

УДК 004.89

doi:

ЗАДАЧИ АНАЛИЗА ТОНАЛЬНОСТИ В НОВОСТНЫХ ТЕКСТАХ

С.О. Уразов (*urazov.msu@gmail.com*)^A

Н.В. Лукашевич (*louk_nat@mail.ru*)^A

^A Московский Государственный Университет имени М. В.
Ломоносова, Москва

В этой статье описывается исследование применения больших языковых моделей (LLM) в различных постановках задачи анализа тональности, как например: поиск мнений конкретного лица или мнений о конкретном лице; классификация (на негативную, нейтральную или позитивную) тональности мнения по отношению к цели; и выделение отношения между двумя сущностями в тексте. Исследования были проведены на датасете русских новостных текстов соревнования RuOpinionNE-2024, с использованием нейросетевых моделей типа BERT и Ruadapt-Qwen2.5. Была проведена серия экспериментов с использованием техники обучения моделей и подсказок (промптов).

Ключевые слова: Анализ тональности, Большие языковые модели, Промпт инжиниринг, Тонкая настройка с помощью техники LoRA.

Введение

Анализ тональности текста является активно развивающейся областью автоматической обработки естественного языка. Целью такого анализа обычно является выявление мнений относительно сущностей и определение общей тональности текста. Например, такой анализ лежит в основе политических исследований в рамках определения настроений в обществе и для отслеживания отношения к конкретным личностям или компаниям [Smetanin, 2020; Семина, 2020].

Анализ тональности в новостных текстах может использоваться для решения разных конкретных задач. Иногда необходимо собрать мнения по разным вопросам какого-то лица (например, недавно избранного на высокую должность). Или, наоборот, важны разные мнения по отношению

к этому лицу. В третьей постановке могут быть важны отношения между людьми, организациями и т. д.

В данной статье рассматриваются различные постановки задач анализа тональности в новостных текстах. Для решения выделенных задач используются как энкодерные модели, такие как BERT [BERT, 2019], так и генеративные (декодерные) модели, например Ruadapt-Qwen2.5¹ [Qwen2.5, 2025], которые адаптированы под русский язык. Рассмотрены различные способы составления запросов для моделей (пром프트-инжиниринг) и влияние подбора различных промптов во время обучения моделей.

Для проведения исследований использовался датасет новостных текстов на русском языке соревнования RuOpinionNE-2024 [RuOpinionNE-2024, 2025], в котором выделены кортежи мнений, включающие источник мнения, объект мнения, тональность мнения и оценочное выражение.

Вклад этой статьи следующий:

- Представлены различные постановки задач анализа тональности новостных текстов;
- Проведены эксперименты с использованием различных подходов к работе с моделями и к технике построения подсказок;
- Проанализированы зависимости, тенденции и различия в результатах использования различных подходов.

Дальнейшая структура статьи следующая. В главе 1 рассматриваются современные подходы, проводится краткий обзор работ по теме. В главе 2 детально описываются постановки задач, которые исследуются в этой статье. В главе 3 дается обзор датасета RuOpinionNE-24. В главе 4 описывается преобразование датасета под особенности каждой формулировки задачи. В главе 5 описываются и обсуждаются полученные результаты. В заключении подводится итог исследования.

1. Обзор работ по теме

Первые методы в области изучения анализа тональности были основаны на составлении правил, словарей, специальном представлении текста (мешок слов, опорные вектора и др.), а также на основе рекуррентных и сверточных нейронных сетей [Effective lstms, 2015]. Серьезный скачок в качестве автоматического анализа произошел при использовании методов, использующих трансформеры, применение которых во многих актуальных задачах дает высокие результаты [BERT, 2019; Golubev et al., 2020; Structured sentiment analysis, 2021; Semeval, 2022].

В настоящее время активно исследуется применение больших генеративных моделей, которые имеют возможности к более глубокому

¹ <https://huggingface.co/collections/RefalMachine/ruadapt-qwen-25-67124a497e75205228348919>

пониманию текста [GPT-4o, 2024] и способности к составлению сложного, структурированного ответа. Были исследованы возможности больших языковых моделей в формате zero-shot и few-shot [Sentiment analysis, 2023] в различных постановках задачи анализа тональности. Это исследование показало, что применение серии моделей GPT-3.5 сравнимо по эффективности с применением тонко настроенной на соответствующую задачу Flan-UL2, несмотря на большую разницу в размерах моделей.

В соревновании RuOpinionNE-24 [RuOpinionNE-24, 2025] лучшие результаты в задаче выделения кортежей мнения среди русских новостных текстов были получены с использованием модели LLaMa-3.3-70B, тонко настроенной с помощью техники QLoRA [LoRA, 2021]. Кортежи мнений включали в себя источник мнения, цель мнения, тональность и выражение, которым описывается мнение источника к цели, и модель была настроена на то, чтобы выделять такие кортежи целиком [Vatolin et al., 2025]. Среди других участников соревнования высокие результаты были получены тоже с применением больших генеративных моделей при помощи тонкой настройки и few-shot формата [Rossyaykin, 2025].

2. Описание постановок задач

2.1. Задача извлечения мнений некоторого лица (источника)

В этой задаче необходимо находить фрагменты текстов (предложения), включающие явно или неявно выраженные мнения, которые были высказаны конкретным лицом, что может помочь составить картину взглядов этого лица, отследить их изменения. Данная задача представляет собой задачу бинарной классификации: есть в предложении мнение или нет.

2.2. Задача извлечения мнений по отношению к некоторой сущности (лицу, организации, событию).

В этой задаче необходимо находить фрагменты текстов (предложения), включающие явно или неявно выраженные мнения разных источников по отношению к заданному лицу. Данная задача также представляет собой задачу бинарной классификации.

Обе задачи возникают, например, при назначении на высокие должности новых лиц, когда требуется представить портрет их убеждений и взаимоотношений с другими людьми.

2.3. Классификация тональности отношения к объекту мнения

Данная задача заключается в определении того, какая именно тональность (нейтральная, позитивная или негативная) обращена к конкретной сущности в тексте, являющейся целью мнения.

Эту задачу также можно решить двумя способами:

- Подойти к задаче напрямую: провести классификацию на три класса;

- Используя результаты первой и второй задач, провести бинарную классификацию тех текстов, в которых есть полярное мнение.

2.4. Классификация отношений между сущностями

Четвертая задача – определить тональность взаимоотношения между сущностями, выделяя в них источник мнения, цель мнения и полярность мнения (негативное, нейтральное или позитивное).

3. Датасет RuOpinionNE-2024 и специализированные датасеты для задач

Датасет RuOpinionNE-2024 представляет собой набор размеченных отрывков новостных текстов, в которых выделены кортежи мнений (источник мнения, цель мнения, тональность мнения, выражение мнения). Источник мнения может отсутствовать ("NULL") или быть автором текста ("AUTHOR").

Мнения могут выражены как явно, так и имплицитно. Например, по предложению «Apple и Samsung нарушали патенты друг друга» должны быть извлечены кортежи:

(Apple, Samsung, NEG, “нарушали патенты друг друга”) и
(Samsung, Apple, NEG, “нарушали патенты друг друга”).

Датасет представлен в виде совокупности предложений, к каждому из которых приписано множество высказанных в них мнений. Предложение может не содержать никаких мнений, и тогда множество будет пустое. Источником и объектом мнения в RuOpinion-2024 могут быть сущности одного из следующих типов: PERSON, ORGANIZATION, PROFESSION, NATIONALITY, COUNTRY, REGION, CITY, IDEOLOGY.

Исходный датасет используется для формирования датасетов для всех конкретных задач. Например, в датасете для первой задачи сущностям заданных типов ставятся в соответствие значения 1 (источник мнения) или 0 (не источник мнения). В датасете второй задачи сущностям заданных типов сопоставляются значения 1 (объект мнения) или 0 (не объект мнения). Для третьей задачи каждой сущности ставится в соответствие одна из трех меток: POS, NEG, NEU.

Датасет четвертой задачи состоит из триплетов: две сущности и тональность отношения между ними. Соответственно, если пара источник-цель присутствовали в кортеже оригинального датасета, то ей ставилась в соответствие метка тональности и оценочное выражение. В противном случае, если такая пара не связана мнением, то ей сопоставлялся кортеж с нейтральной тональностью и пустым текстом (например, {NEU, []}) [Willard et al., 2023].

4. Модели

В качестве методов для решения представленных задач использовались модели на основе архитектуры трансформер. В задачах 1-4 использовалась модель на основе многоязычного энкодера трансформера XLM-RoBERTa-large. Данная модель может принимать на вход два фрагмента текста, разделенные служебным токеном [SEP]. Это позволяет использовать в модели специальный запрос (промпт), который формулирует задачу. Целевое предложение задается как первый фрагмент текста, а промпт как второй фрагмент текста после токена [SEP]. Для тонкой настройки все слои размораживались. Использовалась косинусная кривая обучения с разогревом в 3 эпохи. Процесс тонкой настройки для задач 1-3 проводился на 22 эпохах, для задачи 4 – на 8 эпохах. Размер батча – 16.

В задаче 4 также проводились эксперименты с генеративной моделью Ruadapt-Qwen2.5, которая представляет собой русифицированную версию модели Qwen2.5.

Для выделения именованных сущностей использовалась модель BINDER [Optimizing Bi-Encoder, 2022], обученная извлечению именованных сущностей на датасете NEREL².

5. Результаты экспериментов и обсуждение

5.1. Первая задача определения источника мнения

Для задачи определения источника мнения использовались "запросы" при идентификации источников в тексте.

Среди промптов были использованы следующие:

1. (имя сущности) – источник тональности;
2. (имя сущности) – выражает мнение;
3. (имя сущности) позитивно или негативно относится к сущности в тексте;
4. (имя сущности) – источник позитивного или негативного мнения;
5. ***@(имя сущности)*** – источник позитивного или негативного мнения к объекту текста.

Результаты усреднялись по трем экспериментам. В каждом эксперименте проводилась тонкая настройка с нуля и тестирование модели. Пиковое значение скорости обучения: 1.5e-6. Результаты по разным промптам представлены в Табл. 1.

Табл. 1.

Промпт	Precision	Recall	F1 score
1	0.73 +-0.02	0.86 +-0.01	0.78 +-0.01
2	0.77 +-0.01	0.84 +-0.01	0.80 +-0.01

² <https://github.com/fulstock/binder>

3	0.74 +-0.01	0.84 +-0.01	0.77 +-0.01
4	0.75 +-0.01	0.84 +-0.01	0.78 +-0.01
5	0.76 +-0.01	0.85 +-0.01	0.80-0.01

5.2. Вторая задача определения объекта мнения

Для извлечения объекта мнения использовались следующие промпты:

1. (имя сущности) – это цель тональности;
2. (имя сущности) – это цель мнения;
3. В тексте выражается позитивное или негативное мнение к сущности (имя сущности).

Пиковое значение скорости обучения: 2e-6. Результаты были получены следующие (см. Табл. 2):

Табл. 2.

Промпт	Precision	Recall	F1 score
1	0.77 +-0.02	0.82 +-0.02	0.80 +-0.01
2	0.76 +-0.02	0.80 +-0.02	0.78 +-0.01
3	0.77 +-0.01	0.81 +-0.01	0.79 +-0.01

Таким образом, видно, что в обеих подзадачах бинарной классификации определения источника мнения или объекта мнения качество достигает 80% по F-мере. Наблюдалось, что применении модели в запросе на конкретную сущность, каждая модель склонна отвечать правильно, но иногда путались роли источника и цели.

5.3 Классификация тональности мнения по отношению к сущности

Результаты работы классификации на три класса и на два класса (результаты усреднялись по трем экспериментам, пиковое значение скорости обучения: 3e-6) представлены в Табл. 3:

Табл. 3.

Классификация	Precision	Recall	F1 score
3 класса	0.65 +-0.01	0.70 +-0.01	0.68 +-0.01
2 задача + 2 класса	0.65 +-0.01	0.70 +-0.01	0.67 +-0.01

В работе над бинарной классификацией, для оценки работы были смешаны результаты поиска целей мнения из предыдущей задачи и совмещены с работой классификатора, чтобы провести оценку общей классификации на три класса.

В экспериментах наблюдалось, что в рамках текущего датасета работа модели слабо зависела от предлагаемого ей промпта. При этом лучшим оказался промпт вида: «Тональность (имя сущности) равна».

Результаты трехклассового и бинарного подходов сравнительно схожи и стабильны, что показывает равнозначность рассмотрения разбиения полной задачи на подзадачи. Это важный результат для перехода к рассмотрению третьей задачи, в которой сильная опора идет на результаты первой задачи.

5.4. Задача извлечения тональности отношений между сущностями

Для установления зависимости между источником мнения и целью мнения был использован пакет извлечения отношений OpenNRE [OpenNRE, 2019]. В нашем случае в качестве отношения задается тональность мнения источника к цели.

Аналогично предыдущим задачам, были добавлены следующие подсказки к основному тексту примера, чтобы помочь модели фокусироваться:

- 1. Какое отношение высказывает источник (имя источника) по отношению к цели (имя цели);
- 2. Какую тональность источник (имя источника) выражает к (имя цели);
- 3. Какую тональность источник (имя источника) явно или неявно выражает относительно цели (имя цели);
- 4. Какую тональность источник *@*(имя источника)*@* выражает к >>>(имя цели)<<<.
- 5. Отношение источника (имя источника) к цели (имя цели).

Пиковое значение скорости обучения: 9.5e-6. Результаты по разным запросам представлены в Табл. 4:

Табл. 4.

Модель	Precision	Recall	F1 score
1	0.83 +-0.01	0.62 +-0.01	0.71 +-0.01
2	0.86 +-0.01	0.63 +-0.01	0.73 +-0.01
3	0.87 +-0.01	0.67 +-0.01	0.76 +-0.01
4	0.87 +-0.01	0.67 +-0.01	0.76 +-0.01
5	0.84 +-0.01	0.62 +-0.01	0.71 +-0.01

Оценка работы проводилась по тем примерам, в которых выделен кортеж, соответствующий кортежам оригинального датасета.

Выделение сущностей в самой подсказке (промпт 4) не повлияло на результаты.

Для проведения экспериментов с декодер-моделями была использована модель Ruadapt/Qwen2.5-7B-ext-u48-instruct. Для тонкой настройки этой модели был использован подход LoRA с квантизацией 8bit. Lora rank: 8.

Lora alpha: 32. Lora dropout: 0.1. Шаги накопления градиентов: 4. Пиковое значение скорости обучения: 6e-5.

Во многих примерах модель не была уверена в своем ответе, поэтому могла отвечать по-разному для одного и того же примера. Поэтому был использован прием, описанный в работе участников соревнования RuOpinionNE [Vatolin, 2025] – Most Common N, в котором среди 10 выданных тональностей бралась самая часто выделяемая тональность из них.

Сначала были проведены эксперименты со следующими системными промптами, используя пользовательский промпт 3 из предыдущей задачи:

1. Ты – полезный и умный инструмент для анализа тональности текста.
2. Ты – полезный и умный инструмент для анализа тональности текста. Твоя роль состоит в выявлении отношения в виде структуры {Polarity, Expression}, где Polarity это NEG, POS или NEU, а Expression это часть текста запроса.
3. Ты Qwen, созданный Alibaba Cloud. Ты – полезный помощник.
4. Ты – профессиональный журналист. Твоя роль состоит в выявлении отношения в виде структуры {Polarity, Expression}, где Polarity это NEG, POS или NEU, а Expression это часть текста запроса.
5. Ты – профессиональный психолог. Твоя роль состоит в выявлении отношения в виде структуры {Polarity, Expression}, где Polarity это NEG, POS или NEU, а Expression это часть текста запроса.

Были получены следующие результаты (см. Табл. 5):

Табл. 5.

Промпт	Precision	Recall	F1 score
1	0.98 +0.01	0.58 +0.03	0.72 +0.02
2	0.98 +0.01	0.56 +0.03	0.70 +0.02
3	0.97 +0.01	0.56 +0.02	0.71 +0.03
4	0.98 +0.01	0.58 +0.03	0.72 +0.02
5	0.97 +0.01	0.58 +0.04	0.72 +0.02

Здесь наблюдается точность, практически равная 1 (precision $\approx$ 0.982), но при этом полнота оказывается значительно ниже (recall $\approx$ 0.574). Это говорит о том, что модель не улавливает многие зависимости между сущностями, однако если улавливает, то улавливает правильно.

Оказался интересным тот факт, что лучше всех показал себя достаточно обобщенный системный промпт 1. При этом он показывает результаты немного лучше, чем часто используемый промпт 3.

Эксперименты показали, что при уточнении и расширении промпта 1 (см. промпт 2) наблюдается следующее. Функция потерь на первых шагах

тонкой настройки промпта 1 более чем в два раза выше, однако после нескольких шагов ситуация менялась, и функция потерь промпта 1 в среднем становилась меньше, чем у промпта 2 (см. рис. 1).

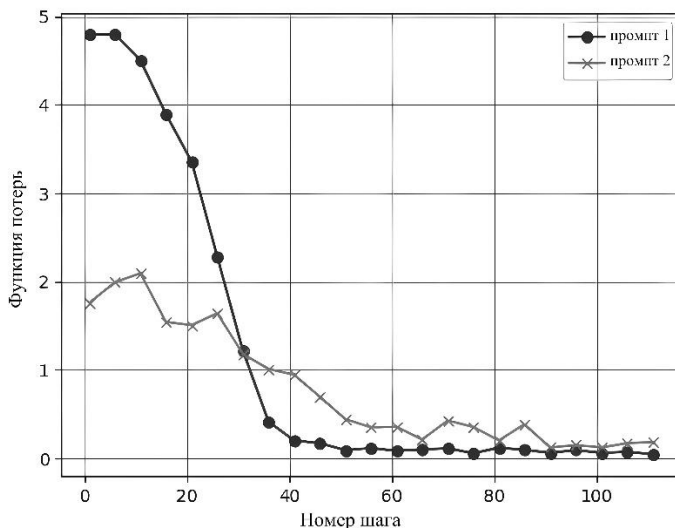


Рис.1. Графики функций зависимости потерь от номера шага промпта 1 и 2

С выбранным лучшим системным промптом были проведены эксперименты со следующими пользовательскими промптами:

1. Какую тональность выражает (имя источника) явно или неявно выражает относительно объекта (имя цели)?;
2. Какое отношение источник (имя источника) явно или неявно выражает относительно объекта (имя цели)? Ответ в виде структуры {(Отношение), (Выражение)};
3. Какое отношение источник (имя источника) явно или неявно выражает относительно объекта (имя цели)? Ответ в виде структуры {(Отношение), (Выражение)}, где (Отношение) это NEG, POS или NEU, а (Выражение) это копия части текста запроса;
4. Какую тональность выражает (имя источника) явно или неявно выражает относительно объекта (имя цели)? Ответ в виде структуры {(Тональность), (Выражение)}, где (Тональность) это NEG, POS или NEU, а (Выражение) это копия части текста запроса;
5. С точки зрения журналистики, какую тональность выражает (имя источника) явно или неявно выражает относительно объекта (имя цели)? Ответ в виде структуры {(Тональность), (Выражение)};

6. С точки зрения психологии, какую тональность выражает (имя источника) явно или неявно выражает относительно объекта (имя цели)? Ответ в виде структуры {(Тональность), (Выражение)}.

На пользовательских промптах были получены следующие результаты (см. Табл. 6):

Табл. 6.

Промпт	Precision	Recall	F1 score
1	0.98 +-0.01	0.58 +-0.03	0.72 +-0.02
2	0.97 +-0.01	0.56 +-0.01	0.70 +-0.01
3	0.98 +-0.01	0.55 +-0.01	0.71 +-0.01
4	0.98 +-0.01	0.58 +-0.03	0.73 +-0.02
5	0.97 +-0.01	0.59 +-0.04	0.73 +-0.04
7	0.97 +-0.01	0.58 +-0.04	0.73 +-0.04

Таким образом, исходя из полученных результатов, лучшие результаты при тонкой настройке достигаются при использовании краткого лаконичного системного и более подробного пользовательского промптов.

При использовании большей модели той же архитектуры (Ruadapt/Qwen2.5-14B-instruct) результаты становятся значительно выше: Recall возрастает до ≈ 0.73 , а общий F1 score – до ≈ 0.84 , что уже превосходит возможности классификации моделей, основанных на моделях-энкодерах.

Заключение

В этой статье были исследованы способы разделения задачи анализа отношения между сущностями на подзадачи в рамках датасета русских новостных текстов соревнования RuOpinionNE-24. Каждая подзадача была проанализирована относительно использования разных моделей и соответствующих промптов.

Было изучено взаимодействие моделей в рамках рассмотренных задач. Рассмотрены различные подходы к тонкой настройке моделей на конкретных задачах.

По результатам исследований каждой подзадачи была получена система составления оценки отношения между сущностями в тексте, которая позволяет достаточно эффективно (F1 score ≈ 0.84 с использованием модели Ruadapt/Qwen2.5-14B-instruct) извлекать тональность отношения между сущностями в предложенном тексте.

Благодарности. Исследование выполнено в рамках государственного задания МГУ имени М. В. Ломоносова.

Список литературы

- [Семина, 2020] Семина Т. А. Анализ тональности текста: современные подходы и существующие проблемы // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6, Языкознание: Реферативный журнал. 2020. Вып. 4. С. 47–64.
- [Structured sentiment analysis, 2021] Structured sentiment analysis as dependency graph parsing. / J. Barnes [et al]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. V. 1. pp. 3387-3402.
- [Semeval, 2022] Semeval 2022 task 10: Structured sentiment analysis. / J. Barnes [et al] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022. pp. 1280-1295.
- [BERT, 2019] Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019. URL: <https://arxiv.org/abs/1810.04805> (дата обращения: 14.06.2025).
- [Golubev et al., 2020] Golubev A., Loukachevitch N. Improving results on Russian sentiment datasets. Conference on artificial intelligence and natural language. 2020. pp. 109-121. URL: <https://arxiv.org/abs/2007.14310> (дата обращения: 14.06.2025).
- [OpenNRE, 2019] OpenNRE: An Open and Extensible Toolkit for Neural Relation Extraction. / X. Han [et al]. 2019. URL: <https://openreview.net/pdf?id=9EAQVEINuum> (дата обращения: 14.06.2025).
- [LoRA, 2021] LoRA: Low-Rank Adaptation of Large Language Models / Hu E. J. [et al]. 2021. URL: <https://arxiv.org/abs/2106.09685> (дата обращения: 14.06.2025).
- [GPT-4o, 2024] GPT-4o System Card. / A. Hurst [et al.]. 2024. URL: <https://arxiv.org/pdf/2410.21276> (дата обращения: 14.06.2025).
- [RuOpinionNE-2024, 2024] RuOpinionNE-2024: Extraction of Opinion Tuples from Russian News Texts. / N. V. Loukachevitch [et al]. 2025. URL: <https://arxiv.org/html/2504.06947v1> (дата обращения: 14.06.2025).
- [Qwen, 2025] Qwen2.5 Technical Report. Qwen [et al]. 2025. URL: <https://arxiv.org/abs/2412.15115> (дата обращения: 14.06.2025).
- [Rosyaykin, 2025] Rosyaykin P. Structured sentiment analysis using few-shot prompting of an ensemble of LLMs. 2025.
- [Smetanin, 2020] Smetanin S. The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives. IEEE Access. 2020. Vol. 8. pp. 110693–110719. URL: <https://api.semanticscholar.org/CorpusID:220078379>.
- [Effective lstms, 2016] Tang D., Qin B., Feng X., Liu T. Effective lstms for target-dependent senti ment classification. // International Conference on Computational Linguistics. 2016. URL: <https://arxiv.org/abs/1512.01100> (дата обращения: 14.06.2025).
- [Vatolin, 2025] Vatolin A. Structured Sentiment Analysis with Large Language Models: A Winning Solution for RuOpinionNE-2024. 2025.
- [Willard et al., 2023] Willard B. T., Louf R. Efficient guided generation for large language models. 2023. URL: <https://arxiv.org/abs/2307.09702> (дата обращения: 14.06.2025).

- [**Optimizing Bi-Encoder, 2022**] Zhang S., Cheng H., Gao J., Poon H. Optimizing Bi-Encoder for Named Entity Recognition via Contrastive Learning. 2022. URL: <https://arxiv.org/pdf/2307.09702> (дата обращения: 14.06.2025).
- [**Sentiment analysis, 2023**] Zhang W., Deng Y., Liu B., Pan S., Bing L. Sentiment analysis in the era of large language models: A reality check. Findings of the Association for Computational Linguistics: NAACL 2024. pp. 3881–3906.

SENTIMENT ANALYSIS TASKS FOR NEWS TEXTS

S.O. Urazov (*urazov.msu@gmail.com*)

N.V. Loukachevitch (*louk_nat@mail.ru*)

Lomonosov Moscow State University, Moscow

This paper is dedicated to our research of the application of large language models (LLM) in various formulations of sentiment analysis task, such as finding opinions of and about a specific person; classifying (as negative, neutral or positive) the sentiment of an opinion towards a target; and extracting the relationship between two entities in a text. The study was conducted on the RuOpinionNE-2024 Russian news texts dataset using neural network models such as BERT and Ruadapt-Qwen2.5. A series of experiments were conducted using model learning techniques and prompts.

Key words: Sentiment analysis, Large language Models, Prompt engineering, LoRA fine tuning.