

УДК 004.85

## ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ LLM ДЛЯ ВЫЯВЛЕНИЯ ТРЕБУЕМЫХ ПРАВИЛ ОФОРМЛЕНИЯ ДОКУМЕНТОВ ИЗ НОРМАТИВНЫХ АКТОВ

Синюков Л.В. (ИТМО), Лаптев Е.И. (ИТМО), Марцинкевич В.И. (ИТМО),  
Терещенко В.В. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Федоров Д.А.  
(ИТМО)

**Введение.** Техническая документация, курсовые и выпускные квалификационные работы должны соответствовать определенным требованиям различных регламентирующих документов. Для того, чтобы определять соответствие стандартам, используется процесс нормоконтроля. На данный момент работа с системой контроля, а именно обработка ГОСТов и преобразование их в данные, является трудоемкой и требует больших затрат по времени, поскольку осуществляется ручным способом [1,2]. Кроме того, выделение требований из регламентирующих документов также проводится вручную. Для того, чтобы уменьшить затраты времени и человеческих усилий, планируется автоматизировать процесс извлечения данных и их последующую обработку [3].

Исследование посвящено проверке эффективности существующих LLM моделей при дополнительной их настройке для выделения и обработки требуемых правил из текстовых документов и их последующем преобразовании в соответствующую базу данных структуру данных.

**Основная часть.** В процессе исследования был проведен эксперимент по настройке LLM моделей для решения поставленной задачи. В качестве основных моделей для их дальнейшей настройки были взяты такие, как YandexGPT, ChatGPT, GigaChat и Theb.Ai. [4]

Принцип обработки нормативных документов:

Входные данные:

- 1) На вход модель получает текстовый документ - описание ГОСТа и/или других установленных требований.
- 2) Далее модель получает запрос, в котором описаны основные правила оформления документов, которые необходимо извлечь из текстового документа.

Выходные данные: На выход модель выдает необходимые правила оформления в соответствии каждому критерию.

Работа с базой данных: Полученные требования по оформлению отправляются в базу данных, которая имеет следующую структуру:

- 1) Список таблиц, содержащих информацию о элементах ГОСТа и их описание.
- 2) Таблица `gost_params`. Содержит связи между ГОСТами, элементами и параметрами.

Включает следующие столбцы:

- a) `id_gost` - ID ГОСТа;
- b) `id_element` - ID элемента;
- c) `id_param` - ID параметра;
- d) `value` - Значение параметра для данного ГОСТа и элемента;
- e) `is_recommended` - Указывает, является ли значение рекомендуемым;
- f) `operator` - Оператор, используемый для сравнения значения с рекомендуемым;
- g) `id` - Уникальный идентификатор записи.

**Выводы.** Предложенный метод обработки с помощью больших языковых моделей регламентирующих документов, заключающих в себе определенные требования по оформлению, позволит сократить время и ресурсы, в отличие от случая, когда данные вводятся ручным способом. Однако, несмотря на значительную экономию ресурсов, результаты получились смешанными. Модели выделили большую часть требуемых свойств и

атрибутов, но возникли ошибки с определением условных свойств, которые возникают при определённых ситуациях. Наиболее эффективной, в сравнении с другими, оказалось модель ChatGPT 3.5 компании OpenAi.

#### **Список использованных источников:**

1. Воронкова П. Н., Французова Ю. В. Обеспечение качества конструкторской документации за счет средств автоматизации нормоконтроля //Известия Тульского государственного университета. Технические науки. – 2017. – №. 8-1. – С. 195-198.

2. Манкевич О. В., Семеняк П. А. Особенности автоматизации нормоконтроля текстовых документов. – 2015.

3. Стариченко Б. Е., Устинов М. А. Программа автоматизации контроля оформления текстовых документов //Педагогическое образование в России. – 2018. – №. 8. – С. 163-168.

4. Блинова С. Д., Пермякова А. А. АВТОМАТИЗАЦИЯ НОРМОКОНТРОЛЯ ВЫПУСКНЫХ КВАЛИФИКАЦИОННЫХ РАБОТ //Общество, наука, инновации (НПК-2014). – 2014. – С. 938-942.