Автоматическая детекция адъективной неопределённости в русском юридическом тексте: датасет, модели, результаты

Алёна Е. Берлин, Ольга В. Блинова

***Аннотация*—Статья посвящена созданию инструмента, способного автоматически классифицировать русские предложения, взятые из нормативных документов (законов), на содержащие и не содержащие правовую неопредённость. Для создания такого инструмента силами лингвистов и юристов разработан обучающий набор данных, содержащий 6000 предложений с метками и локусами неопределённости. Учтены только случаи, когда неопределённость вводится градуируемым прилагательным. В настоящей работе показано, как авторы использовали этот датасет в ряде экспериментов с классическими моделями машинного обучения и с трансформерной моделью. Применение аугментации в эксперименте с RuBERT позволило преодолеть проблему дисбаланса классов и достичь качества с показателем F1=0,89. При анализе языковых признаков, влияющих на предсказания модели, замечена концентрация прилагательных с приставкой не- в классе предложений с отсутствием неопределённости.**

***Ключевые слова* — лингвистическая неопределённость, правовая неопределённость, русский юридический текст, адъективная неопределённость, бинарные классификаторы, RuBERT.**

I. Введение

Настоящая статья посвящена созданию автоматической модели классификации русских предложений на содержащие и не содержащие правовую неопределённость (legal vagueness). Задача, которую мы решаем, сложнее, чем предсказание лингвистической неопределённости (linguistic vagueness). Решение нашей задачи возможно только как результат совместного применения знаний лингвистов и юристов (подробнее см. раздел II-D ниже).

Лингвистическая неопределённость –явление, связанное с градуируемостью границ между значениями и определяемое через «неуверенность в отношении применимости слова к денотату» [1]. Случаи неопределённости можно разделить на «неопределённость критериев» (vagueness in criteria) и «неопределённость степени» (vagueness in degree) [2], [3]. В ситуации с неопределённостью критериев неясно, какие именно критерии или ряд критериев нужно использовать для отнесения денотата к определённому классу, обозначаемому словом (языковым выражением). Например, можно ли назвать помидор «фруктом», а брейк-данс – «видом спорта». В ситуации с неопределённостью степени неясно, при каком значении вполне ясного критерия можно использовать слово (языковое выражение) при характеризации денотата. Например, можно ли назвать семью из пяти человек «большой», новорожденного весом 3,8 кг «крупным», дом высотой в 8 этажей – «многоэтажкой».

В настоящей статье мы рассматриваем один из видов неопределённости степени – адъективную, причём только такую, которую вводят прилагательные, выражающие градуируемые признаки, то есть признаки и свойства, соотносимые со шкалами и способные проявляться в разной степени. Когда какое-то языковое выражение вводит неопределённость, оно называется локусом неопределённости [4]. Например, в предложении (1) локусом неопределённости является градуируемое прилагательное разумный.

(1) Если указанные нарушения влекут за собой разрушение жилого помещения, наймодатель также вправе назначить нанимателю и членам его семьи разумный срок для устранения этих нарушений. [«Жилищный кодекс Российской Федерации» от 29.12.2004 N 188-ФЗ].

Интерпретация высказываний с такими прилагательными контекстно-зависима, вовлекает парадокс Сорита и предполагает существование пограничных случаев, см. об этом в применении к правовым тестам [5].

Вслед за К. Кеннеди [6] градуируемые прилагательные можно разделить на **абсолютные** и **относительные**. Абсолютные градуируемые прилагательные в положительной степени связывают объекты с минимальной или максимальной степенью проявления свойства (то есть с закрытой шкалой). Относительные градуируемые прилагательные в положительной степени связывают объекты с некоторым проявлением свойства на открытой шкале. К абсолютным градуируемым прилагательным относятся, например, *полный, пустой, чистый, грязный*;

Статья получена 25 октября 2025. Работа выполнена при поддержке СПбГУ, шифр проекта 123042000068-8.

Берлин Алёна Евгеньевна, Санкт-Петербургский государственный университет (e-mail: alenakat2000@gmail.com).

Блинова Ольга Владимировна, Санкт-Петербургский государственный университет (e-mail: o.blinova@spbu.ru), НИУ «Высшая школа экономики» (e-mail: ovblinova@hse.ru).

Статья подготовлена по итогам выступления на Международной объединённой конференции «Интернет и современное общество» (IMS-2025).

к относительным – *дорогой, дешёвый, высокий, низкий*. Согласно [6], именно относительные (в терминологии Кеннеди) градуируемые прилагательные вводят неопределённость.

Градуируемые прилагательные в русском относятся к разряду **качественных** и обладают следующими признаками, см. [7]:

1) они сочетаются с показателями степени (*крайне высокий*, *очень высокий*);
2) они имеют формы степеней сравнения (выше, повыше, более высокий, менее высокий, самый высокий);
3) часто они имеют краткие формы (*высок*);
4) от них образуются дериваты с суффиксами субъективной оценки (*высоковатый*, *высоченный*);
5) от них образуются дериваты с приставками и префиксоидами со значением ‘очень’ (*ультравысокий*);
6) их редуплицированное употребление передаёт значение интенсивности (*высокий-высокий*);
7) от них образуются качественные наречия (*высоко*);
8) они могут соотноситься с производными абстрактными существительными (*высота*);
9) для них характерно наличие антонимов и синонимов (*высокий – низкий*; *высокий – рослый – долговязый*);
10) они могут субстантивироваться (*Высокий поздоровался*).

Правовая неопределённость произрастает из лингвистической. Это дефект правового регулирования, выражающийся в недостаточной ясности, точности и однозначности правовых норм. Правовая неопределённость приводит к возможности множественного толкования и произвольного применения норм.

Настоящая статья структурирована следующим образом. В разделе II мы рассказываем о существующих **наборах данных** с разметкой неопределённости и о том, как мы создавали свой (отличающийся от прочих наличием разметки **правовой неопределённости**). Раздел III посвящён организации и результатам **экспериментов по обучению классификаторов**, разработанных для автоматической детекции неопределённых предложений. В разделе IV мы обсуждаем **языковые признаки**, влияющие на предсказания модели. Раздел V содержит выводы и исследовательские планы авторов.

II. Датасет с разметкой по критерию наличия/отсутствия неопределённости

A. Предшествующий опыт

Все существующие аннотированные датасеты, пригодные для решения задач классификации по критерию наличия/отсутствия неопределённости или по типу неопределённости, можно разбить на группы в зависимости от:

1) языка;
2) специализации на каких-то разновидностях текстов (например, жанрах) или её отсутствии;
3) аннотируемых единиц;
4) количества и состава выделяемых классов;
5) экспертности аннотаторов (это могут быть профильные эксперты или краудсорсинговая команда);
6) уровня согласия аннотаторов;
7) размера.

Кроме описанного в [8] и используемого в настоящей статье набора русских предложений с разметкой неопределённости, существующие датасеты преимущественно англоязычны. Авторы [9], впрочем, занимались выделением неопределённых темпоральных выражений в немецком. Кроме того, в работе [10] использовался список английских неопределённых прилагательных и наречий, при этом для элементов списка был получен автоматический перевод на португальский и испанский. Упомянем также исследование [11], в ходе которого был создан набор неопределённых выражений из интервью с российскими чиновниками.

Работы, посвящённые выявлению неопределённости, могут быть нацелены на формирование **общеязыковых** реестров неопределённых выражений [12] или ориентироваться на определённые **разновидности текстов**. Нам известны работы на материале обучающих инструкций, определений в онтологиях, согласий субъектов персональных данных. Например, авторы [13] сопоставили версии английских предложений, взятые из инструкций в составе корпуса wikiHowToImprove, и выявили пары, в которых редактура (замена глагола, напр., “*go*” на “*visit*”, “*get*” на “*perchase*”) конкретизировала фрагмент.

В различных исследованиях метки неопределённости присваивались разным **языковым единицам** (словам, синтаксическим группам, предложениям). Например, Парис с коллегами создали масштабный ресурс по **неопределённым именным группам** на материале Википедии, выделив три категории неопределённости (градуальную, квантитативную и субъективную) [12].

Заметим, что в работе на материале юридических тестов [14] использовали список из **40 терминов**, способных вводить неопределённость, предложенный [15]; список был выверен экспертами в области составления соответствующих документов (согласий субъектов персональных данных), то есть юристами. Мы при аннотировании неопределённых контекстов также применили **двуступенчатую разметку**, выполняемую сначала лингвистами, а затем – юристами.

Количество и состав выделяемых в исследованиях **классов** можно описать следующим образом:

1) выделяется **два класса** выражений (вводящие и не вводящие неопределённость), см. [16];
2) выделяется более двух классов: неопределённые выражения **ранжируются** по степени неопределённости, например, в [10] выделены следующие классы слов: “Definitely Not Vague”, “Partially Not Vague,” “Partially Vague,” “Definitely Vague”;
3) выделяется более двух классов: неопределённые выражения **распределяются** по типам

неопределённости, например, как в [12].

Согласие разметчиков варьирует от среднего до высокого, каппа Коэна [0.64;0.91]. **Объём** наборов данных составляет 2000–4500 размеченных единиц, что обусловлено трудоемкостью ручной аннотации.

B. Получение набора данных: двуступенчатая разметка

Получение набора русских предложений с разметкой неопределённости подробно описан в [8].

Наша цель – создание классификатора, способного диагностировать неопределённость взятого из юридического текста предложения на русском языке. Соответственно, мы использовали схему аннотирования, в которой каждому предложению вручную присваивались метки «нет» («не неопределённое») или «да» («неопределённое»). Далее в настоящей статье мы используем метки **«0»** («нет неопределённости») и **«1»** («есть неопределённость»).

В качестве источника предложений использован корпус “CorCodex” (объём корпуса – 3 млн 227 тыс. токенов). Он состоит из 282 документов и включает 200 федеральных законов РФ, 37 постановлений Правительства РФ, 21 кодекс и некоторые другие **нормативно-правовые документы**.

Выбор для исследования **законов** обусловлен их влиянием на жизнь общества. Наличие в законе неопределённых выражений способно приводить к затруднениям правоприменения, злоупотреблению, невозможности однозначно истолковать тест закона (в том числе – простыми гражданами, неюристами) и другим последствиям.

С помощью словаря «КартаСловСент» [17] мы выбрали из корпуса все предложения, содержащие прилагательные с тональными метками “PSTV” или “NGTV”. Эти предложения были размечены лингвистом, который выделил **локусы неопределённости** (в общей сложности **226 уникальных лемм**). Все предложения, которые эксперт счёл содержащими лингвистическую неопределённость, были отобраны для предъявления юристам.

Мы использовали процедуру параллельного независимого аннотирования. Команда **трёх аннотаторов-юристов** (A1, A2, A3) состояла из одного кандидата юридических наук и двух аспирантов.

Юристы получили **6153 предложения** для ручной разметки. Процесс разметки обсуждался на нескольких семинарах, поэтому подробные письменные инструкции не разрабатывались. Каждый аннотатор заполнял Google-таблицу и выбирал один из трёх ответов: «да» (если неопределённость была обнаружена), «нет» (если неопределённости не было), «неясно». Эксперты размечали не только предложения, но и леммы, которые, по мнению лингвиста, являются локусами неопределённости. Кроме того, у аннотаторов была возможность заполнить поле комментариев.

Анализ результатов разметки в команде юристов показал, что **процент согласия** составил 33%; **коэффициенты согласия** «каппа Флейсса» (-0.13) и «альфа Криппендорфа» (-0.14) отрицательны, согласие между аннотаторами меньше ожидаемого случайного. По-видимому, сложность интерпретации **ограниченного контекста** (предложения) в сочетании со значительной нагрузкой (более 6000 предложений) сделали задачу аннотирования очень сложной.

Учитывая уровень экспертизы, долю отказов от ответа, количество использований метки «неясно» и тщательность анализа языковых примеров (включая количество содержательных комментариев и способность подробно обсуждать примеры на семинарах), мы решили в дальнейшем использовать только ответы аннотатора A1.

C. Набор данных

Описанным образом мы получили набор из 6037 предложений, каждому из которых присвоена метка наличия/отсутствия неопределённости. Кроме того, каждому предложению соположено прилагательное, способное быть локусом неопределённости.

В целом аннотатор A1 выделил 150 лемм как вводящие неопределённость и 171 лемму как не вводящие её; в списках есть пересечения, подробнее см. [8].

Для целей настоящей статьи мы **отредактировали датасет**, выполнив в некоторых случаях дополнительное членение на предложения (подробнее о предобработке см. раздел III–D ниже). Сейчас он состоит из 6126 предложений, из которых к классу «0» (предложения без неопределённости) принадлежит 4165 примеров, к классу «1» (предложения с неопределённостью) – 1961 пример. **Доля предложений миноритарного класса – 32,01%**.

D. Обсуждение результатов разметки и сложность решаемых задач

Мы хотели бы кратко обсудить результаты разметки, чтобы задачи, которые мы решаем, были более ясны для читателя.

Во-первых, мы выяснили, см. [8], что лингвисты и юристы используют при анализе языковых контекстов **разные алгоритмы интерпретации**. Лингвист анализирует узкий контекст, опирается на семантику и признаки градуируемости прилагательных. Юристы учитывают языковые особенности, но полагаются прежде всего **на знание закона (корпуса законов) или наличие** в интерпретируемом предложении **ссылки на какую-то другую статью нормативного акта** (то есть статью, способную прояснить контекст). Сказанное означает, что для пополнения набора данных, пригодного для автоматического выявления правовой неопределённости, привлечение юристов будет необходимо и в дальнейшем.

Во-вторых, мы обнаружили, что **одни и те же леммы могут вносить или не вносить правовую неопределённость**. Соответственно, в дальнейшем необходимо обогащение признаков классификации за счёт дополнительных видов информации, то есть признаков, которые позволят учесть сочетаемостные свойства прилагательных и их синтаксические особенности (как минимум, употребление в атрибутивной или предикативной функции).

III. Экспериментыпообучениюклассификаторов

E. Предшествующий опыт

Задача автоматического выявления неопределённых слов, синтаксических групп или предложений привлекала внимание исследователей из различных предметных областей. Одна из первых попыток классификации была предпринята Алексопулосом и Павлопулосом (2014) [16], которые для обнаружения неопределенных дефиниций в онтологиях применили Наивный байесовский классификатор на данных словарных определений. Исследователи достигли точности в 82%.

Лебанофф и Лю использовали генеративную модель машинного обучения GAN (AC-GAN) [18]. Эта модель позволила создавать синтетические примеры для аугментации данных. В этой работе исследователи достигли значения F1-меры 0,52 на уровне предложений.

Впоследствии Дебнат и Рот (2021) использовали BiLSTM для анализа неопределённости в обучающих инструкциях, достигнув точности 0,74, выросшей до 0,78 при применении эмбеддингов BERT [13].

Малик с коллегами провели масштабное сравнительное исследование архитектур на материале политик конфиденциальности, протестировав Сверточные нейронные сети (CNN) и модели на основе трансформера (BERT, RoBERTa, DistilRoBERTa и специализированный PolicyBERT) [14]. Наилучшие результаты в этом исследовании показала модель DistilRoBERTa с регрессионной постановкой задачи, F1-мера достигла 0,88 для бинарной классификации.

Общей чертой всех исследований стало применение **трансформерных архитектур**, демонстрирующих превосходство над классическими методами машинного обучения. Тем не менее, абсолютные значения метрик качества в большинстве исследований относительно низкие, что свидетельствует о сложности задачи автоматического предсказания лингвистической неопределённости.

F. Основное содержание исследования

Настоящая статья посвящена описанию шагов, направленных на создание **классификатора, способного предсказывать правовую неопределённость** русских предложений, извлечённых из юридических текстов (прежде всего – нормативных документов).

Для достижения этой цели мы провели эксперименты по сравнению шести моделей машинного обучения, принадлежащих разным семействам классификаторов, среди которых следующие.

1) **Логистическая регрессия** (Logistic Regression, LR), используемая нами как интерпретируемый базовый уровень для разреженных представлений текста [19].
2) **Метод опорных векторов** (Support Vector Machine, SVM, с радиально-базисным ядром RBF): этот классификатор улавливает нелинейные границы в высокоразмерных разреженных пространствах, что дает преимущество перед логистической регрессией [20].
3) **Случайный лес** (Random Forest, RF): бэггинг решающих деревьев добавляет в набор нелинейность и устойчивость к выбросам [21].
4) **Градиентный бустинг по деревьям**: XGBoost [22] и CatBoost [23]. XGBoost автоматически выучивает пороговые и взаимодействующие правила. Выбор CatBoost оправдан тем, что упорядоченный бустинг снижает смещение от утечки целевой переменной и стабилизирует обучение на малых и средних выборках. В юридических текстах, где конструкции повторяются, это даёт прирост обобщающей способности.
5) **Стекинг-классификатор** (Stacking) [24]. Мы объединили представителей разных семейств (LR, SVM-RBF, RF, XGBoost и CatBoost) с метамоделью на верхнем уровне. Такой классификатор выигрывает за счёт того, что может обучаться на разнородных ошибках: линейная модель лучше улавливает лексические маркеры, бустинги фиксируют нелинейные композиции атрибутов и контекстов, SVM с ядром RBF — гладкие нелинейные границы. Мета-классификатор учится расставлять повышенные веса более сильным базовым моделям и улучшает финальные предсказания за счёт их согласования.
6) **RuBERT (DeepPavlov)** [25]. Несмотря на существование моделей-трансформеров, обученных специально на юридическом домене (LEGAL-BERT), ни одна из них не поддерживает русский язык, поэтому русскоязычный BERT оказался наиболее подходящим для нашей задачи. Этот трансформер справляется с полисемией и способен улавливать синтаксические признаки (например, согласование с прилагательными в атрибутивных и предикативных конструкциях). Это важно для юридических текстов, в которых синтаксически связанные слова могут быть линейно разнесены.

Выбранная методология — тестирование TF-IDF с линейными моделями и деревьями решений, далее ансамбли и, в итоге, контекстные эмбеддинги RuBERT — **обеспечивает чистоту сравнения**: сначала фиксируется сильная линия базовых результатов, а затем проверяется, насколько глубинные представления способны предсказывать наличие правовой неопределённости с локусом-прилагательным в русских юридических текстах.

G. Метрики оценки качества

Классические модели (LR, SVM с RBF-ядром, Random Forest, XGBoost, CatBoost и стекинг) мы оценивали по стратифицированной 5-кратной кросс-валидации. Для учёта дисбаланса внутри обучающих частей применялся алгоритм синтетической балансировки (SMOTE), встроенный в конвейер предобработки [26]. Метрики приводились как среднее и стандартное отклонение по частям (фолдам).

Трансформерная модель RuBERT оценивалась на валидационной части данных.

Во всех экспериментах считались:

1) F1-мера;
2) Полнота по классу «1»;
3) Точность по классу «1»;
4) Площадь под кривой точность-полнота (PR-AUC).
5) Аккуратность по классу «1».

При дисбалансе классов наиболее информативны F1-мера и площадь под кривой (PR-AUC). Во всех случаях мы выбирали лучшую модель, основываясь на значении F1-меры для класса «1».

H. Эксперименты с классическими моделями

Обучение классических моделей проводилось на оригинальном датасете **с дисбалансом классов 70% – 30%** в сторону класса «0» (предложений без неопределённости).

При **предобработке** удалялись неинформативные элементы (например, «*Статья 254*»), текст приводился к нижнему регистру, предложения токенизировались с помощью NLTK (модуль Punkt для русского) [27], после чего из них удалялись небуквенные токены (цифры, пунктуация).

На этапе **векторизации** применялось удаление русских стоп-слов из библиотеки NLTK, при этом лемматизация и стемминг в базовой конфигурации не использовались: словоформы были сохранены, чтобы не терять морфологические признаки (в частности, синтетические формы степеней сравнения прилагательных).

Векторизация осуществлялась с помощью **TF–IDF** по токенам. Для подавления разреженности было применено усечённое сингулярное разложение (Truncated SVD) на 300 компонент [28].

С целью компенсации дисбаланса классов использовалась **синтетическая балансировка методом SMOTE** с автоматическим подбором стратегии балансировки классов [29]. Разбиение на выборки реализовывалось стратифицированно: обучающая и валидационная части составляли 80% и 20%, соответственно (то есть 4228 и 1057 предложений).

Мы сравнили пять семейств алгоритмов (LR, SVM, RF, XGBoost, CatBoost) в одном и том же конвейере обработки признаков. После того, как был задан пайплайн, проводилась стратифицированная 5-кратная кросс-валидация.

Стекинг-классификатор (Stacking) – это двухуровневый ансамбль: на нижнем уровне работают разнородные базовые модели, на верхнем уровне их объединяет простая Логистическая регрессия (LR) как мета-модель. В нашем случае нижний уровень составляли Метод опорных векторов (SVM), Экстремальный градиентный бустинг (XGBoost), Случайный лес (RF) и Категориальный бустинг (CatBoost). Их задача состояла в том, чтобы предсказать вероятности класса «1» для каждого предложения. Задача верхнего уровня – научиться взвешивать эти вероятности так, чтобы совместное решение было лучше любого одиночного. Сначала для каждого базового алгоритма были получены предсказания на данных, не вошедших в обучающую выборку. Эти оценки агрегировались в матрицу признаков мета-уровня, и поверх неё производилось обучение Логистической регрессии.

I. Результаты экспериментов с классическими моделями

При обучении на TF-IDF представлениях с применением SMOTE для балансировки классов (исходный дисбаланс 70% на 30% в пользу класса «0») **наилучшую производительность среди одиночных классификаторов продемонстрировал метод опорных векторов** (F1=0,65, PR-AUC=0,70), опережая Градиентный бустинг (XGBoost F1=0,62, CatBoost F1=0,61) и существенно превосходя Случайный лес (F1=0,58).

Стекинг всех моделей позволил повысить аккуратность до 0,77 при сохранении F1-меры на уровне лучшего базового классификатора (0,65), что указывает на эффективность ансамблевого подхода для повышения надёжности предсказаний без потери баланса между точностью и полнотой.

J. Эксперимент с трансформером

Следующей моделью для тестирования был выбран RuBERT в реализации DeepPavlov как главный кандидат на сквозное дообучение, поскольку это доступная модель, которая не уступает в качестве более «тяжелым» версиям BERT [30]. Мы использовали публичную контрольную точку RuBERT (ai-forever/RuBERT-base), поддерживаемую на Hugging Face [31].

Было рассмотрено два режима дообучения трансформера:

1) тонкая настройка RuBERT на оригинальном датасете с дисбалансом классов 70% – 30% [32].
2) классическая реализация дообучения RuBERT, но на сбалансированном с помощью аугментации датасете [33].

Разбиение данных проводилось стратифицированно: 20% – тестовая выборка, 80% – обучающая (1226 и 4900 предложений, соответственно).

Использована **стандартная архитектура**: поверх RuBERT добавляется короткая классификационная голова – линейный слой, принимающий эмбеддинг токена и возвращающий два логита для бинарной классификации (класс «0» / класс «1»). Был использован модифицированный класс Trainer с переопределённой функцией потерь, чтобы класс «1» получал больший вес и модель меньше жертвовала полнотой целевого класса.

Контекстуализированные представления RuBERT **обеспечили качественный скачок в решении задачи**: F1-мера составила 0,74 против 0,65 у классических методов, аккуратность – 0,84, площадь под кривой (PR-AUC) – 0,79. Взвешивание классов достаточно эффективно компенсировало дисбаланс без применения синтетической генерации.

Однако полнота (precision) 0,70 указывает на пропуск 30% неопределённых примеров. В связи с этим мы предположили, что **балансировка классов значительно повысит качество классификации**.

Выделение в предложениях локусов

неопределённости (прилагательных) дало возможность для **контекстной аугментации текста**.

Разработанный метод принципиально отличается от синтетического дополнения миноритарного класса (SMOTE). Вместо синтетической генерации мы извлекли из неопределенных предложений подстроки, содержащие локус неопределенности в разных позициях подстроки. Для этого были написаны правила выделения контекстных окон разной длины вокруг первого локуса неопределенности в предложении. Сначала отбирались достаточно длинные предложения миноритарного класса «1», в которых локус неопределённости имеет линейную позицию больше 6. Далее задаются два типа окон для членения: «*правило 5L_10R*»: 5 токенов слева, локус и 10 токенов справа с учётом пунктуации; «правило 2L_to_end», по которому выделялись 2 токена слева, локус и весь контекст до конца предложения. Учёт пунктуации позволил частично сохранять синтаксические группы в подстроках.

Приведём пример полного предложения и выделенных из него подстрок: «В отзыве могут быть указаны номера телефонов, факсов, адреса электронной почты и иные сведения, необходимые для **правильного** и своевременного рассмотрения дела» (полное предложение с локусом неопределённости); «необходимые для **правильного** и своевременного рассмотрения дела» (подстрока 1, правило 5L_10R с обрезкой по пунктуации слева); «сведения, необходимые для **правильного** и своевременного рассмотрения дела» (подстрока 2, правило 2L_to_end).

Указанным образом из одного оригинального предложения выделялось до 4 подстрок.

После применения контекстной аугментации к классу «1» **количество неопределённых примеров увеличилось** с 1961 до 3889.

В результате **модель фиксировала одни и те же маркеры неопределённости в различных контекстных окружениях**, что позволило механизму внимания лучше настроиться на выявление паттернов неопределённости независимо от длины входа.

После предобработки было применено разбиение данных на 6365 примеров в обучающей и 1592 примеров в тестовой выборке. Благодаря аугментации удалось достичь **практически равномерного распределения классов** в обеих выборках: соотношение примеров класса «0» и класса «1» составило приблизительно 51,1% к 48,9%, соответственно.

Для дообучения использовалась та же самая модель RuBERT-base-cased (DeepPavlov) с архитектурой для бинарной классификации. Благодаря достижению баланса классов через аугментацию исчезла необходимость в применении взвешивания классов – модель обучалась на равномерно представленных примерах обеих категорий. Применялась стандартная конфигурация настройки модели [33].

Таблица 1: Метрики качества моделей. Баланс классов в обучающих данных указан в процентах.

	TF-IDF 70/30						**Эмбеддинги**	
	LogReg	**SVM**	**XGBoost**	**CatBoost**	**RF**	**Stacking**	**RuBERT 70/30**	**RuBERT 51/49**
Accuracy	0,71	0,75	0,74	0,74	0,68	0,77	0,84	0,89
Precision	0,54	0,60	0,58	0,59	0,51	0,63	0,78	0,87
Recall	0,68	0,71	0,69	0,68	0,65	0,68	0,70	0,90
F1	0,60	**0,65**	0,62	0,61	0,58	**0,65**	**0,74**	**0,89**
PR-AUC	0,68	0,70	0,66	0,66	0,58	0,68	0,79	0,95

K. Результаты эксперимента с трансформером

Контекстуализированные представления RuBERT **обеспечили качественный скачок в решении задачи** (см. Таблицу 1). Значение F1-меры при обучении на корпусе с дисбалансом классов составило 0,74 против 0,65 у классических методов.

Несмотря на то, что взвешивание классов эффективно компенсировало дисбаланс, после аугментации данных **значение F1-меры возросло с 0,74 до 0,89 на сбалансированном наборе примеров**.

Рис. 1 ниже – **матрица ошибок** обученной на сбалансированном корпусе модели RuBERT.

По вертикали отображен истинный класс, а по горизонтали – предсказанный. Матрица демонстрирует 1 420 верных предсказаний из 1 592 и 172 ошибки: 94 случая, когда пример класса «0» принят за пример класса «1», и 78 – наоборот.

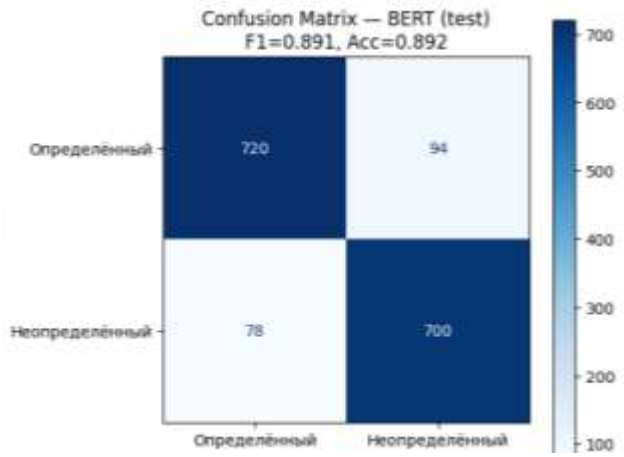

Рисунок 1: Матрица ошибок RuBERT.

Это показывает, что модель склонна незначительно занижать полноту класса «0» и завышать полноту класса «1», сохраняя при этом высокий общий баланс точности и полноты.

IV. Языковые признаки, влияющие на предсказание модели

A. *Способ выделения признаков*

Для интерпретации решений модели RuBERT, обученной на сбалансированном наборе примеров, мы использовали вариант **интегрированных градиентов** — Layer Integrated Gradients, вычисляемый внутри выбранного слоя сети. Метод опирается на работу [34] и реализован через библиотеку Captum.

Атрибуции вычислялись отдельно для каждого из двух классов. Далее атрибуции сворачивались по размерности эмбеддинга до скалярного вклада токена, специальные токены исключались, а значения нормировались по модулю внутри текста с сохранением знака для сопоставимости между примерами. Подтокены объединялись в слова с лемматизацией для русского языка, выполненной с помощью Pymorphy3 [35], формировались униграммы и биграммы, их вес определялся суммой нормализованных вкладов, после чего n-граммы ранжировались по суммарной важности по корпусу отдельно для каждого класса.

Важным для интерпретации лексикона является понимание того факта, что одна и та же n-грамма может одновременно положительно влиять на один логит и отрицательно на другой, или же воздействовать на оба в одном направлении с разной силой. Именно эта особенность требует раздельного вычисления и анализа атрибуций для каждого целевого класса.

Положительная важность n-граммы для класса указывает на то, что её присутствие поддерживает предсказание этого класса. **Отрицательная важность** свидетельствует о противоположном эффекте — n-грамма снижает вероятность выбора класса.

Глобальная важность объединяет силу влияния на предсказание класса и частоту встречаемости. По глобальной важности выбирались наиболее показательные маркеры класса. Высокое значение этого показателя идентифицирует n-граммы, играющие существенную роль на уровне всего набора данных – либо за счёт частого появления с умеренным вкладом, либо благодаря редким, но влиятельным проявлениям.

Локальная важность показывает типичный вклад n-граммы в конкретном предложении.

Указанный метод позволяет выявить как относительно предсказуемые маркеры, так и неожиданные лексические паттерны, специфичные для обучающей выборки или отражающие скрытые корреляции.

B. *Перечень признаков и их первичная интерпретация*

Мы получили 4 списка значимых для модели n-грамм: 76 униграмм для класса «1», 83 униграммы для класса «0», 74 биграммы для класса «1», 67 биграмм для класса «0», см. приложение к настоящей статье по ссылке (https://clck.ru/3Pw55p). Модель разделила лексиконы: пересечения между классами в наборах полнозначных слов практически отсутствуют.

Выяснилось, что среди маркеров класса «1» максимальные значения индекса глобальной важности (здесь и далее мы приводим именно их) получили прилагательные «*уважительный*» (142,10), «*разумный*» (110,79), «*достаточный*» (35,21), «*значительный*» (24,78). Заметим, что среди них два прилагательных, которые юристы называют «эмоционально-оценочными» и одно параметрическое прилагательное.

Отдельный интерес представляют **пары лексических антонимов**, один из которых образован с помощью отрицательной приставки. Вкратце обсудим их распределение между классами.

Пара «*своевременный*» (класс «1», 7,48) и «*несвоевременный*» (класс «0», 8,99) демонстрирует асимметрию: только прилагательное «*несвоевременный*» описывает нарушение некоторых установленных сроков. Члены пары «*уважительный*» (класс «1», 142,10) – «*неуважительный*» (класс «1», 6,91) оказались в одном классе, хотя и с разными весами. В паре «*достоверный*» (класс «0», 3,44) и «*недостоверный*» (класс «0», 18.73) оба прилагательных являются маркерами одного класса, но также обладают разными весами.

В целом концентрация прилагательных с приставкой *не-* в классе «0» (с отсутствием неопределённости) представляет собой не вполне ожидаемую нами закономерность, выявленную моделью, ср.: «*ненадлежащий*» (20,44), «*недостоверный*» (18,73), «*неправомерный*» (16,02), «*неправильный*» (5,24), «*непригодный*» (3,53), «*недоброкачественный*» (2,67).

Более содержательную интерпретацию приведённым наблюдениям может дать анализ значений отдельных прилагательных в терминах шкал с применением аналитического аппарата степенной семантики. Такой анализ является одним из направлений развития нашего исследования.

Обсуждая биграммы, отметим лишь наличие сочетаний с наречием «*заведомо*», ср. «*заведомо ложный*» (класс «1», 19,85) – при том, что это наречие попало в список униграмм – маркеров класса «0» (15,91). Это наречие типично для юридического лексикона; оно означает, что некоторый субъект сознательно осуществил какое-то действие, заранее осознавая его неправомерность. Мы ожидали, что наречие «*заведомо*» будет предсказывать наличие неопределённости.

Лемма «*тяжкий*» попала в список маркеров в двух сочетаниях: в классе «0» это «*тяжкий преступление*» (4,54), в классе «1» это «*тяжкий последствие*» (8,43). Этот результат ожидаем, см. [8]. По-видимому, добавление информации о синтаксических группах и о частотных в юридических текстах именных коллокациях, размеченных по параметру неопределённости, позволит нам улучшить качество предсказаний.

V. Заключение

В настоящей статье представлен опыт создания инструмента классификации русских предложений по параметру наличия/отсутствия правовой неопределённости. Задача создания такого классификатора, как показывают авторы, сложна по нескольким причинам.

Во-первых, для получения обучающего набора данных необходима **двуступенчатая разметка**, выполняемая сначала лингвистами, затем юристами. При этом оценивать согласие между лингвистами и юристами бессмысленно, поскольку первые выделяют случаи лингвистической неопределённости, вторые – правовой неопределённости (а это лишь частично пересекающиеся множества). Кроме того, лингвисты при анализе примеров опираются на языковой контекст, юристы – на знание законов и ссылки, присутствующие в положениях законов.

Во-вторых, одни и те же языковые выражения (слова) могут вводить или не вводить неопределённость, что **не позволяет задать список неопределённых терминов**, но заставляет думать о способах учёта информации о синтаксисе и сочетаемости.

В статье авторы использовали размеченный датасет, состоящий из **6000 предложений с бинарными метками и локусами неопределённости** и предложили сравнительный анализ шести семейств моделей машинного обучения (от классических методов до контекстуализированных трансформерных архитектур). Результаты **подтвердили превосходство предобученных языковых моделей**: RuBERT достиг F1-меры 0,89 на сбалансированном корпусе. Для балансировки классов применялась аугментация за счёт выделения контекстных окон из оригинальных предложений класса «1».

В число перспектив исследования входят:

1) рассмотрение значений градуируемых прилагательных, являющихся маркерами классов, с применением аналитического аппарата степенной семантики;
2) привлечение юристов к интерпретации признаков, влияющих на предсказания модели;
3) расширение набора обучающих данных.

Благодарности

Мы благодарим Ольгу Александровну Митрофанову за ценные советы и обсуждение, повлиявшие на ход описанного в статье исследования. Все ошибки и недочёты остаются на совести авторов.

Разделы II, III, IV подготовлены при поддержке СПбГУ, шифр проекта 123042000068-8. Введение к статье подготовлено в рамках гранта РНФ, проект #22-18-00189 «Структура и функционирование устойчивых неоднословных единиц русской повседневной речи».

Библиография

[1] F. Devos, “Semantic vagueness and lexical polyvalence,” Studia Linguistica, vol. 57, no. 3, 2003, pp. 121–141. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0039-3193.2003.00101.x

[2] F. Devos, “Still fuzzy after all these years: a linguistic evaluation of the fuzzy set approach to semantic vagueness,” *Quaderni di Semantica*, vol. 16, no. 1, pp. 47–82, 1995.

[3] S. Ramotowska, J. Haaf, L. Van Maanen, and J. Szymanik, “Most quantifiers have many meanings,” *Psychonomic Bulletin and Review*, vol. 31, no. 6, pp. 2692–2703, Dec. 2024. [Online]. Available: https://link.springer.com/article/10.3758/s13423-024-02502-7

[4] C. Kennedy, “Ambiguity and vagueness: An overview,” in *Semantics – Lexical Structures and Adjectives*, C. Maienborn, K. von Heusinger, and P. Portner, Eds. Berlin, Boston: De Gruyter Mouton, 2019, pp. 236–271, Available: 10.1515/9783110626391-008.

[5] O. Blinova and S. Belov, “Linguistic ambiguity and vagueness in Russian legal texts,” *Vestnik of Saint Petersburg University. Law*, vol. 11, no. 4, pp. 774–812, Jan. 2020, doi: 10.21638/spbu14.2020.401.

[6] C. Kennedy, “Vagueness and grammar: the semantics of relative and absolute gradable adjectives,” *Linguistics and Philosophy*, vol. 30, no. 1, pp. 1–45, Mar. 2007. [Online]. Available: https://link.springer.com/article/10.1007/s10988-006-9008-0

[7] Г. И. Кустова, “Прилагательное,” *Материалы для проекта корпусного описания русской грамматики*. [Online]. Available: http://rusgram.ru/Прилагательное#111

[8] O. Blinova and A. Berlin, “Creating a Dataset for Automatic Detection of Vague Expressions in Russian Legal Texts,” in *Internet and Modern Society*, vol. 2671, Communications in Computer and Information Science. Cham: Springer Nature Switzerland, 2026, pp. 177–195, Available: 10.1007/978-3-032-04958-2_14.

[9] U. May, K. Zaczynska, J. Moreno-Schneider, and G. Rehm, “Extraction and Normalization of Vague Time Expressions in German,” in *Proc. 17th Conf. Natural Language Processing (KONVENS 2021)*, Düsseldorf, Germany, Sep. 2021, pp. 114–126. [Online]. Available: https://aclanthology.org/2021.konvens-1.10/

[10] B. D. Cruz, B. Jayaraman, A. Dwarakanath, and C. McMillan, "Detecting Vague Words & Phrases in Requirements Documents in a Multilingual Environment," in *2017 IEEE 25th Int. Requirements Engineering Conf. (RE)*, 2017, pp. 233–242, Available: 10.1109/RE.2017.24.

[11] С. В. Чеповецкая, “*Языковая неопределённость в устной речи российских чиновников: корпусное исследование*,” Выпускная квалификационная работа бакалавра филологии, НИУ ВШЭ, Санкт-Петербург, 2025.

[12] P.-H. Paris, S. E. Aoud, and F. M. Suchanek, “The Vagueness of Vagueness in Noun Phrases,” in *Conference on Automated Knowledge Base Construction*, 2021. doi: 10.24432/C5T884.

[13] A. Debnath and M. Roth, “A Computational Analysis of Vagueness in Revisions of Instructional Texts,” arXiv:2309.12107, Sep. 2023. [Online]. Available: http://arxiv.org/abs/2309.12107

[14] G. Malik, S. Yildirim, M. Cevik, and A. Bener, “An Empirical Study on Vagueness Detection in Privacy Policy Texts,” in *Proc. Canadian Conf. Artificial Intelligence*, 2023, Available: 10.21428/594757db.2728303d.

[15] J. Bhatia, T. D. Breaux, J. R. Reidenberg, and T. B. Norton, “A Theory of Vagueness and Privacy Risk Perception,” in *2016 IEEE 24th International Requirements Engineering Conference (RE)*, 2016, pp. 26–35. doi: 10.1109/RE.2016.20.

[16] P. Alexopoulos and J. Pavlopoulos, “A Vague Sense Classifier for Detecting Vague Definitions in Ontologies,” in *Proc. 14th Conf. European Chapter of the Association for Computational Linguistics*, vol. 2: Short Papers, Gothenburg, Sweden, Apr. 2014, pp. 33–37. [Online]. Available: https://aclanthology.org/E14-4007/

[17] D. I. Kulagin, “Publicly available sentiment dictionary for the Russian language KartaSlovSent,” in *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog”* [Komp’yuternaia Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoj Konferentsii “Dialog”], 2021, pp. 1106–1119.

[18] L. Lebanoff and F. Liu, “Automatic Detection of Vague Words and Sentences in Privacy Policies,” in *Proc. 2018 Conf. Empirical Methods in Natural Language Processing*, Brussels, Belgium, Oct.-Nov. 2018, pp. 3508–3517. [Online]. Available: https://aclanthology.org/D18-1387/

[19] S. Wang and C. D. Manning, “Baselines and Bigrams: Simple, Good Sentiment and Topic Classification,” in *Proc. 50th Annual Meeting of the Association for Computational Linguistics*, vol. 2: Short Papers, Jeju Island, Korea, 2012, pp. 90–94. [Online]. Available: https://aclanthology.org/P12-2018/

[20] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, pp. 273–297, 1995. [Online]. Available: https://link.springer.com/article/10.1007/BF00994018

[21] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, Available: 10.1023/A:1010933404324.

[22] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794, Available: 10.1145/2939672.2939785.

[23] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: unbiased boosting with categorical features,” in *Advances in Neural Information Processing Systems* 31, S. Bengio et al., Eds. Curran Associates, Inc., 2018, pp. 6638–6648. [Online]. Available: https://papers.nips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

[24] D. H. Wolpert, “Stacked generalization,” *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, Available: 10.1016/S0893-6080(05)80023-1.

[25] Y. Kuratov and M. Arkhipov, “Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language,” arXiv:1905.07213, 2019. [Online]. Available: https://arxiv.org/abs/1905.07213

[26] G. Lemaitre, F. Nogueira, and C. K. Aridas, “imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning,” *J. Mach. Learn. Res.*, vol. 18, no. 17, pp. 1–5, 2017.

[27] T. Kiss and J. Strunk, “Unsupervised Multilingual Sentence Boundary Detection,” *Computational Linguistics*, vol. 32, no. 4, pp. 485–525, 2006, Available: 10.1162/coli.2006.32.4.485.

[28] S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. Harshman, “Indexing by Latent Semantic Analysis,” *J. American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990, Available: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, Available: 10.1613/jair.953.

[30] T. Shavrina, A. Fenogenova, A. Emelyanov, et al., “RussianSuperGLUE: A Russian Language Understanding Evaluation Benchmark,” in *Proc. 2020 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Online, Nov. 2020, pp. 4717–4726. [Online]. Available: https://aclanthology.org/2020.emnlp-main.381/

[31] T. Wolf, L. Debut, V. Sanh, et al., “Transformers: State-of-the-Art Natural Language Processing,” in Proc. 2020 Conf. Empirical Methods in Natural Language Processing: System Demonstrations, Online, Nov. 2020, pp. 38–45. [Online]. Available: https://aclanthology.org/2020.emnlp-demos.6/

[32] H. He and E. A. Garcia, “Learning from Imbalanced Data,” *IEEE Trans. Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, Available: 10.1109/TKDE.2008.239.

[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1 (Long and Short Papers), Minneapolis, Minnesota, 2019, pp. 4171–4186, Available: 10.18653/v1/N19-1423.

[34] M. Sundararajan, A. Taly, and Q. Yan, “Axiomatic Attribution for Deep Networks,” in *Proc. 34th Int. Conf. Machine Learning*, vol. 70, Sydney, Australia, 2017, pp. 3319–3328. [Online]. Available: https://arxiv.org/abs/1703.01365.

[35] M. Korobov, “Morphological Analyzer and Generator for Russian and Ukrainian Languages,” in *Analysis of Images, Social Networks and Texts*, vol. 542, D. I. Ignatov et al., Eds. Cham: Springer International Publishing, 2015, pp. 320–332, Available: 10.1007/978-3-319-26123-2_31.

Automatic Detection of Adjectival Vagueness in Russian Legal Texts: Dataset, Models, and Results

Alena E. Berlin, Olga V. Blinova

***Abstract* — This paper addresses the automatic classification of Russian sentences from legal documents (laws) into those with and without legal vagueness. A training dataset of 6,000 annotated sentences with vagueness loci was developed through collaboration between linguists and legal experts. The study focuses exclusively on vagueness introduced by gradable adjectives. We evaluate both classical machine learning models and transformer-based architectures. Data augmentation applied to RuBERT successfully resolves class imbalance, achieving an F1-score of 0.89. Analysis of linguistic features reveals, that adjectives with the negative prefix "ne-" predominantly occur in sentences without vagueness.**

***Keywords* — linguistic vagueness, legal vagueness, Russian legal text, adjectival vagueness, binary classifiers, RuBERT.**

REFERENCES

[1] F. Devos, "Semantic vagueness and lexical polyvalence," *Studia Linguistica*, vol. 57, no. 3, 2003, pp. 121–141. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0039-3193.2003.00101.x

[2] F. Devos, "Still fuzzy after all these years: a linguistic evaluation of the fuzzy set approach to semantic vagueness," *Quaderni di Semantica*, vol. 16, no. 1, pp. 47–82, 1995.

[3] S. Ramotowska, J. Haaf, L. Van Maanen, and J. Szymanik, "Most quantifiers have many meanings," Psychonomic Bulletin and Review, vol. 31, no. 6, pp. 2692–2703, Dec. 2024. [Online]. Available: https://link.springer.com/article/10.3758/s13423-024-02502-7

[4] C. Kennedy, "Ambiguity and vagueness: An overview," in Semantics - Lexical Structures and Adjectives, C. Maienborn, K. von Heusinger, and P. Portner, Eds. Berlin, Boston: De Gruyter Mouton, 2019, pp. 236–271, Available: 10.1515/9783110626391-008.

[5] O. Blinova and S. Belov, "Linguistic ambiguity and vagueness in Russian legal texts," *Vestnik of Saint Petersburg University. Law*, vol. 11, no. 4, pp. 774–812, Jan. 2020, doi: 10.21638/spbu14.2020.401.

[6] C. Kennedy, "Vagueness and grammar: the semantics of relative and absolute gradable adjectives," *Linguistics and Philosophy*, vol. 30, no. 1, pp. 1–45, Mar. 2007. [Online]. Available: https://link.springer.com/article/10.1007/s10988-006-9008-0

[7] G. I. Kustova, "Adjective," *Materials for the project of corpus description of Russian grammar*. [Online]. Available: http://rusgram.ru/Прилагательное#111

[8] O. Blinova and A. Berlin, "Creating a Dataset for Automatic Detection of Vague Expressions in Russian Legal Texts," in *Internet and Modern Society*, vol. 2671, Communications in Computer and Information Science. Cham: Springer Nature Switzerland, 2026, pp. 177–195, Available: 10.1007/978-3-032-04958-2_14.

[9] U. May, K. Zaczynska, J. Moreno-Schneider, and G. Rehm, "Extraction and Normalization of Vague Time Expressions in German," in *Proc. 17th Conf. Natural Language Processing (KONVENS 2021)*, Düsseldorf, Germany, Sep. 2021, pp. 114–126. [Online]. Available: https://aclanthology.org/2021.konvens-1.10/

[10] B. D. Cruz, B. Jayaraman, A. Dwarakanath, and C. McMillan, "Detecting Vague Words & Phrases in Requirements Documents in a Multilingual Environment," in *2017 IEEE 25th Int. Requirements Engineering Conf. (RE)*, 2017, pp. 233–242, Available: 10.1109/RE.2017.24.

[11] S. V. Chepovetskaya, "*Linguistic vagueness in oral speech of Russian officials: A corpus study*," Bachelor's thesis, Philology Program, National Research University Higher School of Economics, St. Petersburg, Russia, 2025.

[12] P.-H. Paris, S. E. Aoud, and F. M. Suchanek, "The Vagueness of Vagueness in Noun Phrases," in *Conference on Automated Knowledge Base Construction*, 2021. doi: 10.24432/C5T884.

[13] A. Debnath and M. Roth, "A Computational Analysis of Vagueness in Revisions of Instructional Texts," arXiv: 2309.12107, Sep. 2023. [Online]. Available: http://arxiv.org/abs/2309.12107

[14] G. Malik, S. Yildirim, M. Cevik, and A. Bener, "An Empirical Study on Vagueness Detection in Privacy Policy Texts," in *Proc. Canadian Conf. Artificial Intelligence*, 2023, Available: 10.21428/594757db.2728303d.

[15] J. Bhatia, T. D. Breaux, J. R. Reidenberg, and T. B. Norton, "A Theory of Vagueness and Privacy Risk Perception," in *2016 IEEE 24th International Requirements Engineering Conference (RE)*, 2016, pp. 26–35. doi: 10.1109/RE.2016.20.

[16] P. Alexopoulos and J. Pavlopoulos, "A Vague Sense Classifier for Detecting Vague Definitions in Ontologies," in *Proc. 14th Conf. European Chapter of the Association for Computational Linguistics*, vol. 2: Short Papers, Gothenburg, Sweden, Apr. 2014, pp. 33–37. [Online]. Available: https://aclanthology.org/E14-4007/

[17] D. I. Kulagin, "Publicly available sentiment dictionary for the Russian language KartaSlovSent," in *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog"* [Komp'yuternaia Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoj Konferentsii "Dialog"], 2021, pp. 1106–1119.

[18] L. Lebanoff and F. Liu, "Automatic Detection of Vague Words and Sentences in Privacy Policies," in *Proc. 2018 Conf. Empirical Methods in Natural Language Processing*, Brussels, Belgium, Oct.-Nov. 2018, pp. 3508–3517. [Online]. Available: https://aclanthology.org/D18-1387/

[19] S. Wang and C. D. Manning, "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification," in *Proc. 50th Annual Meeting of the Association for Computational Linguistics*, vol. 2: Short Papers, Jeju Island, Korea, 2012, pp. 90–94. [Online]. Available: https://aclanthology.org/P12-2018/

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995. [Online]. Available: https://link.springer.com/article/10.1007/BF00994018

[21] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, Available: 10.1023/A:1010933404324.

[22] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794, Available: 10.1145/2939672.2939785.

[23] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems* 31, S. Bengio et al., Eds. Curran Associates, Inc., 2018, pp. 6638–6648. [Online]. Available: https://papers.nips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

[24] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, Available: 10.1016/S0893-6080(05)80023-1.

[25] Y. Kuratov and M. Arkhipov, "Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language," arXiv:1905.07213, 2019. [Online]. Available: https://arxiv.org/abs/1905.07213

[26] G. Lemaitre, F. Nogueira, and C. K. Aridas, "imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in

Machine Learning," *J. Mach. Learn. Res.*, vol. 18, no. 17, pp. 1–5, 2017.

[27] T. Kiss and J. Strunk, "Unsupervised Multilingual Sentence Boundary Detection," *Computational Linguistics*, vol. 32, no. 4, pp. 485–525, 2006, Available: 10.1162/coli.2006.32.4.485.

[28] S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. Harshman, "Indexing by Latent Semantic Analysis," *J. American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990, Available: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, Available: 10.1613/jair.953.

[30] T. Shavrina, A. Fenogenova, A. Emelyanov, et al., "RussianSuperGLUE: A Russian Language Understanding Evaluation Benchmark," in *Proc. 2020 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Online, Nov. 2020, pp. 4717–4726. [Online]. Available: https://aclanthology.org/2020.emnlp-main.381/

[31] T. Wolf, L. Debut, V. Sanh, et al., "Transformers: State-of-the-Art Natural Language Processing," in *Proc. 2020 Conf. Empirical Methods in Natural Language Processing: System Demonstrations*, Online, Nov. 2020, pp. 38–45. [Online]. Available: https://aclanthology.org/2020.emnlp-demos.6/

[32] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, Available: 10.1109/TKDE.2008.239.

[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1 (Long and Short Papers), Minneapolis, Minnesota, 2019, pp. 4171–4186, Available: 10.18653/v1/N19-1423.

[34] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," in *Proc. 34th Int. Conf. Machine Learning*, vol. 70, Sydney, Australia, 2017, pp. 3319–3328. [Online]. Available: https://arxiv.org/abs/1703.01365.

[35] M. Korobov, "Morphological Analyzer and Generator for Russian and Ukrainian Languages," in *Analysis of Images, Social Networks and Texts*, vol. 542, D. I. Ignatov et al., Eds. Cham: Springer International Publishing, 2015, pp. 320–332, Available: 10.1007/978-3-319-26123-2_31.