

# ОДНОПРОХОДНОЕ ВЫРАВНИВАНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ ЗАДАЧИ УПРОЩЕНИЯ ЮРИДИЧЕСКИХ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДА ОПТИМИЗАЦИИ ОТНОШЕНИЯ ШАНСОВ

Селенкин Е. Е.

Научный руководитель – канд. физ.-мат. наук, доцент Михайлова Е. Г.  
Университет ИТМО

## Введение

Проблема доступности юридических документов остается одним из ключевых барьеров во взаимодействии граждан и малого бизнеса с правовой системой. Контракты, пользовательские соглашения и нормативные акты традиционно состоят из использования специфической профессиональной лексики и усложненных синтаксических конструкций, что затрудняет их понимание лицами без профильного образования. В последние годы в отечественной и зарубежной практике активно исследуется применение больших языковых моделей (LLM) для автоматического перефразирования и упрощения правовых текстов [1]. Однако прямое использование существующих моделей часто приводит к двум критическим проблемам: генерации фактических ошибок (галлюцинаций), при которых теряются или искажаются важные юридические оговорки, либо к недостаточному уровню упрощения, когда модель лишь незначительно модифицирует исходный текст.

Традиционные методы решения проблемы галлюцинаций требуют сложной процедуры «выравнивания» (Alignment) моделей под предпочтения человека. Зарубежный опыт показывает высокую эффективность подходов RLHF (обучение с подкреплением на основе отзывов человека) и DPO (прямая оптимизация предпочтений) [2]. Тем не менее, данные методы являются крайне ресурсоемкими: они требуют многоэтапных процессов подготовки данных, обучения дополнительных моделей вознаграждения или удержания в памяти графических ускорителей замороженных референсных копий языковой модели, что ограничивает их применимость в локальных исследованиях и малом сегменте LegalTech-индустрии.

## Основная часть

Для решения указанной проблемы в данной работе предлагается использование новейшего метода оптимизации отношения шансов (Odds Ratio Preference Optimization, ORPO) для адаптации легковесных открытых языковых моделей к задаче юридического упрощения. Суть предлагаемого решения заключается в отказе от многоэтапного конвейера выравнивания. Метод ORPO позволяет совместить стадию supervised fine-tuning и процесс выравнивания предпочтений в один вычислительный проход [3]. В процессе обновления весов модели базовая функция потерь комбинируется со штрафом, основанным на логарифмическом отношении шансов, что позволяет модели эффективно усваивать желаемый стиль ответа и одновременно активно подавлять генерацию нежелательных паттернов.

В качестве эмпирической базы исследования был использован и модифицирован корпус текстов коммерческих контрактов Contract Understanding Atticus Dataset (CUAD). Для применения алгоритма ORPO оригинальный датасет был преобразован в набор троек предпочтений, где каждой сложной юридической формулировке сопоставлялись два варианта перевода: корректный (максимально упрощенный текст с полным сохранением юридических сущностей) и отвергаемый (текст, содержащий типичные ошибки языковых моделей, такие как искажение сроков, обязательств сторон или сохранение избыточной сложности).

План проведенного эксперимента включал обучение открытой языковой модели с архитектурой трансформера на сформированном наборе данных. В процессе оценки

качества генерации применялся комплексный подход. Синтаксическая и лексическая простота итоговых текстов измерялась с помощью метрик читаемости (Flesch Reading Ease). Сохранение семантического и юридического смысла проверялось с использованием векторных метрик близости, а также посредством автоматизированной оценки фреймворком «LLM-as-a-Judge» по заранее заданным критериям строгости.

Анализ результатов исследования продемонстрировал, что применение алгоритма ORPO позволяет достичь высоких показателей читаемости выходных текстов, сопоставимых с моделями большего размера, при полном нивелировании эффекта катастрофического забывания. Предложенный подход продемонстрировал существенное снижение частоты галлюцинаций по сравнению с базовым обучением с учителем. Кроме того, ресурсоемкость предложенного решения оказалась кратно ниже, чем при использовании метода DPO, за счет исключения необходимости вычисления логитов референсной модели в процессе оптимизации.

### **Выводы**

Практическая значимость проведенного исследования заключается в разработке экономичного и легко масштабируемого пайплайна создания специализированных лингвистических ИИ-моделей для юридического домена. Полученные результаты открывают возможность локального развертывания высокоточных систем упрощения текстов на аппаратном обеспечении потребительского класса, что критически важно для соблюдения конфиденциальности при обработке NDA и коммерческих договоров. Предложенное решение может быть внедрено в корпоративные системы электронного документооборота, а также использовано в качестве базового модуля в сервисах правовой поддержки (LegalTech) для автоматизированного анализа договоров аренды, трудовых соглашений и кредитных контрактов с целью выявления и объяснения скрытых рисков для потребителей.

### **Литература**

1. Hendrycks D. CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review // *Advances in Neural Information Processing Systems*. 2021. Vol. 34. P. 10428–10440.
2. Rafailov R. Direct preference optimization: Your language model is secretly a reward model // *Advances in Neural Information Processing Systems*. 2023. Vol. 36. P. 53728–53741.
3. Hong J., Lee N., Thorne J. ORPO: Monolithic Preference Optimization without Reference Model [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2403.07691>, свободный. Яз. англ. (дата обращения: 11.03.2024).