



оригинальная статья

<https://elibrary.ru/bjzjcz>

## Влияние соотношения параметров и данных на дообучение LLM в задачах классификации русских текстов

Шамигов Федор Федорович

Национальный исследовательский Томский государственный университет, Россия, Томск

<https://orcid.org/0009-0005-4933-2764>

fshamigov@mail.ru

**Аннотация:** Статья посвящена оптимизации дообучения (fine-tuning) больших языковых моделей (LLM) для задач классификации текстов на русском языке в условиях ограниченных вычислительных ресурсов. Предлагаемый метод основан на балансе между размером модели (числом параметров) и объемом обучающих данных: меньшая модель дообучается на большем датасете и сравнивается с большей моделью, дообученной на меньшем датасете. Цель – установить влияние соотношения параметров моделей и данных для дообучения на качество классификации текстов большими языковыми моделями. Выдвигается гипотеза о том, что «слабая» модель, дообученная на большем объеме данных, может показать близкое или более высокое качество классификации в сравнении с «сильной» моделью, дообученной на меньшем количестве данных. Актуальность исследования обусловлена необходимостью адаптации LLM к русскоязычным данным, где увеличение объема датасета может компенсировать меньший размер модели. Гипотеза проверялась на трех видах классификации: классификация тональности отзывов на фильмы, классификация тональности отзывов на сервисы и классификация новостей по топикам. Эксперименты проводились с использованием мультиязычных моделей: XLM-RoBERTa-comet-small (107 млн параметров) – «слабая» модель и XLM-RoBERTa-base (278 млн параметров) – «сильная» модель, на русскоязычных датасетах. Меньшая модель дообучалась на больших объемах данных (пропорционально разнице в параметрах), большая – на меньших. Сделан вывод о том, что «слабая» модель стабильно превосходит или достигает сопоставимых метрик по сравнению с «сильной» моделью, при этом затрачивая в 2–3 раза меньше вычислительных ресурсов (FLOPs), что демонстрирует практическую ценность подхода для энергоэффективного дообучения в русскоязычном контексте.

**Ключевые слова:** LLM, fine-tuning, XLM-RoBERTa, русскоязычные датасеты, классификация, тональность, топики

**Цитирование:** Шамигов Ф. Ф. Влияние соотношения параметров и данных на дообучение LLM в задачах классификации русских текстов. *Виртуальная коммуникация и социальные сети*. 2025. Т. 4. № 4. С. 344–352.  
<https://doi.org/10.21603/2782-4799-2025-4-4-344-352>

Поступила в редакцию 20.10.2025. Принята после рецензирования 10.11.2025. Принята в печать 11.11.2025.

full article

## Impact of Parameter-To-Data Ratio on LLM Fine-Tuning in Russian Text Classification Tasks

Fedor F. Shamigov

Tomsk State University, Russia, Tomsk

<https://orcid.org/0009-0005-4933-2764>

fshamigov@mail.ru

**Abstract:** This paper addresses the optimization of fine-tuning large language models (LLMs) for Russian-language text classification under constrained computational resources. The proposed approach hinges on balancing the model size (i.e., number of parameters) against the volume of training data: a smaller model is fine-tuned on a larger dataset and compared against a larger model fine-tuned on a smaller dataset. The aim was to establish the impact of different ratios of model parameters and data for further training on the quality

of text classification by large language models. We hypothesized that a weaker (i.e., smaller) model trained on more data could achieve classification performance comparable to or even surpassing that of a stronger (i.e., larger) model trained on less data. This hypothesis was motivated by the need to adapt LLMs to Russian-language tasks, where increased dataset size may compensate for reduced model capacity. The hypothesis was evaluated across three classification tasks: sentiment analysis of movie reviews, sentiment analysis of service reviews, and topic classification of news articles. The experiments were conducted on Russian-language datasets and employed the multilingual models XLM-RoBERTa-comet-small (107M parameters) for the weaker model and XLM-RoBERTa-base (278M parameters) for the stronger model. The smaller model was fine-tuned on proportionally larger datasets (scaled according to the parameter count difference) while the larger model used correspondingly smaller datasets. The weaker model consistently matched or exceeded the performance of the stronger model while requiring 2–3 times fewer computational resources (measured in FLOPs). The result highlights the practical value of this approach for energy-efficient fine-tuning in Russian-language settings.

**Keywords:** LLM, fine-tuning, XLM-RoBERTa, Russian-language datasets, text classification, sentiment, topic classification

**Citation:** Shamigov F. F. Impact of Parameter-To-Data Ratio on LLM Fine-Tuning in Russian Text Classification Tasks. *Virtual Communication and Social Networks*, 2025, 4(4): 344–352. (In Russ.) <https://doi.org/10.21603/2782-4799-2025-4-4-344-352>

Received 20 Oct 2025. Accepted after review 10 Nov 2025. Accepted for publication 11 Nov 2025.

## Введение

Изучение взаимосвязей между количеством параметров модели, объемом обучающих данных и вычислительными ресурсами является одной из ключевых задач в исследованиях больших языковых моделей (LLM). Большие языковые модели представляют собой нейронные сети, предобученные на огромных объемах текстовых данных и имеющие «параметры» – внутренние веса и связи нейронной сети (например, миллиарды числовых значений, которые определяют, как модель обрабатывает входные данные и генерирует предсказания; как правило, чем больше параметров, тем более сложные закономерности может выявить модель и тем качественнее она делает предсказания). Объем данных часто измеряется не просто числом примеров или слов, а количеством токенов – базовых единиц, на которые модель разбивает текст с помощью токенизатора (например, слово *журнал* может быть разбито на 2 токена: *жур* и *нал*). Законы масштабирования (scaling laws), систематизированные в работе [Kaplan et al. 2020], показали, что производительность моделей зависит от трех основных факторов: числа параметров, размера датасета (в токенах) и вычислительного бюджета (compute budget, измеряемого в FLOPs – операции с плавающей запятой, т.е. простые математические действия, которые модель выполняет, чтобы обучиться и дать ответ; чем их больше, тем

более мощная модель, а значит более «тяжелая» и дорогая). Также производительность подчиняется степенным законам, где традиционно считалось, что увеличение параметров (а следовательно, и «тяжести», и дороговизны модели) является основным источником улучшения качества.

Однако работа [Hoffmann et al. 2022] изменила эту парадигму, продемонстрировав, что оптимальное распределение ресурсов требует баланса между размером модели и объемом данных. Авторы показали, что модель Chinchilla (70 млрд параметров), обученная на 1,4 трлн токенов, превзошла по качеству работы более крупную модель Gopher (280 млрд параметров), обученную на 300 млрд токенов, что демонстрирует эффективность меньших моделей при увеличении данных на этапе предобучения (pre-training) – предварительного обучения модели, при котором она «понимает», как устроен язык, и «усваивает» общие знания. Исследователи также вывели оптимальное соотношение параметров к токенам – примерно 1:20. Последующие работы [Sardana et al. 2023] подтвердили стабильность этих закономерностей.

Особое внимание уделяется роли данных: исследование [Brown et al. 2020] показало, что большие объемы данных на этапе предобучения значительно повышают способность модели адаптироваться к узким задачам с минимальным дообучением.

Также Y. Liu и коллеги для модели RoBERTa продемонстрировали, что увеличение датасета компенсирует меньший размер модели, обеспечивая высокую производительность (например, добавление миллионов токенов позволило меньшей модели достичь метрик, сравнимых с крупными аналогами) [Liu et al. 2019].

В контексте дообучения (fine-tuning) – процесса адаптации предобученной модели к конкретной задаче путем обновления параметров модели на целевом датасете – законы масштабирования изучены недостаточно, особенно для русскоязычных данных. Существующие работы по дообучению [Anisuzzaman et al. 2025; Chung et al. 2022; Ding et al. 2023; Lialin et al. 2023; Lu et al. 2024; Pratap et al. 2025; Srinivasan et al. 2024; Wang et al. 2025] охватывают различные его методы, включая параметрически эффективные методы (PEFT), но не проводят систематические сравнения моделей с разным числом параметров и объемами данных на стандартизованных задачах. Исследования [Nikolich, Puchkova 2021; Nikolich et al. 2024; Smetanin, Komarov 2021] фокусируются на задачах, например, анализа тональности или суммаризации, но не анализируют баланс параметров и данных комплексно.

Например, в области классификации тональности Е. С. Плешкова и коллеги сравнивали методы ML (логистическая регрессия, многослойный перцептрон, случайный лес, наивный Байес, KNN, дерево решений, SGD) с векторизацией BoW и TF-IDF на твитах, где SGD показал наилучшие результаты по метрикам ( $F_1 \sim 82\%$ ) [Плешакова и др. 2022].

В области улучшения классификации цифровых документов А. К. Марков и соавторы проводили сравнительный анализ технологий NLP с использованием OCR (оптическое распознавание символов) и моделей глубокого обучения на образовательных документах, достигнув высоких показателей качества классификации ( $F_1 \sim 95\%$ ) [Марков и др. 2024].

В области жанровой классификации литературных текстов П. И. Максименко дообучал модель BERT для многометочной, многоклассовой и бинарной классификации на базе фанфиксов, достигнув качества классификации по метрике  $F_1$  от 53 % до 92 % [Максименко 2025].

В области классификации новостных текстов Э. А. Чельышев и др. сравнивали методы ML (наивный Байес, случайный лес, логистическая регрессия, нейронная сеть), в результате чего достигли значений  $F_1$  при классификации новостей от ~75 % до ~93 % [Чельышев и др. 2022].

В области классификации и анализа тональности текстов из Интернета Т. В. Гальцев и С. А. Нестеров исследовали методы ML (наивный Байес, логистическая регрессия, SVM, дерево решений) и достигли значений  $F_1$  от 80 % до 98% [Гальцева, Нестеров 2024].

Другие ученые [Blinova, Tarasov 2022] также рассматривают классификацию для решения различных задач и применяют разнообразные методы, но не рассматривают комплексно влияние баланса параметров модели и данных для дообучения на эффективность модели.

Настоящее исследование фокусируется на оптимизации дообучения LLM именно для задач классификации текстов на русском языке – актуальной области NLP, которая рассматривается многими учеными, где они стремятся в основном получить более высокие показатели качества работы модели, но не фокусируются на оптимизации ее обучения. В задаче классификации модель присваивает метки входным текстам (например, *позитивный / негативный* для отзывов или *спорт / политика* для новостей).

Цель – установить влияние соотношения параметров моделей и данных для дообучения на качество классификации текстов большими языковыми моделями.

Задачи исследования:

- отобрать русскоязычные данные для задачи классификации;
- предобработать и подготовить данные к дообучению;
- отобрать и настроить модели с различным числом параметров для проведения экспериментов по дообучению;
- провести дообучение выбранных моделей с варьированием объема данных;
- оценить производительность моделей с использованием формальных метрик;
- сравнить результаты экспериментов, проанализировать влияние соотношения параметров и данных на качество работы моделей в задаче классификации текстов.

Новизна исследования заключается в систематическом сравнении производительности моделей с разным числом параметров и объемом данных в контексте дообучения для классификации на русскоязычных данных. Актуальность работы обусловлена необходимостью оптимизации дообучения LLM в условиях ограниченных ресурсов.

Проверяется гипотеза: меньшие модели (с меньшим числом параметров), дообученные на больших объемах данных, могут показывать близкие или

превосходящие результаты в сравнении с крупными моделями, дообученными на меньших датасетах, в задаче классификации на материале русского языка. Для проверки гипотезы были выполнены следующие основные шаги: сбор и обработка данных для дообучения, создание датасетов разных размеров для каждой задачи, дообучение моделей с разным количеством параметров, сравнение результатов работы моделей по метрикам.

## Методы и материалы

В работе были использованы следующие методы: предобработка данных (приведение данных к формату JSON, удаление специальных символов, пробелов и эмодзи, дубликатов и неполных записей, разделение на обучающие и валидационные выборки), дообучение моделей для классификации текстов, сравнительный анализ производительности моделей.

Для проверки гипотезы о влиянии соотношения числа параметров модели и объема обучающих данных на качество дообучения больших языковых моделей была выбрана аналитическая задача классификации в трех вариантах: классификация тональности отзывов на фильмы, классификация тональности отзывов на сервисы и классификация новостей по топикам. Задача классификации заключается в отнесении входной последовательности текста (например, текста отзыва) к одному (или нескольким) предопределенным классам (например, *негативный* или *позитивный*). Выбор нескольких вариантов классификации позволит сделать разностороннюю оценку производительности моделей.

В экспериментах использовались две мультиязычные модели на основе архитектуры RoBERTa с поддержкой русского языка: XLM-RoBERTa-comet-small (107 млн параметров) и XLM-RoBERTa-base (278 млн параметров). Меньшая модель (XLM-RoBERTa-comet-small) дообучалась на больших датасетах, а более крупная (XLM-RoBERTa-base) – на меньших, пропорционально различию в параметрах.

В качестве материала были выбраны три открытых русскоязычных датасета для каждого вида классификации. Для каждой задачи создавались два обучающих датасета: больший – для XLM-RoBERTa-comet-small и меньший – для XLM-RoBERTa-base,

пропорционально разнице в числе параметров (примерно 2,6:1). Для каждой задачи также создавалась тестовая выборка. Все датасеты прошли через предобработку, которая включала в себя приведение данных к формату JSON (структура "ключ": "значение", например, "text": "текст отзыва"), удаление специальных символов, пробелов и эмодзи, дубликатов и неполных записей, а также разделение на выборки. Далее мы рассмотрим каждый из трех видов классификации, выбранных для проверки гипотезы, и разделение их датасетов на выборки для дообучения:

1. Классификация тональности отзывов на фильмы – датасет Kinopoisk's movies reviews<sup>1</sup>, содержащий отзывы пользователей с метками *позитивный* (*pos*), *негативный* (*neg*) и *нейтральный* (*neu*). В этой задаче модели необходимо предсказать, к какому из трех классов относится отзыв.

Разделение на выборки:

- XLM-RoBERTa-comet-small: ~69000 отзывов (~42 млн токенов);
- XLM-RoBERTa-base: ~24000 отзывов (~15 млн токенов);
- Валидационная выборка: ~10000 отзывов (~6,2 млн токенов).

Примеры данных в готовом формате JSON (где поле *text* – отрывок отзыва, поле *label* – метка тональности):

```
{"text": "Не кривя душой, отмечу тот факт, что это одна из лучших спортивных биографий...", "label": "pos"}  
{"text": "Фильм поразил своей глубиной...", "label": "pos"}
```

2. Классификация тональности отзывов на сервисы – Russian Sentiment Dataset<sup>2</sup>, содержащий отзывы с метками 0, 1, 2. В этой задаче (как и в первой) модель должна предсказать, к какому из трех классов относится отзыв.

Разделение на выборки:

- XLM-RoBERTa-comet-small: ~49000 отзывов (~10,5 млн токенов);
- XLM-RoBERTa-base: ~17500 отзывов (~3,7 млн токенов);
- Валидационная выборка: ~7000 отзывов (~1,5 млн токенов).

<sup>1</sup> Kinopoisk's movies reviews. Kaggle. URL: <https://www.kaggle.com/datasets/mikhailklemin/kinopoisks-movies-reviews> (accessed 14 Oct 2025).

<sup>2</sup> Russian sentiment dataset. Kaggle. URL: <https://www.kaggle.com/datasets/senylar/sis-text-class> (accessed 14 Oct 2025).

Примеры данных в готовом формате JSON:

```
{"text": "Месторасположение – прекрасное, лучшее для проживания в Париже. Качество отеля, вот тут вопрос. Номера тесные, шумоизоляция плохая, номера довольно старые, правда, в части номеров уже идет ремонт...", "label": 0}  
{"text": "Это лучший отель, в котором я когда-либо...", "label": 0}
```

3. Классификация новостей по топикам – датасет Lenta.ru<sup>3</sup>. Выбраны 10 популярных топиков: *Россия, Мир, Экономика, Спорт, Культура, Бывший СССР, Наука и техника, Интернет и СМИ, Из жизни, Дом* в равном количестве. В этой задаче модели нужно предсказать, к какому из десяти топиков относится текст новости.

Разделение на выборки:

- XLM-RoBERTa-comet-small: ~75000 новостей (~23 млн токенов);
- XLM-RoBERTa-base: ~30000 новостей (~9 млн токенов);
- Валидационная выборка: ~15000 новостей (~4,6 млн токенов).

Примеры данных в готовом формате JSON:

```
{"text": "Бывший президент СССР Михаил Горбачев в воскресенье получил престижную музыкальную премию Grammy, сообщает Reuters. Михаил Горбачев был награжден в номинации \"Лучший текстовый альбом для детей\" за...", "label": "Культура"}  
{"text": "Московское «Динамо» одержало победу над самарскими «Крыльями Советов» в матче десятого тура.....", "label": "Спорт"}
```

## Дообучение

Дообучение происходило в сервисе Kaggle с использованием графического ускорителя P100. Применялись следующие гиперпараметры:

- Learning rate:  $2 \times 10^{-5}$ ;
- Batch size: 16;
- Количество эпох: 4/5;
- Weight decay: 0,01.

Дообучение XLM-RoBERTa-comet-small для задач классификации тональности отзывов на фильмы, классификации тональности отзывов на сервисы, классификации топиков новостей заняло ~60 минут,

~45 минут, ~80 минут соответственно. Дообучение XLM-RoBERTa-base для тех же задач составило ~117 минут, ~84 минуты, ~150 минут соответственно.

## Метрики

Для оценки результатов работы моделей использовались следующие метрики:

- Accuracy – доля правильно предсказанных меток среди всех предсказаний;
- Precision – доля верно предсказанных положительных классов среди всех предсказанных как положительные;
- Recall – доля верно предсказанных положительных классов среди всех истинно положительных;
- F1 – гармоническое среднее между Precision и Recall, учитывающее баланс между ними.

## Результаты

В рамках исследования было проведено дообучение 6 моделей: 3 модели на базе XLM-RoBERTa-comet-small для классификации тональности отзывов на фильмы, классификации тональности отзывов на сервисы, классификации новостей по топикам, а также 3 модели на базе XLM-RoBERTa-base для тех же задач.

## Классификация тональности отзывов на фильмы

Результаты дообучения моделей для задачи классификации тональности отзывов на фильмы приведены в таблице 1. Верхняя часть таблицы отражает производительность XLM-RoBERTa-comet-small, нижняя – XLM-RoBERTa-base по эпохам. Колонка *Epoch* указывает номер эпохи, *Training Loss* и *Validation Loss* показывают потери при обучении и валидации (чем меньше, тем лучше), а метрики *Accuracy*, *F1*, *Precision* и *Recall* отражают качество предсказаний на валидационной выборке (значения ближе к 1 указывают на более точные предсказания).

Модель XLM-RoBERTa-comet-small демонстрирует стабильное улучшение метрик по мере дообучения, что свидетельствует об эффективной адаптации к задаче благодаря большому объему данных (~42 млн токенов). Напротив, XLM-RoBERTa-base показывает метрики, близкие к случайному угадыванию (~0,33 для трехклассовой задачи). Вычислительный бюджет составил ~9,14e15 FLOPs

<sup>3</sup> News dataset from Lenta.Ru. Kaggle. URL: <https://www.kaggle.com/datasets/yutkin/corpus-of-russian-news-articles-from-lenta> (accessed 14 Oct 2025).

для XLM-RoBERTa-comet-small и ~2,57e16 FLOPs для XLM-RoBERTa-base, что отражает более высокую энергоэффективность меньшей модели при лучшем качестве.

#### Классификация тональности отзывов на сервисы

Результаты дообучения для задачи классификации тональности отзывов на сервисы представлены в таблице 2.

Модель XLM-RoBERTa-comet-small значительно превосходит XLM-RoBERTa-base по всем метрикам. Вычислительные затраты составили ~6,51e15 FLOPs для меньшей модели и ~1,84e16 FLOPs для большей, что подчеркивает энергоэффективность XLM-RoBERTa-comet-small.

#### Классификация новостей по топикам

Обе модели показывают близкие результаты, но XLM-RoBERTa-comet-small немного превосходит XLM-RoBERTa-base при меньших затратах (~9,95e15 FLOPs против ~3,14e16 FLOPs), что делает ее более эффективной (табл. 3).

Результаты дообучения моделей XLM-RoBERTa-comet-small (107 млн параметров; обучены на больших датасетах) и XLM-RoBERTa-base (278 млн параметров; обучены на меньших датасетах) на наборе задач классификации – классификация тональности отзывов на фильмы и сервисы, классификация новостей по топикам – подтвердили гипотезу исследования: дообучение менее мощной модели на большем количестве данных оказалось эффективнее,

Табл. 1. Результаты оценки дообученных моделей (классификация тональности отзывов на фильмы)  
Tab. 1. Evaluation results of fine-tuned models: sentiment classification of movie reviews

XLM-RoBERTa-comet-small (107 млн параметров), ~69000 сэмплов						
Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0,602300	0,603706	0,757696	0,587547	0,655956	0,582659
2	0,558300	0,548053	0,769447	0,638122	0,662220	0,630188
3	0,504400	0,555429	0,775760	0,647897	0,689742	0,624254
4	0,459100	0,571828	0,775954	0,648336	0,691605	0,62335
XLM-RoBERTa-base (278 млн параметров), ~24000 сэмплов						
Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0,883000	0,881431	0,658930	0,264801	0,219643	0,333333
2	0,886300	0,881648	0,658930	0,264801	0,219643	0,333333
3	0,878000	0,877434	0,658930	0,264801	0,219643	0,333333
4	0,903100	0,876557	0,658930	0,264801	0,219643	0,333333

Табл. 2. Результаты оценки дообученных моделей (классификация тональности отзывов на сервисы)  
Tab. 2. Evaluation results of fine-tuned models: sentiment classification of service reviews

XLM-RoBERTa-comet-small (107 млн параметров), ~49000 сэмплов						
Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0,527300	0,547745	0,754989	0,751599	0,745709	0,763574
2	0,473300	0,508967	0,773803	0,763627	0,766510	0,761276
3	0,410600	0,521468	0,777366	0,770718	0,765535	0,777928
4	0,330100	0,541656	0,779504	0,772028	0,771574	0,772500
XLM-RoBERTa-base (278 млн параметров), ~ 17500 сэмплов						
Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	1,062200	1,061831	0,470211	0,213217	0,156759	0,333232
2	1,060100	1,061012	0,470353	0,213261	0,156784	0,333333
3	1,058700	1,059122	0,470353	0,213261	0,156784	0,333333
4	1,058400	1,059291	0,470353	0,213261	0,156784	0,333333

Табл. 3. Результаты оценки дообученных моделей (классификация новостей по топикам)  
Tab. 3. Evaluation results of fine-tuned models: topic classification of news

XLM-RoBERTa-comet-small (107 млн параметров), ~75000 сэмплов						
Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0,455800	0,432171	0,867000	0,864935	0,867601	0,867000
2	0,350900	0,358244	0,893200	0,891638	0,893332	0,893200
3	0,245900	0,345640	0,899133	0,898485	0,898663	0,899133
4	0,187000	0,359772	0,900800	0,899960	0,900352	0,900800

  

XLM-RoBERTa-base (278 млн параметров), ~ 30000 сэмплов						
Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0,518100	0,473979	0,849133	0,844452	0,850760	0,849133
2	0,378500	0,379568	0,880733	0,878761	0,882756	0,880733
3	0,266500	0,358379	0,893000	0,892448	0,893593	0,893000
4	0,158000	0,391681	0,898400	0,897600	0,897759	0,898400

чем дообучение более крупной модели на меньшем количестве данных.

Модель XLM-RoBERTa-comet-small, несмотря на меньшее число параметров и меньшие вычислительные затраты (FLOPs), либо превосходит XLM-RoBERTa-base, либо демонстрирует сопоставимое качество на всех рассмотренных задачах. Это свидетельствует в пользу энергоэффективного подхода, при котором предпочтение отдается меньшей модели и большему количеству данных.

Объективность оценки обеспечивается спецификой задачи классификации: ответы модели однозначно интерпретируемые и точно измеряются формальными метриками, что позволяет проводить надежное сравнение моделей.

Результаты могут иметь прикладное значение для анализа контента в сферах виртуальной коммуникации и социальных медиа. Задачи классификации репрезентативны для реальных сценариев мониторинга пользовательских мнений в соцсетях, комментариях, форумах, платформах отзывов и др. Энергоэффективный подход, основанный на дообучении компактных моделей на расширенных датасетах, позволит развертывать высокоточные системы классификации даже в условиях ограниченных вычислительных ресурсов, что актуально для оперативного мониторинга пользовательского контента в реальном времени. Кроме того, применение подобных моделей может быть адаптировано для отслеживания трендов в медиапространстве, выявления доминирующих нарративов и сегментации аудиторий в цифровой среде, что расширяет возможности контент-аналитики в рамках исследований виртуальной коммуникации.

## Заключение

Проведенное исследование подтвердило гипотезу о том, что модели с меньшим числом параметров, дообученные на больших объемах данных, способны демонстрировать сопоставимую или превосходящую производительность в сравнении с более крупными моделями, дообученными на меньших датасетах. Эксперименты с дообучением моделей XLM-RoBERTa-comet-small и XLM-RoBERTa-base на русскоязычных датасетах для классификации показали, что меньшая модель стабильно достигает высоких результатов при меньших вычислительных затратах (FLOPs). Это согласуется с выводами о законах масштабирования, что подчеркивает их применимость к дообучению.

Результаты демонстрируют практическую значимость для оптимизации процессов дообучения в условиях ограниченных ресурсов.

Перспективы дальнейших исследований:

- проведение экспериментов на мультиязычных датасетах для проверки гипотезы в различных языковых контекстах;
- расширение спектра задач (ответы на вопросы, NER, генерация кода и др.);
- проведение экспериментов с моделями, которые предлагают больше вариаций параметров.

**Конфликт интересов:** Автор заявил об отсутствии потенциальных конфликтов интересов в отношении исследования, авторства и / или публикации данной статьи.

**Conflict of interests:** The author declared no potential conflict of interests regarding the research, authorship, and / or publication of this article.

## Литература / References

- Гальцева Т. В., Нестеров С. А. Классификация и определение тональности текстов, публикуемых в сети Интернет. *Системный анализ в проектировании и управлении: XXVII Междунар. науч.-практ. конф.* (Санкт-Петербург, 13–14 октября 2023 г.) СПб.: ПОЛИТЕХ-ПРЕСС, 2024. Ч. 2. С. 491–498. [Galtseva T. V., Nesterov S. A. Classification and sentiment analysis of texts published on the Internet. *System analysis in design and management: Proc. XXVII Intern. Sci.-Prac. Conf.*, St. Petersburg, 13–14 Oct 2023. St. Petersburg: POLITEKH-PRESS, 2024, pt. 2, 491–498. (In Russ.)] <https://doi.org/10.18720/SPBPU/2/id24-202>
- Максименко П. И. Жанровая классификация литературных текстов с применением нейросетевых методов (на материале русскоязычной электронной базы фэнтези). *Человек: образ и сущность. Гуманистические аспекты. 2025. № 1. С. 184–200.* [Maksimenko P. I. Genre classification of literary texts through neural network methods (based on the Russian-language electronic fanfiction database). *Human being: Image and Essence. Humanitarian Aspects*, 2025, (1): 184–200. (In Russ.)] <https://doi.org/10.31249/chel/2025.01.13>
- Марков А. К., Семеночкин Д. О., Кравец А. Г., Яновский Т. А. Сравнительный анализ применяемых технологий обработки естественного языка для улучшения качества классификации цифровых документов. *International Journal of Open Information Technologies*. 2024. Т. 12. № 3. С. 66–77. [Markov A. K., Semyonochkin D. O., Kravets A. G., Yanovskiy T. A. Comparative analysis of applied natural language processing technologies for improving the quality of digital document classification. *International Journal of Open Information Technologies*, 2024, 12(3): 66–77. (In Russ.)] <https://elibRARY.ru/tubosi>
- Плешакова Е. С., Гатауллин С. Т., Осипов А. В., Романова Е. В., Самбуров Н. С. Эффективная классификация текстов на естественном языке и определение тональности речи с использованием выбранных методов машинного обучения. *Вопросы безопасности*. 2022. № 4. С. 1–14. [Pleshakova E. S., Gataullin S. T., Osipov A. V., Romanova E. V., Samburov N. S. Effective classification of natural language texts and determination of speech tonality using selected machine learning methods. *Security Issues*, 2022, (4): 1–14. (In Russ.)] <https://doi.org/10.25136/2409-7543.2022.4.38658>
- Челышев Э. А., Оцоков Ш. А., Раскатова М. В., Щеголев П. Сравнение методов классификации русскоязычных новостных текстов с использованием алгоритмов машинного обучения. *Вестник кибернетики*. 2022. № 1. С. 63–71. [Chelyshev E. A., Otsokov Sh. A., Raskatova M. V., Shchegolev P. Comparing classification methods for news texts in Russian using machine learning algorithms. *Proceedings in Cybernetics*, 2022, (1): 63–71. (In Russ.)] <https://doi.org/10.34822/1999-7604-2022-1-63-71>
- Anisuzzaman D. M., Malins J. G., Friedman P. A., Attia Z. I. Fine-tuning large language models for specialized use cases. *Mayo Clinic Proceedings: Digital Health*, 2025, 3(1). <https://doi.org/10.1016/j.mcpdig.2024.11.005>
- Blinova O., Tarasov N. A hybrid model of complexity estimation: Evidence from Russian legal texts. *Frontiers in Artificial Intelligence*, 2022, 5. <https://doi.org/10.3389/frai.2022.1008530>
- Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., Agarwal A., Herbert-Voss A., Krueger G., Henighan T., Child R., Ramesh A., Ziegler D. M., Wu J., Winter C., Hesse C., Chen M., Sigler E., Litwin M., Gray S., Chess B., Clark J., Berner C., McCandlish S., Radford A., Sutskever I., Amodei D. Language models are few-shot learners. *arXiv*, 2020. <https://doi.org/10.48550/arXiv.2005.14165>
- Chung H. W., Hou L., Longpre S., Zoph B., Tay Y., Fedus W., Li Y., Wang X., Dehghani M., Brahma S., Webson A., Gu S. S., Dai Z., Suzgun M., Chen X., Chowdhery A., Castro-Ros A., Pellat M., Robinson K., Valter D., Narang S., Mishra G., Yu A., Zhao V., Huang Y., Dai A., Yu H., Petrov S., Chi Ed H., Dean J., Devlin J., Roberts A., Zhou D., Le Q. V., Wei J. Scaling instruction-finetuned language models. *arXiv*, 2022. <https://doi.org/10.48550/arXiv.2210.11416>
- Ding N., Qin Y., Yang G., Wei F., Yang Z., Su Y., Hu S., Chen Y., Chan C.-M., Chen W., Yi J., Zhao W., Wang X., Liu Z., Zheng H.-T., Chen J., Liu Y., Tang J., Li J., Sun M. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 2023, 5: 220–235. <https://doi.org/10.1038/s42256-023-00626-4>
- Hoffmann J., Borgeaud S., Mensch A., Buchatskaya E., Cai T., Rutherford E., de Las Casas D., Hendricks L. A., Welbi J., Clark A., Hennigan T., Noland E., Millican K., Van den Driessche G., Damoc B., Guy A., Osindero S., Simonyan K., Rae J. W., Vinyals O., Sifre L. Training compute-optimal large language models. *arXiv*, 2022. <https://doi.org/10.48550/arXiv.2203.15556>
- Kaplan J., McCandlish S., Henighan T., Brown T. B., Chess B., Child R., Gray S., Radford A., Wu J., Amodei D. Scaling laws for neural language models. *arXiv*, 2020. <https://doi.org/10.48550/arXiv.2001.08361>

- Lialin V., Deshpande V., Yao X., Rumshisky A. Scaling down to scale up: A guide to parameter-efficient fine-tuning. *arXiv*, 2023. <https://doi.org/10.48550/arXiv.2303.15647>
- Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv*, 2019. <https://doi.org/10.48550/arXiv.1907.11692>
- Lu W., Luu R. K., Buehler M. J. Fine-tuning large language models for domain adaptation: Exploration of training strategies, scaling, model merging and synergistic capabilities. *arXiv*, 2024. <https://doi.org/10.48550/arXiv.2409.03444>
- Nikolich A., Korolev K., Bratchikov S., Kiselev I., Shelmanov A. Vikhr: The family of open-source instruction-tuned large language models for Russian. *arXiv*, 2024. <https://doi.org/10.48550/arXiv.2405.13929>
- Nikolich A., Puchkova A. Fine-tuning GPT-3 for Russian text summarization. *arXiv*, 2021. <https://doi.org/10.48550/arXiv.2108.03502>
- Pratap S., Aranha A. R., Kumar D., Malhotra G., Iyer A. P. N., Shylaja S. S. The fine art of fine-tuning: A structured review of advanced LLM fine-tuning techniques. *Natural Language Processing Journal*, 2025, 11. <https://doi.org/10.1016/j.nlp.2025.100144>
- Sardana N., Portes J., Doubov S., Franke J. Beyond chinchilla-optimal: Accounting for inference in language model scaling laws. *arXiv*, 2023. <https://doi.org/10.48550/arXiv.2401.00448>
- Smetanin S., Komarov M. Deep transfer learning baselines for sentiment analysis in Russian. *Information Processing & Management*, 2021, 58(3). <https://doi.org/10.1016/j.ipm.2020.102484>
- Srinivasan K. P. V., Gumpena P., Yattapu M., Brahmbhatt V. H. Comparative analysis of different efficient fine tuning methods of large language models (LLMs) in low-resource setting. *arXiv*, 2024. <https://doi.org/10.48550/arXiv.2405.13181>
- Wang L., Chen S., Jiang L., Pan S., Cai R., Yang S., Yang F. Parameter-efficient fine-tuning in large language models: A survey of methodologies. *Artificial Intelligence Review*, 2025, 58. <https://doi.org/10.1007/s10462-025-11236-4>