

Индекс УДК 004.89:316.77:81'42

Код ГРНТИ 28.23.21

DOI: 10.22204/2587-8956-2025-123-04-56-69



**Ю.В. БАЛАКИНА,  
М.В. ГРИГОРЬЕВА,  
Е.Н. СОКОЛОВА\***

## Применение больших языковых моделей для анализа ценностно-патриотического дискурса русскоязычных пользователей\*\*

В статье рассматриваются возможности применения больших языковых моделей (LLM) для автоматизированного анализа ценностно-патриотического дискурса русскоязычных пользователей социальных медиа. На материале корпуса сообщений из VK, «Одноклассников» и Telegram (2023–2025 гг.) исследуется, насколько результаты автоматической кодировки совпадают с экспертной разметкой по специально разработанной категориальной схеме. Кодбук включает восемь измерений: базовые ценности по Ш. Шварцу, две оси Р. Инглхарта (традиционализм / светскость; выживание / самовыражение), уровни потребностей по А. Маслоу, типы патриотизма (конструктивный / агрессивный) по мотивам К.Д. Ушинского и В.С. Соловьёва, доминирующие типы речевых актов по Дж. Остину, а также бинарные индикаторы эксплицитного патриотизма и гражданской идентичности. Эксперимент проведён на кластере сообщений «Гордость и патриотизм» (N = 456), где плотность ценностных маркеров максимально высока; сопоставление реализовано через матрицы ошибок, accuracy, macro/weighted F1 и коэффициент к Коэна. Показано, что LLM надёжно выделяет эксплицитную патриотическую тематику, но существенно менее согласована с экспертами при многоклассовой и тонкой ценностной классификации (Шварц, Маслоу, шкалы Инглхарта, типы патриотизма, речевые акты Остина), демонстрируя систематические смещения и гипердиагностику отдельных категорий. Сделан вывод, что LLM в текущей конфигурации может использоваться как вспомогательный инструмент предварительной разметки и генерации гипотез, но не как автономный заменитель экспертного контент-анализа ценностного дискурса.

\* **Балакина Юлия Владимировна** — кандидат филологических наук, доцент, профессор департамента фундаментальной и прикладной лингвистики Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ–Нижний Новгород).  
E-mail: [julianaumova@gmail.com](mailto:julianaumova@gmail.com)

**Григорьева Мария Владимировна** — старший преподаватель Школы коммуникаций факультета креативных индустрий НИУ ВШЭ.  
E-mail: [mariya.grigoreva@hse.ru](mailto:mariya.grigoreva@hse.ru)

**Соколова Екатерина Никитична** — кандидат политических наук, заведующая проектно-учебной лабораторией Политических коммуникаций Школы коммуникаций факультета креативных индустрий НИУ ВШЭ.  
E-mail: [e.sokolova@hse.ru](mailto:e.sokolova@hse.ru)

\*\* Статья подготовлена в Национальном исследовательском университете «Высшая школа экономики» в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № ФИ-2025-83 «Современные технологии социально-политической коммуникации»).

**Ключевые слова:** большие языковые модели, LLM, ценности, патриотизм, гражданская идентичность, контент-анализ, Шварц, Инглхарт, Маслоу, речевые акты, социальные медиа, русскоязычный дискурс, автоматическая разметка

Тексты пользователей социальных медиа — от постов в VK и «Одноклассниках» до сообщений в Telegram-каналах — источник информации о ценностных ориентациях общества. Извлечение и анализ ценностно-патриотического контента сопряжены с определёнными трудностями, обусловленными большим объёмом неструктурированных данных, сложностью интерпретации сленга, иронии и культурных аллюзий, а также структурными и формальными особенностями русского языка. В данном контексте возрастает интерес к применению больших языковых моделей (LLM) для автоматизированного и смыслового анализа содержания подобных текстов. В частности, такие модели, как GigaChat, YandexGPT, ChatGPT и DeepSeek, демонстрируют способность автоматически классифицировать тексты в рамках отдельных задач. Тем не менее, на данном этапе рассматривать модели в качестве инструмента социокультурного анализа пока корректно лишь в качестве методологической гипотезы, требующей независимой валидации, репликации и проверки устойчивости результатов.

В настоящей статье рассматриваются возможности использования LLM для анализа русскоязычных пользовательских текстов на темы ценностей, патриотизма и гражданской идентичности. Статья включает теоретический обзор понятия ценностей и патриотизма в языке пользователей, обсуждает лингвистические и методологические возможности LLM при работе с русским языком, описывает методику сбора и обработки корпуса из сообщений социальных медиа (2023–2025 гг.), демонстрирует результаты автоматической атрибуции ценностных категорий с помощью LLM и их сопоставление с экспертной разметкой, а так-

же анализирует выводы из проведённого эксперимента.

### Возможности LLM для анализа русскоязычной речи пользователей

Ценности в социальной психологии определяются как устойчивые убеждения и жизненные принципы, выражающие представления о желаемых целях и выступающие критериями оценивания событий [1].

Традиционно ценности изучались с помощью опросов (опросник Шварца, World Values Survey), однако в последнее десятилетие растёт интерес к анализу ценностей на материале цифровых следов, в том числе на текстах в социальных сетях. Соцмедиа отражают актуальные изменения массового сознания и позволяют исследовать ценностные ориентации в естественном контексте общения. Однако существует проблема дифференциации собственно ценностей и схожих понятий (личностных черт, мнений и пр.) в пользовательских сообщениях. Наряду с этим традиционные методы, основанные на лексических словарях (например, LIWC) и подходе «мешка слов», демонстрируют низкую корреляцию с опросными профилями ценностей (в том числе по методике Шварца) [12].

В ответ на эти вызовы развиваются новые методы, опирающиеся на контекстуальные модели и машинное обучение, применение которых предполагает выявление ценностно-экспрессивных сообщений, а не профилирование личности пользователя по его постам.

Большие языковые модели, обученные на огромных корпусах текстов, демонстрируют определённые способности «понимания» и генерирования текста на разных языках. Применительно к русскоязычному контенту LLM (например, GPT-4)

для корректного анализа содержания сообщений должны учитывать особенности русского языка.

В частности, в процессе обработки русскоязычных текстов модель GPT-4 обнаруживает существенные ограничения, обусловленные спецификой разговорной лексики, идиоматических выражений и некомпозитивных языковых конструкций, а также недостатком контекста и обученности на англоязычной базе, что затрудняет автоматический анализ на уровне ключевых слов, особенно с учётом специфики морфологической системы русского языка [13].

Наряду с этим пользовательские тексты нередко содержат элементы жаргона и сленга, эмодзи, смешение кодов, сарказм и иронию. LLM, будучи обучены на данных из Интернета, способны «расшифровать» многие широко употребляемые жаргонные выражения (модель GPT-4 Turbo), однако в некоторых случаях модели затруднительно уловить сарказм и иронию. Так, если пользователь без явных индикаторов иронизирует, LLM может интерпретировать текст буквально и потому ошибочно определить его эмоциональную окраску. Исследования показывают, что даже продвинутые модели уступают специализированным алгоритмам, обученным на корпусах саркастических примеров [14, 15]. В частности, GPT-4 распознаёт сарказм лучше предшественников и других LLM (она превосходит другие модели в тестах приблизительно на 14% по точности) [14], но всё же заметно отстаёт от человеческого уровня и от узко настроенных моделей, обученных именно для детекции сарказма [16].

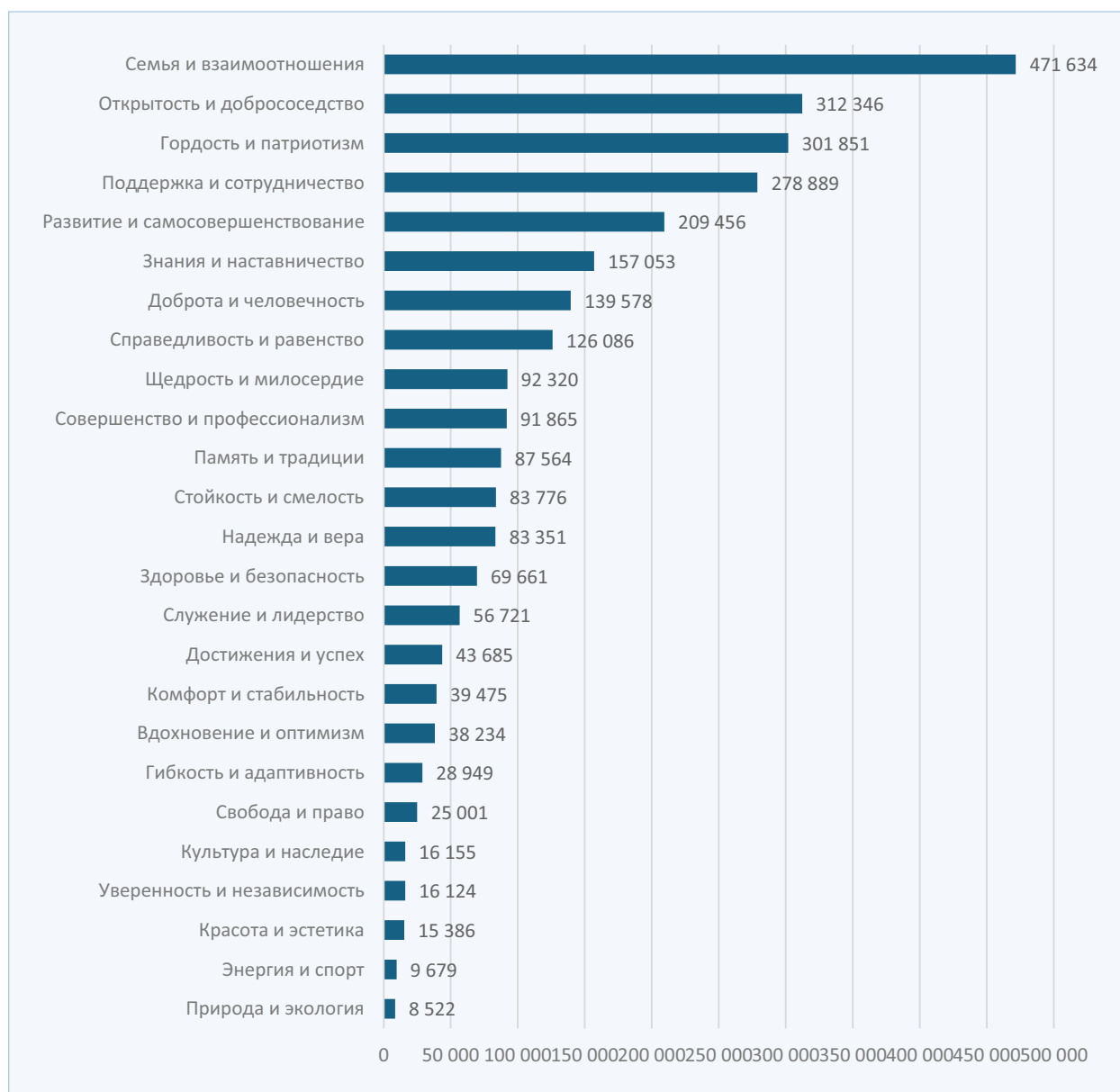
Кроме того, LLM обучены преимущественно на статистике языковых паттернов и могут допускать искажения, унаследованные из данных. Например, если в обучающей выборке широко встречается определённый нарратив, модель может склоняться к такому же пониманию и пропускать альтернативные проявления патриотизма. Во-вторых, классификация

непрозрачна: неясно, по каким признакам ChatGPT относит сообщение к той или иной категории. Одновременно признаётся чувствительность LLM к формулировке запроса и настройкам генерации, что требует дополнительных протоколов контроля и сопоставления результатов с ручной экспертной разметкой [17].

### Методология исследования

В качестве исходного массива использован корпус из более чем двух миллионов публичных сообщений российского сегмента социальных сетей, сформированный средствами системы Brand Analytics. В корпус вошли сообщения на русском языке, опубликованные в 2023–2024 гг. во «ВКонтакте», «Одноклассниках», Telegram и на открытых веб-площадках, доступных без авторизации и без обхода технических ограничений. Географический фокус задавался привязкой пользователей и сообществ к России. Отбор сообщений осуществлялся по 38 формульным лексико-семантическим маркерам (например, конструкции типа «для меня важно», «я горжусь», «главное в жизни» и др.), что позволило целенаправленно выделить тексты с эксплицитной артикуляцией ценностей, а также минимизировать долю бытового и новостного шума.

Предобработка включала удаление технических элементов, полную деидентификацию (ники, персональные ссылки, контакты, идентифицирующие URL), исключение точных и почти точных дубликатов и коммерческого спама, лемматизацию (PyMystem3/PyMorphy2), расширенный стоп-список и отсеечение крайне редких лексем при сохранении ценностно значимой лексики. Структурирование ценностного поля выполнялось в два этапа: тематическое моделирование LDA с завышенным числом тем ( $K = 5000$ ) для выделения тонких латентных компонент; кластеризация агрегированных тематических представлений на основе эмбедингов Word2Vec методом k-means ( $k = 25$ ) с многократными перезапусками и выбором решения



Ил. 1. Распределение ценностных кластеров в российском сегменте соцмедиа (число сообщений, в которых присутствует соответствующая ценностная тематика)

по минимальной внутрикластерной дисперсии. Полученные кластеры прошли независимую экспертную проверку: для каждого были заданы устойчивые наборы ключевой лексики и содержательные определения.

Структура ценностного пространства, достаточная для понимания дальнейших шагов, обобщена на ил. 1.

Кластер «Гордость и патриотизм» трактуется как совокупность сообщений с явно позитивным или нормативным отношением к России, её истории, символике, армии,

государственным институтам и коллективной идентичности, с риторикой гордости, долга, благодарности, защиты и служения. Кластер выбран для анализа как фрагмент с высокой концентрацией явных ценностных маркеров, что позволяет оценить не просто наличие смысла, а качество его ценностной атрибуции.

Из данного кластера случайным образом отобраны 456 сообщений пользователей и публичных сообществ без дополнительных идеологических фильтров, что снижает селекционное смещение и сохраняет ре-

альный спектр патриотического дискурса в этом сегменте. Анализ имеет формат контролируемого локального эксперимента и служит исключительно для сопоставления двух процедур атрибуции: независимой экспертной разметки по формализованному кодировочному справочнику и автоматической разметки большой языковой моделью (ChatGPT GPT-5 PRO) при фиксированных инструкциях и параметрах<sup>1</sup>. Обе разметки выполнялись независимо: эксперты не видели результатов модели, модель не имела доступа к человеческим меткам.

### Схема категориальной разметки

Для сопоставления экспертной и LLM-аннотации использовалась единая категориальная схема с восемью измерениями; по каждому измерению каждому сообщению присваивалось одно значение из фиксированного перечня.

- 1. Базовая ценность по Шварцу (Schwartz).** Доминирующая ценность: Власть, Достижения, Гедонизм, Стимуляция, Самонаправленность, Универсализм, Благожелательность, Традиции, Конформность, Безопасность, Нет. Операционализация основана на переработанной модели Ш Шварца; значение присваивается при наличии явного ценностного акцента, при его отсутствии — «Нет».
- 2. Традиционные и секулярно-рациональные ценности по Р. Инглхарту (Inglehart\_trad\_vs\_secular).** Два состояния: traditional (явные ссылки на религию, «традиционные ценности», авторитет государства/армии/семьи, иерархию, послушание) и neutral (отсутствие выраженного традиционализма, в том числе светско-рациональные высказывания). Фиксируется только явно выраженный традиционалистский полюс.
- 3. Ценности выживания и самовыражения по Р. Инглхарту (Inglehart\_survival\_vs\_selfexpr).** Три значения: survival (угрозы, базовая безопасность, порядок любой ценой, материальное выживание), self-expression (самореализация, права, свободы, участие, протест, креативность, экологическая и гражданская повестка), neutral (нет достаточных признаков обоих полюсов).
- 4. Уровни потребностей по А. Маслоу.** Доминирующий уровень потребностей: Физиологические, Безопасность, Принадлежность/любовь, Уважение/признание, Самореализация. Схема соответствует классической модели А. Маслоу; выбор уровня основан на ведущем мотиве сообщения.
- 5. Тип патриотического высказывания по модели «Ушинский–Соловьёв».** Тип патриотического высказывания: конструктивный (нравственно мотивированная любовь к стране, ответственность, ориентация на общее благо при признании ценности других народов), агрессивный (шовинизм, ксенофобия, противопоставление «своих» и «чужих», оправдание насилия), нет (патриотическая проблематика отсутствует или выражена недостаточно).
- 6. Тип речевого акта по Дж. Остину (Austin).** Доминирующий тип речевого акта: вердиктивы — verdictive (оценки, «суждения» и заключения), экзерситивы — exercitive (призывы, требования, выражение воли), бегабитивы — behabitive (выражения отношения, одобрения, порицания, уважения/неуважения), экспозитивы — expositive (объяснение, интерпретация, оформление аргументации). Класс определяется по преобладающей интенции высказывания.
- 7. Патриотизм.** Бинарный индикатор эксплицитной патриотической тематики: ИСТИНА — при прямых упоминаниях России, Родины, Отечества, национальных символов, армии, Победы, территориальной целостности; ЛОЖЬ — при их отсутствии. Поле фиксирует наличие темы, а не её оценку.
- 8. Гражданская идентичность.** Бинарный индикатор гражданской идентичности:

<sup>1</sup> Полный текст промпта представлен в Приложении.

ИСТИНА — при апелляциях к правам и обязанностям гражданина, участию в выборах и институтах, формуле «мы, граждане», ответственности за государство и общество; ЛОЖЬ — при отсутствии таких маркеров.

### Интерпретация результатов

Анализируемый корпус сообщений представляет собой целостный идеологический контур, воспроизводящий официально-патриотический дискурс. Ядро составляют государственные символы (флаг, герб, гимн), героизированное прошлое, поддержка действующей власти, армии и СВО, а также нормативная эмоция гордости за великую страну. В координатах теории ценностей Шварца и модернизационной схемы Инглхарта доминируют кластеры «традиции — конформизм — безопасность» с выборочной интеграцией мотивов достижения и самореализации, но строго в русле одобряемых государством практик (служба, участие в грантах, патриотические акции). В терминах Маслоу преобладают уровни принадлежности и безопасности: субъекту предлагается идентичность, основанная на лояльности и коллективной сплочённости перед внешними и внутренними угрозами.

Анализ речевых актов (в рамках типологии Дж. Остина) демонстрирует, что значительная часть высказываний носит не только экспозитивный (объясняющий), но также бехабитивный и экзерситивный характер: через поздравления, декларации гордости, призывы к участию в акциях, поддержку армии, изменение онлайн-статусов и воспроизведение ритуалов пользователи вовлекаются в перформативное подтверждение лояльности. При этом декларируемый патриотизм почти повсеместно присутствует как позитивная норма, тогда как собственно гражданская идентичность (как осознание прав, ответственности и субъектности по отношению к государству) артикулируется слабо. Таким образом, формируется модель «правильного» гражданина как эмоционально

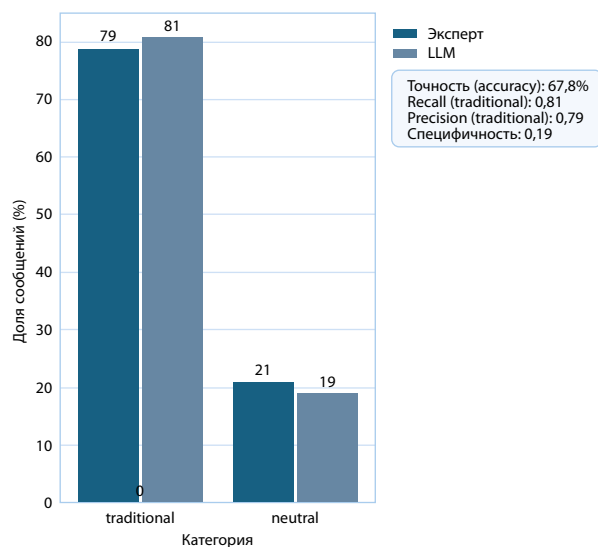
преданного и ценностно традиционного участника коллективного «мы», при этом маргинализируются элементы критической рефлексии.

Экспертная разметка по ценностям Шварца была сопоставлена с разметкой LLM по следующим критериям: доля точных совпадений, распределение классов, матрица ошибок, precision, recall и F1-мера по каждому классу, а также интегральные показатели (macro- и weighted-F1) и коэффициент к Коэна.

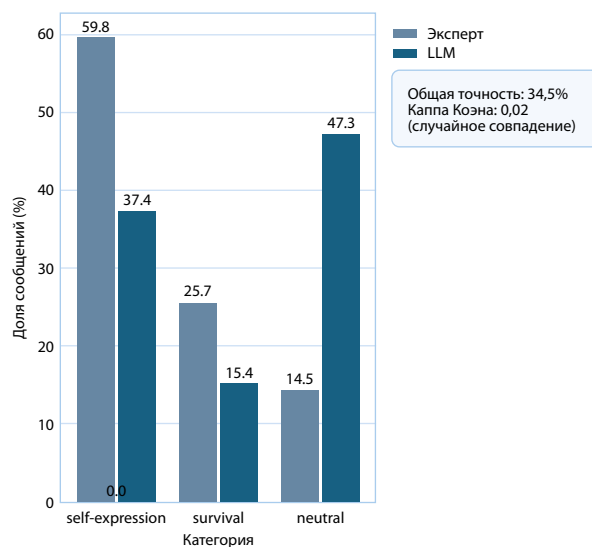
Полученные результаты показали, что точность совпадений LLM с экспертом составила 0,408 при  $\kappa = 0,172$ , а macro-F1 = 0,318 и weighted-F1 = 0,485; что позволяет зафиксировать недостаточную согласованность модели с экспертным эталоном. Анализ распределений и матрицы ошибок выявил систематические смещения. В экспертной разметке явно доминирует ценность «Традиции», тогда как модель существенно занижает её долю, одновременно завышая частоту меток «Власть» и ряда других ценностей, которые у эксперта либо редки, либо практически не встречаются. Для класса «Традиции» модель демонстрирует относительно высокую точность при пониженной полноте ( $F1 \approx 0,61$ ), однако для «Достижений» и особенно «Власти» наблюдается выраженное ухудшение показателей за счёт большого числа ложных срабатываний. Сравнение показывает, что без дополнительной настройки языковая модель не воспроизводит структуру экспертных ценностных категорий, недооценивает доминирующие для корпуса ориентации и вводит «избыточные» ценностные интерпретации.

При проведении сравнительного анализа в рамках бинарной кодировки выявляются существенные различия в распределении традиционных и секулярно-рациональных ценностей согласно концептуальной модели Р. Инглхарта (*Inglehart\_trad\_vs\_secular*): сообщения, отнесённые экспертом к категории *traditional*, противопоставлялись всем остальным высказываниям, отнесённым к *neutral*. По экспертной разметке

Inglehart\_trad\_vs\_secular: распределение по категориям



Inglehart\_survival\_vs\_selfexpr: распределение по категориям



Ил. 2. Разметка по двум осям модели Инглхарта–Вельцеля (традиционные-светско-рациональные; выживание-самовыражение)

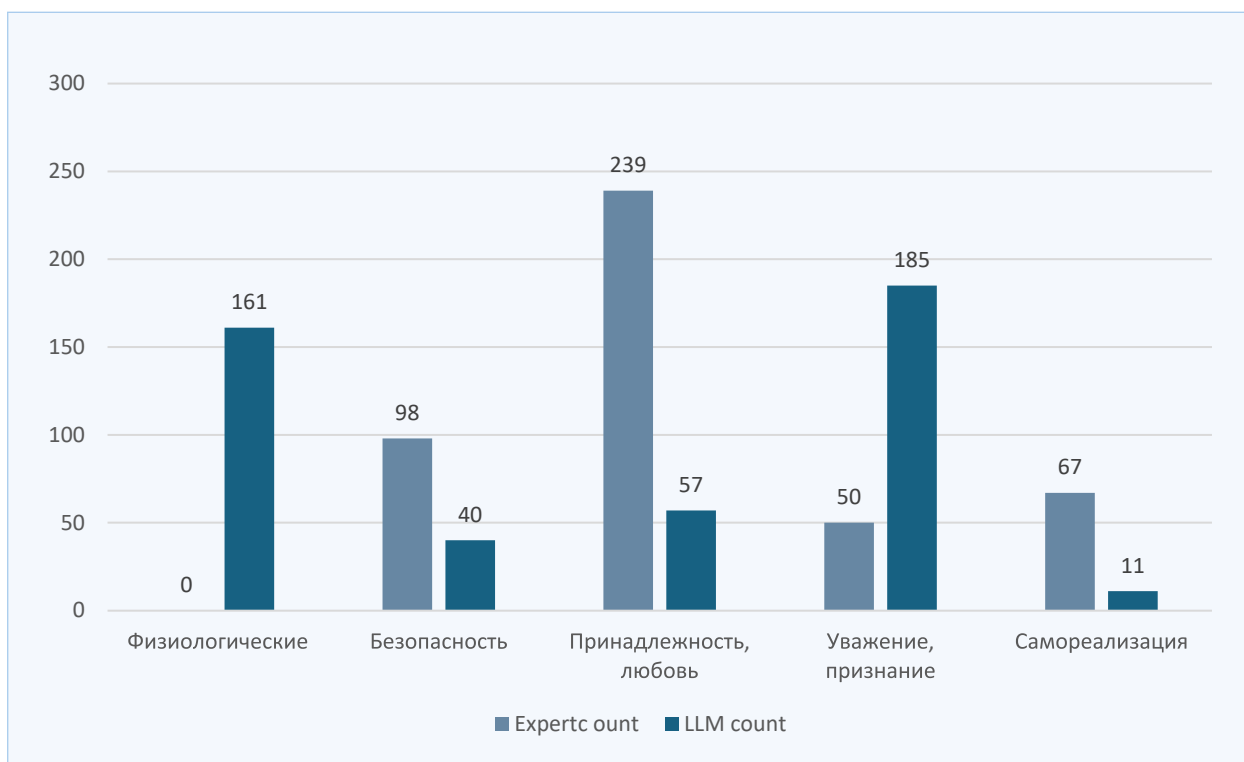
ке 360 сообщений (79%) получили метку *traditional*, 96 (21%) – *neutral*; по разметке LLM 369 сообщений (81%) были классифицированы как *traditional* и 87 (19%) как *neutral*. Матрица ошибок имеет вид: TP = 291, FN = 69, FP = 78, TN = 18.

Полученные метрики показывают, что при учёте дисбаланса классов согласованность LLM с экспертом по данной шкале минимальна. Номинальная точность составляет 67,8%, при этом чувствительность (recall) к традиционалистским высказываниям высока – 0,81, а точность (precision) – 0,79; однако специфичность крайне низка (0,19), что отражает выраженную тенденцию LLM «перепредсказывать» традиционализм. При выделении традиционалистского полюса модель не воспроизводит экспертную логику дифференциации между *traditional* и *neutral*, а достигаемая точность обусловлена главным образом общей доминантой традиционалистских высказываний в корпусе, а не качественным совпадением критериев.

В процессе разметки по параметру выживания и самовыражения согласно концептуальной модели Р. Инглхарта (*Inglehart\_survival\_vs\_selfexpr*) языковая модель также продемонстрировала

ряд методологически некорректных решений. Экспертная разметка показала, что по шкале *Inglehart\_survival\_vs\_selfexpr* корпус сообщений характеризуется преобладанием ценностей самовыражения: к категории *self-expression* отнесено 59,8% текстов, к *survival* – 25,7%, к нейтральным высказываниям – 14,5%. Разметка LLM дала значимо иное распределение: только 37,4% сообщений были отнесены к *self-expression* и 15,4% – к *survival*, при этом доля *neutral* была завышена до 47,3%. Общая точность составила 34,5%, коэффициент каппа Коэна – 0,02, что указывает на согласие лишь на уровне случайного совпадения. Результаты наглядно представлены на ил. 2.

По доминирующему уровню потребностей (модель А. Маслоу) итоговая точность сопоставления составила 16,3%, коэффициент к Коэна – 0,034, что лишь незначительно превышает случайное совпадение. Развёрнутый анализ показал, что LLM систематически искажает распределение уровней потребностей по сравнению с экспертной оценкой. Категория «Физиологические потребности» в экспертной разметке фактически не использовалась (что логично для текстов с аксиологиче-



Ил. 3. Доминирующий уровень потребностей в сообщениях корпуса (модель А. Маслоу)

ской нагрузкой), тогда как модель присвоила её 161 сообщению, что говорит о существенном смещении в сторону более «низких» уровней. Для уровней «Безопасность» и «Принадлежность/любовь» были получены умеренные значения точности при крайне низкой полноте (например, для «Безопасности» precision = 0,575 при recall = 0,235; для «Принадлежность/любовь» precision = 0,561 при recall = 0,134), что указывает на выборочную, фрагментарную чувствительность модели к релевантным сигналам. Для «Уважение/признание» и «Самореализации» полнота также остаётся низкой (0,240 и 0,104 соответственно) при точности 0,126 и 0,636, а значительная часть сообщений с экспертной меткой «Принадлежность/любовь» и «Самореализация» модель склонна относить к уровню «Уважение/признание». Результаты визуализированы на ил. 3.

Совпадение предсказаний модели по типологии патриотических высказываний (модель «Ушинский–Соловьёв») с экспертной разметкой по совокупности классов составило 90,6%. Основной вклад в данный

показатель вносит класс «конструктивный патриотизм»: при значительном преобладании этой категории в выборке модель демонстрирует высокие значения полноты и точности именно по этому классу, но практически не различает «агрессивный» патриотизм и отсутствие патриотической проблематики, систематически смещая их в сторону «конструктивного» патриотизма.

Аналогичное сравнение разметки доминирующего типа речевого акта по схеме Остина показало низкую степень совпадения между LLM и экспертом. Общая точность составила 29,9%, коэффициент к Коэна — около 0,04, что указывает на согласие. При этом модель систематически переиспользует класс бехабитивов, куда попадает значительная доля случаев, которые эксперт относит к экзерситивам и экспозитивам; реже распознаются экспозитивы (высокая точность при крайне низкой полноте) и практически не воспроизводится класс вердиктивов.

Сравнение разметки доминирующего типа речевого акта по схеме Остина про-

водилось между экспертной аннотацией и аннотацией LLM. В анализ были включены только те сообщения, для которых обе разметки использовали допустимые классы Austin: вердиктивы, экзерситивы, комиссивы, бехабитивы, экспозитивы (одна строка с некорректной меткой в экспертной разметке исключена из расчётов). На этой выборке из 455 сообщений была построена матрица несоответствий, рассчитаны точность по каждому классу (precision, recall, F1), общая доля совпадений и коэффициент согласия Коэна.

Сравнение метки бинарного индикатора «патриотизм» дало высокую степень совпадения (что ожидаемо в корпусе данных из ценностного кластера «Гордость и патриотизм»): при доминирующем удельном весе патриотической тематики в корпусе экспертная разметка отнесла к патриотическим 94,7% сообщений, тогда как LLM — 97,8%. Рассчитанные интегральные показатели подтверждают высокую согласованность: точность классификации составила 93,4%, полнота по положительному классу — 98,1%, точность (precision) — 95,1%, F1-мера — 96,6%. Коэффициент к Коэна равен 0,22, что отражает «умеренное» согласие крайне высокой априорной вероятности положительного класса и иллюстрирует эффект смещения распределения: почти полная заполненность корпуса патриотическими упоминаниями делает любую редкую дивергенцию статистически чувствительной.

В то же время экспертная проверка по бинарному индикатору гражданской идентичности показала существенное расхождение между разметкой LLM и экспертной разметкой. По итогам сопоставления были зафиксированы 35 совпадений в случаях присутствия гражданской идентичности (TP) и 239 совпадений в случаях её отсутствия (TN); при этом модель допустила 173 ложных срабатывания (FP), некорректно маркируя высказывания как отражающие гражданскую идентичность, и пропустила девять случаев, отмеченных экспертом (FN). Совокупно это дало точность классифика-

ции 0,60 при высокой полноте по позитивному классу (recall = 0,80) и крайне низкой точности предсказаний гражданской идентичности (precision = 0,17; F1 = 0,28). Полученные результаты позволяют заключить, что LLM демонстрирует выраженную тенденцию к гипердетекции гражданской идентичности, охотно маркируя широкий круг политико-патриотических высказываний, и потому может использоваться лишь на этапе предварительного отбора с обязательной последующей экспертной валидацией.

### Заключение

Сравнительный анализ разметки, осуществлённой в рамках различных теоретических моделей, выявляет статистически значимый разрыв между простыми и сложными типами кодировки.

На бинарном индикаторе патриотизма LLM показывает высокие показатели при умеренном, но систематическом завышении числа патриотических меток; модель надёжно улавливает эксплицитные лексические сигналы («Россия», «Родина», флаг, армия, Победа и т.п.). По мере усложнения задач, связанных с семантической дифференциацией, отмечается прогрессирующее снижение показателей согласованности между автоматизированными выводами и экспертными заключениями.

В итоге по результатам эксперимента можно заключить, что в рамках задач автоматизированного анализа ценностных и патриотических смыслов в русскоязычном пользовательском дискурсе без специальной настройки LLM воспроизводит главным образом общие, поверхностно маркированные семантические категории, но плохо различает более тонкие, теоретически детализированные ориентации и интенции высказываний. На данном этапе выводы модели на сложных шкалах не могут рассматриваться как эквивалент экспертной разметки, что свидетельствует о необходимости разработки и эмпирической проверки специализированных код-буков и процедур человеко-машинной валидации.

## Приложение «Промпт для LLM»

Твоя задача — выполнять стандартизированную категориальную разметку одного сообщения по фиксированным полям. Работай строго по тексту сообщения. Не используй внешние знания, политические предпочтения, сведения об авторе или площадке.

**Общие правила**

1. Для каждого поля выбери **ровно одно** значение из допустимого списка.
2. Если предусмотрено значение *neutral*, нет, ИСТИНА/ЛОЖЬ, используй только их, никаких других маркеров отсутствия/наличия.
3. Не изобретай новые категории, не меняй регистр и написание.
4. При неоднозначности выбирай наиболее обоснованную категорию; если ось допускает *neutral*/нет, используй её.
5. Сарказм и иронию учитывай только при явно выраженных маркерах.

**ОПРЕДЕЛЕНИЯ ПОЛЕЙ****1. Schwartz**

Основная ценность, выраженная в сообщении. Выбери одну:

- Власть — сила, влияние, доминирование, статус через контроль.
- Достижения — успех, результат, эффективность, признание заслуг.
- Гедонизм — удовольствие, комфорт, наслаждение.
- Стимуляция — новизна, риск, поиск впечатлений.
- Самонаправленность — автономия, свобода выбора и мнения.
- Универсализм — забота о человечестве, правах, мире, экологии, толерантности.
- Благожелательность — забота о близких и «своих», помощь окружению.
- Традиции — религия, обычаи, «предки», историческая преемственность.
- Конформность — послушание, соблюдение правил и норм.
- Безопасность — порядок, стабильность, защита от угроз.
- Нет — явный ценностный акцент отсутствует.

**2. Inglehart\_trad\_vs\_secular**

Ось «традиционные vs секулярно-рациональные». Выбери одну:

- *traditional* — подчеркнуты религия, традиционные ценности, авторитет государства/армии/семьи, иерархия, долг.
- *neutral* — нет выраженных традиционных индикаторов (и нет устойчивого светско-рационального противопоставления).

(Светско-рациональный полюс здесь не выделяется отдельно, всё нетрадиционное маркируется как *neutral*.)

**3. Inglehart\_survival\_vs\_selfexpr**

Ось «выживание vs самовыражение». Выбери одну:

- *survival* — акцент на выживании, базовой безопасности, страхе, бедности, жёстком порядке.
- *self-expression* — акцент на самореализации, правах, свободе, участии, протесте, креативности, экологии.
- *neutral* — нет достаточных признаков ни выживания, ни самовыражения.

**4. Maslow**

Доминирующий уровень потребностей. Выбери одну:

- Физиологические — еда, физическое выживание, здоровье как базовое условие.

- Безопасность — физическая, экономическая, социальная безопасность, стабильность, защита.
- Принадлежность/любовь — семья, друзья, община, «мы», сплочённость.
- Уважение/признание — статус, уважение, гордость, признание заслуг, в т.ч. гордость за страну.
- Самореализация — смысл, идеалы, долг, развитие, служение, высшие ценности.

Всегда выбирай один из пяти, исходя из доминирующего мотива текста.

### 5. Ushinsky\_Solovyov

Тип патриотизма по мотивам Ушинского и Соловьёва. Выбери одну:

- конструктивный — любовь к стране, гордость, ответственность, моральная аргументация, без вражды к другим народам.
- агрессивный — шовинизм, ненависть/унижение других, культ силы, противопоставление «мы против всех».
- нет — патриотическая риторика отсутствует или не выражена достаточно ясно.

### 6. Austin

Классификация доминирующего речевого акта по Остину. Выбери одну:

- *verdictive* — вынесение оценок и «судов»: осуждение, одобрение, признание, утверждения с характером решения/оценки.
- *exercitive* — высказывания, выражающие волеизъявление: приказы, призывы, рекомендации, требования, «надо/должны».
- *behabitve* — выражения отношения, установок и реакций: похвала, благодарность, осуждение, выражение уважения/неуважения, поздравления. (В нашей схеме сюда относятся эмоционально-оценочные реакции, связанные с нормами поведения и отношением к другим.)
- *expositve* — высказывания, которые поясняют, аргументируют, структурируют рассуждение, «объясняют позицию», описывают, как понимается ситуация или текст.

Всегда выбери один тип, который лучше всего характеризует форму высказывания. Если сообщение сложное, бери преобладающий.

### 7. Патриотизм

Индикатор наличия патриотической тематики.

- ИСТИНА — если сообщение явно связано с темой Родины, России, Отечества, национальных символов, армии, Победы и т.п.
- ЛОЖЬ — если такой тематики нет.

### 8. Civic\_identity

Индикатор гражданской идентичности.

- Истина — если автор говорит как гражданин: права/обязанности, участие в выборах, институтах, «мы, граждане», ответственность за страну/общество.
- ЛОЖЬ — если таких элементов нет.

(Используй именно ИСТИНА для столбца «Патриотизм» и Истина для civic\_identity, как различающиеся значения.)

Текст сообщения; []

## ЛИТЕРАТУРА

1. Schwartz S.H. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries // *Advances in Experimental Social Psychology*. 1992. Vol. 25. Pp. 1–65. [https://doi.org/10.1016/S0065-2601\(08\)60281-6](https://doi.org/10.1016/S0065-2601(08)60281-6).
2. Соловьёв В.С. Патриотизм // *Энциклопедический словарь Брокгауза и Ефрона*. Т. XXIII: Патенты на изобретения – Петропавловский. СПб.: Тип.-лит. И.А. Эфрона, 1898. С. 36–38.
3. Соловьёв В.С. О народности и народных делах в России (О соединении Церквей). 1884.
4. Соловьёв В.С. Национальный вопрос в России. Вып. 1. СПб., 1891.
5. Козлова Н.Н., Рассадин С.В. Сетевой дискурс патриотических онлайн-сообществ в современной России: проблемное поле и аксиологические модусы // *Вестник Российского университета дружбы народов. Серия: Политология (RUDN Journal of Political Science)*. 2025. Т. 27, № 3. С. 494–506. DOI: 10.22363/2313-1438-2025-27-3-494-506.
6. Анкудинов И.А. Патриотический дискурс в Рунете: до и после 24 февраля 2022 г. // *Мониторинг общественного мнения: экономические и социальные перемены*. 2024. № 2. С. 153–177. DOI: 10.14515/monitoring.2024.2.2515. URL: <https://doi.org/10.14515/monitoring.2024.2.2515>.
7. Petrovska I. Typology of civic identity // *Current Issues in Personality Psychology*. 2022. Vol. 11, № 2. Pp. 150–161. DOI: 10.5114/cipp.2022.116324.
8. World Values Survey Association. Inglehart–Welzel Cultural Map. 2023 version. <https://www.worldvaluessurvey.org/WVSNewsShow.jsp?ID=467>.
9. Inglehart R., Welzel C. *Modernization, Cultural Change, and Democracy*. Cambridge: Cambridge University Press, 2005. DOI: 10.1017/CBO9780511790881.
10. Austin J.L. *How to Do Things with Words*. Oxford: Clarendon Press, 1962. (2nd ed. / Ed. by J.O. Urmson, M. Sbisà. Oxford: Clarendon Press, 1975).
11. Searle J.R. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press, 1969. DOI: 10.1017/CBO9781139173438.
12. Milkova M., Rudnev M., Okolskaya L. Detecting value-expressive text posts in Russian social media: Препринт, arXiv:2312.08968. 2023. DOI: 10.48550/arXiv.2312.08968.
13. Plotnikov T. Analyzing GPT-4 Misinterpretations of Russian Grammatical Constructions // *Linguística. Revista de Estudos Linguísticos da Universidade do Porto*. 2024. Vol. 19. Pp. 157–182. DOI: 10.21747/16466195/ling19a7.
14. Zhang Y., Zou C., Lian Z., Tiwari P., Qin J. SarcasmBench: Towards Evaluating Large Language Models on Sarcasm Understanding: Препринт, arXiv:2408.11319, v1 от 21.08.2024. 2024. DOI: 10.48550/arXiv.2408.11319.
15. Zhou J. An Evaluation of State-of-the-Art Large Language Models for Sarcasm Detection: Препринт, arXiv:2312.03706, v1 от 07.10.2023. 2023. DOI: 10.48550/arXiv.2312.03706.
16. Bojić L., Zagovora O., Zelenkauskaite A. et al. Comparing large language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm // *Scientific Reports*. 2025. Vol. 15. Art. 11477. <https://doi.org/10.1038/s41598-025-96508-3>.
17. Gilardi F., Alizadeh M., Kubli M. ChatGPT outperforms crowd workers for text-annotation tasks // *Proceedings of the National Academy of Sciences*. 2023. Vol. 120, № 30. e2305016120. DOI: 10.1073/pnas.2305016120. URL: <https://www.pnas.org/doi/10.1073/pnas.2305016120>.

## Application of Large Language Models for the Analysis of Value-Patriotic Discourse of Russian-Speaking Users

**Yulia Vladimirovna Balakina** – Candidate of Sciences in Philology, Associate Professor, Professor of the Department of Fundamental and Applied Linguistics of the National Research University Higher School of Economics (HSE–Nizhny Novgorod).

E-mail: [julianaumova@gmail.com](mailto:julianaumova@gmail.com)

**Maria Vladimirovna Grigoryeva** – senior lecturer at the School of Communications at the Faculty of Creative Industries at the Higher School of Economics.

E-mail: [mariya.grigoreva@hse.ru](mailto:mariya.grigoreva@hse.ru)

**Ekaterina Nikitichna Sokolova** – Candidate of Sciences in Political Science and the head of the Political Communications Project and Training Laboratory at the School of Communications at the Faculty of Creative Industries at the Higher School of Economics.

E-mail: [e.sokolova@hse.ru](mailto:e.sokolova@hse.ru)

The article explores the potential of using large language models (LLMs) for the automated analysis of value-laden patriotic discourse among Russian-speaking social media users. Drawing on a corpus of messages from VK, Odnoklassniki, and Telegram (2023–2025), it investigates the degree of alignment between automated coding results and expert annotations based on a specially developed categorical scheme. The codebook includes eight dimensions: Sh. Schwartz's basic values; R. Inglehart's two axes (traditionalism/secularism and survival/self-expression); A. Maslow's hierarchy of needs; types of patriotism (constructive/aggressive), drawing on the concepts of K.D. Ushinsky and V.S. Solovyov; dominant speech act types per J. Austin; and binary indicators for explicit patriotism and civic identity. The experiment was conducted on the Pride and Patriotism message cluster (N = 456), where the density of value markers is highest; the comparison was implemented through error matrices, accuracy, macro/weighted F1, and Cohen's  $\kappa$  coefficient. It was shown that while the LLM reliably identifies explicit patriotic themes, its agreement with experts is significantly lower in multi-class and fine-grained value classification (Schwartz, Maslow, Inglehart scales, types of patriotism, Austin's speech acts). The model demonstrated systematic biases and a tendency to over-diagnose certain categories. It is concluded that LLMs in their current configuration can serve as auxiliary tools for preliminary markup and hypothesis generation but cannot function as an autonomous substitute for expert-led content analysis of value discourse.

**Keywords:** large language models, LLM, values; patriotism, civic identity, content analysis, Schwartz, Inglehart, Maslow, speech acts, social media, Russian-language discourse, automatic markup

## REFERENCES

1. Schwartz S.H. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries // *Advances in Experimental Social Psychology*. 1992. Vol. 25. Pp. 1–65. [https://doi.org/10.1016/S0065-2601\(08\)60281-6](https://doi.org/10.1016/S0065-2601(08)60281-6).
2. Solov'yov V.S. Patriotizm // *Entsiklopedicheskii slovar' Brokgauza i Efrona*. T. XXIII: Patenty na izobreteniya – Petropavlovskii. SPb.: Tip.-lit. I.A. Efrona, 1898. S. 36–38 (in Russian).
3. Solov'yov V.S. O narodnosti i narodnykh delakh v Rossii (O soedinenii Tserkvei). 1884 (in Russian).
4. Solov'yov V.S. Natsional'nyi vopros v Rossii. Vyp. 1. SPb., 1891 (in Russian).
5. Kozlova N.N., Rassadin S.V. Setevoi diskurs patrioticheskikh onlain-soobshchestv v sovremennoi Rossii: problemnoe pole i aksiologicheskie modusy // *Vestnik Rossiiskogo universiteta družby*

- narodov. Seriya: Politologiya (RUDN Journal of Political Science). 2025. T. 27, № 3. S. 494–506. DOI: 10.22363/2313-1438-2025-27-3-494-506 (in Russian).
6. Ankudinov I.A. Patrioticheskii diskurs v Runete: do i posle 24 fevralya 2022 g. // Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny. 2024. № 2. S. 153–177. DOI: 10.14515/monitoring.2024.2.2515. URL: <https://doi.org/10.14515/monitoring.2024.2.2515> (in Russian).
  7. Petrovska I. Typology of civic identity // Current Issues in Personality Psychology. 2022. Vol. 11. № 2. Pp. 150–161. DOI: 10.5114/cipp.2022.116324.
  8. World Values Survey Association. Inglehart–Welzel Cultural Map. 2023 version. <https://www.worldvaluessurvey.org/WVSNewsShow.jsp?ID=467>.
  9. Inglehart R., Welzel C. Modernization, Cultural Change, and Democracy. Cambridge: Cambridge University Press, 2005. DOI: 10.1017/CBO9780511790881.
  10. Austin J.L. How to Do Things with Words. Oxford: Clarendon Press, 1962. (2<sup>nd</sup> ed. / Ed. by J.O. Urmson, M. Sbisà. Oxford: Clarendon Press, 1975).
  11. Searle J.R. Speech Acts: An Essay in the Philosophy of Language. Cambridge: Cambridge University Press, 1969. DOI: 10.1017/CBO9781139173438.
  12. Milkova M., Rudnev M., Okolskaya L. Detecting value-expressive text posts in Russian social media: Preprint, arXiv:2312.08968. 2023. DOI: 10.48550/arXiv.2312.08968.
  13. Plotnikov T. Analyzing GPT-4 Misinterpretations of Russian Grammatical Constructions // Linguística. Revista de Estudos Linguísticos da Universidade do Porto. 2024. Vol. 19. Pp. 157–182. DOI: 10.21747/16466195/ling19a7.
  14. Zhang Y., Zou C., Lian Z., Tiwari P., Qin J. SarcasmBench: Towards Evaluating Large Language Models on Sarcasm Understanding: Preprint, arXiv:2408.11319, v1 ot 21.08.2024. 2024. DOI: 10.48550/arXiv.2408.11319.
  15. Zhou J. An Evaluation of State-of-the-Art Large Language Models for Sarcasm Detection: Preprint, arXiv:2312.03706, v1 ot 07.10.2023. 2023. DOI: 10.48550/arXiv.2312.03706.
  16. Bojić L., Zagovora O., Zelenkauskaitė A. et al. Comparing large language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm // Scientific Reports. 2025. Vol. 15. Art. 11477. <https://doi.org/10.1038/s41598-025-96508-3>.
  17. Gilardi F., Alizadeh M., Kubli M. ChatGPT outperforms crowd workers for text-annotation tasks // Proceedings of the National Academy of Sciences. 2023. Vol. 120. No. 30. e2305016120. DOI: 10.1073/pnas.2305016120. URL: <https://www.pnas.org/doi/10.1073/pnas.2305016120>.