

Научная статья
УДК 004.853
doi:10.37614/2949-1215.2025.16.3.005

ПРЕДИКТИВНОЕ МОДЕЛИРОВАНИЕ СОЦИАЛЬНЫХ РЕАКЦИЙ ДЛЯ РЕГИОНАЛЬНОГО УПРАВЛЕНИЯ НА БАЗЕ МЕТОДОВ ОБЪЯСНИМОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Владимир Витальевич Диковицкий[✉]

*Институт информатики и математического моделирования имени В. А. Путилова
Кольского научного центра Российской академии наук, Апатиты, Россия,
dikovitsky@gmail.com[✉], <https://orcid.org/0000-0003-0329-9979>*

Аннотация

Данная работа представляет модульную архитектуру, использующую современные методы искусственного интеллекта. Подход базируется на четырех компонентах: каузальных агентных моделях для имитации динамики общественных реакций и понимания механизмов формирования мнений; RAG-системах, привязывающих рассуждения к историческим фактам; динамическом графе знаний, связывающем социальные сигналы и события и деонтической логике для обеспечения объяснения решений. Представленная архитектура преодолевает ограничения ретроспективных методов, предоставляя инструмент для понимания причин возникновения трендов, учета нормативной среды и генерации конкретных сценариев действий в реальном времени, что способствует прозрачному использованию методов объяснимого искусственного интеллекта в управлении.

Ключевые слова:

объяснимый искусственный интеллект, социальная сеть, граф знаний

Благодарности:

исследование выполнено в рамках государственного задания ИИММ КНЦ РАН Министерства науки и высшего образования РФ, темы научно-исследовательской работы «Методы и технологии создания интеллектуальных информационных систем для поддержки развития сложных динамических систем с региональной спецификой в условиях неопределённости и риска» (шифр темы FMEZ-2025-0053).

Для цитирования:

Диковицкий В. В. Предиктивное моделирование социальных реакций для регионального управления на базе методов объяснимого искусственного интеллекта // Труды Кольского научного центра РАН. Серия: Технические науки. 2025. Т. 16, № 3. С. 71–79. doi:10.37614/2949-1215.2025.16.3.005.

Original article

PREDICTIVE MODELING OF SOCIAL REACTIONS FOR REGIONAL MANAGEMENT BASED ON EXPLAINABLE ARTIFICIAL INTELLIGENCE METHODS

Vladimir V. Dikovitsky^{1✉}

*Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre
of the Russian Academy of Sciences, Apatity, Russia, v.dikovitsky@ksc.ru[✉],
<https://orcid.org/0000-0003-0329-9979>*

Abstract

This paper presents a modular architecture utilizing modern artificial intelligence methods. The approach is based on four components: causal agent-based models for simulating the dynamics of public reactions and understanding the mechanisms of opinion formation; RAG systems that anchor reasoning in historical facts; a dynamic knowledge graph linking social signals and events; and deontic logic to provide rationale for decisions. The presented architecture overcomes the limitations of retrospective methods by providing a tool for understanding the causes of trends, taking into account the regulatory environment, and generating specific action scenarios in real time, thereby facilitating the transparent use of explainable artificial intelligence methods in governance.

Keywords:

explainable artificial intelligence, social network, knowledge graph

Acknowledgments:

The study was carried out within the framework of the state assignment of the Institute of Mechanics and Mathematics of the Kola Science Center of the Russian Academy of Sciences of the Ministry of Science and Higher Education of the Russian Federation, research topic “Methods and technologies for creating intelligent information systems to support the development of complex dynamic systems with regional specificity in conditions of uncertainty and risk” (topic code FMEZ-2025-0053).

For citation:

Dikovitsky V. V. Predictive modeling of social reactions for regional management based on explainable artificial intelligence methods. *Trudy Kol'skogo nauchnogo centra RAN. Seriya: Tekhnicheskie nauki* [Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences], 2025, Vol. 16, No. 3, pp. 71–79. doi:10.37614/2949-1215.2025.16.3.005.

Введение

В условиях динамичной цифровизации общества и повсеместного внедрения искусственного интеллекта (ИИ) в информационные системы, традиционные подходы к мониторингу и прогнозированию общественного мнения, основанные на опросах и статическом анализе данных, сталкиваются с серьезными ограничениями, такими как снижение репрезентативности и невозможность улавливать быстро меняющуюся динамику онлайн-дискурса [1]. Алгоритмы, предназначенные для предсказания социальных реакций на основе анализа социальных сетей, часто вызывают опасения как инструменты контроля, а не как средства для развития гражданского диалога [2]. Некорректное применение методов социального инжиниринга может привести к результатам, противоположным ожидаемым. Системы, стремящиеся к управлению общественными настроениями, рискуют подорвать доверие к институтам [3].

Современные парадигмы ИИ претерпевают значительные изменения, смещая акцент с чистого прогнозирования на поддержку и сопровождение рассуждений [4]. Целью становится не создание систем, которые «думают за человека», а разработка нормативно-ориентированных агентов, которые помогают человеку принимать решения с помощью объяснимого ИИ [5]. Этот переход отражает стремление к более прозрачному, подотчетному и этичному использованию ИИ в управлении, где системы не только предсказывают, но и объясняют свои выводы [6].

Настоящая статья представляет архитектуру Human-in-the-Loop [7], предназначенную для совместного управления, в которой ИИ функционирует не как автономный субъект решения, а как совместный советник. Такой подход позволяет интегрировать человеческое суждение и ценности в процесс принятия решений, используя ИИ для повышения эффективности и этичности [8]. Эта архитектура базируется на четырех взаимосвязанных компонентах:

1) каузальные агентные модели, имитирующие динамику общественных реакций. Эти модели, основанные на взаимодействии автономных агентов, позволяют исследовать возникновение таких явлений, как консенсус и поляризация, путем моделирования причинно-следственных связей [9]. Использование таких моделей позволяет не только прогнозировать, но и понимать механизмы формирования общественного мнения, а также оценивать влияние различных факторов и интервенций [10];

2) RAG-системы (Retrieval-Augmented Generation), привязывающие рассуждения к историческим фактам [11];

3) динамический граф знаний, связывающий социальные сигналы и события [12];

4) деонтическая логика, предоставляющая формальный аппарат для выражения нормативных концепций, что позволяет ИИ-системам не только соблюдать этические нормы, но и объяснять свои решения в терминах дозволенного, обязательного и запрещенного [13]. Это способствует распределению моральной ответственности между человеком и ИИ [14]. Применение деонтической логики в государственном управлении позволяет повысить прозрачность, подотчетность и доверие к автоматизированным процессам и обеспечивает соответствие действий ИИ установленным нормам и ценностям [15].

Представленный подход отличается от предыдущих работ, которые преимущественно сосредоточены на выявлении сложившихся тенденций путем анализа статистики публикаций и различных вариаций тематического моделирования. Эти ретроспективные методы, несмотря на свою техническую состоятельность, позволяют обнаруживать уже произошедшие события, сложившиеся темы и доминирующие сообщения [16]. Они не приближают к пониманию причин возникновения трендов, норм и ценностей, которые их провоцировали, групп, инициировавших дискуссию, и потенциальных внешних вмешательств, способных изменить ее направление. Кроме того, модели ранжирования контента, вовлеченности или тематического моделирования не эффективны для анализа причин и не функционируют в реальном времени. Они не взаимодействуют с нормативной средой, не учитывают юридические ограничения, а также часто игнорируют семантику и контекст, что препятствует генерации конкретных сценариев и действий [17].

Предыдущая работа

В прошлых работах были предложены методы построения предиктивных моделей для определения популярности публикаций и определения нормативных высказываний. В работе [18] проведена оценка возможности современных LLM (Large Language Model) для моделирования трендов в социальных сетях. Исследование показало, что LLM способны эффективно анализировать контент социальных медиа для прогнозирования популярности публикаций, что является шагом к пониманию распространения информации и формированию общественного мнения. В работе [19] исследуется возможность автоматического извлечения компонентов деонтических высказываний. Методология, предложенная в этой работе, обеспечивает преобразование слабоструктурированных правовых документов в формализованные деонтические выражения, что является ключевым для интеграции нормативных знаний в ИИ-системы.

Обзор существующих подходов к моделированию социальных реакций

В условиях постоянно меняющейся цифровой среды, социальные сети трансформировались из простых коммуникационных платформ в общественные арены, оказывающие значительное влияние на коллективные настроения и поведение граждан [20]. В сфере регионального управления это создает как новые возможности для прямого взаимодействия с населением, так и серьезные вызовы, в первую очередь связанные с необходимостью оперативного мониторинга и прогнозирования социальных настроений [18]. Традиционные модели предиктивного анализа социальных сетей, ориентированные на раннее выявление зарождающихся очагов напряжения, оценку эффективности информационных кампаний и предотвращение кризисов [21], часто демонстрируют ограниченный охват, ретроспективный характер анализа и уязвимость результатов к искажениям.

В таких условиях мониторинг не соответствует требованиям управления, которое нуждается в инструментах, способных не только предсказывать всплески негативных высказываний, призывы к действию или массовую поддержку инициатив, но и формировать нормативно обоснованный контекст для вмешательства, а также проводить анализ возможных сценариев. Фрагментарность и несистемность существующих систем мониторинга подчеркивают необходимость перехода к методам прогнозирования и управления, основанным на больших данных и алгоритмических подходах. Ключевым требованием становится не только способность к предсказанию, но и прозрачная объяснимость: система должна аргументировать, почему ожидается та или иная реакция и какие последствия могут иметь различные управленческие решения [22].

Государственные и муниципальные органы активно используют мониторинг социальных сетей для решения широкого круга задач — от обеспечения общественной безопасности и управления кризисами до анализа общественного мнения, выявления угроз и повышения эффективности коммуникации [23]. Например, раннее обнаружение слухов о загрязнении воды через анализ пользовательских сообщений позволяет органам власти оперативно реагировать и предотвращать панику [24]. В период пандемии COVID-19 мониторинг тональности обсуждений помог правительствам учитывать общественную реакцию на локдауны и кампании вакцинации, адаптируя стратегии донесения сообщений [18]. В правоохранительной практике методы предиктивной аналитики применяются для выявления закономерностей преступной активности и более рационального распределения ресурсов, хотя прямые причинно-следственные эффекты от таких вмешательств еще недостаточно обоснованы [25]. В совокупности социальные медиа выступают как ценный источник обратной связи, предоставляющий возможность корректировать политику и повышать качество взаимодействия органов власти с населением в режиме реального времени. Тем не менее для перехода от мониторинга к обоснованному управлению, способному включать причинно-следственные рассуждения и обеспечивать объяснимость решений, необходима интеграция передовых методов ИИ и алгоритмов, которые не просто выявляют тренды, но и интерпретируют их в рамках правовых и этических норм, а также моделируют последствия различных управленческих сценариев.

Большие языковые модели являются современным фундаментом для анализа текстовых сигналов и прогнозирования социальных процессов благодаря своей способности обрабатывать обширные объемы неструктурированного текста и выявлять скрытые взаимосвязи [26]. В основе работы LLM лежит концепция векторизации: слова, фразы, предложения и документы преобразуются

в векторы в многомерном пространстве, где семантически близкие элементы располагаются ближе друг к другу. Ранние прообразы таких подходов, например модели GloVe, продемонстрировали, как контекстуальные представления слов могут отражать семантические связи [27]. С появлением архитектуры трансформеров векторизация приобрела контекстуальный характер: значение слова определяется не его статическим представлением, а зависит от всего предложения или документа, что позволяет учитывать полисемию и тончайшие смысловые оттенки. Размерность таких эмбеддингов существенно возросла и достигла в продвинутых системах тысяч измерений (например, до 12288 в некоторых конфигурациях GPT) [28]. Более того, современные LLM способны векторизовать не только текст, но и код, векторную графику (SVG), числовые последовательности и иные типы данных, расширяя свою сферу применения за пределы классической обработки естественного языка.

Для повышения достоверности и адаптивности моделей широко применяется архитектура Retrieval-Augmented Generation (RAG), в которой генеративный модуль LLM дополняется внешним хранилищем фрагментов знаний в виде векторной базы данных [29]. Это хранилище содержит эмбеддинги релевантных текстов: исторические данные, нормативные документы, фрагменты графов знаний. При поступлении запроса модель сначала осуществляет поиск релевантных фрагментов, затем интегрирует их в контекст и генерирует ответ с учетом как исходного запроса, так и дополнительной информации. Такой подход снижает необходимость переобучения модели и уменьшает риск «галлюцинаций», повышая достоверность прогнозов. Однако в динамичных предметных областях, таких как анализ социальных реакций, этот метод требует адаптации, поскольку его прямое применение может привести к одновременному добавлению противоречивых и/или утративших актуальность фрагментов данных.

Система предиктивного моделирования социальных реакций Crowdsearch

Система представляет собой модульный фреймворк, объединяющий обработку больших данных, современные методы представления текста и графов, каузальное агентное моделирование и формализацию нормативных требований. Система ориентирована на поддержку органов регионального и муниципального управления путем предоставления инструментов для проактивного понимания общественных настроений, проведения контрфактического анализа последствий управленческих решений и оптимизации коммуникационных стратегий. Для достижения этих целей архитектура системы должна обеспечить надежное, масштабируемое и объяснимое представление как содержательных, так и структурных аспектов коммуникативного пространства. Это достигается сочетанием контекстуальных текстовых эмбеддингов, графовых эмбеддингов, динамического графа знаний и каузального модуля (рис. 1).

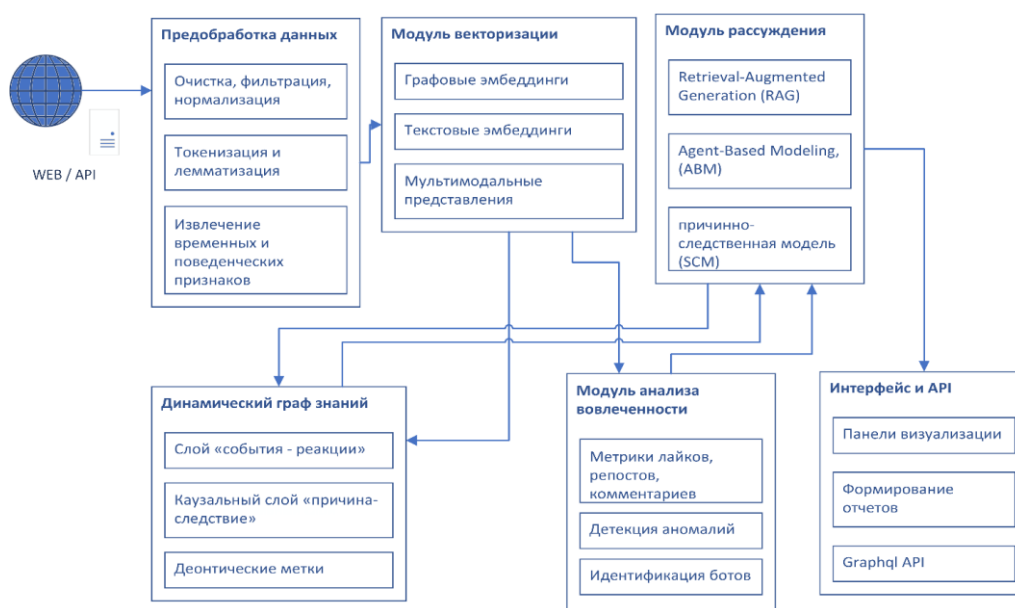


Рис. 1. Компоненты системы предиктивного моделирования социальных реакций Crowdsearch

Для учета нормативного контекста применяется технология обнаружения нормативных высказываний как компонентов деонтической логики, описанная в работе [19]. Деонтическая логика — раздел модальной логики, формализующий такие понятия, как обязательство, разрешение и запрет, служит инструментом формализации нормативных ожиданий и ограничений [13]. Ее операторы $O(\phi)$ («обязательно, чтобы ϕ »), $P(\phi)$ («разрешено, что ϕ ») и $F(\phi)$ («запрещено, что ϕ ») позволяют задавать нормы вроде «обязательно соблюдать правила платформы». Деонтическая логика применяется для анализа политических высказываний, оценки их нормативной нагрузки и возможного воздействия на аудиторию [30]. В юридической логике она используется для проверки корректности аргументации и соответствия вывода нормативной базе. Однако реальная среда, в которой действуют пользователи, сочетает как формализованные, так и неформализованные законы, правила платформ, общественные ожидания и т. п. Поэтому применена гибридная модель: деонтическая логика интегрирована с графом знаний, что позволяет формально моделировать конфликт норм и контекстуальные исключения. Технически это реализуется путем добавления в граф знаний слоя с деонтическими метками. Начальной базой для извлечения норм и правил служат правила платформы. Система может проверять соответствие прогнозируемого сообщения существующим нормам, генерировать рекомендации с учетом нормативных ограничений и обосновывать свои решения. Для анализа динамики норм и их внедрения в общественное пространство применяется алгоритмическая модель распространения: агенты (пользователи) в сети получают бинарные состояния $s_u \in \{0,1\}$, где 1 — соблюдает норму, 0 — нет. Распространение регулируется пороговыми механизмами: агент принимает норму, если доля соседей, ее соблюдающих, превышает индивидуальный порог θ_u . Также допускается затухание нормы с вероятностью p_{drop} . При каждом временном шаге измеряются проникновение нормы $P_{total}(t)$, скорость ее распространения и образование кластеров сопротивления. Модель учитывает неоднородность порогов, влияние лидеров мнений и структуру сети. Алгоритмическое моделирование позволяет моделировать динамику норм, выявлять точки сопротивления или узлы-распространители, что важно при проектировании коммуникации.

Модуль векторизации и семантического представления выполняет три параллельные функции: генерация эмбеддингов для постов и комментариев, построение графовых эмбеддингов для сетевых структур и формирование мультимодальных представлений при наличии изображений или других медиа. Для текстовой части применяется предобученная трансформер-модель RuBERT, адаптированная под русский язык [31]. Для векторизации графа знаний используется метод представлений узлов GraphSAGE — метод агрегации соседних признаков, позволяющий строить эмбеддинги для новых (ранее невидимых) узлов в динамически растущих графах. GraphSAGE позволяет обучать генератор эмбеддингов (функцию агрегации), что принципиально для социально генерируемого контента с постоянно добавляющимися узлами и источниками [32]. Практическая реализация этих компонентов эффективно поддерживается библиотекой PyTorch Geometric, предоставляющей инструменты для обработки больших графов и механизмы масштабирования на мульти-GPU серверах [33].

Динамический граф знаний с каузальным слоем является ядром архитектуры и содержит события, представленные в виде кортежей:

$$E = (P, T_{pub}, K, G_{rep}, \Phi),$$

где P — текст поста-триггера; T_{pub} — время публикации; K — набор тем; G_{rep} — множество групп-репостеров (с порогом $|G_{rep}| \geq 2$ для сигнала межгруппового охвата); Φ — деонтическая метка. Реакция A формализуется через вектор метрик активности $M(P, \Delta t) = (L, R, C)$ и сравнение с базовой статистикой аналогичных постов: фиксируется всплеск, если для какого-либо типа метрик выполняется $M_{type} > \bar{M}_{100,type} + \alpha \sigma_{100,type}$ (порог α настраиваем). Тональность реакции оценивается агрегированием классификаций сентимента отдельных комментариев $S(c_j)$ и последующей интерпретацией суммарной оценки $S_{overall}$. Одно из ключевых отличий архитектуры — интеграция графовых эмбеддингов и каузального слоя с Retrieval-Augmented Generation (RAG) для обеспечения объяснимости и опоры на источники. В рабочих сценариях LLM не генерирует рекомендации «из головы»: перед генерацией RAG-модуль извлекает из векторного индекса релевантные фрагменты (исторические кейсы, части графа знаний), которые затем становятся контекстом для генеративного шага. Такой подход улучшает

фактическую точность и обеспечивает возможность указывать на источники, снижая риск «галлюцинаций». Однако использование RAG без адаптации индекса влечет проблему множественности контекстов, когда в векторном хранилище содержатся релевантные, но противоречивые и/или утратившие актуальность фрагменты данных. Для решения этой проблемы применяется процедура отбора и ранжирования фрагментов данных с учетом версионирования индекса [34].

Для причинно-следственного вывода использованы методы из теории каузальности: модельные подходы и графы причинно-следственных связей, позволившие формализовать интервенции [35]. Практически это реализуется через сочетание причинно-следственных графов с агентно-ориентированными моделями. Граф причинно-следственных связей осуществляет идентифицируемое представление события и его зависимостей, агентно-ориентированная модель позволяет симулировать механизмы распространения и эмпирически проверять гипотезы о поведении агентов в сетевом окружении.

Компонент обнаружения аномалий сочетает статистическую детекцию всплесков, признаки сетевой активности и набор моделей классификации на основе анализа сетевых метрик и контента. Для оценки достоверности предложенных интервенций последние дополняются разделом «обоснование» в виде ссылок на извлеченные фрагменты RAG-хранилища и «дерево причин» каузального графа.

Источниками данных являются публичные группы и страницы социальной сети «ВКонтакте» за длительный период (2020 — н. в.), что позволяет учитывать как фоновую активность, так и кризисные всплески. Для оперативной аналитики система использует инкрементальное извлечение метрик с временными метками и шагом обновления каждые 20 минут; хранится история метрик для каждого поста, что служит базой для детекции аномалий и расчета нормализованных порогов всплесков.

Заключение

Предложенный архитектурный фреймворк сочетает проверенные алгоритмические блоки (GraphSAGE, RAG, LLM, каузальный аппарат и агентно-ориентированные модели) и опирается на современные библиотечные экосистемы (PyTorch Geometric, Milvus для векторных индексов). Такой набор обеспечивает практическую реализуемость, масштабируемость и объяснимость решений, необходимых для нормативно-ориентированного рассуждения и контрфактического анализа в региональном управлении. Ведутся работы по экспериментальной проверке эффективности отдельных предложенных модулей. Для задач дообучения LLM, построения динамических графов и верификации деонтических и каузальных предсказаний значительная часть текстов (публикации и комментарии) аннотируются в полуавтоматическом режиме. Процесс аннотации включает: категоризацию социальных реакций (негативные комментарии, призывы к действию, поддержку инициатив, репосты), оценку тональности реакций, аннотацию деонтических характеристик. Для создания обучающего набора каузальных связей и управленческих воздействий промаркированы пары «вмешательство — реакция» и «событие — реакция» путем учета наличия официальной реакции, например «всплеск негативных комментариев — разъяснение уполномоченного представителя». Эти данные являются основой (обучающей выборкой) для каузального слоя графа знаний и для валидации контрфактических сценариев.

Список источников

1. Kim R. M., Veselovsky V., Anderson A. Capturing dynamics in online public discourse: A case study of universal basic income discussions on reddit // Proceedings of the International AAAI Conference on Web and Social Media. 2025. Vol. 19. P. 1021–1037.
2. Bennett C. J., Lyon D. Data-driven elections: implications and challenges for democratic societies // Internet Policy Review. 2019. 8 (4). doi: 10.14763/2019.4.1433.
3. Карпова И. В., Воронкова И. Е., Бирюкова Е. А. PR-технологии как средство манипулирования политической активностью молодежи // Реклама и PR в координатах социума, бизнеса и медиапространства. 2020. С. 18–73.
4. Revel M., Pénigaud T. AI-Enhanced Deliberative Democracy and the Future of the Collective Will. 2025.
5. Argyle L. P. et al. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale // Proceedings of the National Academy of Sciences. 2023. Vol. 120, No. 41. P. e2311627120.
6. Evkoski B., Pollak S. XAI in Computational Linguistics: Understanding Political Leanings in the Slovenian Parliament //arXiv preprint arXiv:2305.04631. 2023.

7. Mosqueira-Rey E. et al. Human-in-the-loop machine learning: a state of the art // *Artificial Intelligence Review*. 2023. Vol. 56, No. 4. P. 3005–3054.
8. Chen X. Ethical Governance of AI: An Integrated Approach via Human-in-the-Loop Machine Learning // *Comput. Sci. Math. Forum*. 2023. 8. 29. <https://doi.org/10.3390/cmsf2023008029>.
9. Antosz P., Szczepanska T., Bouman L., Gareth Polhill J., Jager W. Sensemaking of causality in agent-based models // *International Journal of Social Research Methodology*. 2022. 25:4. P. 557–567. doi: 10.1080/13645579.2022.2049510.
10. Arnold K. F. et al. DAG-informed regression modelling, agent-based modelling and microsimulation modelling: a critical comparison of methods for causal inference // *International journal of epidemiology*. 2019. Vol. 48, No. 1. P. 243–253.
11. Lewis P. et al. Retrieval-augmented generation for knowledge-intensive nlp tasks // *Advances in neural information processing systems*. 2020. Vol. 33. P. 9459–9474.
12. Yang R., Salim F. D., Xue H. Sstk: Simple spatio-temporal knowledge graph for interpretable and versatile dynamic information embedding // *Proceedings of the ACM Web Conference 2024*. 2024. P. 551–559.
13. Benz Müller C., Parent X., van der Torre L. Designing Normative Theories for Ethical and Legal Reasoning: LogiKEY Framework // *Methodology, and Tool Support*. 2020.
14. Rao S. et al. Deontic Temporal Logic for Formal Verification of AI Ethics // *arXiv preprint arXiv:2501.05765*. 2025.
15. Priya T. V., Rao S. Deontic Temporal Logic for Formal Verification of AI Ethics // *arXiv preprint arXiv:2501.05765*. 2025.
16. Tong, Yangfan, Sun, Wei. Multimedia Network Public Opinion Supervision Prediction Algorithm Based on Big Data // *Complexity*. 2020. 6623108, 11 p. <https://doi.org/10.1155/2020/6623108>.
17. Yukhno A. Digital transformation: Exploring big data governance in public administration // *Public Organization Review*. 2024. Vol. 24, No. 1. P. 335–349.
18. Shishaev M., Dikovitsky V. Predicting the Popularity of Social Network Publications Based on Content Analysis Using the Transformer Language Model // *International Scientific and Practical Conference Digital and Information Technologies in Economics and Management*. Cham: Springer Nature Switzerland, 2023. P. 180–191.
19. Dikovitsky V. V., Shishaev M. G. Automated extraction of deontological statements through a multilevel analysis of legal acts // *Proceedings of the Computational Methods in Systems and Software*. Cham: Springer International Publishing, 2018. P. 102–110.
20. Yang G. R. et al. A method of predicting and managing public opinion on social media: An agent-based simulation // *Information Sciences*. 2024. Vol. 674. P. 120722.
21. Li H. et al. Detecting early-warning signals for social emergencies by temporal network sociomarkers // *Information Sciences*. 2023. Vol. 627. P. 189–204.
22. Wang X. et al. Impact of social participation types on depression in the elderly in China: an analysis based on counterfactual causal inference // *Frontiers in Public Health*. 2022. Vol. 10. P. 792765.
23. Akopova T. S., Tikhonova A. V. Citizens' activity in social networks: factors of influence. 2020. *Креативная экономика*, 18 (12), 3805–3826. doi: 10.18334/ce.18.12.122155
24. Xiong J., Hswen Y., Naslund J. A. Digital surveillance for monitoring environmental health threats: a case study capturing public opinion from Twitter about the 2019 Chennai water crisis // *International journal of environmental research and public health*. 2020. Vol. 17, No. 14. P. 5077.
25. Lee Y., Bradford B., Posch K. The effectiveness of big data-driven predictive policing: Systematic review // *Justice Evaluation Journal*. 2024. Vol. 7, No. 2. P. 127–160.
26. Rusnachenko N. L. Language Models Application in Sentiment Attitude Extraction Task // *Trudy ISP RAN/Proc. ISP RAS*. 2021. Vol. 33, issue 3. P. 199–222. doi: 10.15514/ISPRAS-2021-33(3)-14.
27. Shishaev M., Dikovitsky V., Pimeshkov V., Kuprikov N., Kuprikov M., Shkodyrev V. Extracting relations from texts using vector language models and a neural network classifier // *PeerJ Computer Science*. 2023. 9. e1636.
28. Wang Z. et al. History, development, and principles of large language models: an introductory survey // *AI and Ethics*. 2025. Vol. 5, No. 3. P. 1955–1971.
29. Huwiler D., Stockinger K., Fürst J. VersionRAG: Version-Aware Retrieval-Augmented Generation for Evolving Documents // *arXiv preprint arXiv:2510.08109*. 2025.
30. Сытько А. В. Говорящий как субъект деонтики в политической речи (на материале немецкого и русского языков) // *Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика*. 2019. Т. 10, № 4. С. 1003–1020. doi: 10.22363/2313-2299-2019-10-4-1003-1020.
31. RAS F.ruSciFact: Open Benchmark for Verifying Scientific Facts in Russian // *Proceedings of the International Conference “Dialogue 2025”*. 2025.
32. Huang K., Chen C. Subgraph generation applied in GraphSAGE deal with imbalanced node classification // *Soft Computing*. 2024. Vol. 28, No. 17. P. 10727–10740.
33. Fey M., Lenssen J. E. Fast graph representation learning with PyTorch Geometric // *arXiv preprint arXiv:1903.02428*. 2019.

34. Lewis P. et al. Retrieval-augmented generation for knowledge-intensive nlp tasks // *Advances in neural information processing systems*. 2020. Vol. 33. P. 9459–9474.
35. Pearl J. *Causality*. 2nd ed. Cambridge University Press, 2009.

References

1. Kim R. M., Veselovsky V., Anderson A. Capturing dynamics in online public discourse: A case study of universal basic income discussions on Reddit. *Proceedings of the International AAAI Conference on Web and Social Media*, 2025, Vol. 19, pp. 1021–1037.
2. Bennett C. J., Lyon D. Data-driven elections: implications and challenges for democratic societies. *Internet Policy Review*, 2019, 8 (4). doi: 10.14763/2019.4.1433.
3. Karpova I. V., Voronkova I. E., Biryukova E. A. PR technologies as a means of manipulating the political activity of young people. *Advertising and PR in the coordinates of society, business, and media space*, 2020, pp. 18–73. (In Russ.).
4. Revel M., Pénigaud T. *AI-Enhanced Deliberative Democracy and the Future of the Collective Will*, 2025.
5. Argyle, Lisa P et al. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences of the United States of America*, 2023, vol. 120, 41, p. e2311627120. doi:10.1073/pnas.2311627120.
6. Evkoski B., Pollak S. XAI in computational linguistics: understanding political leanings in the Slovenian Parliament. 2023. Available at: <https://arxiv.org/pdf/2305.04631> (accessed 20.11.2025).
7. Mosqueira-Rey E. et al. Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, 2023, 56 (4), pp. 3005–3054.
8. Chen X. Ethical governance of AI: an integrated approach via human-in-the-loop machine learning. *Comput. Sci. Math. Forum*, 2023, 8, 29. <https://doi.org/10.3390/cmsf2023008029>.
9. Antosz P., Szczepanska T., Bouman L., Polhill J. G., Jager W. Sensemaking of causality in agent-based models. *International Journal of Social Research Methodology*, 2022, 25 (4), pp. 557–567.
10. Arnold K. F. et al. DAG-informed regression modeling, agent-based modeling and microsimulation modeling: a critical comparison of methods for causal inference. *International Journal of Epidemiology*, 2019, 48 (1), pp. 243–253.
11. Lewis P. et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 2020, Vol. 33, pp. 9459–9474.
12. Yang R., Salim F.D., Xue H. Sstkg: simple spatial-temporal knowledge graph for interpretable and versatile dynamic information embedding. *Proceedings of the ACM Web Conference 2024*, 2024, pp. 551–559.
13. Benz Müller C., Parent X., van der Torre L. Designing normative theories for ethical and legal reasoning: LogiKey framework. *Methodology and Tool Support*, 2020.
14. Rao S. et al. Deontic Temporal Logic for Formal Verification of AI Ethics. 2025.
15. Priya T. V., Rao S. Deontic Temporal Logic for Formal Verification of AI Ethics. 2025. Available at: <https://arxiv.org/pdf/2501.05765> (accessed 19.11.2025).
16. Tong Y., Sun W. Multimedia network public opinion supervision prediction algorithm based on big data. *Complexity*, 2020, 6623108. doi: 10.1155/2020/6623108.
17. Yukhno A. Digital transformation: exploring big data governance in public administration. *Public Organization Review*, 2024, 24 (1), pp. 335–349.
18. Shishaev M., Dikovitsky V. Predicting the popularity of social network publications based on content analysis using the transformer language model. *Digital and Information Technologies in Economics and Management*. Cham, Springer, 2023, pp. 180–191.
19. Dikovitsky V. V., Shishaev M. G. Automated extraction of deontological statements through multilevel analysis of legal acts. *Computational Methods in Systems and Software*. Cham, Springer, 2018, pp. 102–110. (In Russ.)
20. Yang G.R. et al. A method of predicting and managing public opinion on social media: an agent-based simulation. *Information Sciences*, 2024, 674. doi: 10.1016/j.ins.2024.120722
21. Li H. et al. Detecting early-warning signals for social emergencies by temporal network sociomarkers. *Information Sciences*, 2023, 627, pp. 189–204.
22. Wang X. et al. Impact of social participation types on depression in the elderly in China: an analysis based on counterfactual causal inference. *Frontiers in Public Health*, 2022, 10, 792765. doi:10.3389/fpubh.2022.792765.
23. Akopova T. S., Tikhonova A. V. Citizens' activity in social networks: influence factors, 2020. *Creative Economy*, 18 (12), 3805–3826. (In Russ.). doi: 10.18334/ce.18.12.122155.

24. Xiong J., Hsuen Y., Naslund J. A. Digital surveillance for monitoring environmental health threats: a case study capturing public opinion from Twitter about the 2019 Chennai water crisis. *International Journal of Environmental Research and Public Health*, 2020, 17 (14), 5077. doi: 10.3390/ijerph17145077.
25. Lee Y., Bradford B., Posch K. The effectiveness of big data-driven predictive policing: systematic review. *Justice Evaluation Journal*, 2024, 7 (2), pp. 127–160.
26. Rusnachenko N. L. Language models application in sentiment attitude extraction task. *Proc. ISP RAS*, 2021, 33 (3), pp. 199–222.
27. Shishaev M., Dikovitsky V., Pimeshkov V., Kuprikov N., Kuprikov M., Shkodyrev V. Extracting relations from texts using vector language models and a neural network classifier. *PeerJ Computer Science*, 2023, 9, e1636.
28. Wang Z. et al. History, development, and principles of large language models: an introductory survey. *AI and Ethics*, 2025, 5 (3), pp. 1955–1971.
29. Huwiler D., Stockinger K., Fürst J. VersionRAG: Version-aware retrieval-augmented generation for evolving documents. 2025. Available at: <https://arxiv.org/pdf/2510.08109> (accessed 25.10.2025).
30. Sytko A. V. The speaker as a subject of deontics in political speech (based on the German and Russian languages). *RUDN Journal of Language Theory. Semiotics. Semantics*, 2019, 10 (4), pp. 1003–1020.
31. RAS F. ruSciFact: open benchmark for verifying scientific facts in Russian. *Proceedings of the International Conference “Dialogue”*, 2025.
32. Huang K., Chen C. Subgraph generation applied in GraphSAGE to deal with imbalanced node classification. *Soft Computing*, 2024, 28 (17), pp. 10727–10740.
33. Fey M., Lenssen J.E. Fast graph representation learning with PyTorch Geometric. Available at: <https://arxiv.org/pdf/1903.02428> (accessed 20.10.2025).
34. Lewis P. et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 2020, Vol. 33, pp. 9459–9474.
35. Pearl J. *Causality*. Cambridge, Cambridge University Press, 2009.

Информация об авторе

В. В. Диковицкий — кандидат технических наук, старший научный сотрудник.

Information about the author

V. V. Dikovitsky — Candidate of Science (Tech.), Senior Research Fellow.

Статья поступила в редакцию 23.10.2025; одобрена после рецензирования 30.10.2025; принята к публикации 03.11.2025.
The article was submitted 23.10.2025; approved after reviewing 30.10.2025; accepted for publication 03.11.2025.