

**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**  
**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**  
**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**  
**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**  
**(РУТ (МИИТ))**



Рабочая программа дисциплины (модуля),  
как компонент образовательной программы  
базового высшего образования  
по направлению подготовки  
09.03.02 Информационные системы и технологии,  
утвержденной первым проректором РУТ (МИИТ)  
Тимониным В.С.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

**Генеративный искусственный интеллект**

Направление подготовки: 09.03.02 Информационные системы и технологии

Направленность (профиль): Технологии искусственного интеллекта в транспортных системах

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)  
ID подписи: 5665  
Подписал: заведующий кафедрой Нутович Вероника Евгеньевна  
Дата: 01.09.2026

## 1. Общие сведения о дисциплине (модуле).

Дисциплина формирует у студентов-бакалавров фундаментальные и прикладные компетенции в области создания программных продуктов на базе генеративных нейросетевых моделей. В условиях растущего кадрового дефицита инженеров по интеграции искусственного интеллекта курс обеспечивает глубокое понимание внутренней математики больших языковых моделей, вероятностной природы генерации и современных методов локального инференса. Студенты осваивают полный жизненный цикл разработки прикладных ИИ-приложений – от инжиниринга неструктурированных данных и продвинутого промпт-инжиниринга до программного конструирования цепочек с использованием специализированных фреймворков. Особое внимание уделяется технологиям retrieval-augmented generation и параметро-эффективной тонкой настройке моделей для работы в закрытых корпоративных контурах с соблюдением требований информационной безопасности и импортозамещения.

Целью освоения дисциплины является формирование у обучающихся способности самостоятельно проектировать, разрабатывать и развертывать локальные программные системы на базе генеративного искусственного интеллекта, применяя современные методы промпт-инжиниринга, доменной адаптации моделей и интеграции с векторными базами знаний для решения прикладных инженерных задач.

Для достижения поставленной цели в рамках дисциплины решается комплекс задач, направленных на формирование у обучающихся способности: анализировать архитектурные особенности генеративных моделей и механизмы самовнимания, применять методы квантования весов для оптимизации инференса на ограниченных аппаратных ресурсах, конструировать сложные программные цепочки и шаблоны промптов с использованием специализированных фреймворков, а также реализовывать механизмы защиты от галлюцинаций и несанкционированного доступа к данным.

## 2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

**ПК-6** - Способен проектировать и разрабатывать модели и системы искусственного интеллекта с применением современных библиотек, фреймворков и методологий машинного обучения.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

**Знать:**

- архитектуры и форматы представления неструктурированных документов, включая специфику синтаксиса и влияние шума на извлечение данных;
- методы и алгоритмы семантического разбиения текстового корпуса с учетом алгоритмов токенизации и лимитов контекстного окна;
- математические основы векторных представлений текста, стратегии пулинга и метрики близости в многомерном пространстве;
- алгоритмы индексации и приближенного поиска в векторных базах данных, включая графовые структуры и продуктивное квантование;
- архитектуру декодерных трансформеров, эволюцию механизма самовнимания и вероятностную природу предсказания токенов;
- математические и алгоритмические методы пост-тренировочного квантования весов нейросетей и форматы сжатия моделей;
- архитектурные принципы работы высокопроизводительных локальных серверов инференса и механизмы управления кэшем;
- теоретические основы и таксономию промпт-инжиниринга, включая когнитивные архитектуры и паттерны изоляции инструкций;
- принципы программного конструирования цепочек вызовов и управления памятью диалога в специализированных фреймворках;
- механизмы обеспечения информационной безопасности приложений, классификацию уязвимостей и концепцию многоступенчатой фильтрации;
- эволюцию архитектурных паттернов retrieval-augmented generation и стратегии гибридного поиска с реранжированием;
- математические и алгоритмические основы параметро-эффективной тонкой настройки больших языковых моделей на предметно-ориентированных корпусах;
- систему объективных метрик оценки качества генеративных систем и парадигму автоматизированного тестирования;
- принципы контейнеризации и оркестрации распределенных приложений для воспроизводимого развертывания в отечественных операционных системах;
- этические, правовые и регуляторные аспекты разработки генеративного интеллекта, включая защиту корпоративных данных и предотвращение смещений.

**Уметь:**

- уметь извлекать и структурировать текстовые и табличные данные из неструктурированных документов при помощи специализированных библиотек в условиях работы с зашумленными сканами технической документации;

- уметь проектировать стратегии семантического разбиения текстового корпуса при помощи алгоритмов фреймворков для сохранения логического контекста при ограничении размера окна модели;

- уметь генерировать и индексировать векторные представления текстов при помощи локальных моделей и векторных субд в условиях строгого запрета на передачу данных во внешние сервисы;

- уметь разворачивать и оптимизировать открытые большие языковые модели при помощи инструментов локального инференса с применением методов квантования для работы на оборудовании с ограниченной видеопамятью;

- уметь конструировать системные промпты и программные цепочки при помощи специализированных фреймворков в условиях жесткого требования к изоляции инструкций и управлению памятью диалога;

- уметь внедрять механизмы защиты от уязвимостей и фильтрации нежелательного контента при помощи выходных парсеров и эвристических фильтров в условиях соблюдения корпоративных политик безопасности;

- уметь собирать продвинутые пайплайны поиска с гибридным ранжированием при помощи векторных и классических индексов для минимизации галлюцинаций генеративных систем;

- уметь проводить доменную адаптацию открытых языковых моделей при помощи методов параметро-эффективной тонкой настройки и квантования в условиях ограниченных вычислительных ресурсов;

- уметь рассчитывать метрики качества генерации при помощи специализированных фреймворков оценки на основе размеченного тестового датасета для количественной валидации ответов системы;

- уметь разрабатывать пользовательские интерфейсы и упаковывать итоговый программный продукт при помощи фреймворков визуализации и систем контейнеризации для развертывания на серверах под управлением отечественных ос.

### **Владеть:**

- навыками настройки изолированных сред разработки и конвейеров обработки неструктурированных данных для нужд машинного обучения;

- методами программного управления генерацией и конструирования сложных шаблонов запросов с использованием современных фреймворков;

- инструментарием локального развертывания, квантования и доменного дообучения больших языковых моделей на потребительском оборудовании;

- подходами к оценке фактологической точности, внедрению защитных контуров и финальной упаковке прикладных ИИ-приложений в изолированные контейнеры.

### 3. Объем дисциплины (модуля).

#### 3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 3 з.е. (108 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Семестр №7
Контактная работа при проведении учебных занятий (всего):	64	64
В том числе:		
Занятия лекционного типа	32	32
Занятия семинарского типа	32	32

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 44 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

### 4. Содержание дисциплины (модуля).

#### 4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	<p><b>Введение в генеративный искусственный интеллект и эволюция архитектур</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- историческая ретроспектива развития генеративных моделей и переход к парадигме фундаментальных моделей;</li> <li>- классификация генеративных систем по модальностям и анализ современного рыночного спроса на инженеров;</li> <li>- концепция локальных импортозамещенных ИИ-решений в корпоративном секторе.</li> </ul>
2	<p><b>Архитектуры генеративных моделей: от вариационных автокодировщиков к состязательным сетям</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- математический аппарат вариационных автокодировщиков, латентное пространство и функция потерь;</li> <li>- архитектура генеративно-состязательных сетей и методы решения проблемы коллапса мод;</li> <li>- применение моделей для генерации синтетических данных и аугментации обучающих выборок.</li> </ul>
3	<p><b>Диффузионные модели генерации данных: математический аппарат и процессы шума</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- теория стохастических дифференциальных уравнений и процесс прямого зашумления данных;</li> <li>- архитектуры нейросетей для предсказания шума и алгоритмы обратного сэмплирования;</li> <li>- механизмы кондиционирования генерации и латентные диффузионные модели.</li> </ul>
4	<p><b>Архитектура декодерных трансформеров и вероятностная природа генерации</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- компонентная структура декодерного трансформера, позиционное кодирование и маскирование контекста;</li> <li>- математика механизма масштабированного точечного самовнимания и эволюция архитектур;</li> <li>- вероятностная природа предсказания токенов и влияние параметров сэмплирования на детерминированность.</li> </ul>
5	<p><b>Методы пост-тренировочного квантования и оптимизации весов нейронных сетей</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- теоретические основы квантования тензоров весов и компромисс между точностью и размером модели;</li> <li>- алгоритмы пост-тренировочного квантования и современные форматы сжатия моделей;</li> <li>- методы адаптивной тонкой настройки для эффективного дообучения на ограниченных ресурсах.</li> </ul>
6	<p><b>Архитектурные принципы локального инференса и управление памятью</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- проблема узкого горлышка памяти при генерации текста и концепция непрерывного батчинга;</li> <li>- алгоритмы эффективного управления кэшем в видеопамяти сервера;</li> <li>- сравнительный анализ локальных рантаймов для инференса и их применимость в корпоративном секторе.</li> </ul>
7	<p><b>Инжиниринг неструктурированных данных: парсинг и иерархия документов</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- внутренняя структура файлов и специфика извлечения текста из многоколоночных технических документов;</li> <li>- методы обработки зашумленных сканов и сохранение семантической иерархии заголовков;</li> <li>- стратегии нормализации текста и извлечение табличных данных.</li> </ul>
8	<p><b>Методы семантического разбиения текстового корпуса с учетом лимитов окна</b></p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- влияние алгоритмов токенизации на стратегию нарезки текста на смысловые фрагменты;</li> </ul>

№ п/п	Тематика лекционных занятий / краткое содержание
	- классификация сплиттеров и проектирование перекрытий для сохранения глобального контекста; - методы сохранения иерархии документов при векторизации.
9	<b>Математические основы векторных представлений текста и метрики близости</b> Рассматриваемые вопросы: - архитектура бидирекциональных трансформеров и процесс контрастивного обучения эмбеддеров; - стратегии агрегации токенов в единый вектор документа; - математический аппарат метрик расстояния и их влияние на качество поиска.
10	<b>Алгоритмы индексации и приближенного поиска в векторных базах данных</b> Рассматриваемые вопросы: - проблема исчерпывающего поиска и концепция приближенного поиска ближайших соседей; - алгоритмы на основе графов и инвертированных файлов с квантованием; - архитектурные отличия in-memory хранилищ от расширений для реляционных СУБД.
11	<b>Теоретические основы и таксономия промпт-инжиниринга</b> Рассматриваемые вопросы: - структура системного промпта и механизмы изоляции инструкций от пользовательского ввода; - паттерны проектирования промптов и их влияние на детерминированность выдачи; - когнитивные архитектуры промптинга для решения многошаговых логических задач.
12	<b>Программный промптинг и конструирование цепочек во фреймворках</b> Рассматриваемые вопросы: - архитектура фреймворков для программного связывания промптов и управления памятью диалога; - механизмы динамической подстановки контекста и работы с шаблонами; - стратегии обработки исключений и маршрутизации внутри программных цепочек.
13	<b>Механизмы обеспечения безопасности: уязвимости и строгий парсинг</b> Рассматриваемые вопросы: - таксономия уязвимостей языковых моделей и атаки типа инъекций; - архитектура многоступенчатой фильтрации на уровне входа и выхода нейросети; - методы строгого парсинга структурированных ответов и защиты от галлюцинаций.
14	<b>Эволюция архитектурных паттернов retrieval-augmented generation</b> Рассматриваемые вопросы: - ограничения базового пайплайна и проблемы потери релевантности при длинном контексте; - концепция продвинутого поиска с предретривальным анализом и пост-ретривальным реранжированием; - гибридный поиск и маршрутизация запросов между специализированными индексами.
15	<b>Доменная адаптация моделей методами параметро-эффективной тонкой настройки</b> Рассматриваемые вопросы: - математический аппарат низкоранговых адаптеров и их внедрение в слои трансформера; - алгоритмы квантования для обучения на потребительском оборудовании; - стратегии формирования инструктивных датасетов и предотвращения катастрофического забывания.
16	<b>Оценка качества генерации, метрики и основы инженерной эксплуатации</b> Рассматриваемые вопросы: - компонентная оценка пайплайна и метрики качества ретривера и генератора; - парадигма автоматизированного тестирования и проектирование оценочных промптов; - принципы контейнеризации и версионирования артефактов для воспроизводимого развертывания.

#### 4.2. Занятия семинарского типа.

##### Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	<p><b>Развертывание среды разработки и первичный парсинг нормативных документов</b>            Студент настраивает изолированное виртуальное окружение и устанавливает необходимые зависимости для работы с текстом. Обучающийся пишет скрипт на языке Python для извлечения сырого текста из предоставленных сканов технических регламентов. В процессе работы анализируется структура полученных страниц и выявляются артефакты распознавания. Результаты парсинга сохраняются в промежуточный текстовый формат для последующей очистки.</p>
2	<p><b>Очистка текстового корпуса и извлечение табличных данных</b>            Обучающийся применяет методы регулярных выражений и эвристические алгоритмы для удаления служебных символов и шума из сырого текста. Студент разрабатывает модуль для поиска и структурирования простых таблиц внутри нормативных документов с сохранением их в формат JSON. На этом этапе формируется первичный датасет, пригодный для дальнейшего семантического анализа. Готовый очищенный корпус экспортируется в локальное хранилище.</p>
3	<p><b>Проектирование стратегии семантического разбиения текста</b>            Студент реализует алгоритмы нарезки очищенного корпуса на смысловые фрагменты с учетом ограничений контекстного окна нейросети. Обучающийся экспериментирует с различными параметрами перекрытия соседних чанков для сохранения логических связей между предложениями. Итоговая конфигурация сплиттера фиксируется, а размеченные фрагменты сохраняются в структурированный файл. Данный массив данных выступает фундаментом для построения векторного индекса.</p>
4	<p><b>Генерация векторных представлений локальными моделями</b>            Обучающийся загружает предобученную локальную модель трансформера-энкодера для преобразования текстовых чанков в числовые векторы. Студент пишет конвейер пакетной обработки данных, оптимизируя использование оперативной памяти кафедры. В ходе работы анализируется размерность получаемых эмбеддингов и время инференса на тестовой выборке. Результирующие векторы подготавливаются к загрузке в специализированную базу данных.</p>
5	<p><b>Индексация данных и настройка векторной базы знаний</b>            Студент развертывает локальный экземпляр векторной базы данных в режиме сохранения на диск. Обучающийся конфигурирует метрику косинусного расстояния и загружает сгенерированные ранее эмбеддинги вместе с метаданными исходных документов. Проводится тестирование скорости индексации и проверяется целостность сохраненных записей. Настроенное хранилище становится центральным узлом для поиска релевантной информации.</p>
6	<p><b>Локальный деплой и квантование большой языковой модели</b>            Студент устанавливает локальный рантайм и загружает квантованную версию открытой языковой модели на рабочий компьютер. Обучающийся настраивает параметры генерации через локальный интерфейс, изучая влияние температуры и топ-сэмплирования на результат. В процессе работы измеряется потребление видеопамати и скорость генерации токенов на доступном оборудовании. Развернутая модель проверяется на способность отвечать на простые технические вопросы.</p>
7	<p><b>Базовый промпт-инжиниринг и изоляция системных инструкций</b>            Обучающийся конструирует базовые и системные промпты для управления стилем и форматом ответов языковой модели. Студент реализует механизм динамической подстановки извлеченного из базы контекста в тело запроса без нарушения синтаксиса инструкций. Проводится серия экспериментов по выявлению уязвимостей промптов к простым инъекциям. Оптимальные шаблоны запросов сохраняются в отдельный конфигурационный файл проекта.</p>
8	<p><b>Продвинутый промптинг и когнитивные архитектуры запросов</b>            Студент применяет паттерны многошагового рассуждения для решения сложных логических задач на основе извлеченных регламентов. Обучающийся экспериментирует с передачей нескольких примеров в контекст для повышения детерминированности выдачи модели. Проводится сравнительный анализ качества ответов при использовании различных когнитивных шаблонов. Успешные конфигурации интегрируются в общий модуль обработки запросов.</p>

№ п/п	Наименование лабораторных работ / краткое содержание
9	<p><b>Программные цепочки и шаблоны во фреймворке LangChain</b></p> <p>Обучающийся изучает архитектуру фреймворка для программного конструирования сложных цепочек вызовов модели. Студент реализует механизм динамической подстановки извлеченного контекста в тело запроса с использованием строгих шаблонов. Проводится серия экспериментов по изоляции системных инструкций от пользовательского ввода. Оптимальные конфигурации цепочек сохраняются в отдельный модуль проекта.</p>
10	<p><b>Управление памятью диалога и историей контекста</b></p> <p>Студент интегрирует механизмы сохранения истории переписки для поддержки многоходового взаимодействия с системой. Обучающийся настраивает стратегии ограничения объема памяти, чтобы история диалога не превышала лимиты контекстного окна модели. Тестируются сценарии, требующие от системы обращения к ранее упомянутым фактам из регламентов. Настроенный модуль памяти связывается с основной программной цепочкой.</p>
11	<p><b>Строгий парсинг структурированных ответов и Guardrails</b></p> <p>Обучающийся интегрирует выходные парсеры для строгой проверки сгенерированного текста на соответствие заданному формату JSON. Студент настраивает эвристические фильтры, которые принудительно прерывают генерацию и возвращают стандартизированный отказ при отсутствии релевантной информации. Тестируется реакция системы на провокационные вопросы, выходящие за рамки загруженных технических регламентов. Защитный контур документируется и фиксируется в коде приложения.</p>
12	<p><b>Сборка Advanced RAG-пайплайна с гибридным поиском</b></p> <p>Студент объединяет векторный поиск с классическим полнотекстовым ранжированием для повышения точности выдачи. Обучающийся программирует логику кросс-энкодерного реранжирования найденных фрагментов перед их передачей в языковую модель. Проводится тестирование системы на способность давать ответы исключительно на основе предоставленного контекста со ссылками на источники. Базовая версия продвинутого пайплайна выводится в консоль для отладки.</p>
13	<p><b>Доменная адаптация языковой модели методом параметро-эффективной тонкой настройки</b></p> <p>Студент формирует из извлеченных нормативных документов инструктивный датасет в формате пар запроса и эталонного ответа. Обучающийся конфигурирует параметры адаптеров и запускает процесс дообучения локальной модели с использованием квантования. В ходе работы анализируется динамика функции потерь и потребление видеопамати при обновлении весов. Полученная доменно-адаптированная модель сохраняется для интеграции в итоговый пайплайн.</p>
14	<p><b>Разработка пользовательского веб-интерфейса</b></p> <p>Обучающийся проектирует и реализует интерактивное веб-приложение на базе специализированного фреймворка для взаимодействия с RAG-системой. Студент программирует компоненты интерфейса для ввода запросов, отображения истории диалога и вывода кликабельных сносков на исходные страницы документов. Настраивается обработка состояний загрузки и отображение сообщений об ошибках при сбоях генерации. Готовый интерфейс связывается с бэкендом и тестируется в локальном браузере.</p>
15	<p><b>Оценка качества генерации и расчет объективных метрик</b></p> <p>Студент формирует тестовый датасет из пар вопрос-эталонный ответ на основе нормативных документов. Обучающийся применяет автоматизированные подходы для оценки фактологической точности и релевантности ответов сгенерированной системы. Результаты измерений визуализируются в виде простых отчетов для выявления узких мест в поиске или генерации. На основе полученных метрик студент вносит точечные корректировки в параметры ретривера.</p>
16	<p><b>Контейнеризация, документирование и защита портфолио</b></p> <p>Студент пишет оптимизированный конфигурационный файл для упаковки кода приложения и его зависимостей в изолированный образ. Обучающийся проводит комплексное приемочное тестирование развернутого приложения на соответствие критериям конфиденциальности и</p>

№ п/п	Наименование лабораторных работ / краткое содержание
	отсутствия галлюцинаций. Финализируется структура репозитория и пишется исчерпывающая документация с инструкциями по запуску проекта. Подготовленный программный продукт и отчет представляются преподавателю для итоговой аттестации.

#### 4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение рекомендованной литературы.
2	Подготовка к лабораторным работам.
3	Подготовка к промежуточной аттестации.
4	Подготовка к текущему контролю.

#### 5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Душкин, Р. В. Генеративный искусственный интеллект : руководство / Р. В. Душкин. — Москва : ДМК Пресс, 2025. — 228 с. — Текст : электронный	ЭБС Лань : [сайт]. — URL: <a href="https://e.lanbook.com/book/514902">https://e.lanbook.com/book/514902</a> (дата обращения: 10.06.2026)
2	Лекун, Я. Как учится машина: Революция в области нейронных сетей и глубокого обучения / Я. Лекун. — Москва : Альпина Паблишер. — Текст : электронный	ЭБС Лань : [сайт]. — URL: <a href="https://e.lanbook.com/book/213980">https://e.lanbook.com/book/213980</a> (дата обращения: 10.06.2026)
3	Глубокое обучение с TensorFlow, Keras и PyTorch. — Текст : электронный	ЭБС Лань : [сайт]. — URL: <a href="https://e.lanbook.com/book/514946">https://e.lanbook.com/book/514946</a> (дата обращения: 10.06.2026)
4	Паттерсон, Дж. Глубокое обучение с точки зрения практика / Дж. Паттерсон, А. Гибсон. — Текст : электронный	ЭБС Лань : [сайт]. — URL: <a href="https://e.lanbook.com/book/116122">https://e.lanbook.com/book/116122</a> (дата обращения: 10.06.2026)

#### 6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

Электронно-библиотечная система «Лань» (официальный доступ к учебным и монографическим изданиям).

Официальная документация открытых фреймворков для глубокого обучения и трансформеров (Hugging Face).

Официальная документация фреймворков для конструирования цепочек промптов и управления памятью (LangChain / LlamaIndex).

Фундаментальные руководства по таксономии промпт-инжиниринга и когнитивным архитектурам (Prompt Engineering Guide).

Документация локальных рантаймов для инференса и квантования моделей (Ollama / vLLM).

Официальные руководства по созданию интерактивных веб-интерфейсов для Data Science приложений (Streamlit).

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

операционные системы: отечественные дистрибутивы linux, включенные в единый реестр российского ПО и поддерживающие работу с контейнеризацией;

офисные пакеты: российские решения для подготовки технической документации, отчетов и презентаций в соответствии с требованиями ГОСТ;

среды разработки: популярные IDE и редакторы кода с поддержкой расширений для data science и изолированных виртуальных окружений;

технологический стек ИИ: открытые фреймворки для глубокого обучения, векторного поиска, параметро-эффективной тонкой настройки и конструирования программных цепочек;

системы управления базами данных: реляционные СУБД с поддержкой векторных расширений и специализированные in-memory хранилища эмбеддингов;

инструменты контейнеризации: платформы для изоляции зависимостей, упаковки итоговых приложений и их воспроизводимого развертывания.

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Учебные аудитории для проведения учебных занятий, оснащенные компьютерной техникой и наборами демонстрационного оборудования.

Для лабораторных занятий – наличие персональных компьютеров вычислительного класса.

9. Форма промежуточной аттестации:

Зачет в 7 семестре.

## 10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

старший преподаватель кафедры  
«Цифровые технологии управления  
транспортными процессами»

И.С. Разживайкин

Согласовано:

Заведующий кафедрой ЦТУТП

В.Е. Нутович

Председатель учебно-методической  
комиссии

Н.А. Андриянова