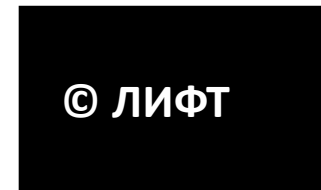
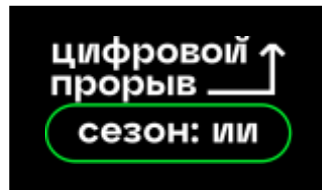


# Контроль и управление изменениями в тендерных закупках

---

Цифровой прорыв 2024  
Международный хакатон, Калининград



# ЛИФТ

состав команды



## Руслан Латипов

Full Stack Developer, Зеленодольск

✉@rus\_lat116

## Юрий Дон

Data Science, Краснодар

✉@Yuriy\_Nikitich

## Алексей Верт-Миллер

Data Science, Архангельск

✉@alexwert3

## Татьяна Моисеева

Full Stack Developer, Москва

✉@Estochka



# Задача

Задача комплаенс контроля широко известна и востребована в индустриях автомобилестроения, авиастроения, при создании сложных программно-аппаратных систем в железнодорожной, энергетической и других отраслях.

Участникам хакатона необходимо разработать алгоритм для осуществления комплаенс контроля. Необходимо производить качественный анализ документации на соответствие, а также устанавливать уровень соответствия по предоставленной шкале.

# Комплаенс контроль

разработка алгоритма ИИ



# Данные

обзор данных

## Документация на дизайн (HMI)

12 элементов спецификации, описывающих  
сценарий интерфейсной работы системы

## Технические спецификации (SSTS)

11 элементов технической документации,  
описывающий работу блока системы



## Отчёт

анализ и выявленная зависимость (уровень  
расхождения) документов с выводами

# EDA

анализ данных

1

## Наличие примеров, где отсутствует спецификация SSTS

отсутствие возможности оценки реализации разработанного дизайна

### Рекомендация

возрат на доработку/разработка критериев  
необходимости разработки технической  
спецификации

2

## Нет описания критерия Compliance Level

высокий риск человеческого фактора при оценке уровня  
соответствия от эталонного, отсутствие возможности  
итерпритации результата соответствия при использовании  
алгоритмов ИИ

### Рекомендация

разработать описания критерия оценки  
соответствия, а также методические  
рекомендации для работников

# EDA

анализ данных

3

## Разнородный формат текста в документах

затрудняет реализацию автоматизации обработки документов, а так же человеческое восприятие при чтении документа

### Рекомендация

разработать требование к формату документов (шрифт, стилистика, нумерация разделов) — образец типовой формы документа

4

## Отсутствие информации в отдельных разделах

затрудняет полноценную оценку документов, например раздел **Description**: не заполнен

### Рекомендация

В случае отсутствия информации раздел не включать в документацию или пометить раздел - «**требования не предъявляются**»

# EDA

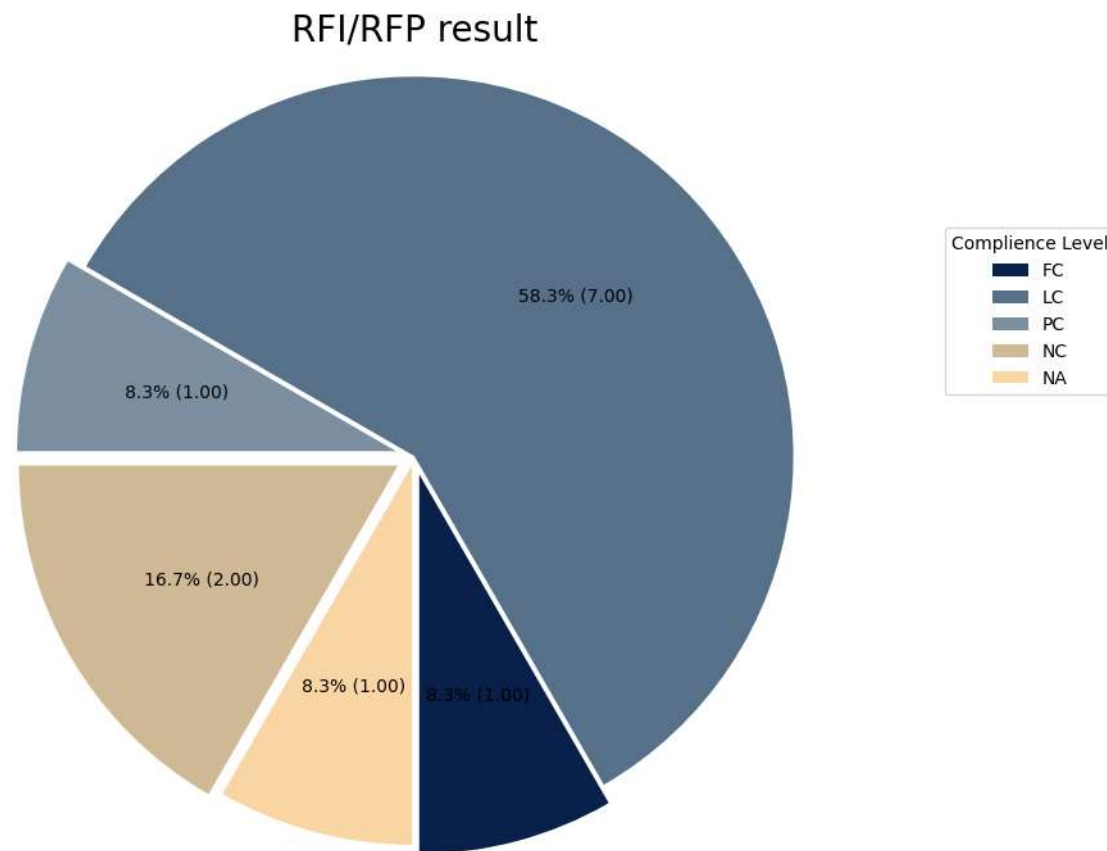
5

## Отсутствие обобщающей характеристики RFI/RFP

отсутствие возможности принятия решения по результатам  
общего анализа документации

### Рекомендация

разработать критерии (допустимую долю) в  
разрезе Compliance Level





# FE

генерация признаков

1

## Извлечение текстов из документов

данный признак позволяет формировать токены для обучения модели ИИ

2

## Признак наличия спецификации SSTS

данный признак позволяет определить уровень расхождения без использования алгоритмов ИИ

3

## Удаление не релевантных слов из описания

данный признак позволяет улучшить качество сопоставления документов

## Качественное сравнение

BartForConditionalGeneration  
(с начальными весами facebook/bart-base)

1

Bart использует стандартную архитектуру seq2seq/машинного перевода с двунаправленным кодером (типа BERT) и декодером слева направо (типа GPT).

2

BART особенно эффективен при тонкой настройке для генерации текста, но также хорошо подходит для задач на понимание. Он соответствует производительности RoBERTa с сопоставимыми учебными ресурсами на GLUE и SQuAD, достигает новых передовых результатов в ряде задач абстрактного диалога, ответов на вопросы и резюмирования с приростом до 6 ROUGE.

## Количественное сравнение

### SentenceTransformer + Cosine Similarity

(с весами roberta-base-nli-stsb-mean-tokens)

1

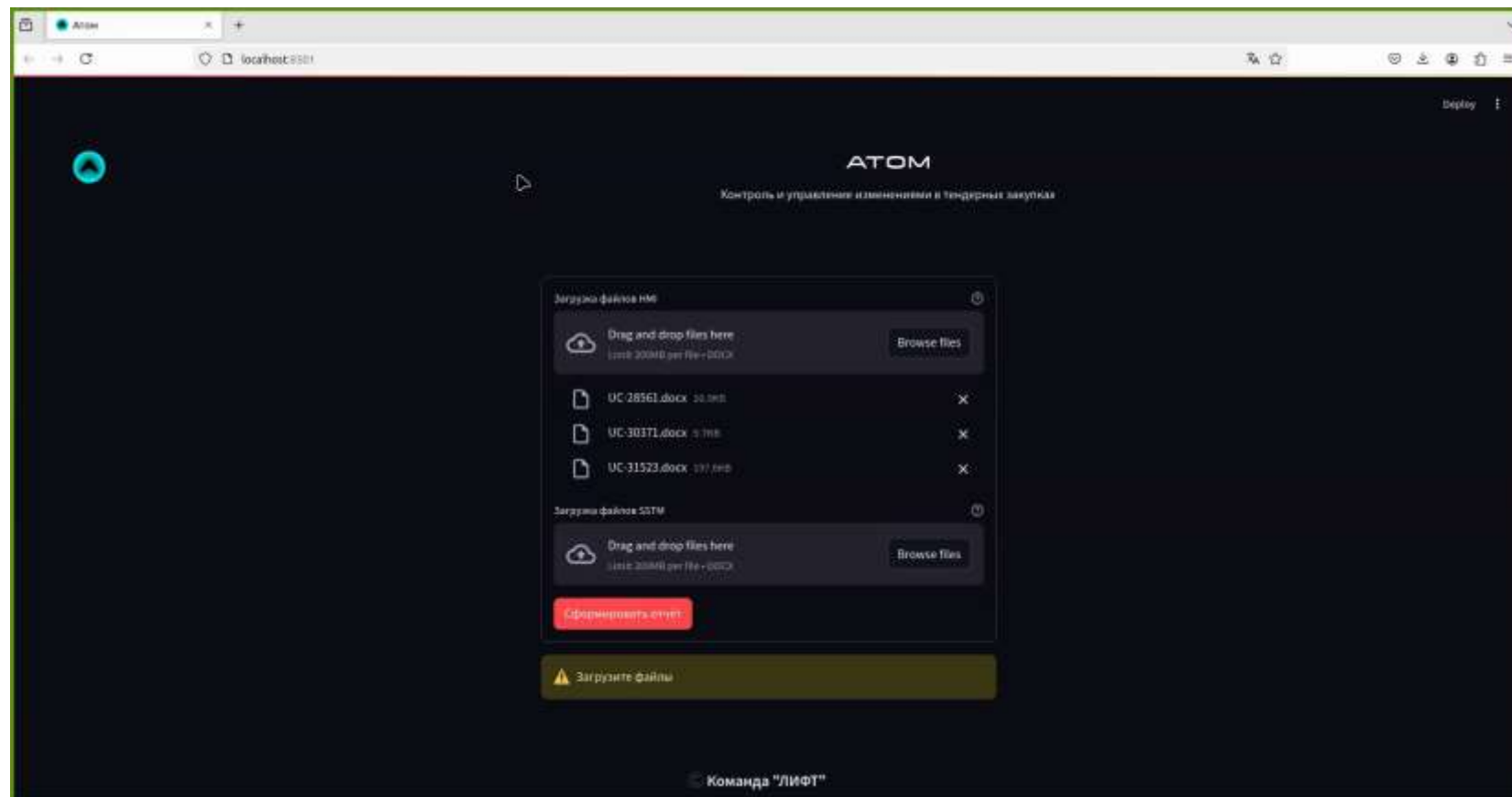
Косинусное сходство — это мера сходства между двумя ненулевыми векторами, широко применяемая во многих приложениях машинного обучения и анализа данных. Фактически она измеряет косинус угла между двумя векторами. В результате дается представление о том, насколько далеко два вектора указывают в одном направлении независимо от их величин.

2

SentenceTransformer — формирование эмбедингов для сравнения похожести Differences и Description

# Сервис

Web интерфейс



# Дальнейшее развитие проекта

---

идеи

1

## Расширение набора документов

fine-tuning используемых архитектур моделей на большем количестве документов

2

## Внедрения единого формата документов HMI и SSTs

использованием ML для приведения в соответствии единому стандарту или выявлению примеров не соответствия стандартам. Что позволит более эффективно в будущем обучать ML модели для текущей и смежных задач.

# Дальнейшее развитие проекта

идеи

3

## Критерии шкалы соответствия документов

разработка более четких критериев шкалы соответствия документов, с утвержденными метриками по конкретным критериям оценки отсутствия или наличия расхождений.

4

## Разделение функционала на ключевой и дополнительный

установить предел/долю каждого из функционала по количественной характеристике



### Вывод

всеобъемлющий тотальный контроль и управление изменениями в тендерных закупках

# Преимущества

почему нужно выбрать нашу модель ИИ

## Web-интерфейс

- загрузка документов
- формирование отчёта
- выгрузка отчёта в файл



## Потенциал

- идеи развития
- возможность дообучения модели

## Контроль

- качественный анализ
- количественный анализ
- интерпритация результатов
- конкретные шаги использования




## Скорость работы

- время обработки  
одного файла не более 3 секунд

# БЛИЖЕ ЧЕМ КАЖЕТСЯ

---

контроль и управление изменениями в тендерных закупках

цифровой  
прорыв   
сезон: ии



АТОМ

© ЛИФТ