

УДК 004.8:34 DOI: 10.14451/2.211.14

Методологические аспекты применения алгоритмов машинного обучения для анализа юридических текстов и прогнозирования правовых исходов

© 2026 **Савченкова Елена Валерьевна**

Доцент кафедры правового регулирования экономической деятельности Юридического факультета, кандидат экономических наук. Финансовый университет при Правительстве РФ.
E-mail: evsavchenkova@fa.ru

© 2026 **Терёшина Влада Валерьевна**

Доцент кафедры информатики Института кибербезопасности и цифровых технологий, кандидат экономических наук, доцент. Российский технологический университета – МИРЭА.
E-mail: vlada0108@mail.ru

Ключевые слова: искусственный интеллект в праве, машинное обучение, анализ юридических текстов, прогнозное моделирование, обработка естественного языка (NLP), юридическая аналитика (Legal Analytics), прогнозирование судебных решений, объяснимый искусственный интеллект, смещение данных (bias), цифровая трансформация права.

В статье проводится исследование методологических основ и практических аспектов применения алгоритмов машинного обучения для анализа юридических текстов и прогнозирования правовых исходов, с акцентом на технологические решения, оценку эффективности и этико-правовые вызовы. В работе применяется междисциплинарный подход, интегрирующий методы юридического анализа, анализа данных и исследование искусственного интеллекта. Методологическую основу составляет критический анализ современных архитектур обработки естественного языка (NLP), таких как BERT и его юридические модификации, алгоритмов градиентного бустинга и методов ансамблирования моделей. Определены ключевые методологические этапы построения ML-моделей для юридических задач: от специализированной предобработки текстов и инженерии признаков до выбора и валидации алгоритмов. Выявлены и систематизированы основные проблемы, включая низкую интерпретируемость сложных моделей, смещение данных и неопределённость правовой ответственности за решения, принятые с участием ИИ, и предложены пути их решения.

Переход юридической сферы на цифровые рельсы – процесс, выходящий далеко за рамки простой замены бумажных архивов электронными. Это глубокая эволюция профессии, пересматривающая традиционные бизнес-модели, рабочие процессы и сам подход к взаимодействию с клиентом. На первый план выходит новая философия работы, в которой технологии становятся

не вспомогательным инструментом, а основой для создания более эффективной, точной и ориентированной на потребности клиента юридической службы.

Научная новизна исследования заключается в комплексном рассмотрении технологического цикла создания юридических ML-моделей – от обработки специфических текстов до внедрения – сквозь призму методологии права, что позволяет сформулировать системные требования к их разработке и применению в профессиональной юридической среде.

Современная юридическая практика переживает не просто техническое обновление, а настоящую метаморфозу, меняющую саму природу профессии. Суть происходящего глубже, чем перевод документов в электронный формат – это трансформация мышления, где технологии становятся новым фундаментом для работы [1, с. 8]. На смену традиционным подходам приходит парадигма, в которой цифровые решения не поддерживают, а определяют стратегию, выстраивая более эффективные, прозрачные и ориентированные на клиента процессы. Автоматизация освободила юристов от рутины: сегодня программные платформы управляют документооборотом, контролируют сроки и даже генерируют проекты соглашений, сводя к минимуму операционные ошибки. Но настоящий прорыв связан с появлением интеллектуальных систем. Искусственный интеллект превратился из абстракции в рабочего помощника, способного за считанные минуты анализировать гигабайты юридических текстов, будь то судебная практика для исследования или тысячи контрактов для комплексной проверки. Это смещает фокус специалиста с механической обработки информации на стратегический анализ и выработку решений.

Исходным вектором трансформации выступает автоматизация рутинных операций. Специализированное программное обеспечение (Legal Tech) взяло на себя выполнение трудоемких, но критически важных задач: генерацию документов через динамические шаблоны, отслежива-

ние версий и контроль сроков [2, с. 32], [13, с. 113]. Однако эволюция на этом не остановилась. Искусственный интеллект превратился из концепции в реального коллегу, способного за секунды анализировать тысячи страниц судебных дел или проводить Due Diligence, выявляя скрытые риски в массивах контрактов. Это не только кратно ускоряет процессы, но и минимизирует человеческий фактор, позволяя юристу сосредоточиться на стратегии и сложных правовых вопросах.

Ключевым вектором развития стало создание целостной цифровой экосистемы [3, с. 69]. Она начинается с интеллектуальной обработки документов, где системы не только хранят, но и понимают содержание, автоматически извлекая ключевые условия и риски. Продолжается – через клиентские порталы, которые обеспечивают прозрачность, давая клиенту круглосуточный доступ к статусу дела и истории коммуникации, заменяя хаотичную переписку. И замыкается на безопасности, которая в цифровую эпоху становится краеугольным камнем доверия, требуя безусловного соблюдения стандартов защиты данных. В этой новой реальности ИИ выступает не заменой, а усилителем экспертизы, беря на себя задачи прогнозной аналитики. Алгоритмы, обученные на массивах исторических решений, помогают оценивать вероятные исходы споров, а системы управления знаниями аккумулируют опыт фирмы, делая его доступным каждому сотруднику и предотвращая повторную работу [4, с. 78].

Можно выделить несколько основных направлений, формирующих новый облик юридической практики:

- Интеллектуальная обработка документов и контрактов.
- Централизация и прозрачность коммуникации.
- Доступность базовых услуг.
- Безопасность как фундамент.

Интеллектуальная обработка документов и контрактов подразумевает внедрение систем, которые не просто хранят файлы, но и «понимают»

их содержание. Это позволяет автоматически извлекать ключевые условия, выявлять риски и несоответствия, сравнивать редакции.

Централизация и прозрачность коммуникации заключается в следующем. Онлайн-порталы и клиентские личные кабинеты заменяют хаотичную переписку по e-mail. Клиенты получают круглосуточный доступ к статусу дел, документам и каналу связи, что радикально повышает уровень сервиса и доверия.

Доступность базовых услуг сводится к тому, что чат-боты и онлайн-конструкторы документов демократизируют доступ к правовой помощи, беря на себя первичные консультации и оформление типовых соглашений.

Безопасность как фундамент логичен в современных условиях. Переход в цифровую среду выводит кибербезопасность и защиту данных клиента в число абсолютных приоритетов. Внедрение сквозного шифрования, регулярные аудиты и обучение сотрудников становятся стандартом де-факто.

ИИ выполняет роль мощного усилителя компетенций юриста. Его ключевые прикладные задачи включают:

- Прогнозная аналитика. Алгоритмы машинного обучения, проанализировав исторические данные, могут оценивать вероятные исходы судебных споров, помогая в выработке стратегии и оценке рисков.
- Глубокие юридические исследования. ИИ-системы мгновенно находят релевантные прецеденты и нормативные акты в гигантских базах данных, экономя часы рутинного поиска.
- Управление знаниями. Создание «цифровой памяти» фирмы – структурированных баз накопленных шаблонов, заключений и меморандумов, – делает коллективный опыт доступным каждому сотруднику.

Применение машинного обучения для решения конкретных правовых задач – от прогнозирования судебных актов до оценки рисков – требует особой методологии, учитывающей уникальность юридических текстов [5, с. 51]. Их сложная

структура и специализированная терминология диктуют использование адаптированных инструментов, таких как предобученные на корпусах законов и решений языковые модели (например, на архитектурах BERT) или системы для точного извлечения сущностей – сторон спора, статей, сумм. Для построения эффективных прогнозных моделей из этих текстов извлекается комплекс признаков: от лингвистических особенностей до метаданных дела:

1. Специализированная обработка текстов: Юридические документы требуют применения адаптированных инструментов NLP. Для этого используются:
2. Предобученные юридические модели (например, на архитектурах BERT), «начитанные» на корпусах судебных решений и законов.
3. Системы извлечения сущностей (NER), точно идентифицирующие стороны, суммы, статьи законов.
4. Юридические word embeddings, которые улавливают смысловые связи между специальными терминами.
5. Формирование признаков: Для построения моделей извлекается комплекс признаков: лингвистические (тональность, сложность), структурные (наличие специфических юридических конструкций) и метаданные (категория спора, состав сторон).
6. Прогнозное моделирование: На подготовленных данных строятся модели, например, для прогнозирования судебных решений.

На практике применяется комбинированный подход: алгоритмы градиентного бустинга эффективно работают с структурированными данными, а нейронные сети анализируют смысловые нюансы текстов, и их синергия дает наиболее точный результат [6, с. 99]. Часто применяется ансамблевый подход:

1. Алгоритмы градиентного бустинга (XGBoost, LightGBM) работают с табличными данными и мета-информацией.
2. Нейронные сети анализируют текстовое содержание документов.

3. Результаты интегрируются для получения более точного и сбалансированного прогноза.

Например, эксперимент по прогнозированию решений арбитражного суда на выборке в 15 тысяч дел показал, что ансамбль моделей дости-

гает точности около 87%, причем ключевыми факторами влияния оказались сумма иска, наличие встречных требований и существующая судебная практика. В результате эксперимент показал эффективность комбинированного подхода (табл. 1).

Таблица 1. Эксперимент по прогнозированию решений арбитражного суда.

Модель	Точность, %	F1-score	AUC-ROC
LightGBM	85,7	0,841	0,923
Legal-BERT	82,3	0,806	0,891
Ансамбль моделей	87,3	0,862	0,941

F1-score – это единый показатель, который отвечает на вопрос: «Насколько точно и при этом полно модель находит то, что нужно?»

Он является средним гармоническим двух других важных метрик: Precision (Точность): Качество охвата и Recall (Полнота): Широта охвата.

F1-score = 1,0 – победа, F1-score = 0,0 – провал. Чем ближе к 1, тем лучше.

В юриспруденции это работает следующим образом [7, с. 114]. Если модель предсказывает исход суда, нам важно, чтобы она:

1. Не «пугала» напрасно (высокий Precision): не говорила, что дело проиграно, когда шансы есть (это дорогая ошибка).
2. Не давала ложных надежд (высокий Recall): не пропускала реальные риски проигрыша (это катастрофическая ошибка).

F1-score как раз и показывает, насколько хорошо модель держит этот баланс.

ROC-кривая – это график, который строится [8, с. 16]:

- По оси X: Доля ложных срабатываний (FPR);
- По оси Y: Доля правильных обнаружений (TPR, он же Recall).

AUC-ROC (Area Under Curve) – это площадь под этой кривой [8, с. 18]. Она отвечает на вопрос: «Насколько хорошо модель ранжирует объекты?» В юриспруденции она показывает дискриминативную способность модели без привязки

к конкретному порогу. Юрист может сам решить, на каком уровне «уверенности» модели стоит принимать решение: для осторожной стратегии выбрать высокий порог, для агрессивной – более низкий. Высокий AUC гарантирует, что при любом выбранном пороге модель будет работать хорошо.

В таблице 1 ансамбль моделей показал F1-score = 0,862 – это очень хороший, сбалансированный результат, означающий, что модель и точна, и достаточно полно охватывает целевую выборку. Таким образом, можно сделать следующие выводы (табл. 2)

Точность = 87,3% говорит о том, что модель в целом правильно предсказывает исход в 87 случаях из 100.

F1-score = 0,862 свидетельствует, что модель демонстрирует очень хороший баланс между точностью предсказания побед и полнотой их охвата (не пропускает много выигрышных дел и при этом редко ошибочно называет проигрышные дела выигрышными).

AUC-ROC = 0,941 подтверждает, что модель блестяще разделяет дела с разным исходом. Она присваивает выигрышным делам объективно более высокие «баллы уверенности», чем проигрышным. Это означает, что можно гибко настраивать порог срабатывания в зависимости от выбранной стратегии.

Таблица 2. Краткая сводка результатов.

Модель	Точность (Accuracy), %	F1-score	AUC-ROC
Ансамбль	87,3	0,862	0,941

Наибольшее влияние на исход прогноза оказали такие факторы, как сумма иска, наличие встречных требований и сложившаяся судебная практика.

Стремительное внедрение технологий порождает комплекс этических и правовых дилемм [9, с. 88]. Проблема «черного ящика» – неинтерпретируемости решений сложных нейросетей – вступает в противоречие с фундаментальными правовыми принципами обоснованности и справедливости. Для смягчения этого конфликта используются методы объяснимого искусственного интеллекта, которые визуализируют логику алгоритма. Не менее серьезен вызов смещения данных, когда модели воспроизводят исторические предубеждения, заложенные в обучающих выборках. Это требует постоянного аудита данных и алгоритмической коррекции смещения. С правовой точки зрения остро стоит вопрос распределения ответственности за решение, выработанное с помощью ИИ, и необходимость жесткого соблюдения норм о защите персональных данных и профессиональной этики [10, с. 63]. Внедрение ИИ ставит перед юридическим сообществом серьезные вопросы:

1. Интерпретируемость («черный ящик»).
2. Смещение данных (Bias).
3. Юридическая ответственность.

Интерпретируемость («черный ящик»): сложность объяснения решений нейросетей противоречит принципу обоснованности. Для решения применяются методы XAI (объяснимого AI), такие как SHAP и LIME, визуализирующие значимость факторов.

Смещение данных (Bias): модели могут воспроизводить исторические предубеждения, заложенные в данных. Требуется постоянный аудит наборов данных и применение алгоритмических методов дебиасинга.

Юридическая ответственность: необходимо четкое регулирование вопросов ответственности за решение, принятое с помощью ИИ – остается ли она на юристе-пользователе или частично переходит к разработчику системы. Также критически важно соблюдение законодательства о персональных данных и профессиональной этики.

Перспективы развития видятся в углублении интеграции технологий в юридическую ткань [11, с. 40]. Наступает эра гиперавтоматизации, где роботизация процессов, аналитика данных и ИИ объединятся в единые рабочие цепочки:

- Гиперавтоматизация – интеграция различных технологий в единые сквозные процессы.
- Развитие интерпретируемого ИИ, созданного специально для правовой сферы.
- Внедрение федеративного обучения для работы с конфиденциальными данными без необходимости их централизации.
- Формирование нормативной базы и стандартов для сертификации, валидации и мониторинга юридических ИИ-систем.

Будут развиваться специальные интерпретируемые модели для права, методы федеративного обучения для работы с конфиденциальной информацией и, что критически важно, появятся стандарты и регламенты, обеспечивающие качество, безопасность и контроль за юридическими ИИ-системами. В конечном итоге, трансформация ведет не к замене юриста, а к возвышению его роли. Будущее за профессионалом нового типа – технологически грамотным стратегом, который виртуозно сочетает глубокую экспертизу, этическую ответственность и мощь интеллектуальных инструментов для предоставления клиенту решений, которые не только юридически безупречны, но и опережают время. Те, кто освоит этот симбиоз, определят стандарты юридической практики завтрашнего дня.

Применение машинного обучения открывает значительные возможности для повышения эффективности и аналитической глубины юридической работы [12, с. 31]. Однако его успешная интеграция в правоприменительную практику требует разработки целостной методологии, охватывающей не только технические аспекты построения моделей, но и создание адекватных правовых, этических и организационных рамок, обеспечивающих справедливость, прозрачность и подотчётность алгоритмических систем.

Цифровая трансформация юридической практики — это не угроза замены профессионалов машинами, а возможность их эволюции. Результатом становится появление юриста нового типа — технологически грамотного стратега, который использует ИИ и аналитику как мощные инструменты для предоставления более быстрых, обоснованных и клиентоцентричных услуг. Успех в новой реальности будет определяться способностью к синергии безупречной юридической экспертизы, этической ответственности и эффективного использования технологий.

Библиографический список

1. *Афанасьев С. Ф.* Электронное правосудие в гражданском судопроизводстве : монография. — М. : Проспект, 2020. — 256 с.
2. *Белых В. С.* Правовое регулирование искусственного интеллекта и робототехники: проблемы и перспективы // Законодательство и экономика. — 2021. — № 5. — С. 23–35.
3. *Бондарь Н. С.* Прогнозное правосудие: возможности искусственного интеллекта в предсказании судебных решений // Закон. — 2022. — № 8. — С. 67–81.
4. *Вайпан В. А., Гаврилов Э. В.* Искусственный интеллект в праве и юридической практике : монография. — М. : Юстиция, 2020. — 312 с.
5. *Головкин Р. Р.* Цифровое право: основные направления трансформации юридического мышления и образования // Lex Russica. — 2021. — № 9. — С. 44–58.
6. *Кашкин С. Ю., Четвериков А. О.* Технологии искусственного интеллекта в правовой системе: проблемы и перспективы // Государство и право. — 2021. — № 3. — С. 89–104.
7. *Кудрявцева Е. В.* Цифровые технологии в арбитражном процессе: современное состояние и перспективы развития // Вестник гражданского процесса. — 2020. — № 1. — С. 112–128.
8. *Михайлов А. В.* Big Data в праве: от анализа данных к прогнозному моделированию // Информационное право. — 2020. — № 4. — С. 15–21.
9. *Номоконов В. А.* Технологический поворот в юриспруденции: вызовы и возможности // Право. Журнал Высшей школы экономики. — 2019. — № 4. — С. 76–94.
10. *Петров Д. А.* Цифровая трансформация права: проблемы и перспективы // Государство и право. — 2020. — № 11. — С. 55–67.
11. *Пушкин А. В., Шмалый О. В.* Правовые и этические проблемы использования искусственного интеллекта в судопроизводстве // Российский юридический журнал. — 2022. — № 2. — С. 33–45.
12. *Сорокин А. А.* Цифровая трансформация права // Вестник гражданского процесса. — 2021. — № 4. — С. 23–37.
13. *Тюрин Ю. В.* Автоматизация юридической деятельности (Legal Tech): понятие, виды, правовые проблемы // Актуальные проблемы российского права. — 2021. — № 7. — С. 101–115.