

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

АННОТАЦИЯ К
РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Обработка естественного языка

Направление подготовки: 09.03.02 – Информационные системы и технологии

Направленность (профиль): Технологии искусственного интеллекта в транспортных системах

Форма обучения: Очная

Общие сведения о дисциплине (модуле).

Дисциплина посвящена методам обработки естественного языка как направлению прикладного искусственного интеллекта. В ходе изучения рассматриваются подготовка текстовых корпусов, очистка и нормализация текста, токенизация, морфологический и синтаксический анализ, векторные представления слов и документов, классификация текстов, извлечение именованных сущностей, тематическое моделирование, архитектура Transformer, языковые модели, поиск по смысловой близости, RAG и оценка качества текстовых решений. На лабораторных занятиях обучающиеся последовательно создают воспроизводимое программное решение на Python для анализа русскоязычных текстов и оформляют техническую документацию.

Целью освоения дисциплины является формирование способности разрабатывать, проверять и документировать программные решения обработки естественного языка, обеспечивающие подготовку текстовых данных, извлечение признаков, классификацию, поиск сведений, работу с

языковыми моделями и оценку качества результата в прикладных задачах искусственного интеллекта.

Для достижения поставленной цели в рамках дисциплины решается комплекс задач, направленных на формирование у обучающихся способности – анализировать прикладную задачу обработки текста, собирать и подготавливать текстовый корпус, выполнять нормализацию и морфологический анализ, строить числовые представления текстов, обучать модели классификации и извлечения сущностей, применять архитектуру Transformer и языковые модели, реализовывать поиск по смысловой близости, проектировать RAG, оценивать качество текстового решения и готовить техническую документацию.

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