

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук  
Образовательная программа «Программная инженерия»

**СОГЛАСОВАНО**

Магистр наук о данных, страшний  
разработчик X5 Tech

\_\_\_\_\_ B. Е. Козловский

«\_\_\_\_» \_\_\_\_\_ 2025 г.

**УТВЕРЖДЕНО**

Академический руководитель  
образовательной программы  
"Программная инженерия",  
старший преподаватель департамента  
программной инженерии

\_\_\_\_\_ Н. А. Павловев

«\_\_\_\_» \_\_\_\_\_ 2025 г.

**РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОГО АССИСТЕНТА ДЛЯ  
ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЙ.**

**Техническое задание**

**ЛИСТ УТВЕРЖДЕНИЯ**

**RU.17701729.05-09 ТЗ 01-1-ЛУ**

Исполнители:

Студент группы БПИ244

\_\_\_\_\_ / Л. Д. Субботин /

«\_\_\_\_» \_\_\_\_\_ 2025 г.

Студент группы БПИ236

\_\_\_\_\_ / С. Е. Растворов /

«\_\_\_\_» \_\_\_\_\_ 2025 г.

**2025**

Инв.№ подп	Подп. и дата	Взам. инв.№	Инв.№ дубл.	Подп. и дата

УТВЕРЖДЕН

RU.17701729.05-09 ТЗ 01-1-ЛУ

**РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОГО АССИСТЕНТА ДЛЯ  
ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЙ.**

**Техническое задание**

**RU.17701729.05-09 ТЗ 01-1**

**Листов 30**

Инв.№ подп	Подп. и дата	Взам. инв.№	Инв.№ дубл.	Подп. и дата

**2025**

## АННОТАЦИЯ

Техническое задание – это основной документ, оговаривающий набор требований и порядок создания программного продукта, в соответствии с которым производится разработка программы ее тестирование и приемка.

Настоящее Техническое задание на разработку «Разработка интеллектуального ассистента для образовательных учреждений» содержит следующие разделы: «Введение», «Основания для разработки», «Назначение разработки», «Требования к программе», «Требования к программной документации», «Технико-экономические показатели», «Стадии и этапы разработки», «Порядок контроля и приемки», приложения [7].

В разделе «Введение» указано наименование и краткая характеристика области применения программы.

В разделе «Основания для разработки» указан документ, на основании которого ведется разработка, и наименование темы разработки.

В разделе «Назначение разработки» указано функциональное и эксплуатационное назначение создаваемого программного продукта.

Раздел «Требования к программе» содержит указание на основные требования к функциональным характеристикам программы, к её надежности и к условиям эксплуатации, к составу и параметрам технических средств, к информационной и программной совместимости, к маркировке и упаковке, к транспортировке и хранению, а также специальные требования.

Раздел «Требования к программным документам» содержит указание на предварительный состав программной документации и специальные требования к ней.

Раздел «Технико-экономические показатели» содержит информацию об ориентировочной экономической эффективности разработки, экономические преимущества разработки программы.

Раздел «Стадии и этапы разработки» содержит информацию о стадиях разработки, этапах и содержании работ.

В разделе «Порядок контроля и приемки» указаны общие требования к приемке работы.

Настоящий документ разработан в соответствии с требованиями:

1. ГОСТ 19.101-77 [1]: Виды программ и программных документов.
2. ГОСТ 19.102-77 [2]: Стадии разработки.
3. ГОСТ 19.103-77 [3]: Обозначения программ и программных документов.
4. ГОСТ 19.104-78 [4]: Основные надписи.
5. ГОСТ 19.105-78 [5]: Общие требования к программным документам.

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

6. ГОСТ 19.106-78 [6]: Требования к программным документам, выполненным печатным способом.
7. ГОСТ 19.201-78 [7]: Техническое задание. Требования к содержанию и оформлению.

Изменения к данному Техническому заданию оформляются согласно ГОСТ 19.603-78 [12], ГОСТ 19.604-78 [13].

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ .....	6
1.1. Наименование программы .....	6
1.2. Краткая характеристика области применения программы .....	6
2. ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ .....	7
2.1. Документ(ы), на основании которого(ых) ведется разработка .....	7
2.2. Наименование темы разработки .....	7
3. НАЗНАЧЕНИЕ РАЗРАБОТКИ .....	8
3.1. Функциональное назначение .....	8
3.2. Эксплуатационное назначение .....	8
4. ТРЕБОВАНИЯ К ПРОГРАММЕ .....	9
4.1. Требования к функциональным характеристикам .....	9
4.1.1. Требования к составу выполняемых функций .....	9
4.1.2. Требования к организации входных данных .....	9
4.1.3. Требования к организации выходных данных .....	10
4.1.4. Требования к временным характеристикам .....	11
4.1.5. Требования к интерфейсу .....	12
4.2. Требования к надежности .....	12
4.3. Условия эксплуатации .....	13
4.3.1. Климатические условия эксплуатации .....	13
4.3.2. Требования к видам обслуживания .....	13
4.4. Требования к составу и параметрам технических средств .....	14
4.4.1. Программные средства .....	14
4.4.2. Технические средства .....	15
4.5. Требования к информационной и программной совместимости .....	16
4.5.1. Требования к информационным структурам и методам решения .....	16
4.5.2. Требования к программным средствам, используемым программой .....	17
4.5.3. Требования к исходным кодам и языкам программирования .....	18
4.5.4. Требования к защите информации и программы .....	19
4.6. Требования к маркировке и упаковке .....	20
4.7. Требования к транспортированию и хранению .....	21
4.8. Специальные требования .....	21

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ .....	22
5.1. Состав программной документации .....	22
5.2. Специальные требования к программной документации .....	22
6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ .....	23
6.1. Предполагаемая потребность .....	23
6.2. Целевая аудитория .....	23
6.3. Преимущества перед аналогами .....	24
7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ .....	25
7.1. Стадии разработки, этапы и содержание работ .....	25
8. ПОРЯДОК КОНТРОЛЯ И ПРИЕМКИ .....	26
8.1. Виды испытаний .....	26
8.2. Общие требования к приемке работы .....	26
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ .....	28
ПРИЛОЖЕНИЕ. ССЫЛКИ НА АНАЛОГИ .....	29

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## 1. ВВЕДЕНИЕ

### 1.1. Наименование программы

Наименование программы – «Разработка интеллектуального ассистента для образовательных учреждений».

Наименование программы на английском языке – «Development of an Intelligent Assistant for Educational Institutions».

### 1.2. Краткая характеристика области применения программы

Программа предназначена для использования в образовательных учреждениях в качестве интеллектуального ассистента для студентов и преподавателей.

Основное назначение системы — предоставление оперативных и корректных ответов на вопросы пользователей, связанных с учебным процессом, нормативными документами и внутренними регламентами вуза.

Программа обрабатывает и структурирует информацию из различных источников: учебные планы, внутренние регламенты, веб-ресурсы университета и справочные материалы. Результат работы — упорядоченные и готовые к использованию данные, которые обеспечивают функционирование чат-бота на базе RAG-архитектуры.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## 2. ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ

### 2.1. Документ(ы), на основании которого(ых) ведется разработка

Разработка ведется на основании учебного плана подготовки бакалавров по направлению 09.03.04 «Программная инженерия» и утвержденной академическим руководителем программы темы курсового проекта.

### 2.2. Наименование темы разработки

Разработка интеллектуального ассистента для образовательных учреждений с использованием RAG-архитектуры для обработки и структурирования внутренних данных университета.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

### **3. НАЗНАЧЕНИЕ РАЗРАБОТКИ**

#### **3.1. Функциональное назначение**

Разработка предназначена для создания интеллектуального ассистента (чат-бота), который обеспечивает пользователей образовательного учреждения оперативной и корректной информацией на основе внутренних данных университета.

Основные функции системы:

Сбор, предобработка и структурирование данных из различных источников (PDF-документы, внутренние веб-ресурсы, базы данных).

Организация данных в формате, пригодном для использования в RAG-архитектуре.

Обеспечение возможности поиска и предоставления ответов на вопросы студентов и преподавателей.

Поддержка взаимодействия с пользователями через удобный интерфейс (Telegram-бот).

Обеспечение актуальности и согласованности информации для корректной работы интеллектуального ассистента.

#### **3.2. Эксплуатационное назначение**

Программа предназначена для использования в образовательных учреждениях (университетах, колледжах) в качестве интеллектуального ассистента для студентов и преподавателей.

Эксплуатация системы предполагает:

Работа в составе информационной инфраструктуры университета с доступом к внутренним данным (учебные планы, нормативные документы, справочные материалы).

Поддержка онлайн-взаимодействия с пользователями через веб-интерфейс или мессенджеры (Telegram).

Регулярное обновление и корректировку базы данных для обеспечения актуальности информации.

Обеспечение надежного и безопасного хранения и обработки данных университета в соответствии с действующими требованиями по информационной безопасности.

Система обеспечивает повышение эффективности работы пользователей и сокращение времени на поиск информации внутри образовательного учреждения.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## 4. ТРЕБОВАНИЯ К ПРОГРАММЕ

### **4.1. Требования к функциональным характеристикам**

#### **4.1.1. Требования к составу выполняемых функций**

Тимлид, ответственный за сбор, очистку и структурирование данных, должен обеспечивать выполнение следующих функций:

**Сбор данных**

Идентификация и систематизация источников информации внутри университета (учебные планы, нормативные документы, внутренние сайты, базы данных).

Организация процесса извлечения данных с минимальными потерями и сбоев.

**Очистка данных**

Удаление дублирующихся, некорректных или устаревших записей.

Корректировка ошибок форматирования и синтаксиса текстов.

**Предобработка и нормализация**

Приведение данных к единому формату, необходимому для обработки системой интеллектуального ассистента.

Стандартизация структурированных и неструктурированных данных (таблицы, тексты, PDF).

**Классификация и структурирование данных**

Разделение информации на категории (учебные планы, нормативные акты, справочные материалы).

Формирование базы знаний для интеграции с RAG-пайплайном чат-бота.

**Обеспечение качества данных**

Проверка полноты, актуальности и согласованности информации.

Подготовка отчетов о состоянии базы данных и выявленных проблемах.

**Поддержка обновлений и модификаций**

Организация процесса регулярного обновления и расширения базы данных без потери структуры и качества.

#### **4.1.2. Требования к организации входных данных**

Для эффективной работы интеллектуального ассистента необходимо обеспечить корректную организацию и подготовку входных данных.

**Источники данных**

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Подп. и дата		Взам. Инв. №	Инв. № дубл.	Подп. и дата

Учебные планы, программы дисциплин, нормативные документы университета.

Внутренние веб-ресурсы и базы данных университета.

Дополнительные справочные материалы (PDF, Word, таблицы, HTML-страницы).

Форматы данных

Поддержка текстовых и табличных форматов (PDF, DOCX, CSV, XLSX, HTML).

Возможность обработки неструктурированных текстовых данных с последующей нормализацией.

Требования к целостности и корректности

Данные должны быть полными, актуальными и соответствовать официальным источникам.

Исключение дубликатов, ошибок форматирования и устаревшей информации.

Предварительная обработка

Приведение данных к единому стандарту и структуре для последующего использования в RAG-пайплайне.

Обеспечение возможности автоматизированного обновления базы данных без потери качества.

Документирование источников и изменений

Ведение реестра используемых источников данных.

Фиксация всех изменений и обновлений данных для контроля качества и воспроизводимости.

#### **4.1.3. Требования к организации выходных данных**

Выходные данные должны быть подготовлены в форме, обеспечивающей корректную работу интеллектуального ассистента и удобство дальнейшей интеграции:

Структура данных

Информация должна быть разделена по категориям: учебные планы, нормативные документы, справочные материалы.

Данные должны быть представлены в стандартизованных форматах (JSON, CSV, XML), пригодных для использования в RAG-пайплайне.

Качество данных

Выходные данные должны быть полными, актуальными и непротиворечивыми.

Исключение дубликатов, некорректных или устаревших записей.

Документирование

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Каждая единица данных должна сопровождаться информацией о происхождении (источник, дата получения/обновления).

Ведение журналов изменений и логов обработки данных для обеспечения прозрачности и контроля качества.

#### Совместимость и интеграция

Данные должны быть готовы к непосредственной интеграции в систему чат-бота.

Обеспечение возможности масштабирования и обновления базы данных без нарушения структуры.

#### Формат предоставления

Структурированные файлы данных предоставляются в виде отдельных модулей или пакетов для загрузки в систему.

Возможность автоматизированного экспорта данных для использования различными компонентами интеллектуального ассистента.

#### **4.1.4. Требования к временным характеристикам**

##### Сроки подготовки данных

Сбор, очистка и структурирование первичной базы данных должны быть выполнены в течение установленного этапа проекта (например, 4–6 недель с момента утверждения ТЗ).

##### Скорость обработки данных

Инструменты и процессы обработки данных должны обеспечивать возможность обработки больших объёмов документов (PDF, таблицы, веб-страницы) в разумные сроки без снижения качества.

##### Периодичность обновлений

Обновление и корректировка базы данных должны выполняться регулярно, согласно установленному графику (например, ежемесячно или по мере появления новых документов).

##### Временные показатели интеграции

Структурированные данные должны быть готовы к интеграции в систему интеллектуального ассистента без задержек, обеспечивая своевременную работу чат-бота.

##### Временные требования к отчетности

Генерация отчетов о состоянии базы данных и качестве обработанных данных должна выполняться в течение рабочего дня после запроса.

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

#### **4.1.5. Требования к интерфейсу**

##### Взаимодействие с системой

Тимлид должен иметь доступ к инструментам для сбора, очистки и структурирования данных через удобный графический интерфейс или командную строку.

Интерфейс должен обеспечивать возможность импорта данных из различных источников (PDF, HTML, базы данных, таблицы).

##### Управление процессами обработки

Интерфейс должен позволять запускать, контролировать и мониторить этапы предобработки данных.

Возможность отслеживания статуса обработки, обнаружения ошибок и повторного запуска процессов.

##### Просмотр и редактирование данных

Предоставление возможности просмотра, редактирования и аннотирования отдельных элементов данных до их интеграции в систему интеллектуального ассистента.

Подсветка ошибок и дублирующихся записей для облегчения очистки данных.

##### Отчётность и логирование

Интерфейс должен предоставлять отчеты о состоянии базы данных, полноте и актуальности данных.

Ведение логов действий пользователя для контроля качества и воспроизводимости процессов.

#### **4.2. Требования к надежности**

##### Целостность данных

Все входные и выходные данные должны сохранять целостность в процессе обработки, исключая потерю или искажение информации.

Любые ошибки или сбои в обработке должны фиксироваться и быть доступными для анализа.

##### Устойчивость к сбоям

Процессы сбора, очистки и структурирования данных должны корректно завершаться даже при неполных или некорректных входных данных.

Система должна предусматривать возможность восстановления обработки с точки сбоя без повторного выполнения всех операций.

##### Верификация и контроль качества

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Должны быть реализованы механизмы проверки полноты, согласованности и корректности данных на каждом этапе обработки.

Тимлид должен иметь возможность контролировать качество данных с помощью отчетов и логов.

#### Надёжность процессов интеграции

Структурированные данные должны гарантированно быть пригодны для интеграции в систему интеллектуального ассистента без потери функциональности.

#### Сохранение истории изменений

Все изменения, внесённые в базу данных, должны фиксироваться для возможности восстановления предыдущих версий данных и аудита процессов.

### 4.3. Условия эксплуатации

#### 4.3.1. Климатические условия эксплуатации

##### Температурный диапазон

18–25 °C для комфортной работы пользователей и надежной работы оборудования.

##### Влажность

Относительная влажность воздуха 30–60 % без конденсации влаги.

##### Защита от пыли и загрязнений

Помещение должно быть защищено от избыточной пыли, жидкости и прямого воздействия солнечных лучей на рабочие места и оборудование.

##### Электропитание и стабильность сети

Электропитание с напряжением 220 В ±10 %, с возможностью подключения к источникам бесперебойного питания (UPS) для критически важных серверов или рабочих станций.

#### 4.3.2. Требования к видам обслуживания

##### Техническое обслуживание программного обеспечения

Регулярное обновление инструментов для обработки данных (скрипты, библиотеки, модули).

Контроль исправности всех компонентов системы, используемых для сбора, очистки и структурирования данных.

##### Поддержка пользователей

Обеспечение помощи тимлиду и другим сотрудникам проекта при возникновении ошибок или сложностей в процессе обработки данных.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Документирование решений и инструкций для повторного использования.

Мониторинг и аудит процессов

Регулярная проверка полноты и корректности собранной информации.

Контроль за соблюдением стандартов качества данных и процедур обновления.

Обслуживание базы данных

Поддержка структуры базы данных в актуальном состоянии.

Устранение дублирующихся или некорректных записей.

Резервное копирование данных для предотвращения их потери.

Обновление и масштабирование

Обновление базы данных по мере появления новых документов и информации.

Масштабирование хранилища данных при увеличении объема обрабатываемой информации.

#### **4.4. Требования к составу и параметрам технических средств**

##### **4.4.1. Программные средства**

Для выполнения задач по сбору, очистке и структурированию данных необходимы следующие программные средства:

Языки программирования и библиотеки

Python с библиотеками для обработки данных (pandas, numpy, openpyxl, PyPDF2, pdfplumber, BeautifulSoup, requests).

Библиотеки для работы с NLP и RAG-пайплайном (transformers, langchain, sentence-transformers).

СУБД и системы хранения данных

PostgreSQL, MySQL или аналогичные СУБД для хранения структурированных данных.

Возможность работы с временными файловыми хранилищами для промежуточной обработки документов.

Инструменты для автоматизации и ETL-процессов

Apache Airflow, Prefect или аналогичные системы для организации процессов сбора и предоб-работки данных.

Интерфейсы для работы с данными

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Веб-интерфейс или локальные GUI-инструменты для мониторинга и управления процессами обработки данных.

Возможность интеграции с мессенджерами (например, Telegram) для проверки работы чат-бота с подготовленными данными.

#### Прочее ПО

Средства контроля версий и совместной работы (Git, GitHub/GitLab).

Средства документирования процессов и логирования (Jupyter Notebook, Confluence, Google Docs).

#### **4.4.2. Технические средства**

Для выполнения задач по сбору, очистке и структурированию данных необходимы следующие технические средства:

##### Рабочие станции

Компьютеры с современными процессорами (Intel i5/i7 или AMD Ryzen 5/7), не менее 16 ГБ оперативной памяти и SSD-накопителями для быстрого доступа к данным.

Возможность установки всех необходимых программных средств и библиотек для обработки данных.

##### Серверное оборудование

Сервер для хранения и обработки больших объемов данных (желательно с 32 ГБ и более оперативной памяти, SSD/HDD RAID-хранилище).

Доступ к GPU (при необходимости для работы с моделями NLP и RAG-пайплайнами).

##### Сетевое оборудование

Стабильное подключение к локальной сети университета и интернету для доступа к внутренним и внешним источникам данных.

Возможность безопасного VPN-доступа для работы с защищенными ресурсами.

##### Средства резервного копирования

NAS или облачные хранилища для регулярного резервного копирования базы данных и промежуточных результатов обработки.

Средства восстановления данных в случае сбоев.

##### Дополнительное оборудование

Мониторы с разрешением Full HD или выше для удобного просмотра и редактирования данных.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Принтеры и сканеры при необходимости оцифровки бумажных документов.

#### **4.5. Требования к информационной и программной совместимости**

##### **Информационные структуры**

Данные должны храниться в структурированном виде, разделенные на категории: учебные планы, нормативные документы, справочные материалы.

Для хранения использовать реляционные базы данных (PostgreSQL, MySQL) или документные хранилища (JSON, MongoDB) для неструктурированных данных.

Каждая запись должна содержать метаданные: источник, дата получения, автор, статус обработки.

##### **Методы обработки данных**

Сбор данных из различных источников с помощью скриптов и инструментов автоматизации.

Очистка данных: удаление дубликатов, исправление ошибок, нормализация формата текстов и таблиц.

Структурирование данных для интеграции в RAG-пайплайн интеллектуального ассистента.

Классификация данных по типам и категориям для обеспечения быстрого поиска и выдачи ответов.

Контроль качества данных с автоматическим логированием ошибок и предупреждений.

##### **Методы решения задач**

Использование алгоритмов ETL (Extract, Transform, Load) для извлечения, преобразования и загрузки данных.

Применение методов NLP (обработка естественного языка) для структурирования неструктурированных текстов.

Обеспечение возможности масштабируемой обработки больших объемов данных с минимальными временными задержками.

Поддержка обновлений и модификаций данных без нарушения целостности информационной структуры.

#### **4.5.1. Требования к информационным структурам и методам решения**

##### **Информационные структуры**

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Данные должны быть организованы в структурированном виде с чётким разделением по категориям: учебные планы, нормативные документы, справочные материалы, внутренние регламенты университета.

Для хранения использовать реляционные базы данных (PostgreSQL, MySQL) для структурированных данных и документные хранилища (JSON, MongoDB) для неструктурированных данных.

Каждая запись должна содержать метаданные: источник информации, дату получения, статус обработки, автора и версию.

Должна обеспечиваться возможность быстрого поиска и фильтрации данных по ключевым полям и категориям.

#### Методы обработки и предобработки данных

Автоматизированный сбор данных из различных источников (PDF, DOCX, таблицы, веб-страницы, внутренние базы данных).

Очистка данных: удаление дубликатов, исправление ошибок, нормализация форматов текстов и таблиц.

Классификация и структурирование информации для последующего использования в RAG-пайpline интеллектуального ассистента.

Верификация данных на полноту, согласованность и актуальность с возможностью логирования ошибок.

#### Методы решения задач

Применение ETL-процессов (Extract, Transform, Load) для извлечения, преобразования и загрузки данных.

Использование методов обработки естественного языка (NLP) для работы с неструктурированными текстами и извлечения ключевой информации.

Обеспечение масштабируемой обработки больших объёмов данных без потери качества.

Возможность регулярного обновления базы данных и внесения изменений без нарушения целостности информационной структуры.

#### **4.5.2. Требования к программным средствам, используемым программой**

Для эффективного выполнения задач по сбору, очистке и структурированию данных интеллектуального ассистента необходимо использовать следующие программные средства:

#### Языки программирования и библиотеки

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Python с библиотеками для работы с данными: pandas, numpy, openpyxl, PyPDF2, pdfplumber, BeautifulSoup, requests.

Библиотеки для обработки естественного языка и интеграции с RAG-пайплайном: transformers, langchain, sentence-transformers.

Системы управления базами данных (СУБД)

Реляционные СУБД: PostgreSQL, MySQL или аналогичные для структурированных данных.

Документные хранилища (JSON, MongoDB) для хранения неструктурированных данных.

Инструменты автоматизации и ETL-процессы

Apache Airflow, Prefect или аналогичные средства для организации процессов извлечения, трансформации и загрузки данных.

Интерфейсы для взаимодействия с данными

Веб-интерфейс или локальный GUI для мониторинга и управления процессами обработки данных.

Возможность интеграции с мессенджерами (например, Telegram) для тестирования работы чат-бота с подготовленными данными.

Средства контроля версий и документации

Git, GitHub/GitLab для совместной работы и управления версиями скриптов.

Jupyter Notebook, Confluence или Google Docs для документирования процессов обработки данных и логирования изменений.

#### **4.5.3. Требования к исходным кодам и языкам программирования**

Языки программирования

Основной язык разработки: Python.

Дополнительно могут использоваться скриптовые языки для автоматизации процессов (Bash, PowerShell).

Стиль и структура исходного кода

Код должен быть структурированным, читабельным и соответствовать стандартам PEP8 (для Python).

Модули и функции должны быть логически разделены по функционалу: сбор данных, очистка, нормализация, структурирование, подготовка для RAG-пайплайна.

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Все ключевые алгоритмы и процедуры должны сопровождаться комментариями и документацией.

#### Документирование кода

Использование встроенной документации (docstrings) для функций и классов.

Ведение отдельного описания архитектуры и структуры кода (README, Confluence).

#### Контроль версий и управление изменениями

Все изменения исходного кода должны контролироваться через систему управления версиями (Git, GitHub/GitLab).

Обеспечение возможности отката к предыдущим версиям кода без потери функциональности.

#### Совместимость и переносимость

Код должен быть совместим с основными операционными системами, используемыми в университете (Windows, Linux).

Возможность запуска кода как на локальных рабочих станциях, так и на серверной инфраструктуре с минимальными изменениями.

### **4.5.4. Требования к защите информации и программы**

#### Конфиденциальность данных

Все данные университета, включая учебные планы, нормативные документы и внутренние регламенты, должны обрабатываться с соблюдением политики конфиденциальности.

Доступ к данным должен предоставляться только уполномоченным сотрудникам проекта.

#### Аутентификация и контроль доступа

Все пользователи системы должны иметь индивидуальные учетные записи с разграничением прав доступа.

Тимлид должен иметь расширенные права для управления данными и контроля качества обработки.

#### Защита от потери и повреждения данных

Регулярное резервное копирование базы данных и промежуточных файлов.

Возможность восстановления данных в случае сбоя оборудования или программного обеспечения.

#### Защита от несанкционированного вмешательства

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Использование безопасных каналов передачи данных (HTTPS, VPN) при работе с внутренними и внешними источниками.

Мониторинг активности пользователей и логирование всех операций по обработке данных.

Соответствие нормативным требованиям

Обеспечение соответствия требованиям законодательства РФ по защите персональных и служебных данных.

Соблюдение внутренних регламентов университета по информационной безопасности.

Защита программного обеспечения

Защита исходного кода и скриптов от несанкционированного доступа и модификаций.

Использование систем контроля версий для отслеживания изменений и обеспечения безопасности кода.

#### **4.6. Требования к маркировке и упаковке**

Маркировка данных и файлов

Все файлы с исходными и обработанными данными должны иметь уникальные имена, отражающие их содержание и дату создания/обновления.

Каждая версия базы данных и модуль скриптов должны содержать метаданные: источник данных, ответственный сотрудник, статус обработки, версия.

Структурирование и организация хранения

Данные должны храниться в логически структурированных каталогах по категориям: учебные планы, нормативные документы, справочные материалы.

Скрипты и программные модули должны быть упакованы в отдельные директории с README-файлами и документацией по использованию.

Упаковка для передачи и интеграции

Выходные данные должны быть предоставлены в стандартизованных форматах (JSON, CSV, XML) для интеграции с системой интеллектуального ассистента.

Все пакеты данных должны быть снабжены контрольными суммами для проверки целостности при передаче или резервном копировании.

Контроль версий

Каждая поставка данных или скриптов должна сопровождаться версионной информацией для отслеживания изменений и возможности восстановления предыдущих версий.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

#### **4.7. Требования к транспортированию и хранению**

##### **Хранение данных**

Все данные университета должны храниться в защищенных каталогах серверов или рабочих станций с ограниченным доступом.

Структура хранения должна быть логичной и соответствовать категориям: учебные планы, нормативные документы, справочные материалы.

Должны быть предусмотрены регулярные резервные копии на отдельные носители или облачные хранилища для предотвращения потери данных.

##### **Транспортирование данных**

Передача данных между рабочими станциями и серверами должна осуществляться только через защищенные каналы (VPN, HTTPS).

При необходимости физической передачи данных (например, на внешнем носителе) файлы должны быть зашифрованы и снабжены контрольными суммами для проверки целостности.

##### **Сроки и условия хранения**

Все данные должны храниться столько, сколько требуется для проекта, с учетом возможности последующего анализа и интеграции в систему интеллектуального ассистента.

Должны соблюдаться условия температуры и влажности для надежной работы серверного и офисного оборудования (18–25 °C, 30–60 % влажности).

##### **Документирование и контроль**

Каждое перемещение или копирование данных должно фиксироваться в журнале с указанием даты, ответственного лица и целей перемещения.

Регулярные проверки целостности и актуальности данных должны проводиться согласно графику обслуживания.

#### **4.8. Специальные требования**

Отсутствуют

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## 5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

### 5.1. Состав программной документации

1. «Разработка интеллектуального ассистента для образовательных учреждений». Техническое задание (ГОСТ 19.201-78 [7]).
2. «Разработка интеллектуального ассистента для образовательных учреждений». Пояснительная записка (ГОСТ 19.404-79 [10]).
3. «Разработка интеллектуального ассистента для образовательных учреждений». Программа и методика испытаний (ГОСТ 19.301-79 [8]).
4. «Разработка интеллектуального ассистента для образовательных учреждений». Текст программы (ГОСТ 19.401-78 [9]).
5. «Разработка интеллектуального ассистента для образовательных учреждений». Руководство оператора (ГОСТ 19.505-79 [11]).

### 5.2. Специальные требования к программной документации

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## 6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

### **6.1. Предполагаемая потребность**

Образовательные учреждения

Университеты и колледжи, заинтересованные в автоматизации поддержки студентов и преподавателей.

Факультеты и кафедры, которым требуется оперативная информационная помощь и быстрый доступ к внутренним нормативным документам и учебным материалам.

Административные подразделения

Службы учебного процесса и отделы учебно-методической работы для облегчения работы с планами занятий, расписаниями и нормативными документами.

Отделы ИТ и поддержки пользователей, заинтересованные в уменьшении нагрузки на коллекции и сокращении времени обработки запросов.

Студенты и преподаватели

Пользователи, которым необходим быстрый доступ к учебным материалам, внутренним регламентам и справочной информации через удобный интерфейс чат-бота.

### **6.2. Целевая аудитория**

Студенты

Получают быстрый доступ к учебным планам, расписаниям, нормативным документам и справочной информации.

Используют чат-бота для решения повседневных вопросов, связанных с учебным процессом.

Преподаватели

Имеют возможность оперативно получать информацию о программах дисциплин, внутренних регламентах и других образовательных материалах.

Используют систему для поддержки учебного процесса и коммуникации со студентами.

Административные подразделения университета

Службы учебного процесса, отделы учебно-методической работы, ИТ-отделы.

Применяют интеллектуального ассистента для оптимизации обработки запросов и автоматизации рутинных информационных процессов.

Руководство образовательного учреждения

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Получает аналитику по использованию системы, эффективности предоставления информации и загруженности сотрудников.

### **6.3. Преимущества перед аналогами**

Во ВШЭ аналогов нет.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## 7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ

### 7.1. Стадии разработки, этапы и содержание работ

Таблица 2 – Стадии и этапы разработки

Стадия разработки	Этап работ	Содержание работ	Сроки выполнения
Техническое задание	Обоснование необходимости разработки	Постановка задачи	13.11.25
		Сбор исходных теоретических материалов	13.11.25
	Научно-исследовательский этап разработки	Определение структуры входных и выходных данных	15.11.25 – 03.12.25
		Предварительный выбор методов решения задач	15.11.25 – 03.12.25
		Определение требований к техническим и программным средствам	15.11.25 – 03.12.25
		Обоснование возможности решения поставленной задачи	15.11.25 – 03.12.25
		Определение требований к программному продукту	15.11.25 – 03.12.25
		Выбор языков программирования	15.11.25 – 03.12.25
		Разработка и согласование технического задания с научным руководителем	15.11.25 – 03.12.25
		Загрузка согласованного технического задания в SmartLMS	16.12.25

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## 8. ПОРЯДОК КОНТРОЛЯ И ПРИЕМКИ

Контроль и приемка разработки осуществляются в соответствии с документом «Программа и методика испытаний» (ГОСТ 19.301–79 [8]).

### **8.1. Виды испытаний**

#### Функциональные испытания

Проверка корректности работы всех функций системы: сбор, очистка и структурирование данных, поиск информации по запросам пользователей, интеграция с RAG-пайплайном.

Проверка работы веб-интерфейса и мессенджер-бота (например, Telegram) для взаимодействия с пользователями.

#### Испытания надежности и устойчивости

Проверка корректной работы системы при различных нагрузках (объемы данных, количество запросов).

Оценка устойчивости к ошибкам ввода и некорректным данным.

Тестирование восстановления после сбоев или прерывания работы.

#### Испытания безопасности и защиты информации

Проверка разграничения прав доступа и аутентификации пользователей.

Тестирование защищённости передачи данных (HTTPS, VPN) и резервного копирования.

Контроль целостности и конфиденциальности данных при хранении и обработке.

#### Приемо-сдаточные испытания

Проверка соответствия системы требованиям технического задания и потребностям целевой аудитории.

Тестирование работы системы на фокус-группе (студенты, преподаватели, административный персонал).

Подтверждение готовности к внедрению в образовательное учреждение.

### **8.2. Общие требования к приемке работы**

#### Соответствие требованиям ТЗ

Разработка принимается, если выполнены все требования, указанные в техническом задании, включая сбор, очистку и структурирование данных, а также интеграцию с RAG-пайплайном и веб-интерфейсом чат-бота.

#### Качество выполнения работ

Иzm.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

Все этапы разработки должны быть выполнены с надлежащим качеством, без существенных дефектов и ошибок в программном обеспечении и обработанных данных.

Данные должны быть актуальными, корректными и структурированными, готовыми к использованию интеллектуальным ассистентом.

#### Документальное оформление

Все результаты работы должны сопровождаться необходимой документацией: отчеты о сборе и обработке данных, схемы и структуры информационных модулей, инструкции по эксплуатации и обслуживанию системы.

Техническое задание и его согласованные изменения должны быть оформлены и загружены в SmartLMS.

#### Испытания и тестирование

Приемка работы проводится на основании результатов функциональных, надежностных, безопасностных и приемо-сдаточных испытаний, подтверждающих соответствие системы требованиям.

В ходе приемки проверяется работа системы с точки зрения конечного пользователя: студента, преподавателя и административного персонала.

#### Принятие и подпись акта

Результаты приемки фиксируются в акте приемки работы, подписываемом всеми участниками процесса (팀리더, научный руководитель, представители заказчика).

Работа признается принятой только после подтверждения выполнения всех требований и устранения выявленных замечаний.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. ГОСТ 19.101-77: Виды программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.102-77: Стадии разработки. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77: Обозначения программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78: Основные надписи. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78: Общие требования к программным документам. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78: Требования к программным документам, выполненным печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.201-78: Техническое задание. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.301-79: Программа и методика испытаний. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.401-78: Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ 19.404-79: Пояснительная записка. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
11. ГОСТ 19.505-79: Руководство оператора. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
12. ГОСТ 19.603-78: Общие правила внесения изменений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
13. ГОСТ 19.604-78: Правила внесения изменений в программные документы, выполненные печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

**ПРИЛОЖЕНИЕ. ССЫЛКИ НА АНАЛОГИ**

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05-09 Т3 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

## **ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ**