

Федеральное государственное унитарное предприятие
«Главный радиочастотный центр»
(ФГУП «ГРЧЦ»)

Дербеневская набережная, д. 7, стр. 15, Москва, 117997, тел.: (495) 748 38 98, факс: (499) 230 15 31, <http://www.grfc.ru>,
E-mail: grfc@grfc.ru, ОКПО 56562879, ОГРН 1027739334479, ИНН/КПП 7706228218/772501001, ОКУД 0200000

ПАСПОРТ ПРОЕКТА
«Разработка единого модуля анализа»

Москва, 2021

Название проекта:	Разработка единого модуля анализа
Руководитель проекта:	Уткин Никита Викторович
Дата (DD/MM/YYYY):	01.10.2021

Версии (строки добавляются по необходимости):		
Версия	Дата (DD.MM.YYYY)	Комментарий
1.0	01.10.2021	Первоначальное формирование проекта Паспорта проекта

Термины, определения и сокращения

Термин	Определение
АРМ	Автоматизированное рабочее место
АС	Автоматизированная система
ЕМА	Единый модуль анализа
ДМСМК	Департамент организации мониторинга СМК
ДРЗИ	Департамент ведения реестров запрещённой информации
СМК	Средства массовой коммуникации
Роскомнадзор	Федеральная служба по надзору в сфере связи, информационных технологий и массовых коммуникаций
ДЦТ	Департамент цифровой трансформации
ФГУП «ГРЧЦ»	Федеральное государственное унитарное предприятие «Главный радиочастотный центр»
АС МАСМ	Автоматизированная система мониторинга и анализа социальных сетей
АС МАВР	Автоматизированная система мониторинга аудио - визуальных ресурсов
АС МТРВ	Автоматизированная система мониторинга теле- и радиовещания
АС ЧИ	Автоматизированная система «Чистый интернет»
ЗИ	Запрещённая информация

1. Краткое описание проекта

1.1. Название проекта

Полное наименование проекта: Разработка единого модуля анализа.

Краткое наименование проекта: ЕМА.

1.2. Суть проекта

Проект нацелен на разработку прототипа программного обеспечения (далее - модуль), которое позволит для получаемого на входе текста определять признаки наличия в нем нарушений по фиксированному набору тематик (Приложение 1). Данный перечень тематик является приоритетным ввиду особой опасности контента, содержащего нарушения по данным тематикам. Расширение перечня выявляемых нарушений возможно в рамках новых проектов. Модуль должен обеспечивать при вводе в промышленную эксплуатацию минимальную производительность, достаточную для обработки не менее 200 тысяч текстовых материалов в сутки.

В рамках проекта будут применены методы искусственного интеллекта и машинного обучения. Так будет создана модель, способная (в границах, указанных в разделе 1.4) определять, содержит ли тот или иной текст признаки нарушений.

Разработанный модуль должен иметь механизмы интеграции со сторонними АС, которые будут источником анализируемых текстов и потребителем результатов анализа.

1.3. Бизнес-окружение проекта

Проект инициирован в связи с необходимостью оптимизации направления деятельности ФГУП «ГРЧЦ» по мониторингу выполнения Интернет-ресурсами требований Федерального закона от 27.07.2006 №149 «Об информации, информационных технологиях и о защите».

В настоящее время анализ и оценка информации, опубликованной на Интернет-ресурсах, с целью выявления ЗИ, осуществляется с использованием анализа по словарям (регулярным выражениям). Указанный метод обладает недостаточной точностью и требует высоких трудозатрат ввиду существенного количества материалов, которые необходимо проверять операторам ДМСМК ФГУП «ГРЧЦ».

Вследствие большого количества анализируемой информации, а также постоянного роста ее объёмов в перспективе выявление ЗИ предполагает существенное и регулярное увеличение трудозатрат и привлечения достаточного количества операторов с соответствующей подготовкой по использованию методики выявления нарушений. Ввиду существующих ограничений на рост штатной численности подразделений (как бюджетного, так и административно-организационного характера), выполняющих задачи по мониторингу СМК, а также учитывая экономическую целесообразность, было принято решение об автоматизации операций, связанных с выявлением признаков нарушений в текстовом контенте Интернет-ресурсов, с использованием технологий искусственного интеллекта.

Предполагается использование результатов проекта ЕМА для снижения трудозатрат:

- при анализе контента социальных сетей (проект АС МАСМ);
- при анализе транскрибированной речи в рамках выявления нарушений в аудиовизуальных произведениях (проект АС МАВР).

Также планируется использование данного модуля при реализации перспективной системы АС ЧИ в будущем – при реализации иных перспективных АС, таких как АС МТРВ.

Разработка ЕМА инициирована в соответствии с Распоряжением от 01.09.2021 № 163 «О создании во ФГУП «ГРЧЦ» автоматизированных систем, предназначенных для обеспечения осуществления мониторинга средств массовой информации и массовых коммуникаций, информационно-телекоммуникационных сетей и информационного пространства в части соблюдения законодательства Российской Федерации в сфере связи, информационных технологий и массовых коммуникаций».

1.4.Цели проекта

Повысить точность анализа текстовых материалов, содержащих нарушения, путём создания модуля, способного после достаточного обучения определять нарушения (согласно перечню в Приложении 1). Точность определения нарушений в текстовых материалах по итогам проекта должна быть следующей:

1. Процент подтверждённых оператором текстов, содержащих нарушения в общем объёме обработанных оператором (точность_1, precision_1) - до 55%;
2. Процент правильно определённых текстов с нарушением из всех текстов, прошедших анализ системой и содержащих нарушения (полнота_1, recall_1) - до 55%;
3. Для текстов, содержащих нарушение и правильно определённых моделью (п.1), среднее значение процента:
 - 3.1. Подтверждённых оператором тематик нарушения из общего объёма обработанных оператором (точность_2, precision_2) – до 75%;
 - 3.2. Правильно определённой тематики нарушения из всех текстов, которые модель отнесла к данной тематике (полнота_2, recall_2) - до 75%.

Указанные цели могут быть достигнуты при следующих условиях:

1. Количество примеров по каждому нарушению – не менее 15 000;
2. Для каждого из типов нарушений, должно быть собрано не менее 20 000 текстов, не являющихся запрещёнными, в которых используются слова, словосочетания и сленг, свойственные текстам соответствующего нарушения;
3. Должна быть осуществлена частичная ручная проверка текстов, не содержащих ЗИ, экспертами ДМСМК ФГУП «ГРЧЦ».

5. Риски проекта

Положительные риски:

1. Реализация проекта быстрее плана;
2. Достижение показателей лучше плановых.

Отрицательные риски:

1. Появление новых заинтересованных сторон и требований по ходу проекта;
2. Недостаточный объём материалов требуемых классов в датасете;
3. Уход членов команды с проекта;
4. Длительная болезнь членов команды;
5. Невозможность разработки интеграции со стороны смежных систем;
6. Невозможность реализации оперативного обучения модели со стороны смежных систем;
7. Длительная интеграция со сторонними системами;
8. Невозможность интеграции со сторонними системами ввиду отсутствия их разработчиков;
9. Невозможность интеграции со сторонними системами ввиду технических ограничений.

2. Описание продукта и поставок

2.1. Продуктом проекта является (перечень поставок)

Модуль проверки текстов на нарушения, развёрнутый на серверных мощностях ФГУП «ГРЦЦ» (ДЦТ).

2.2. Главными требованиями к продукту проекта являются

1. Выявление в текстовой информации признаков наличия нарушений по согласованному списку (Приложение 1). Результатом работы программного обеспечения (далее – ПО) для каждого поступившего на вход текста должен являться перечень типов нарушений, признаки которых были выявлены ПО, и оценка вероятности наличия каждого нарушения в %.
2. Обеспечение возможности подключения существующих и перспективных АС СМК к разработанному ПО через программный интерфейс взаимодействия, который обеспечивает доступ к функционалу, реализующему основные пользовательские требования. Формат эксплуатации ПО – «как сервис».
3. Выполнена интеграция с АС (при условии готовности интеграции со стороны указанных АС):
 - a. АС ЧИ;
 - b. АС МАСМ;
 - c. Модуль транскрибирования и АС МТРВ.
4. Возможность поэтапного наращивания перечня нарушений без комплексной переработки ПО.
5. Возможность увеличения производительности ПО за счет масштабирования аппаратного обеспечения.

При необходимом аппаратном обеспечении вычислительной мощности ПО способно обеспечить следующую производительность при подключении через API иных АС: 100 миллионов текстовых материалов в сутки.

Нефункциональные требования, а также не ключевые функциональные требования могут быть скорректированы по ходу проекта ввиду технических и технологических ограничений по согласованию в рабочем порядке с заинтересованными лицами проекта.

2.3. Требованиями к продукту НЕ являются (продукт не включает)

- пользовательский интерфейс для регулярных пользователей;
- интеграция с другими АС, включая, но не ограничиваясь АС МСМК;
- функциональность по выявлению признаков нарушений в метаданных, описывающих объект проверки;
- выявление нарушений в медиаконтенте (графические, видео-, аудиофайлы; потоковые видео и аудио);
- инструкции пользователя и прочая пользовательская документация;
- обеспечение пропускной способности более чем в 200 000 текстов в день.

2.4. Правила приемки поставок

Прием поставок осуществляется путем проверки соответствия результатов проекта утвержденному перечню целей в присутствии представителей заказчика и иных заинтересованных лиц.

3. Ограничения проекта

3.1. Сроки реализации проекта

Длительность проекта – 3 месяца.

В случае выявления дополнительных требований, противоречащих п.2 Паспорта проекта, либо расширяющих набор требований со стороны заинтересованных лиц после старта проекта, работы замораживаются до момента пересогласования Паспорта проекта, плана работ и показателей успешности выполнения проекта всеми заинтересованными сторонами.

3.2. Даты начала и завершения проекта

Дата начала проекта: 01.10.2021 г.

Дата завершения проекта: 30.12.2021 г.

4. Руководитель проекта и его полномочия

4.1. Назначенный руководитель проекта

Уткин Никита Викторович — главный специалист машинного обучения отдела машинного обучения ДЦТ ФГУП «ГРЧЦ».

4.2. Полномочия руководителя проекта

- влияние на состав команды в рамках бюджета;
- распределение задач в команде проекта и требование их исполнения в срок, в рамках согласованного выделения времени у членов команды;
- инициация изменения содержания проекта;
- сбор совещания с привлечением команды проекта, представителей заказчика и ключевых пользователей результатов проекта;
- поднятие вопроса об утверждении обоснованных изменений плана, требование выделения времени у любого из участника команды проекта для любых целей в рамках реализации проекта;
- выход с инициативой о приостановке или досрочном прекращении проекта.

5. Заинтересованные лица и ресурсы

5.1. Заказчик проекта

Субботин Вадим Алексеевич – заместитель руководителя Роскомнадзора.

5.2. Полномочия заказчика проекта

- назначение своих представителей;
- внедрение измененных процессов, после утверждения изменений у куратора проекта со стороны заказчика;
- согласование совещаний, требований, планов, изменений бизнес-процессов и другой проектной документации с куратором проекта со стороны заказчика и со всеми заинтересованными лицами проекта со стороны заказчика;
- вынесение предложений по составу команды управления проектом;
- требование от руководителя проекта отчетности по проекту;
- выход с инициативой о смене руководителя проекта;
- выход с инициативой о приостановке или досрочном прекращении проекта;
- выход с инициативой по корректировке содержания работ проекта.

5.3. Другие заинтересованные лица проекта

- Мосенз Кристина Евгеньевна – заместитель руководителя ДМСМК ФГУП «ГРЧЦ»;
- Шарапов Игорь Родионович – руководитель ДМСМК ФГУП «ГРЧЦ»;
- Распутько Дмитрий Александрович – и. о. руководителя ДРЗИ ФГУП «ГРЧЦ»;
- Иванов Игорь Александрович – директор в сфере СМК ФГУП «ГРЧЦ»;
- Буланов Константин Алексеевич – директор по цифровым технологиям ФГУП «ГРЧЦ».

5.4. Полномочия заинтересованных лиц проекта

- выход с инициативой по корректировке содержания работ проекта;
- участие в совещаниях по проекту или иных работах в рамках проекта;
- запрос от руководителя проекта информации о ходе работ по проекту;
- согласование результатов совещаний и рабочих встреч.

5.5. Спонсор проекта

Нестеренко Руслан Васильевич – врио генерального директора ФГУП «ГРЧЦ».

5.6. Полномочия спонсора проекта

Открытие проекта за счет собственных ресурсов.

5.7. Куратор проекта со стороны исполнителя

Волкова Анастасия Александровна – начальник Управления внутренних проектов Проектного офиса ДЦТ ФГУП «ГРЧЦ».

5.8. Полномочия куратора проекта со стороны исполнителя

- вынесение предложений по составу команды проекта со стороны исполнителя;
- инициация изменения содержания проекта;
- приостановка или досрочное прекращение проекта;
- требование от руководителя проекта отчетности по проекту;
- выход с инициативой по корректировке содержания работ проекта.

5.9. Куратор проекта со стороны заказчика

Зайцев Евгений Юрьевич – начальник Управления контроля и надзора в сфере электронных коммуникаций Роскомнадзора.

5.10. Полномочия куратора проекта со стороны заказчика

- вынесение предложений по составу команды проекта со стороны заказчика;
- инициация изменения содержания проекта;

- утверждения измененных процессов для их внедрения;
- приостановка или досрочное прекращение проекта;
- требование от руководителя проекта отчетности по проекту;
- выход с инициативой по корректировке содержания работ проекта.

5.11. Команда проекта

Уткин Никита Викторович – главный специалист машинного обучения отдела машинного обучения ДЦТ ФГУП «ГРЧЦ»;

Белая Виктория Сергеевна – специалист по тестированию отдела тестирования ДЦТ ФГУП «ГРЧЦ».

5.12. Полномочия команды проекта

- инициация изменения содержания проекта;
- сбор совещания с привлечением команды проекта, представителей заказчика и ключевых пользователей результатов проекта;
- поднятие вопроса об утверждении обоснованных изменений плана, требование выделения времени у любого из участника команды проекта для любых целей в рамках реализации проекта;
- инициация смены руководителя проекта.

6. Закупки проекта

Для реализации проекта закупки не требуются.

7. Инфраструктура

Продукт проекта будет развёрнут на имеющемся оборудовании ФГУП «ГРЧЦ».

Набор тематик для выявления в текстовых материалах

1. Текстовые материалы с порнографическими изображениями несовершеннолетних и (или) объявлений о привлечении несовершеннолетних в качестве исполнителей для участия в зрелищных мероприятиях порнографического характера (детская порнография).
2. Информации о способах, методах разработки, изготовления и использования наркотических средств, психотропных веществ и их прекурсоров, новых потенциально опасных психоактивных веществ, местах их приобретения, способах и местах культивирования наркосодержащих растений.
3. Информации о способах совершения самоубийства, а также призывов к совершению самоубийства.
4. Экстремизм и терроризм, включая:
 - a. Призывы к осуществлению террористической деятельности;
 - b. Информация, оправдывающая терроризм и иную террористическую деятельность;
 - c. Призывы к массовым беспорядкам;
 - d. Призывы к осуществлению экстремистской деятельности;
 - e. Призывы к насильственному изменению основ конституционного строя и нарушению целостности Российской Федерации;
 - f. Информация, возбуждающая социальную, расовую, национальную или религиозную рознь + призывы;
 - g. Пропаганда исключительности, превосходства либо неполноценности человека по признаку его социальной, расовой, национальной, религиозной или языковой принадлежности, или отношения к религии + призывы.