

Научная статья  
УДК 004.891  
doi: 10.17586/2713-1874-2025-1-57-68

## ПРИМЕНЕНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ ЮРИДИЧЕСКИХ ЭКСПЕРТИЗ

*Максим Валерьевич Улизко<sup>1</sup>, Александра Сергеевна Ватьян<sup>2</sup>,  
Наталья Федоровна Гусарова<sup>3</sup>, Наталья Викторовна Добренко<sup>4</sup>*

<sup>1,2,3,4</sup>Университет ИТМО, Санкт-Петербург, Россия  
<sup>1</sup>contrey1337@gmail.com<sup>✉</sup>, <https://orcid.org/0009-0001-2374-8025>  
<sup>2</sup>asvatian@itmo.ru, <https://orcid.org/0000-0002-5483-716X>  
<sup>3</sup>nfgusarova@itmo.ru, <http://orcid.org/0000-0002-1361-6037>  
<sup>4</sup>graziokisa@yandex.ru, <https://orcid.org/0000-0001-6206-8033>  
Язык статьи – русский

**Аннотация:** В статье рассматривается применение больших языковых моделей (LLM) для анализа юридических кейсов. Исследуются их возможности и ограничения при обработке нормативно-правовых документов, а также влияние немонотонной и деонтической логики на корректность выводов. Проведено тестирование моделей GPT-4o, GigaChat и YandexGPT с целью оценки точности обработки правовой информации. Выявлены основные проблемы, включая галлюцинации, некорректные ссылки на законодательство и сложность интерпретации логических парадоксов. Для повышения качества предложены меры по адаптации моделей: использование системного промпта, RAG-хранилища нормативных документов и оптимизация параметров генерации текста. Результаты исследования показали, что доработанная модель позволяет получать более точные юридические консультации и значительно ускоряет процесс анализа правовых вопросов.

**Ключевые слова:** большие языковые модели, БЯМ, деонтическая логика, немонотонная логика, юриспруденция, чат-бот, GPT, Llama

**Ссылка для цитирования:** Улизко М. В., Ватьян А. С., Гусарова Н. Ф., Добренко Н. В. Применение больших языковых моделей для юридических экспертиз // Экономика. Право. Инновации. 2025. № 1. С. 57–68. <http://dx.doi.org/10.17586/2713-1874-2025-1-57-68>.

## APPLICATION OF LARGE LANGUAGE MODELS FOR LEGAL EXPERTISE

*Maxim V. Ulizko<sup>1</sup>, Aleksandra S. Vatian<sup>2</sup>, Natalia F. Gusarova<sup>3</sup>, Natalya V. Dobrenko<sup>4</sup>*

<sup>1,2,3,4</sup>ITMO University, Saint Petersburg, Russia  
<sup>1</sup>contrey1337@gmail.com<sup>✉</sup>, <https://orcid.org/0009-0001-2374-8025>  
<sup>2</sup>asvatian@itmo.ru, <https://orcid.org/0000-0002-5483-716X>  
<sup>3</sup>nfgusarova@itmo.ru, <http://orcid.org/0000-0002-1361-6037>  
<sup>4</sup>graziokisa@yandex.ru, <https://orcid.org/0000-0001-6206-8033>  
Article in Russian

**Abstract:** The article examines the application of Large Language Models (LLMs) in legal case analysis. Their capabilities and limitations in processing legal documents are analyzed, along with the impact of non-monotonic and deontic logic on inference accuracy. GPT-4o, GigaChat, and YandexGPT were tested to assess their precision in handling legal information. Key issues were identified, including hallucinations, incorrect legal references, and difficulties in interpreting logical paradoxes. To enhance performance, several adaptation measures were proposed, including the use of a system prompt, a RAG repository of legal documents, and optimization of text generation parameters. The results demonstrate that the improved model provides more accurate legal insights and significantly speeds up the process of legal analysis.

**Keywords:** deontic logic, GPT, Large Language Models, Legal Case Analysis, Llama non-monotonic logic

**For citation:** Ulizko M. V., Vatian A. S., Gusarova N. F., Dobrenko N. V. Application of Large Language Models for Legal Expertise. *Ekonomika. Pravo. Innovacii*. 2025. No. 1. pp. 57–68. (In Russ.). <http://dx.doi.org/10.17586/2713-1874-2025-1-57-68>.

**Введение.** Внедрение программных решений, основанных на больших языковых моделях (БЯМ), для оптимизации и ускорения работы рабочих и научных предприятий сегодня имеет большой потенциал. БЯМ успешно справляются с большинством рутинных задач, связанных с обработкой естественного языка. В том числе, широкое применение технология находит в правовых задачах. Сегодня БЯМ умеют анализировать договоры, юридические кейсы, соглашения и другие нормативные документы [1–3].

Однако существует ряд проблем, связанных с обработкой таких данных. Во-первых, БЯМ путаются в правовых системах. Специалисты, пытаясь применить модели в реалиях российских законов, сталкиваются со значительным количеством галлюцинаций и неточностей. Под галлюцинациями понимается выдумывание моделью несуществующей информации или искажение текущей. Во-вторых, актуален вопрос обработки логических парадоксов немонотонной и деонтической логики, которые присущи всем правовым системам.

Немонотонная логика – это форма логики, в которой отношения между умозаключениями не подчиняются монотонности. Иначе, немонотонные логики создаются для фиксации и представления опровергающих выводов. Это означает, что в таких умозаключениях может возникнуть предварительный вывод, который впоследствии можно отменить на основании новых, дополнительно полученных данных [4]. Деонтическая логика, или логика норм, является разделом модальной логики, которая изучает структуру и связи нормативных высказываний. Существует множество систем деонтической логики, различающихся символическими средствами, однако они исследуют аналогичные нормы с одинаковой структурой: характер (обязывает, разрешает или запрещает), содержание (действие, которое должно быть выполнено), условия применения и субъект (лицо или группа лиц, для которых предназначена норма) [5].

Исследование больших языковых моделей в контексте немонотонной логики в литературе не представлено.

В связи с этим в статье авторы ставят следующие задачи.

1) Исследовать актуальные большие языковые модели на предмет обработки юридических кейсов, в том числе, работу с деонтической логикой и логическими парадоксами.

2) Разработать программное решение, которое позволит адаптировать БЯМ под рамки российского законодательства, учитывая ее особенности.

**Литературный обзор.** Применение БЯМ в качестве помощников и ассистентов для решения юридических кейсов находит широкое применение на сегодняшний день. В литературе представлен ряд примеров.

В [6] авторы статьи представляют бенчмарк (единый стандарт оценки) для правовых решений в БЯМ. Авторы применяют структуру исследования, которое состоит из вопроса, правила, применения и заключения. Работа акцентирует внимание на потенциале моделей с точки зрения вычислительного инструмента для юриспруденции. Однако исследование не имеет прикладного характера и не рассматривает вопрос логических парадоксов.

В статье [7] авторы представляют сравнение между автоматизированными решениями и традиционным подходом – ручной работой младших ассистентов. Авторы исследуют, могут ли БЯМ превзойти людей по точности, скорости и экономической эффективности при проверке правовых документов. Результатом исследования стал вывод, что БЯМ почти не уступают точности реальных экспертов и снижают затраты на 99,97%.

В работе [8] рассматривается использование БЯМ для предоставления юридических консультаций. Авторы провели консультации с двадцатью юридическими экспертами и рассмотрели конкретные запросы, анализируя технические и юридические ограничения, что позволило сформировать набор рекомендаций для разработчиков моделей. Эксперты рекомендовали помогать пользователям формулировать правильные вопросы и находить релевантную информацию, а не давать окончательные юридические решения.

В исследовании [9] рассматривается случай судопроизводства в Великобритании. Авторы анализируют два подхода: традиционный метод на основе обработки естественного языка с использованием ключевых слов и логических операторов и применение

модели Claude 2, использующей контент-ориентированные подсказки. На основе анализа 356,011 судебных решений (Кембриджский корпус права) модель Claude 2 достигла F1-метрики 0.94 против 0.78 у метода ключевых слов.

В работе [10] предлагается система LawLuo – многоагентная модель для многократных китайских юридических консультаций. Она включает четыре агента: агент-рецепционист, оценивающий намерения пользователя и выбирающий юриста; агент-юрист, взаимодействующий с пользователем; агент-секретарь, который организует записи разговоров и генерирует отчеты, и агент-босс, оценивающий работу юриста и секретаря для достижения оптимальных результатов. Взаимодействие агентов имитирует работу реальных юридических фирм.

Исследование парадоксов немонотонной логики в БЯМ представлено лишь одной работой. В статье [11] для решения этой проблемы авторы оценивают способность БЯМ к логическому рассуждению на 25 различных схемах рассуждений, включая немонотонную логику. Авторы представляют фреймворк Logic Bench – набор данных для оценки на основе вопросов и ответов. Результаты исследования показали, что существующие модели недостаточно хорошо справляются с Logic Bench. В особенности они испытывают трудности с комплексными рассуждениями и отрицаниями, а также иногда отдают предпочтение параметрическим знаниям перед

контекстной информацией, упуская правильную цепочку рассуждений.

Таким образом, в ходе анализа было выявлено, что на сегодняшний день, несмотря на широкое применение БЯМ в качестве помощников по работе с юридическими вопросами, почти не представлены работы, которые ставят цель анализировать принципы принятия решений большой языковой моделью в контексте юридических вопросов логических парадоксов.

**Материалы и методы исследования. Анализ актуальных БЯМ и деонтической логики.** В рамках исследования моделей необходимо протестировать их на устойчивость к деонтическим парадоксам. Деонтический парадокс представляет собой истинное, на первый взгляд, утверждение, которое приводит к противоречию или интуитивно непостижимой ситуации [12]. Несмотря на это, в рамках деонтической логики такие утверждения считаются логически верными.

В данной работе рассмотрен ряд парадоксов деонтической логики. Для каждого из этих парадоксов были отобраны реальные юридические кейсы сфере разработки программного обеспечения. В таблице 1 представлены соответствующие примеры. Использование деонтической логики SDL (Standard Deontic Logic) предполагает применение следующих операторов при формулировании утверждений: *P* – разрешение, *O* – обязательство, *F* – запрет, “->” – следование, “-” оператор дизъюнкции.

Таблица 1

**Данные для тестирования**

Источник: составлено авторами

Парадокс	Формула (SDL)	Ситуация	Проблема	Решение юриста
Chrisholm paradox	$O(p)$ $O(p \rightarrow q)$ $\neg p \rightarrow O(\neg q)$ $\neg p$	Компания обязана обеспечить безопасность данных пользователей. Компания обязана либо обеспечить безопасность данных, либо временно удалить их.	Парадокс возникает из-за того, что последнее утверждение кажется некорректным с точки зрения здравого смысла: намерение было обязать соблюдать безопасности данных, а не дать альтернативу их уничтожение.	Компания обязана произвести минимизацию ущерба и понести соразмерное наказание по итоговому ущербу

Парадокс	Формула (SDL)	Ситуация	Проблема	Решение юриста
<p>The Gentle Murderer Paradox</p>	<p><math>O(\neg p)</math>  <math>p \rightarrow O(q)</math>  <math>p</math></p>	<p>Компания X (Исполнитель), специализирующаяся на разработке программного обеспечения, заключает договор с корпорацией Y (Заказчик) на разработку системы для автоматизации обработки платежей. Сумма проекта: 10 миллионов рублей. Условие в договоре:          «Общая ответственность Исполнителя по настоящему договору ограничивается суммой, равной общей стоимости полученной оплаты за выполненные работы. Исполнитель не несет ответственности за косвенные убытки, упущенную выгоду или вред, причиненный из-за использования программного обеспечения».</p> <p>Развитие ситуации: Система оплаты была разработана и внедрена. Через 2 месяца Заказчик обнаружил серьезный баг в ПО, из-за которого произошли ошибки в обработке платежей, что привело к финансовым потерям на сумму 25 млн руб. X обратилась к Y с требованием компенсировать убытки. Однако X ссылается на ограничительное условие договора, утверждая, что их</p>	<p>Суть парадокса заключается в утверждении, что, если совершение противоправного действия неизбежно, оно должно быть совершено максимально «хорошо», что звучит противоречиво с точки зрения морали. То есть, хотя нарушение конфиденциальности всё еще считается недопустимым, в случае его неизбежности минимизация ущерба становится профессиональным и юридическим «долгом» компании.</p>	<p>Компании должны провести переговоры и в соответствии с возможной минимизацией ущерба выделить ответственность обеих сторон.</p>

Парадокс	Формула (SDL)	Ситуация	Проблема	Решение юриста
		максимальная ответственность составляет 10 миллионов рублей (уплаченная сумма).		
The Good Samaritan Paradox	$O(p \wedge q)$ $O(q)$	Модуль системы, созданный для анализа случаев мошенничества, оправдывает своё существование только при наличии данных о реальных случаях мошенничества. Если мошенничество станет редким или будет искоренено, существование системы станет менее полезным. Это может вызвать непреднамеренный интерес со стороны разработчиков к поддержанию статистического уровня мошенничества для сохранения значимости и финансирования системы.	Поддерживая развитие инструмента, мы невольно "оправдываем" существование угроз, с которыми он борется.	Намеренное допущение багов в системе разработчиками является строгим нарушением и развитие продукта таким образом оправдать нельзя.
Ross Paradox	$O(p)$ $O(p \vee q)$	Система управления пользователями требует от администратора удаления пробелов в данных (например, удаление лишних символов из строк). Изначальное требование: «Администратор обязан удалить все пробелы из строки». Логическое следствие: «Администратор обязан удалить все пробелы или не предпринимать никаких действий».	Если система допускает меньшее действие, вместо строгого выполнения требований, это может привести к сбоям в обработке данных и нарушению бизнес-логики.	Решение должно быть принято на основании должностной инструкции и дополнительных данных о ценностях каждого этапа бизнес-процесса. Если экспертами будет оценено, что оценено, что требование из пункта 1 имеет критический приоритет, то логическое следствие 3 не верно.

Для тестирования был выбран ряд наиболее распространенных и применяемых БЯМ, среди которых GigaChat MAX от Сбера [13], Yandex GPT 3 от Яндекса [14], флагманская большая языковая модель GPT-4o от OpenAI, далее – GPT-4o [15]. Модель для тестирования работала в исходном виде, без применения дополнительных

RAG-данных, системных промтов и файн-тюнинга. Для тестирования был сформулирован промт:

«Выступай в роли юридического консультанта. В компании произошла ситуация X. Опираясь на следующие законы: Y принятые в компании, какое решение должен принять юрист?».

Таблица 2

**Результаты тестирования**  
 Источник: составлено авторами

Модель	№ кейса (таблица 1)	Комментарий эксперта
GPT-4o	1	Совпадает с мнением юристов, но не ссылается на документы
	2	“Если баг был случайным и не связан с грубой небрежностью <math>\diamond</math> или нарушением условий договора, то юрист должен сосредоточиться на защите интересов Исполнителя” – модель ошибается, считая, что баг может быть “случайным”. На практике такое решение однозначно принято не будет.
	3	Модель дает нечеткий ответ, ссылаясь на то, что необходим анализ всего процесса. Конкретных решений не предлагает.
	4	Модель дает решение схожее с юристом, но в большей части уходит в рассуждение про документирование процессов, однако запрос к модели имел несколько другой смысл.
GigaChat	1	Модель частично повторяет мнение юриста, ссылается на корректный закон (152 ФЗ), но большую часть ответа посвящает рекомендациям по отслеживанию дальнейших утечек.
	2	Модель частично повторяет мнение юриста, ссылается на статьи ГК РФ, верно сослался в 2 из 4 случаев, отрицательные можно отнести к галлюцинациям.
	3	Модель дает слишком общий ответ, неверно ссылается на статьи ГК и УК РФ.
	4	Модель делает заключение, что администратор прав в этой ситуации, т. е. модель некорректно обрабатывает логический парадокс.
YandexGPT	1	Модель дает общий ответ без конкретики и не ссылается на законы РФ.
	2	Модель дает ответ, в котором просит принять сторону компании Y. Конкретных аргументов в эту пользу не дает. Примечание: в этом случае модель дает ограничение в 1000 символов для промта, поэтому он был сокращен на 113 символов.
	3	Модель дает общий ответ без конкретики и не ссылается на законы РФ.
	4	Модель дает общий ответ без конкретики и не ссылается на законы РФ,

Исходя из результатов в таблице 2, можно сделать вывод, что на данный момент ни одна модель в исходном виде не дает удовлетворительное качество для работы с правовыми вопросами. При этом лучший результат показала GPT-4o. Из основных недостатков можно отметить отсутствие ссылок на нормативно-правовые российские документы. Худший результат показала Yandex GPT, которая на большинство вопросов давала слишком общий ответ. Для дальнейшей адаптации выбрана модель от Open AI.

**Доработка модели.** Согласно результатам проведенного тестирования, качество, которое выдают исходные модели, является недостаточным. За основу для доработки принято решение использовать модель GPT-4o версии за январь 2025 года. Был предложен следующий состав доработок для исходной большой языковой модели.

1. Разработка системного промпта для модели. Под системным промптом подразумевается текстовое указание для настройки и направления работы модели. Как правило, он задает контекст, рамки, цели или контекст. В качестве структуры промпта будет использована рекомендованная президентом OpenAI Греггом Брокманом структура [16]:

цель: для чего используется чат-бот, возвращаемый формат: требования к форматированию и содержанию итогового ответа, предупреждения: описывают ограничения, с которыми должна работать модель, контекст: дополнительный блок информации, который содержит описание предметной области.

2. Настройка параметров модели была выполнена в соответствии с недостатками, обнаруженными в первой части исследования, результаты которого представлены в таблице 3.

Таблица 3

**Параметры**

*Источник: составлено авторами*

Параметр	Описание	Значение	Целевое значение	Комментарий
max_tokens	Ограничивает количество токенов, генерируемых моделью.	от 1 до бесконечности	1024	По результатам тестирования, 1024 токенов достаточно для полного ответа. При увеличении параметра возникает риск галлюцинаций.
temperature	Определяет, насколько «творческой» будет модель. Больше значение – креативнее ответы.	от 0 до 2	0.2	Так как модель разрабатывается для юристов, необходимо обеспечить максимальную точность ответа и исключить творчество.
top_p	Ограничивает разнообразие возможных токенов в последовательности по процентам. Больше значение – более вариативный ответ.	от 0 до 1	0.9	Необходимо обеспечить умеренное разнообразие в выборе лексики.
top_k	Ограничивает разнообразие возможных токенов в последовательности по количеству.	от 1 до бесконечности	50	Необходимо обеспечить умеренное разнообразие в выборе лексики.

Параметр	Описание	Значение	Целевое значение	Комментарий
	Больше значение – более вариативный ответ.			
frequency penalty	Сокращает количество повторяющихся токенов, делая ответ более структурным.	от -2 до 2	0.1	Повторяющиеся токены уместны для юридических текстов, поэтому берется среднее значение по этому параметру.
presence penalty	Принудительная генерация новых токенов.	от -2 до 2	0.1	Повторяющиеся токены уместны для юридических текстов, поэтому берется среднее значение по этому параметру.

RAG – метод, используемый для улучшения работы больших языковых моделей для специфического контента [17]. RAG состоит из двух компонентов: Retrieval – отвечает за поиск релевантной информации из внешних баз данных или информационных хранилищ; Generation – компонент формирует информацию согласно полученным данным. RAG-хранилище для доработанной модели состоит из нормативно-правовых документов, как общих (УК, ГК РФ), так и специфичных (устав компании, должностные инструкции). На рисунке 1 представлена архитектура модели.

Юрист (целевой пользователь) взаимодействует с интерфейсной частью приложения (чат-бот) и вводит свой промпт. Запрос уходит в компонент большой языковой модели, которая обогащена параметрами и системным промптом.

На первом шаге модель определяет, какие документы из RAG-хранилища будут релевантны для запроса пользователя. Модель собирает необходимые документы и возвращает их пользователю для валидации. Юрист подтверждает, что действительно хочет работать с документами которая предложила модель и возвращает ответ. Большая языковая модель обрабатывает промпт с соответс-

твующими документами и формирует финальный ответ.

**Результаты исследования.** Для оценки эффективности предложенного решения используется сравнение текстов исходной модели и предложенного решения по метрике перплексии, экспертная оценка разработанного инструмента и оценка его экономической эффективности.

1) Перплексия (Perplexity) – распространенная метрика для оценки производительности языковых моделей [18]. Она измеряет, насколько хорошо модель предсказывает последовательность текста. Меньшее значение перплексии указывает на лучшую производительность модели. Для токенизированной последовательности  $X = (x_0, x_1, \dots, x_t)$  перплексия ( $PPL$ ) от  $X$  равна:

$$PPL(X) = \exp \left\{ -\frac{1}{t} \sum_i^t \log p\theta(x_i | x_{<i}) \right\}, \quad (1)$$

где  $\log p\theta(x_i | x_{<i})$  – логарифмическая вероятность  $i$ -го токена, обусловленная предыдущим токеном  $x_{<i}$  в соответствии с моделью.

В таблице 4 представлено сравнение метрики  $PPL$  для исходной и доработанной модели.



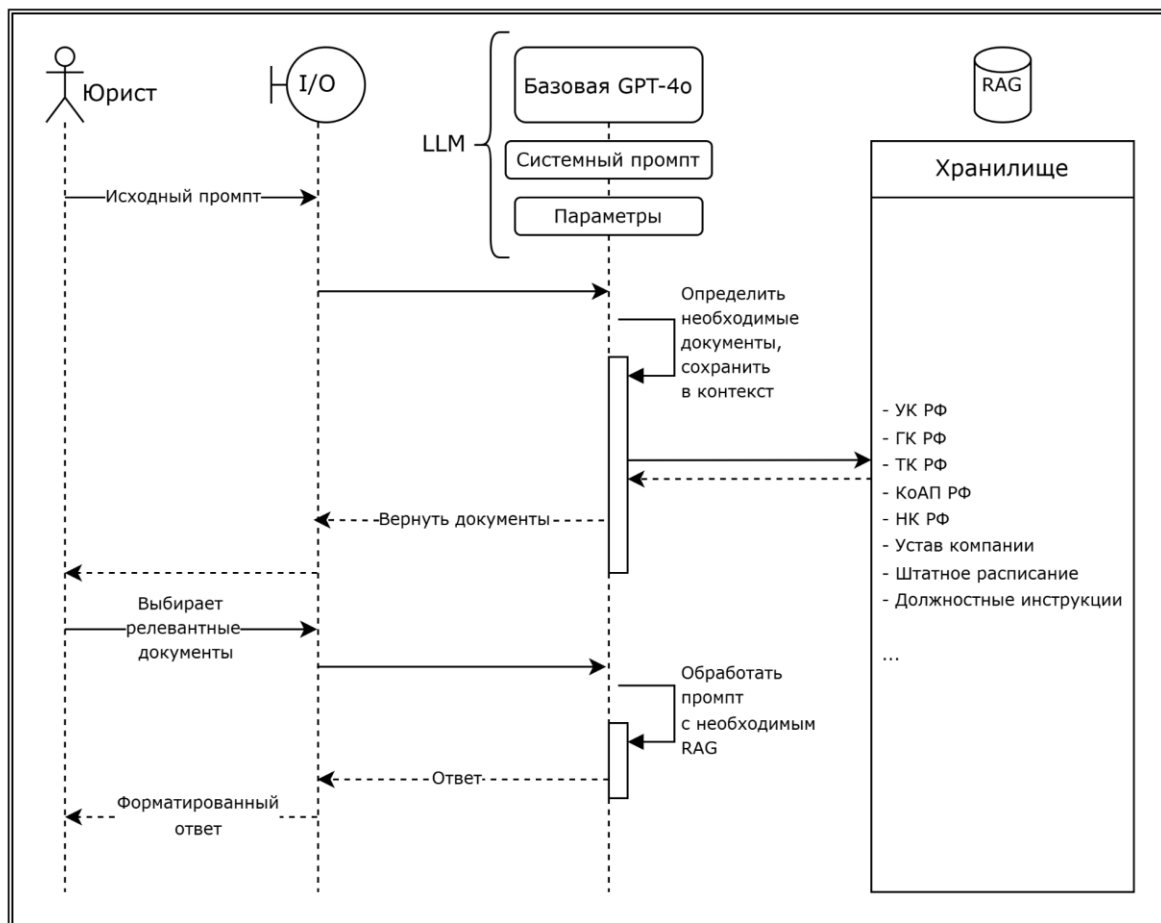


Рисунок 1 – Архитектура доработанной модели

Источник: составлено авторами

Таблица 4

**Оценка перплексии (меньше – лучше)**

Источник: составлено авторами

№ кейса	GPT-4o	GPT-4o доработанная
1	4.45	2.33
2	5.31	1.18
3	8.11	2.43
4	4.00	1.22

2) Экспертная оценка (Human evaluation, HE) проводилась согласно методике [19]. В качестве экспертов выступали три юриста предприятия, которые являются профессионалами в отрасли и в целевом процессе являются конечными пользователями решения. Пяти экспертам (старшим юристам) компании было предложено оценить ответ модели по параметрам: релевантность (R), качество речи (Q) и согласованность (C) по шкале от 0

до 10 баллов (0 – неудовлетворительно, 10 – ответ превосходит ожидания).

Экспертная оценка, представленная в таблице 5, для кейса 1–4 от эксперта вычисляется по формуле:

$$HE(n) = \left( \frac{R + Q + C}{3} * 100 \right), \quad (2)$$

где R – релевантность, Q – качество речи, C – согласованность.

3) Экономическая эффективность. Для обоснования внедрения нового инструмента на предприятии потребовалось провести оценку его экономической эффективности. В рамках этой оценки решено было измерить время, необходимое для первичного разбора и оценки юридического кейса (кейсы 1–4). В тестовую группу вошли

старшие и младшие юристы, которым было предложено воспользоваться тремя разными методами: исходным GPT-4o, доработанным GPT-4o и традиционным методом без использования больших языковых моделей. Время выполнения задания фиксировалось в минутах. Результат представлен в таблице 6.

Таблица 5

**Экспертная оценка (больше – лучше)**

*Источник: составлено авторами*

№ кейса	GPT-4o	GPT-4o доработанная	Комментарий эксперта (доработанная модель)
1	90%	100%	Полностью совпадает с мнением юристов, корректно ссылается на документы.
2	56,7%	94%	Совпадает с мнением юристов, есть не критические неточности со ссылками на устав компании.
3	40%	94%	Совпадает с мнением юристов, есть некритические неточности, связанные со структурой ответа.
4	34,5%	88%	Совпадает с мнением юристов, указан неверный пункт в ссылке на должностную инструкцию.

Таблица 6

**Временные затраты**

*Источник: составлено авторами*

Роль	Среднее время на документ (минуты)
Старший юрист	23,5
Младший юрист	34
Старший юрист и GPT-4o исходный	15
Младший юрист и GPT-4o исходный	24
Старший юрист и GPT-4o доработанный	4,5
Младший юрист и GPT-4o доработанный	6

4) Авторы статьи [20] предлагают метрику близости готового документа к качественному юридическому заявлению. Из предложенных авторами метрик были выбраны актуальные для исследования позиции представленные в таблице 7, 8.

**Выводы.** В ходе исследования было показано, что большие языковые модели успешно решают задачи обработки

естественного языка, включая анализ договоров, юридических кейсов и нормативной документации, но сталкиваются с рядом проблем, которые затрудняют их применение в контексте российского законодательства. В исходном виде модели обладают галлюцинациями и низким уровнем контекстного понимания, а также сталкиваются с парадоксами деонтической логики.

Таблица 7

**Метрика качества юридических текстов***Источник: составлено авторами на основе [20]*

Метрика	Формула	Описание
Полнота обязательных реквизитов заявления (R)	$R_i / R$ (0, 1]	$R_i$ – количество обязательных реквизитов в заявлении $R$ – количество обязательных реквизитов заявления
Ссылки на законодательство (L)	$L_i / L$ (0, 1]	$L_i$ – количество правильно использованных ссылок на законодательство $L$ – общее количество необходимых ссылок на законодательство
Правильность использования терминов (T)	$T_i / T$ (0, 1]	$T_i$ – количество неправильно использованных терминов $T$ – общее количество терминов

Таблица 8

**Результаты оценки качества юридических текстов***Источник: составлено авторами на основе [20]*

№ кейса	R	L	T
1	1,00	1,00	1,00
2	0,75	0,75	0,75
3	1,00	0,5	1,00
4	1,00	0,75	1,00

Авторами статьи предложена усовершенствованная архитектура модели, которая позволила разработать чат-бот-помощник для юристов. Это позволило избежать ошибок при работе с немонотонной логикой и значительно уменьшило количество галлюцинаций. Для метрики перплексии качество ответа выросло в среднем в 4 раза, а экспертная группа оценила доработанную модель на 38% лучше, чем исходную. Представлено

доказательство экономической эффективности: использование решения позволяет экономить от 20 до 25 минут в зависимости от класса специалиста.

В качестве дальнейших исследований планируется разработка программного решения, которое позволит в реальном времени адаптироваться специфике различных правовых систем, включая локальные законодательные законы и судебную практику.

**Список источников**

1. Wang X., Zhang X., Hoo V., Shao Z., Zhang X. LegalReasoner: A Multi-Stage Framework for Legal Judgment Prediction via Large Language Models and Knowledge Integration // *IEEE Access*. 2024. Т. 12. (In Eng.). DOI: 10.1109/ACCESS.2024.3496666.
2. Gutiérrez J. D. Critical Appraisal of Large Language Models in Judicial Decision-Making // *Elgaronline*. 2024. (In Eng.). DOI: 10.4337/9781803922171.00033.
3. Nelson J. The Other ‘LLM’: Large Language Models and the Future of Legal Education // *European Journal of Legal Education*. 2024. Т. 5. № 1. С. 127–155. (In Eng.).
4. Antonelli G. A., Strasser C. Non-Monotonic Logic // *Stanford Encyclopedia of Philosophy* [Электронный ресурс]. – Режим доступа: <https://plato.stanford.edu/entries/logic-nonmonotonic/> (In Eng.).
5. Wright G. H. von. Norm and Action: A Logical Inquiry. – Routledge & Kegan Paul, 1963. – 214 С. (In Eng.).

**References**

1. Wang X., Zhang X., Hoo V., Shao Z., Zhang X. LegalReasoner: A Multi-Stage Framework for Legal Judgment Prediction via Large Language Models and Knowledge Integration. *IEEE Access*. 2024. Vol. 12. DOI: 10.1109/ACCESS.2024.3496666.
2. Gutiérrez J. D. Critical Appraisal of Large Language Models in Judicial Decision-Making. *Elgaronline*. 2024. DOI: 10.4337/9781803922171.00033.
3. Nelson J. The Other ‘LLM’: Large Language Models and the Future of Legal Education. *European Journal of Legal Education*. 2024. Vol. 5. No. 1. pp. 127–155.
4. Antonelli G. A., Strasser C. Non-Monotonic Logic. *Stanford Encyclopedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/logic-nonmonotonic/>
5. Wright G. H. von. Norm and Action: A Logical Inquiry. *Routledge & Kegan Paul* 1963. 214 P. (In Eng.).

6. Guha N., Nyarko J., Ho D., Ré C., Chilton A., Chohlas-Wood A., Li Z. LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2308.11462> (In Eng.).
7. Martin L., Whitehouse N., Yiu S., Catterson L., Perera R. Better Call GPT: Comparing Large Language Models Against Lawyers [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2401.16212> (In Eng.).
8. Cheong I., Xia K., Feng K. K., Chen Q. Z., Zhang A. X. I Am Not a Lawyer, But...: Engaging Legal Experts Towards Responsible LLM Policies for Legal Advice [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2402.01864> (In Eng.).
9. Izzidien A., Sargeant H., Steffek F. LLM vs. Lawyers: Identifying a Subset of Summary Judgments in a Large UK Case Law Dataset [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2403.04791> (In Eng.).
10. Sun J., Dai C., Luo Z., Chang Y., Li Y. LawLuo: A Chinese Law Firm Co-Run by LLM Agents [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2407.16252v1> (In Eng.).
11. Parmar M., et al. Towards Systematic Evaluation of Logical Reasoning Ability of Large Language Models [Электронный ресурс]. – Режим доступа: <https://arxiv.org/html/2404.15522v1> (In Eng.).
12. Von Wright G. H. Deontic Logic // *Mind*. Т. 60. № 237. С. 1–15. (In Eng.).
13. Русскоязычная нейросеть от Сбера // GigaChat [Электронный ресурс]. – Режим доступа: <https://giga.chat/>
14. БЯМ от Яндекса // YandexGPT 3 [Электронный ресурс]. – Режим доступа: <https://ya.ru/ai/gpt-3>
15. GPT-4o // OpenAI [Электронный ресурс]. – Режим доступа: <https://openai.com/index/gpt-4/>
16. Recommendations for Creating System Prompts for LLMs [Электронный ресурс]. – Режим доступа: <https://x.com/gdb/status/1878489681702310392?mx=2> (In Eng.).
17. Non-Monotonic Logic // *Stanford Encyclopedia of Philosophy* [Электронный ресурс]. – Режим доступа: <https://plato.stanford.edu/entries/logic-nonmonotonic/> (In Eng.).
18. Retrieval-Augmented Generation (RAG) Method // Hugging Face [Электронный ресурс]. – Режим доступа: <https://huggingface.co/docs/transformers/perplexity> (In Eng.).
19. Li H., Dong Q., Chen J., Su H., Zhou Y., Ai Q., Liu Y. LLMs-as-Judges: A Comprehensive Survey on LLM-Based Evaluation Methods [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2412.05579> (In Eng.).
20. Юрьев Р. Н., Бессмертный И. А., Груздева А. С. Метрики качества юридических текстов // *Экономика. Право. Инновации*. 2024. № 2. С. 81–93.
6. Guha N., Nyarko J., Ho D., Ré C., Chilton A., Chohlas-Wood A., Li Z. LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models. Available at: <https://arxiv.org/pdf/2308.11462>
7. Martin L., Whitehouse N., Yiu S., Catterson L., Perera R. Better Call GPT: Comparing Large Language Models Against Lawyers. Available at: <https://arxiv.org/pdf/2401.16212>
8. Cheong I., Xia K., Feng K. K., Chen Q. Z., Zhang A. X. I Am Not a Lawyer, But...: Engaging Legal Experts Towards Responsible LLM Policies for Legal Advice. Available at: <https://arxiv.org/pdf/2402.01864>
9. Izzidien A., Sargeant H., Steffek F. LLM vs. Lawyers: Identifying a Subset of Summary Judgments in a Large UK Case Law Dataset. Available at: <https://arxiv.org/pdf/2403.04791>
10. Sun J., Dai C., Luo Z., Chang Y., Li Y. LawLuo: A Chinese Law Firm Co-Run by LLM Agents. Available at: <https://arxiv.org/pdf/2407.16252v1>
11. Parmar M., et al. Towards Systematic Evaluation of Logical Reasoning Ability of Large Language Models. Available at: <https://arxiv.org/html/2404.15522v1>
12. Von Wright G. H. Deontic Logic. *Mind*. Vol. 60. No. 237. pp. 1–15.
13. Russian-Language Neural Network from Sber. *GigaChat*. Available at: <https://giga.chat/> (In Russ.).
14. BYAM from Yandex. *YandexGPT 3*. Available at: <https://ya.ru/ai/gpt-3> (In Russ.).
15. GPT-4o. *OpenAI*. Available at: <https://openai.com/index/gpt-4/>
16. Recommendations for Creating System Prompts for LLMs. Available at: <https://x.com/gdb/status/1878489681702310392?mx=2>
17. Non-Monotonic Logic. *Stanford Encyclopedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/logic-nonmonotonic/>
18. Retrieval-Augmented Generation (RAG) Method. *Hugging Face*. Available at: <https://huggingface.co/docs/transformers/perplexity>
19. Li H., Dong Q., Chen J., Su H., Zhou Y., Ai Q., Liu Y. LLMs-as-Judges: A Comprehensive Survey on LLM-Based Evaluation Methods. Available at: <https://arxiv.org/pdf/2412.05579>
20. Iurev R. N., Bessmertny I. A., Gruzdeva A. S. Quality Metrics of Legal Texts. *Ekonomika. Pravo. Innovacii*. 2024. No. 2. pp. 81–93. (In Russ.).