confidence</i>	<u>Fine-tuned ruGPT-3:</u> <i>a person who wants to learn how to work with digital assets begins to think that he can easily manage them.</i>	3 from 4
	<u>Original ruGPT-3:</u> <i>people are not able to check how reliable and correct their statements are.</i>	3 from 4

References

- Khalid Al Khatib, Lukas Trautner, Henning Wachsmuth, Yufang Hou, and Benno Stein. 2021. Employing Argumentation Knowledge Graphs for Neural Argument Generation. // *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, P 4744–4754.
- Yonatan Bilu and Noam Slonim. 2016. Claim Synthesis via Predicate Recycling. // *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL) - Short Papers*, P 525–530.
- Yonatan Bilu, Ariel Gera, Daniel Hershcovich, Benjamin Sznajder, Dan Lahav, and et al. 2020. Argument Invention from First Principles. // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, P 1013–1026.
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, and et al. 2020. Language Models are Few-Shot Learners. // *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, P 1877–1901.
- Giuseppe Carenini and Johanna D. Moore. 2006. Generating and evaluating evaluative arguments. *Artificial Intelligence*, 170(11):925–952.
- Roxanne El Baff, Henning Wachsmuth, Khalid Al-Khatib, Manfred Stede, and Benno Stein. 2019. Computational Argumentation Synthesis as a Language Modeling Task. // *Proceedings of the 12th International Conference on Natural Language Generation*, P 54–64.
- William Fedus, Barret Zoph, and Noam Shazeer. 2021. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity.
- Irina Fishcheva and Evgeny Kotelnikov. 2019. Cross-Lingual Argumentation Mining for Russian Texts. // *Proceedings of the 8th International Conference “Analysis of Images, Social networks and Texts” (AIST 2019), Lecture Notes in Computer Science*, P 134–144.
- Irina Fishcheva, Valeriya Goloviznina, and Evgeny Kotelnikov. 2021. Traditional Machine Learning and Deep Learning Models for Argumentation Mining in Russian Texts. // *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog-2021”*, P 246–258.

- Shai Gretz, Yonatan Bilu, Edo Cohen-Karlik, and Noam Slonim. 2020. The workweek is the best time to start a family – A Study of GPT-2 Based Claim Generation. // *Findings of the Association for Computational Linguistics (EMNLP)*, P 528–544.
- Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, and et al. 2021. Pre-Trained Models: Past, Present and Future. *AI Open*.
- Christopher Hidey and Kathleen McKeown. 2019. Fixed that for you: Generating contrastive claims with semantic edits. // *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, volume 1, P 1756–1767.
- Xinyu Hua, Zhe Hu, and Lu Wang. 2020. Argument generation with retrieval, planning, and realization. // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, P 2661–2672.
- D Ilina, I Kononenko, and E Sidorova. 2021. On Developing a Web Resource to Study Argumentation in Popular Science Discourse. // *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog-2021”*, P 318–327.
- Nitish Shirish Keskar, Bryan McCann, Lav R. Varshney, Caiming Xiong, and Richard Socher. 2019. CTRL: A Conditional Transformer Language Model for Controllable Generation.
- Yuri Kuratov and Mikhail Arkhipov. 2019. Adaptation of deep bidirectional multilingual transformers for Russian language. // *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog-2019”*, P 333–340.
- J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.
- Andreas Peldszus and Manfred Stede. 2015. An Annotated Corpus of Argumentative Microtexts. // *Proceedings of the 1st European Conference on Argumentation*, P 801–815.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2020. Language Models are Unsupervised Multitask Learners. *OpenAI Technical report*.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, and et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21:1–67.
- Paul Reisert, Naoya Inoue, Naoaki Okazaki, and Kentaro Inui. 2015. A Computational Approach for Generating Toulmin Model Argumentation. // *Proceedings of the 2nd Workshop on Argumentation Mining*, P 45–55.
- ruGPT-3. 2020. pretrained language model. <https://sbercloud.ru/ru/datahub/rugpt3family/rugpt-3-large>.
- N. V. Salomatina, I. S. Kononenko, E. A. Sidorova, and I. S. Pimenov. 2021. Identification of connected arguments based on reasoning schemes “from expert opinion”. *Journal of Physics: Conference Series*, 1715.
- SBERT. 2020. pretrained language model. <https://developers.sber.ru/portal/services/sbert>.
- Benjamin Schiller, Johannes Daxenberger, and Iryna Gurevych. 2021. Aspect-Controlled Neural Argument Generation. // *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, P 380–396.
- Tal Schuster, Roei Schuster, Darsh J. Shah, and Regina Barzilay. 2020. The Limitations of Stylometry for Detecting Machine-Generated Fake News. *Computational Linguistics*, 46(2):499–510.
- Maria Skeppstedt, Andreas Peldszus, and Manfred Stede. 2018. More or less controlled elicitation of argumentative text: Enlarging a microtext corpus via crowdsourcing. // *Proceedings of the 5th Workshop on Argument Mining*, P 155–163.
- Christian Stab and Iryna Gurevych. 2014. Annotating argument components and relations in persuasive essays. // *Proceedings of the International Conference on Computational Linguistics*, P 1501–1510.
- Christian Stab, Tristan Miller, Pranav Schiller, Benjamin Rai, and Iryna Gurevych. 2018. Cross-topic argument mining from heterogeneous sources using attention-based neural networks. // *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, P 3664–3674.
- Manfred Stede and Jodi Schneider. 2018. Argumentation Mining. *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers.

Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, and et al. 2021. ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation.

Henning Wachsmuth, Manfred Stede, Roxanne El Baff, Khalid Al-Khatib, Maria Skeppstedt, and Benno Stein. 2018. Argumentation Synthesis following Rhetorical Strategies. // *Proceedings of the 27th International Conference on Computational Linguistics*, P 3753–3765.

Douglas N. Walton and Thomas F. Gordon. 2012. The Carneades Model of Argument Invention. *Pragmatics & Cognition.*, 20(1):1–31.

Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against Neural Fake News. // *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, P 9054–9065.