

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В АНАЛИЗЕ ДОГОВОРОВ: ГРАНИЦЫ ФОРМАЛИЗАЦИИ ЮРИДИЧЕСКОГО СМЫСЛА

*Асанди М.Г., студент группы ИУ6-22Б
Московский государственный технический университет им. Н.Э. Баумана*

*Научный руководитель: Шалдунова Т.Н., кандидат исторических наук,
доцент кафедры «Информационная аналитика и политические технологии»*

Аннотация: В работе рассматриваются возможности и ограничения применения искусственного интеллекта для анализа договорных текстов. Исследуются типовые юридические риски, возникающие в договорной практике, включая отсутствие существенных условий, неопределённость формулировок и противоречие условий договора требованиям законодательства. Была предложена концептуальная модель интеллектуальной системы проверки договоров и реализован программный прототип, предназначенный для выявления потенциально проблемных положений и формирования рекомендаций.

Ключевые слова: искусственный интеллект, договорная практика, юридические риски, интеллектуальная система, программный прототип.

В условиях стремительного развития цифровых технологий и увеличения объёма документооборота особую актуальность приобретает автоматизация анализа юридических документов. Договоры являются основным инструментом регулирования гражданско-правовых отношений и содержат большое количество условий, которые могут создавать юридические риски, противоречить законодательству или содержать формальные ошибки. Проверка таких документов традиционно осуществляется юристами и требует значительных временных затрат, высокого уровня квалификации и внимательности.

Современные методы искусственного интеллекта и машинного обучения открывают новые возможности для автоматизации анализа текстов, в том числе юридических документов. Нейросетевые модели способны выявлять закономерности в больших массивах текстовых данных, распознавать потенциальные риски, анализировать структуру договора и формировать рекомендации по его улучшению. Внедрение подобных технологий позволяет повысить эффективность правовой экспертизы документов, снизить вероятность пропуска ошибок и ускорить процесс подготовки договоров.

Настоящий проект посвящён разработке интеллектуальной нейросетевой системы анализа договоров для выявления юридических рисков и формирования рекомендаций. В работе рассматриваются теоретические аспекты применения искусственного интеллекта в правовой сфере,

особенности правового регулирования договорных отношений в российском праве, а также практическая реализация программной системы, предназначенной для автоматизированной проверки договоров.

Несмотря на значительные достижения в области обработки естественного языка, остаётся открытым вопрос о степени формализуемости юридического смысла. В отличие от обычных текстов, договор представляет собой не только совокупность формулировок, но и инструмент правового регулирования, требующий интерпретации с учётом контекста, судебной практики и принципов права. В этой связи возникает вопрос: в какой мере анализ договоров может быть автоматизирован, и где проходят границы применения искусственного интеллекта в юридической сфере.

Целью проекта является исследование возможностей и границ применения искусственного интеллекта в анализе договоров, а также разработка программного прототипа системы, демонстрирующей возможность автоматизированного выявления типовых юридических рисков.

В процессе работы над проектом были изучены теоретические основы применения искусственного интеллекта и методов обработки естественного языка в юридической сфере; проанализированы особенности договорного права Российской Федерации и типовые юридические риски, возникающие в договорной практике; рассмотрены существующие подходы к автоматизации правового анализа документов. Была разработана концептуальная модель интеллектуальной системы анализа договоров.

В рамках проведённого исследования были выделены ключевые категории юридических рисков, наиболее часто встречающиеся в договорной практике. К ним относятся отсутствие существенных условий договора, неопределённость или двусмысленность формулировок, противоречие положений договора действующему законодательству, а также отсутствие механизмов ответственности сторон и порядка разрешения споров.

Особую значимость представляют формулировки, допускающие различное толкование, такие как «в разумный срок», «по согласованию сторон», «в установленном порядке». Подобные конструкции создают правовую неопределённость и могут приводить к конфликтам между сторонами, а также затрудняют судебное разрешение споров.

Кроме того, значительный риск связан с неполнотой договора, когда в тексте отсутствуют обязательные элементы (например, сроки исполнения, порядок оплаты или условия расторжения). Выявление подобных недостатков может быть частично формализовано и автоматизировано, что делает их удобной целью для применения интеллектуальных систем анализа.

Разработанная система анализа договоров построена по модульному принципу и представляет собой многоуровневый программный конвейер. Каждый модуль отвечает за выполнение определённой функции и может быть независимо модифицирован или расширен.

На первом уровне осуществляется загрузка документа и его предварительная обработка. Далее следует модуль структурного анализа, который выполняет сегментацию текста на логические фрагменты (разделы, пункты, предложения).

Следующий уровень включает блок анализа юридических рисков, реализованный в виде гибридной системы, объединяющей правилые методы и нейросетевую модель. Правильный компонент отвечает за выявление формальных нарушений и отсутствующих элементов, тогда как нейросеть анализирует семантическое содержание текста.

Финальный модуль системы формирует отчёт, содержащий структурированную информацию о выявленных рисках, включая их описание, уровень значимости и рекомендации по устранению. Такая архитектура обеспечивает масштабируемость, гибкость и возможность дальнейшего развития системы.

В системе применяется комплекс методов обработки естественного языка и машинного обучения, направленных на анализ юридических текстов.

Правильный подход основан на использовании заранее определённых шаблонов, регулярных выражений и словарей юридических конструкций. Он позволяет эффективно выявлять формальные ошибки, такие как отсутствие ключевых условий или использование неопределённых формулировок.

Нейросетевой компонент реализован на основе трансформерной архитектуры BERT, адаптированной для русского языка. Модель обучается на размеченном корпусе договорных фрагментов и выполняет задачу классификации текста по типам юридических рисков.

Дополнительно используются методы предварительной обработки текста, включая токенизацию, нормализацию и разбиение на смысловые единицы. Комбинация правилогого и нейросетевого подходов позволяет повысить точность анализа и компенсировать ограничения каждого из методов по отдельности.

Практическая реализация системы выполнена на языке программирования Python, который обладает широкими возможностями для разработки интеллектуальных приложений и обработки текстовых данных. Архитектура программного прототипа реализована в виде набора взаимосвязанных модулей, каждый из которых отвечает за отдельный этап анализа договора.

На этапе предобработки используются библиотеки для работы с различными форматами документов. В частности, библиотека `python-docx` применяется для извлечения текста из файлов формата DOCX, а для обработки PDF-документов могут использоваться инструменты, такие как `pdfplumber` или `PyMuPDF`. Полученный текст очищается от служебных символов, лишних пробелов и форматирования с помощью регулярных выражений, после чего приводится к нормализованному виду.

Далее реализуется модуль структурного анализа, который разбивает текст договора на логические фрагменты. Для этого используются эвристические правила, основанные на длине строк, наличии заголовков и ключевых слов. Такой подход позволяет выделить отдельные пункты договора, которые затем анализируются независимо друг от друга.

Ключевым элементом системы является модуль анализа рисков. В нём реализован гибридный подход:

- 1) правилый анализ выявляет заранее определённые паттерны юридических ошибок;
- 2) нейросетевая модель, построенная на основе библиотеки `transformers` и фреймворка `PyTorch`, классифицирует текстовые фрагменты по типам рисков.

Обучение модели осуществляется на специально подготовленном датасете договорных фрагментов, размеченных по категориям рисков. В ходе работы модель получает текст фрагмента и возвращает предсказанный класс риска вместе с оценкой уверенности.

Результаты анализа передаются в модуль формирования отчёта. Для генерации итогового документа используется библиотека `python-docx`, позволяющая создавать структурированные отчёты в формате DOCX. В отчёте автоматически формируются разделы с общей информацией о документе, перечнем выявленных рисков, описанием проблемных фрагментов и рекомендациями по их исправлению.

Таким образом, практическая реализация демонстрирует возможность создания полноценной интеллектуальной системы, способной выполнять комплексный анализ договорных документов в автоматическом режиме.

Результаты проведённого исследования и экспериментальной реализации системы показывают, что использование методов искусственного интеллекта позволяет эффективно выявлять типовые юридические риски в договорных текстах.

Нейросетевая модель продемонстрировала высокую точность классификации на тестовых данных, что свидетельствует о её способности распознавать характерные признаки проблемных формулировок. Однако при использовании синтетически сгенерированных данных на-

блюдается эффект переобучения, что указывает на необходимость дальнейшего расширения и усложнения обучающего корпуса.

Практическое применение системы подтверждает её эффективность в качестве инструмента предварительного анализа договоров. Она позволяет значительно сократить время проверки документов и снизить вероятность пропуска типовых ошибок.

Вместе с тем установлено, что автоматизированный анализ имеет ограничения, связанные с невозможностью полной формализации юридического смысла. Сложные правовые конструкции, контекстуальные зависимости и интерпретационные аспекты требуют участия квалифицированного специалиста.

Выполненное исследование было посвящено анализу возможностей и границ применения искусственного интеллекта в договорной экспертизе. В ходе работы были рассмотрены теоретические основы договорного права Российской Федерации, выделены типовые юридические риски, характерные для договорной практики, и исследован вопрос о степени формализуемости юридического смысла.

На основе проведённого анализа была предложена и реализована концепция гибридной интеллектуальной системы, сочетающей правиловые методы и нейросетевую классификацию фрагментов текста. Практическая реализация в виде программного прототипа показала, что автоматизация анализа договоров действительно возможна и даёт прикладной результат. Система способна:

- 1) извлекать текст из документа;
- 2) выделять структурные блоки договора;
- 3) обнаруживать отсутствие существенных условий;
- 4) фиксировать неопределённые формулировки;
- 5) выявлять отдельные типы рисков с использованием нейросетевой модели;
- 6) формировать итоговый аналитический отчёт с рекомендациями.

Вместе с тем результаты практического эксперимента показали, что юридический смысл не может быть полностью сведен к набору формальных шаблонов или статистических признаков. Наиболее сложные элементы правового анализа связаны не только с текстовой формой, но и с контекстом сделки, балансом интересов сторон, судебной практикой и профессиональной интерпретацией нормы. Именно в этих аспектах проявляются границы применения искусственного интеллекта в юридической сфере.

Таким образом, основная гипотеза исследования подтвердилась: искусственный интеллект может эффективно использоваться в договорном анализе, но прежде всего как инструмент предварительной или вспомогательной экспертизы, а не как полностью автономный заменитель юриста.

Практическая значимость проекта заключается в демонстрации возможности построения работоспособного программного прототипа, который способен автоматизировать часть рутинных операций при проверке договоров и повысить эффективность предварительной юридической оценки документов.

Перспективы дальнейшего развития работы заключаются в:

- 1) расширении корпуса обучающих данных за счёт реальных договорных текстов;
- 2) улучшении качества нейросетевой модели на основе более репрезентативной разметки;
- 3) расширении перечня типов юридических рисков;
- 4) учёте отраслевой специфики договоров;
- 5) интеграции анализа судебной практики и нормативных ограничений;
- 6) создании пользовательского интерфейса для практического применения системы.

Следовательно, разработанный прототип представляет собой не завершённое универсальное решение, а исследовательскую и прикладную основу для дальнейшего развития интеллектуальных систем в области правового анализа договоров.

Полная версия исходного кода программного прототипа вынесена за пределы основного текста и размещена в репозитории GitHub автора.

Литература и источники:

1. Гражданский кодекс Российской Федерации (часть первая) от 30.11.1994 № 51-ФЗ.
2. Гражданский кодекс Российской Федерации (часть вторая) от 26.01.1996 № 14-ФЗ.
3. Противодействие фальсификации истории великой отечественной войны / Бочарников И.В., Суздалева Т.Р., Федоров К.В., Криворучко А.А., Петренко А.И., Зеленков М.Ю., Кандыбович С.Л., Разина Т.В., Овсянникова О.А., Трипольский В.Б. Москва, 2020.
4. Ремарчук В.Н. Информационно-аналитическая деятельность: проблемы и перспективы // Вестник Академии военных наук. 2023. № 1 (82). С. 31–35.
5. Bird S., Klein E., Loper E. Natural Language Processing with Python. Sebastopol: O'Reilly Media, 2009.
6. Chollet F. Deep Learning with Python. 2nd ed. Shelter Island: Manning, 2021.
7. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. 2019.
8. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge: MIT Press, 2016.
9. Jurafsky D., Martin J. H. Speech and Language Processing. 3rd ed. Draft. Stanford University, 2023.

10. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson, 2021.
11. Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need // Advances in Neural Information Processing Systems. 2017.
12. Python Software Foundation. Python Documentation. // <https://www.python.org/doc/>.
13. PyTorch Foundation. PyTorch Documentation. // <https://docs.pytorch.org/docs/stable/index.html/>.
14. Hugging Face. Transformers Documentation. // <https://huggingface.co/docs/transformers/index/>.
15. Scikit-learn developers. Scikit-learn Documentation. // <https://sklearn.org/stable/index.html/>.
16. GitHub-репозиторий проекта анализа договоров [Электронный ресурс]. – URL:<https://github.com/Maks0Ny/AiDocumentsAnalyse.git>.