

ФОРМИРОВАНИЕ И РАЗМЕТКА КОРПУСА РУССКОЯЗЫЧНЫХ НОВОСТНЫХ ТЕКСТОВ ДЛЯ АВТОМАТИЗИРОВАННОГО ВЫЯВЛЕНИЯ ПОЛИТИЧЕСКИХ МАНИПУЛЯЦИЙ

Н. Л. Кулюлина^[0009-0006-1715-1114]

Московский физико-технический институт, г. Долгопрудный, Россия

kuliulina.nl@phystech.edu

Аннотация

Исследована проблема создания специализированных корпусных ресурсов для задач автоматизированного анализа политических манипуляций в русскоязычных текстах. Несмотря на активное развитие методов семантического и вычислительного анализа текстов, существующие корпусные ресурсы и схемы разметки в основном ориентированы на англоязычные данные и плохо учитывают языковую и контекстуальную специфику русскоязычных новостных средств массовой информации (СМИ).

Целями исследования были создание специализированного корпуса русскоязычных новостных текстов и разработка схемы разметки, ориентированной на автоматизированный анализ политических манипуляций с учетом особенностей русскоязычного медиапространства.

В рамках проведенного исследования сформирован корпус фраз, извлеченных из русскоязычных новостных текстов и опубликованных в период 2010–2019 гг., и разработана схема разметки манипулятивных техник. В основе разметки лежит адаптация международных классификаций манипулятивных стратегий, сведенных к ограниченному числу интерпретируемых техник, релевантных для анализа русскоязычных новостных текстов. Предлагаемая схема охватывает эмоциональные, аргументативные и контекстуальные формы манипулятивного воздействия.

Полученные корпус и схема разметки могут использоваться в качестве эмпирической основы для разработки и тестирования методов автоматизированного анализа политических манипуляций в русскоязычных новостных СМИ, а также дальнейших исследований политических и медиа-текстов.

Ключевые слова: медиа-манипуляции, русскоязычные СМИ, корпус текстов, разметка данных, манипулятивные техники, политическая коммуникация, семантический анализ, вычислительный дискурс-анализ.

ВВЕДЕНИЕ

В современных новостных медиа политические манипуляции нередко реализуются не через прямое искажение фактов, а через контекстуальные механизмы отбора тем, расстановки акцентов и подбора терминологии, формирующие интерпретацию политических событий при сохранении видимости нейтрального информирования [1, 2]. Такие механизмы проявляются в устойчивых способах текстового и контекстуального оформления сообщений и не сводятся к отдельным лексическим маркерам, что делает их анализ центральной задачей исследований политической коммуникации [3, 4].

Эмпирическое изучение политических манипуляций требует сопоставления теоретических подходов с наблюдаемыми единицами текста, которые могут быть систематически выделены и описаны. Традиционные качественные методы контент- и дискурс-анализа обеспечивают такую связь теории и эмпирики, однако их высокая трудоемкость и ограниченная масштабируемость существенно затрудняют анализ больших массивов новостных данных. При этом в условиях цифровизации медиaprостранства при отсутствии специализированных корпусов новостных текстов с единой схемой разметки манипулятивных приемов применение автоматизированных методов анализа политических манипуляций также оказывается затруднительным [5].

Несмотря на развитие корпусных ресурсов и вычислительных методов анализа, направленных на выявление манипулятивных стратегий в новостных текстах, большинство существующих решений ориентировано на англоязычные данные и в ограниченной степени применимо к русскоязычным новостным текстам [6, 7]. Недостаток специализированных корпусов и схем разметки, учитывающих языковую и контекстуальную специфику русскоязычных СМИ, существенно ограничивает возможности автоматизированного анализа политических манипуляций [8, 9].

В связи с этим целью настоящей статьи является представление корпуса

фраз, извлеченных из русскоязычных новостных текстов, и схемы разметки манипулятивных техник, разработанной с учетом особенностей русскоязычных новостных текстов. Корпус и схема разметки рассматриваются как эмпирическая основа для решения задачи автоматизированного выявления манипулятивных техник в русскоязычных новостных текстах на уровне предложения. Задача сформулирована как двухэтапная: на первом этапе осуществляется бинарная классификация предложений на манипулятивные и неманипулятивные, а на втором этапе — классификация манипулятивных предложений по типам используемых техник.

ОБЗОР БЛИЗКИХ ПО ТЕМАТИКЕ ИССЛЕДОВАНИЙ

Одним из наиболее известных корпусных ресурсов в рамках задачи выявления политических манипуляций является англоязычный Media Frames Corpus, предложенный Д. Кардом и соавторами [6], в котором новостные тексты размечены по тематическим категориям, описывающим типичные способы представления и объяснения одних и тех же событий в новостных медиа. Этот корпус заложил основу для последующих исследований, демонстрирующих, что наличие качественной разметки позволяет выявлять манипулятивные стратегии в больших массивах новостных данных и отслеживать их динамику во времени и по источникам. В дальнейшем подход был расширен за счет интеграции метаданных документов, что повысило точность и интерпретируемость автоматизированного анализа [10].

Отдельное направление исследований связано с выявлением манипулятивных техник на уровне фрагментов текста. В работах Дж. Да Сан Мартино и соавторов была предложена типология из 18 манипулятивных техник и создан корпус новостных текстов с разметкой на уровне фраз [7, 11]. Эти исследования показали, что переход от документного уровня анализа к более мелким текстовым единицам позволяет точнее зафиксировать конкретные приемы манипулятивного воздействия, в том числе с использованием автоматизированных методов анализа. При этом сами авторы подчеркивают, что устойчивость моделей в значительной степени определяется качеством и репрезентативностью разметки.

Применение аналогичных методологий к русскоязычным новостным тек-

стам остается ограниченным. Существующие исследования показали, что модели и схемы разметки, разработанные для английского языка, плохо переносятся на русскоязычные тексты из-за лингвистических различий (например, в русском языке активно используются эвфемизмы, конструкторы двойного значения, грамматические конструкции с оценочной окраской) [9], а также из-за особенностей медиасреды (различные медиа используют принципиально разные коммуникативные стратегии) [8]. Отсутствие открытых и специализированных размеченных корпусов для анализа политических манипуляций в русскоязычных новостных СМИ существенно сдерживает развитие воспроизводимых и масштабируемых исследований в данной области.

МЕТОДЫ

В рамках проведенного исследования был сформирован корпус фраз на материале русскоязычных новостных публикаций, находящихся в открытом доступе в Интернете. В корпус включены тексты, опубликованные в период с 2010 по 2019 г., что позволяет снизить влияние краткосрочных информационных кампаний и отдельных политических событий на структуру данных и соответствует практике корпусных исследований медиа [6, 8]. Источниками данных выступили русскоязычные новостные СМИ, различающиеся по редакционным стратегиям и позиционированию: *Газета.Ру*, *Лента.ру*, *RT (Россия сегодня)* и др., что позволяет учитывать вариативность новостных текстов и потенциально способствует устойчивости моделей к источниковому сдвигу.

Масштаб корпуса делает невозможной полную ручную разметку текстов, особенно при переходе к фразовому уровню анализа, где количество наблюдений возрастает кратно. В связи с этим в работе использована стратегия ограниченной экспертной разметки репрезентативной выборки единиц с возможностью последующего расширения данных с помощью генерации синтетических данных.

В качестве единицы анализа в настоящем исследовании использован фразовый фрагмент (предложение), извлеченный из новостного текста. Выбор фразового уровня анализа обусловлен спецификой манипулятивных техник: большинство из них реализуется на уровне отдельных высказываний через подбор

слов и структуру формирования конкретных предложений, которые задают аргументативные конструкции и контекстуальные акценты [7]. Анализ на уровне целых документов, напротив, нередко затрудняет выявление конкретных приемов воздействия и приводит к смешению различных стратегий в рамках одного наблюдения.

Сегментация текстов на предложения осуществлялась автоматически с применением стандартных правил пунктуационной сегментации. Дополнительная фрагментация предложений не проводилась, поскольку большинство манипулятивных техник реализуется на уровне завершеного высказывания, а сохранение целостности предложения повышает интерпретируемость разметки. Для каждой фразовой единицы в корпусе сохраняется информация об источнике, дате публикации, заголовке статьи и полной ссылке на исходный материал, что позволяет при необходимости восстанавливать расширенный контекст и анализировать пограничные случаи.

Схема разметки корпуса основана на определении понятия политической манипуляции как *формы скрытого информационного воздействия, реализуемого через конкретные языковые и контекстуальные техники при сохранении видимости нейтрального информирования* [14].

В качестве базовой типологии манипулятивных техник использована международная классификация SemEval [11], разработанная для задач автоматизированного выявления манипуляций в новостных текстах. При адаптации этой типологии к русскоязычным текстам отбор техник осуществлялся на основе трех критериев: (1) регулярность и воспроизводимость реализации техники на фразовом уровне, (2) интерпретируемость разметки без привлечения широкого контекстуального знания и (3) применимость техники к разнородным новостным тематикам и источникам. В результате исходный набор из 18 техник был редуцирован до 6 основных техник, релевантных для последующего корпусного и автоматизированного анализа. Выделенные техники соотносятся с тремя укрупненными типами манипулятивного воздействия: эмоциональным, аргументативным и контекстуальным. Итоговая схема включает следующие манипулятивные техники:

- **эмоционально заряженный язык** (Loaded language) – использование лексики с выраженной эмоциональной коннотацией;

- **навешивание ярлыков** (Name calling / labeling) – использование оценочных «ярлыков» по отношению к актерам или группам;
- **запугивание** (Appeal to fear / prejudice) – апелляция к страхам, угрозам или социальным предубеждениям;
- **апелляция к авторитету** (Appeal to authority) – ссылка на авторитет в качестве аргумента без критической проверки;
- **ложная причинность** (Causal oversimplification) – чрезмерное упрощение причинно-следственных связей;
- **ложная дихотомия** (Black-and-white fallacy) – представление ситуации в бинарной оппозиции.

Для формирования выборки предложений, подлежащих ручной разметке, применялся автоматический предварительный отбор кандидатов на основе эвристического анализа текста предложения без использования внешнего контекста. В качестве признаков потенциальной манипулятивности учитывались лексико-дискурсивные маркеры (апелляции к источникам и авторитетам, оценочная и эмоционально окрашенная лексика, обобщающие формулы), аргументативные конструкции (причинно-следственные связки, бинарные противопоставления), а также апелляции к угрозам и негативным последствиям. По совокупности этих признаков предложения ранжировались, после этого итоговая выборка для ручной разметки формировалась как комбинация предложений с высокой оценкой по итогам ранжирования и случайно отобранных нейтральных фрагментов.

Табл. 1. Распределение классов в размеченной выборке.

Класс	Число предложений	Доля
Бинарная классификация		
Манипулятивные	358	44.7%
Нейтральные	442	55.3%
Классификация по типам манипулятивных техник		
Эмоционально заряженный язык	94	26.3%
Апелляция к авторитету	70	19.6%
Навешивание ярлыков	69	19.3%

Ложная причинность	45	12.5%
Запугивание	43	12.0%
Ложная дихотомия	37	10.3%

Ручная разметка предложений осуществлялась экспертно с использованием разработанной схемы манипулятивных техник. В размеченной выборке наблюдался умеренный дисбаланс между манипулятивными и неманипулятивными предложениями (см. табл. 1); данное распределение не требует балансировки на уровне бинарной классификации. В то же время анализ распределения манипулятивных техник в размеченной выборке показал выраженный дисбаланс между классами.

Исходя из полученного распределения, можно утверждать, что балансировка данных на уровне бинарной классификации не является критически необходимой, тогда как для задачи классификации по типам манипулятивных техник требуется дополнительная работа с редкими классами. Реализация этого этапа планируется посредством дополнительного целевого отбора кандидатов из неразмеченного корпуса с последующей ручной валидацией, только после этого возможно использование данной разметки в качестве основы для генерации синтетических данных; качество и корректность синтетически сгенерированных данных предлагается оценивать посредством выборочной экспертной валидации, включающей проверку соответствия разметки исходной схеме манипулятивных техник, а также анализ типичных ошибок и пограничных случаев.

Кроме того, для осуществления предварительных экспериментов в рамках бинарной классификации был реализован этап генерации синтетических данных. Синтетические примеры были сформированы на основе исходных вручную размеченных предложений с использованием слабых преобразований текста, сохраняющих метку. В итоговом наборе данных (см. табл. 2) оригинальные и синтетические примеры хранились совместно и сопровождались явной маркировкой, указывающей на их происхождение.

Табл. 2. Распределение классов в размеченной выборке
(с дополнением синтетическими данными).

Класс\тип данных	Оригинальные	Синтетические	Итого
Манипулятивные	358	558	916 (44.2%)
Нейтральные	442	713	1155 (55.8%)
Итого	800	1271	2071

ПРИМЕРЫ И ЭКСПЕРИМЕНТЫ

Для демонстрации логики предложенной схемы разметки приведем иллюстративные примеры фраз из корпуса с указанием выявленных манипулятивных техник. Примеры отобраны из русскоязычных новостных публикаций Газета.Ru за 2010 год и отражают различные типы манипулятивного воздействия на фразовом уровне.

Пример 1 (Ложная дихотомия): *«Проблемы порнобизнеса ничем не отличаются от проблем любого другого бизнеса»*. В данном фрагменте сложное и социально нагруженное явление представлено в виде упрощенной бинарной аналогии. Формулировка исключает альтернативные интерпретации и нивелирует специфические этические и правовые аспекты рассматриваемого феномена, тем самым формируя однозначную интерпретационную рамку.

Пример 2 (Апелляция к авторитету): *«Так, координатор государственно-патриотического движения заявил, что происходящее является прямым следствием внешнего давления»*. В данном случае манипулятивное воздействие реализовано через апелляцию к неконкретизированному институциональному авторитету. Ссылка на позицию представителя организации использована как аргумент, не сопровождаемый независимым обоснованием или альтернативными точками зрения.

Пример 3 (Запугивание): *«Эксперты предупреждают, что дальнейшее развитие ситуации может привести к резкому росту преступности и дестабилизации обстановки в стране»*. В данном фрагменте использована апелляция к неопределенной, но социально значимой в российском контексте угрозе (ассоциация с кризисом 90-х годов XX в.). Формулировка апеллирует к страху перед негативными последствиями, не предоставляя конкретных механизмов или

проверяемых оснований для подобного прогноза, что способствует формированию тревожной интерпретационной рамки у читателя.

Пример 4 (неманипулятивный фрагмент): «*Будущие молодожены прибыли к собору по отдельности: 32-летний Вестлинг приехал вместе с братом Викторией, а саму принцессу на старинном автомобиле привез ее отец, король Карл XVI Густав*». Данный фрагмент содержит информационно-статистические данные о ходе мероприятия без использования методов воздействия на искажение восприятия читателя.

Под пограничными случаями в рамках данного корпуса понимаются предложения, которые содержат отдельные лексико-дискурсивные маркеры, ассоциируемые с манипулятивными техниками, однако в конкретном контексте они не выполняют функции скрытого воздействия и потому были размечены как неманипулятивные.

Пример 5 (неманипулятивный фрагмент): «*Как пишет Bild, в результате аварии тяжелые травмы получили несколько человек*». Фраза содержит маркер источника, который формально может напоминать технику Appeal to Authority, а также устойчивое выражение «тяжелые травмы», по форме соответствующее эмоционально окрашенному эпитету (Loaded language). Вместе с тем данный фрагмент представляет собой нейтральное информационное сообщение с указанием источника и без выраженной оценочной нагрузки.

Пример 6 (неманипулятивный фрагмент): «*Лидер оппозиционной Либерально-демократической партии заявил, что не будет участвовать в голосовании*». Фраза содержит ссылку на политического актора и его заявление, а также предоставляет потенциально конфликтный контекст. В то же время само по себе высказывание не содержит интерпретации либо оценки действий политика, а также не является аргументом для поддержания авторской точки зрения.

Анализ пограничных случаев показывает, что наличие формальных лексических или синтаксических маркеров само по себе не является достаточным основанием для отнесения предложения к манипулятивным. Он также демонстрирует ключевой критерий экспертной разметки — функцию высказывания в контексте (используется ли соответствующая конструкция для изменения интерпретации или давления на мнение читателя). Однако, отдельные пограничные слу-

чаи могут допускать альтернативные трактовки, что в дальнейшем предполагается корректировать с помощью привлечения дополнительных валидирующих разметку экспертов.

Для оценки влияния синтетических данных на задачу бинарной классификации манипулятивных высказываний были проведены предварительные эксперименты с использованием линейных моделей. В качестве базового подхода использовалась линейная модель логистической регрессии с TF-IDF-представлением текста. Сравнивались два варианта обучения модели: обучение только на данных, размеченных вручную, и обучение на тех же данных с добавлением синтетических примеров, сгенерированных для бинарной классификации. Во всех экспериментах разбиение на обучающую и тестовую выборки выполнялось исключительно по оригинальным данным, синтетические примеры не включались в тестовую выборку, что исключало утечки информации. Качество классификации оценивалось с использованием метрики macro-F1 .

Первоначальные эксперименты с одиночным разбиением на обучающую и тестовую выборки показали лишь незначительные различия между базовой моделью и моделью, обученной с использованием синтетических данных. В частности, значение macro-F1 для базовой модели составило 0.85, тогда как при добавлении синтетических примеров прирост оказался минимальным ($\Delta \approx +0.001$), что укладывается в пределы статистического шума. Однако применение стратифицированной кросс-валидации с пятью фолдами позволило выявить устойчивый положительный эффект от использования синтетических данных. Среднее значение macro-F1 для базовой модели составило 0.85 ± 0.04 , тогда как модель, обученная на расширенной обучающей выборке, продемонстрировала значение 0.90 ± 0.04 . Улучшение наблюдалось во всех фолдах кросс-валидации, что указывает на систематический характер эффекта. Полученные результаты свидетельствуют о том, что использование синтетических данных может способствовать повышению обобщающей способности моделей, по крайней мере бинарной классификации в условиях ограниченного объема данных. Отметим, что данные эксперименты носят предварительный характер и направлены прежде всего на оценку целесообразности применения синтетических данных в рамках дальнейших этапов исследования.

ЗАКЛЮЧЕНИЕ

Рассмотрены формирование специализированного корпуса предложений русскоязычных новостных текстов и разработка схемы разметки манипулятивных техник, ориентированная на задачу классификации по типам политических манипуляций на уровне предложений в русскоязычных новостных СМИ. В отличие от существующих корпусных ресурсов, разработанных преимущественно для англоязычных данных, предложенный подход учитывает языковую и контекстуальную специфику русскоязычных новостных медиа и адаптирован к особенностям реализации манипулятивных стратегий в данном контексте.

Предложенный корпус и схема разметки могут рассматриваться как эмпирическая основа для последующей разработки и тестирования методов автоматизированного выявления политических манипуляций в русскоязычных новостных СМИ. В дальнейшем развитие работы может быть связано с расширением корпуса, валидацией схемы разметки с участием нескольких экспертов, а также с применением предложенного ресурса в задачах автоматизированного анализа новостных текстов. Полученные результаты и разработанные инструменты могут быть использованы как в академических исследованиях, так и в прикладных задачах анализа медиаконтента.

СПИСОК ЛИТЕРАТУРЫ

1. *Entman R.M.* Framing: Toward clarification of a fractured paradigm // *Journal of Communication*. 1993. Vol. 43, No. 4, P. 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
2. *Chong D., Druckman J.N.* Framing theory // *Annual Review of Political Science*. 2007. Vol. 10. P. 103–126. <https://doi.org/10.1146/annurev.polisci.10.072805.103054>
3. *Mejias U.A., Vokuev N.E.* Disinformation and the media: The case of Russia and Ukraine // *Media, Culture & Society*. 2017. Vol. 39, No. 7. P. 1027–1042. <https://doi.org/10.1177/0163443716686672>
4. *Rozenas A., Stukal D.* How autocrats manipulate economic news: Evidence from Russia's state-controlled television // *The Journal of Politics*. 2019. Vol. 81, No. 3. P. 982–996. <https://dx.doi.org/10.2139/ssrn.3023254>
5. *Lazer D.M.J., Pentland A., Adamic L. et al.* Computational social science:

Obstacles and opportunities // *Science*. 2020. Vol. 369, No. 6507. P. 1060–1062.

<https://doi.org/10.1126/science.aaz8170>

6. *Card D., Boydstun A.E., Gross J.H., Resnik P., Smith N.A.* The media frames corpus: Annotations of frames across issues // *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*. 2015. P. 438–444. <https://doi.org/10.3115/v1/P15-2072>

7. *Da San Martino G., Barrón-Cedeño A., Wachsmuth H., Nakov P.* Fine-grained analysis of propaganda in news articles // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019. P. 5636–5646. <https://doi.org/10.18653/v1/D19-1565>

8. *Field A., Atanasov P., Stukal D., Tucker J.A., Guess A.* Framing and agenda-setting in Russian news: A computational analysis of intricate political strategies // *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2018. P. 3570–3580. <https://doi.org/10.18653/v1/D18-1393>

9. *Bhatia V., Chhaya N., Pala K., Bhargava P.* OpenFraming: Open-sourced tool for computational framing analysis of multilingual data // *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021. P. 242–250. <https://doi.org/10.18653/v1/2021.emnlp-demo.28>

10. *Card D., Paul M.J., Smith N.A.* Neural models for documents with metadata // *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2018. P. 2031–2040. <https://doi.org/10.18653/v1/P18-1189>

11. *Da San Martino G., Yu S., Barrón-Cedeño A., et. al.* SemEval-2020 Task 11: Detection of propaganda techniques in news articles // *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020)*. 2020. P. 1377–1414. <https://doi.org/10.18653/v1/2020.semeval-1.186>

12. *Kwak H., An J., Jing E.M., Ahn Y.* A systematic media frame analysis of 1.5 million New York Times articles from 2000 to 2017 // *Proceedings of the 12th ACM Conference on Web Science*. 2020. P. 305–314. <https://doi.org/10.1145/3394231.3397921>

13. *Kwak H., An J., Jing E.M., Ahn Y.* FrameAxis: Characterizing microframe bias and intensity with word embedding // *PeerJ Computer Science*. 2021. Vol. 7, Article e644. <https://doi.org/10.7717/peerj-cs.644>

14. Entman R.M. Framing bias: Media in the distribution of power // Journal of Communication. 2007. Vol. 57, No. 1. P. 163–173.
<https://doi.org/10.1111/j.1460-2466.2006.00336.x>

CONSTRUCTION AND ANNOTATION OF A RUSSIAN-LANGUAGE NEWS CORPUS FOR AUTOMATED DETECTION OF POLITICAL MANIPULATION

N. L. Kulyulina^[0009-0006-1715-1114]

Moscow Institute of Physics and Technology, Dolgoprudny, Russia

kuliulina.nl@phystech.edu

Abstract

This paper addresses the challenge of developing specialized corpus resources for the automated analysis of political manipulation in Russian-language media discourse. Although semantic text analysis and computational discourse analysis have advanced substantially in recent years, most existing corpora and annotation schemes are designed for English-language data and do not adequately capture the linguistic and discursive characteristics of Russian-language news media. The objective of this study is to construct a specialized corpus of Russian-language news texts and to develop an annotation scheme tailored to the automated analysis of political manipulation, with explicit consideration of the linguistic and discursive features of the Russian-language media environment. The study introduces a corpus of sentence-level fragments extracted from Russian-language news texts published between 2010 and 2019, together with an annotation scheme for manipulative techniques. The scheme is based on an adaptation of established international classifications of manipulative strategies and is reduced to a limited set of interpretable techniques relevant to Russian-language news discourse. The proposed framework covers emotional, argumentative, and contextual forms of manipulative influence. The resulting corpus and annotation scheme provide an empirical foundation for the development and evaluation of automated methods for analyzing political manipulation in Russian-language news media and may also support further research in media and political discourse.

Keywords: *media manipulation, Russian media, text corpus, data annotation, manipulative techniques, political communication, semantic analysis, computational discourse analysis.*

REFERENCES

1. *Entman R.M.* Framing: Toward clarification of a fractured paradigm // *Journal of Communication*. 1993. Vol. 43, No. 4. P. 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
2. *Chong D., Druckman J.N.* Framing theory // *Annual Review of Political Science*. 2007. Vol. 10. P. 103–126. <https://doi.org/10.1146/annurev.polisci.10.072805.103054>
3. *Mejias U.A., Vokuev N.E.* Disinformation and the media: The case of Russia and Ukraine // *Media, Culture & Society*. 2017. Vol. 39, No. 7. P. 1027–1042. <https://doi.org/10.1177/0163443716686672>
4. *Rozenas A., Stukal D.* How autocrats manipulate economic news: Evidence from Russia's state-controlled television // *The Journal of Politics*. 2019. Vol. 81, No. 3. P. 982–996. <https://dx.doi.org/10.2139/ssrn.3023254>
5. *Lazer D.M.J., Pentland A., Adamic L. et al.* Computational social science: Obstacles and opportunities // *Science*. 2020. Vol. 369, No. 6507. P. 1060–1062. <https://doi.org/10.1126/science.aaz8170>
6. *Card D., Boydston A.E., Gross J.H., Resnik P., Smith N.A.* The media frames corpus: Annotations of frames across issues // *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*. 2015. P. 438–444. <https://doi.org/10.3115/v1/P15-2072>
7. *Da San Martino G., Barrón-Cedeño A., Wachsmuth H., Nakov P.* Fine-grained analysis of propaganda in news articles // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019. P. 5636–5646. <https://doi.org/10.18653/v1/D19-1565>
8. *Field A., Atanasov P., Stukal D., Tucker J.A., Guess A.* Framing and agenda-setting in Russian news: A computational analysis of intricate political strategies // *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2018. P. 3570–3580. <https://doi.org/10.18653/v1/D18-1393>

9. *Bhatia V., Chhaya N., Pala K., Bhargava P.* OpenFraming: Open-sourced tool for computational framing analysis of multilingual data // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2021. P. 242–250. <https://doi.org/10.18653/v1/2021.emnlp-demo.28>
10. *Card D., Paul M.J., Smith N.A.* Neural models for documents with metadata // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL). 2018. P. 2031–2040. <https://doi.org/10.18653/v1/P18-1189>
11. *Da San Martino G., Yu S., Barrón-Cedeño A. et al.* SemEval-2020 Task 11: Detection of propaganda techniques in news articles // Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020). 2020. P. 1377–1414. <https://doi.org/10.18653/v1/2020.semeval-1.186>
12. *Kwak H., An J., Jing E.M., Ahn Y.* A systematic media frame analysis of 1.5 million New York Times articles from 2000 to 2017 // Proceedings of the 12th ACM Conference on Web Science. 2020. P. 305–314. <https://doi.org/10.1145/3394231.3397921>
13. *Kwak H., An J., Jing E.M., Ahn Y.* FrameAxis: Characterizing microframe bias and intensity with word embedding // PeerJ Computer Science. 2021. Vol. 7, Article e644. <https://doi.org/10.7717/peerj-cs.644>
14. *Entman R.M.* Framing bias: Media in the distribution of power // Journal of Communication. 2007. Vol. 57, No. 1. P. 163–173. <https://doi.org/10.1111/j.1460-2466.2006.00336.x>

СВЕДЕНИЯ ОБ АВТОРЕ



КУЛЮЛИНА Нина Леонидовна – студентка 2 курса магистерской программы Московского физико-технического института (национальный исследовательский университет) по направлению «Науки о данных»; научный сотрудник Центра изучения стабильности и рисков Национального исследовательского университета «Высшая школа экономики». Научные интересы включают применение передовых вычислительных методов анализа данных в сфере социальных наук, включая методы машинного обучения и большие языковые модели.

Nina Leonidovna KULYULINA – second-year student in the Master's program in Data Science at the Moscow Institute of Physics and Technology and a research fellow at the Center for Stability and Risk Analysis at the HSE University. Her research interests include the application of advanced computational methods of data analysis in social sciences, including machine learning methods and large-scale language models.

email: kuliulina.nl@phystech.edu

ORCID: 0009-0006-1715-1114

Материал поступил в редакцию 15 апреля 2026 года