

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 22.05.2026 14:55:10

Уникальный программный ключ:

ca953a01204891083f939673078ef1a989dae18a

Федеральное государственное автономное образовательное учреждение высшего образования

«Российский университет дружбы народов имени Патриса Лумумбы»

Факультет искусственного интеллекта

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

АНАЛИЗ ЕСТЕСТВЕННОГО ЯЗЫКА С ПОМОЩЬЮ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

(наименование дисциплины/модуля)

Рекомендована МССН для направлений подготовки:

**02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ;**

09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗРАБОТКА И ОБУЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Анализ естественного языка с помощью методов искусственного интеллекта» входит в программу бакалавриата «Искусственный интеллект: разработка и обучение интеллектуальных систем» по направлениям подготовки 02.03.02 Фундаментальная информатика и информационные технологии и 09.03.03 Прикладная информатика, и изучается в 5 семестре 3 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 4 разделов и 34 тем и направлена на изучение продвинутых методов обработки естественного языка на основе глубокого обучения: нейросетевых архитектур для текстовых данных (RNN, LSTM, Transformer), предобученных языковых моделей (BERT, GPT, T5) и методов их дообучения (fine-tuning, adapter-tuning, prompt-tuning), решения прикладных задач NLP (классификация текстов, извлечение именованных сущностей, вопросно-ответные системы, суммаризация, машинный перевод, генерация текста), методов оценки качества NLP-моделей по специализированным метрикам (BLEU, ROUGE, BERTScore, F1-NER), а также практик развёртывания и критической оценки NLP-систем с учётом ограничений, предвзятостей и этических аспектов.

Целью освоения дисциплины является формирование у студентов практических навыков решения задач обработки естественного языка с использованием нейросетевых моделей и предобученных трансформеров, включая способность выбирать и дообучать языковые модели для конкретных прикладных задач, подготавливать и аннотировать текстовые данные, оценивать качество NLP-моделей по специализированным метрикам, анализировать ошибки и предвзятости, формулировать критерии приёмки NLP-систем, а также критически оценивать целесообразность и ограничения применения ИИ к задачам анализа текста.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Анализ естественного языка с помощью методов искусственного интеллекта» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ПК-3	Способен разрабатывать и реализовывать стратегии тестирования и контроля качества программного обеспечения систем ИИ	ПК-3.1 Верифицирует требования к ПО систем ИИ, определяет требования к тестам и критерии приёмки;
DL-1	Способен применять и (или) разрабатывать архитектуры глубоких нейронных сетей	DL-1.5 Разрабатывает и оптимизирует специализированные архитектуры для работы с последовательностями, учитывая их уникальные свойства;
DL-4	Способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка	DL-4.1 Применяет (проводя выбор и эксперименты) известные алгоритмы и библиотеки для обработки естественного языка, предобученные глубокие нейросетевые модели для прикладных задач анализа текстов, при необходимости дообучая и валидируя на собственных наборах данных; DL-4.2 Определяет стек технологий, методов и алгоритмов для построения продуктов с обработкой естественного языка (диалоговые системы, вопросно-ответные системы, рекомендательные системы и т.д.); DL-4.3 Имплементирует известные алгоритмы, архитектуры и модели обработки естественного языка на реальных данных,

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
		строит пайплайны обучения моделей и развертывания NLP-сервисов в продуктивной среде;
FC-2	Способен проводить передовые исследования в области фундаментальных и генеративных моделей	FC-2.1 Исследует и разрабатывает большие языковые модели (LLM) и другие модели для символьных данных;
SS-3	Способен к критическому анализу, метарефлексии и переносу знаний при работе с системами ИИ	SS-3.1 Учитывает в работе когнитивные искажения человека и примеры их проявления при работе с данными и ИИ, выявляет предвзятости систем ИИ, аргументированно оценивает надежность данных и выдачи ИИ, применяет базовые принципы критического мышления (оценка источников, проверка аргументов, отличие факта от интерпретации);

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Анализ естественного языка с помощью методов искусственного интеллекта» относится к блоку по выбору блока образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Анализ естественного языка с помощью методов искусственного интеллекта».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ПК-3	Способен разрабатывать и реализовывать стратегии тестирования и контроля качества программного обеспечения систем ИИ	Теория вероятностей и математическая статистика; Этика и безопасность использования искусственного интеллекта; Статистические методы и первичный анализ данных; Программирование на языке Python; Методы разработки решений на основе искусственного интеллекта (Git, Docker); Технологическая (проектно-технологическая) практика (учебная); Эксплуатационная практика (учебная);	Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная); Методы машинного обучения; Нейронные сети; Безопасность систем искусственного интеллекта; MLOps и промышленная разработка систем искусственного интеллекта; Проектирование и разработка систем компьютерного зрения; Практикум по обработке естественного языка (NLP); Оптимизация моделей машинного обучения; Практическая подготовка на проектах отраслевых промышленных партнеров;
SS-3	Способен к критическому анализу, метарефлексии и переносу знаний при	Эксплуатационная практика (учебная); Теория вероятностей и	Эксплуатационная практика (производственная); Технологическая (проектно-

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	работе с системами ИИ	математическая статистика; Искусственный интеллект и когнитивная психология; Этика и безопасность использования искусственного интеллекта; Статистические методы и первичный анализ данных; Правоведение; Введение в искусственный интеллект; Введение в компьютерное зрение;	технологическая) практика (производственная); Преддипломная практика; Методы машинного обучения; Нейронные сети; Безопасность систем искусственного интеллекта; <i>Вайб-кодиг**</i> ; Оптимизация моделей машинного обучения; MLOps и промышленная разработка систем искусственного интеллекта; Практическая подготовка на проектах отраслевых промышленных партнеров; Проектирование и разработка систем компьютерного зрения; Практикум по обработке естественного языка (NLP); <i>Основы робототехники**</i> ; <i>Цифровые двойники**</i> ; <i>Информационный поиск**</i> ; <i>Рекомендательные системы**</i> ; <i>Обработка сигналов**</i> ; <i>Анализ временных рядов**</i> ; <i>Большие языковые модели**</i> ;
DL-1	Способен применять и (или) разрабатывать архитектуры глубоких нейронных сетей		Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Основы глубокого обучения; Нейронные сети; Практическая подготовка на проектах отраслевых промышленных партнеров; Проектирование и разработка систем компьютерного зрения; <i>Генеративные модели**</i> ; Практикум по обработке естественного языка (NLP); <i>Большие языковые модели**</i> ; MLOps и промышленная разработка систем искусственного интеллекта;
DL-4	Способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка	Лингвистические основы анализа естественного языка;	Эксплуатационная практика (производственная); Практикум по обработке естественного языка (NLP);
FC-2	Способен проводить передовые исследования в области фундаментальных	Лингвистические основы анализа естественного языка; Введение в компьютерное зрение;	Преддипломная практика; Технологическая (проектно-технологическая) практика

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	и генеративных моделей		(производственная); Практикум по обработке естественного языка (NLP); <i>Большие языковые модели</i> **; Основы глубокого обучения; Нейронные сети; Проектирование и разработка систем компьютерного зрения; <i>Генеративные модели</i> **; Практическая подготовка на проектах отраслевых промышленных партнеров;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Анализ естественного языка с помощью методов искусственного интеллекта» составляет «4» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			5
<i>Контактная работа, ак.ч.</i>	68		68
Лекции (ЛК)	0		0
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	68		68
<i>Самостоятельная работа обучающихся, ак.ч.</i>	49		49
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
Общая трудоемкость дисциплины	ак.ч.	144	144
	зач.ед.	4	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы *	Формируемые индикаторы
Раздел 1	Нейросетевые представления текста и базовые архитектуры NLP	1.1	Обзор задач NLP и нейросетевых подходов	Карта задач NLP: классификация, NER, POS-tagging, sentiment analysis, QA, summarization, MT, generation. Эволюция подходов: правилые → статистические → нейросетевые → предобученные трансформеры. Экосистема инструментов: Hugging Face Transformers, spaCy, NLTK, datasets. Бенчмарки: GLUE, SuperGLUE, SQuAD, RuBenchmark. Обзор курса и практических проектов	СЗ	DL-4.1, DL-4.2
		1.2	Word embeddings: обучение и использование	Повторение: Word2Vec (CBOW, Skip-gram), GloVe. Fasttext: субсловные эмбединги, OOV-слова. Загрузка предобученных эмбедингов (gensim, Naves для русского). Визуализация: t-SNE, UMAP. Оценка качества: аналогии, similarity. Ограничения статических эмбедингов: полисемия. Практика: загрузка, сравнение, визуализация эмбедингов для русского и английского	СЗ	DL-4.1, FC-2.1
		1.3	Подготовка текстовых данных для нейросетей	Токенизация: пословная, субсловная (BPE, WordPiece, SentencePiece). Построение словаря. Padding и truncation. torchtext и Hugging Face tokenizers. Создание Dataset и DataLoader для текстов. Батчирование текстов переменной длины (pad_sequence, collate_fn). Практика: подготовка датасета для классификации текстов	СЗ	DL-1.5, DL-4.1
		1.4	Классификация текстов: RNN и LSTM	nn.Embedding → nn.LSTM → Linear. Bidirectional LSTM. Стратегии агрегации: last hidden, mean pooling, attention pooling. Обучение на задаче sentiment analysis. Метрики: accuracy, F1, classification report. Практика: классификация отзывов (IMDB / русскоязычный датасет), сравнение RNN/LSTM/GRU	СЗ	DL-1.5, DL-4.2
		1.5	CNN для текстов и TextCNN	1D свёртки для текста: nn.Conv1d. TextCNN (Kim, 2014): множественные размеры фильтров, max-over-time pooling. Сравнение с LSTM: скорость vs. quality. Практика: реализация TextCNN для классификации, сравнение с LSTM по метрикам и скорости обучения	СЗ	DL-1.5, DL-4.2

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы *	Формируемые индикаторы
		1.6	Sequence labeling: NER с BiLSTM-CRF	Задача sequence labeling: POS-tagging, NER, chunking. Форматы разметки: BIO, BIOES. BiLSTM для последовательностей. Conditional Random Field (CRF) как выходной слой: моделирование зависимостей между тегами. Метрики: entity-level F1 (seqeval). Практика: NER с BiLSTM-CRF на CoNLL-2003 / русскоязычном датасете	СЗ	DL-4.2, DL-4.1
		1.7	Seq2seq и механизм внимания для NLP	Seq2seq: encoder-decoder с LSTM. Проблема bottleneck. Attention (Bahdanau): alignment scores, context vector. Teacher forcing. Beam search vs. greedy decoding. Практика: реализация seq2seq с attention для простой задачи (date normalization / transliteration)	СЗ	DL-4.1, DL-1.5
		1.8	Оценка качества NLP-моделей: метрики	Метрики классификации: accuracy, precision, recall, F1 (macro/micro/weighted). NER: entity-level F1 (seqeval). Генерация: BLEU, ROUGE (ROUGE-1, ROUGE-2, ROUGE-L), METEOR, BERTScore. Перплексия для языковых моделей. Человеческая оценка: fluency, adequacy, factual consistency. Практика: вычисление всех метрик на учебных примерах	СЗ	DL-4.2, ПК-3.1
		1.9	Мини-проект: классификация текстов с нейросетями	Сквозная задача: выбор датасета (тональность, тематика, спам) → предобработка → обучение нескольких моделей (LSTM, TextCNN) → оценка по метрикам → анализ ошибок → формулирование критериев приёмки. Документирование: Model Card, описание датасета	СЗ	DL-4.2, ПК-3.1, DL-1.5
Раздел 2	Трансформеры и предобученные языковые модели	2.1	Архитектура Transformer для NLP	Повторение: self-attention, multi-head attention, positional encoding. Transformer encoder и decoder. Предобучение: MLM (Masked Language Modeling), CLM (Causal Language Modeling), Seq2Seq. Семейства моделей: encoder-only (BERT), decoder-only (GPT), encoder-decoder (T5, BART). Размеры моделей: base, large, количество параметров	СЗ	DL-4.1, FC-2.1
		2.2	BERT: архитектура, предобучение, fine-tuning	BERT: архитектура, предобучение (MLM + NSP). Tokenization: WordPiece, special tokens ([CLS], [SEP], [MASK]). Fine-tuning: добавление classification head поверх [CLS]. Hugging Face: AutoTokenizer, AutoModelForSequenceClassification. Trainer API. Практика: fine-tuning BERT для классификации текстов	СЗ	FC-2.1, DL-4.2

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы *	Формируемые индикаторы
		2.3	Fine-tuning BERT для NER и token classification	AutoModelForTokenClassification. Выравнивание токенов и меток при субсловной токенизации (label alignment). Fine-tuning на NER-датасете. Метрики: entity-level F1. Практика: fine-tuning BERT для NER на CoNLL / русскоязычном датасете. Сравнение с BiLSTM-CRF	СЗ	DL-4.2, FC-2.1
		2.4	Модели для русского языка	RuBERT, DeepPavlov BERT, XLM-RoBERTa, multilingual BERT. Сравнение моноязычных и мультиязычных моделей. Русскоязычные датасеты и бенчмарки: RuBenchmark, RuSentiment, NEREL, RuCoLA. Практика: fine-tuning русскоязычной модели на задаче классификации/NER, сравнение моделей	СЗ	DL-4.2, DL-4.1
		2.5	Вопросно-ответные системы: extractive QA	Задача extractive QA: найти span ответа в контексте. SQuAD: формат данных, метрики (Exact Match, F1). AutoModelForQuestionAnswering. Fine-tuning. Постобработка: выбор лучшего span. Практика: fine-tuning BERT для extractive QA на SQuAD / SberQuAD	СЗ	DL-4.2, FC-2.1
		2.6	Семантическое сходство и sentence embeddings	Sentence-BERT (SBERT): siamese architecture, mean pooling. Sentence-transformers library. Применения: semantic search, clustering, paraphrase detection. Cosine similarity. FAISS для эффективного поиска (обзор). Практика: semantic search по корпусу документов с помощью SBERT	СЗ	DL-4.2, FC-2.1
		2.7	Parameter-efficient fine-tuning: LoRA и adapters	Проблема: fine-tuning большой модели требует много ресурсов. Методы: adapter layers, LoRA (Low-Rank Adaptation), prefix-tuning, prompt-tuning (обзор). Библиотека PEFT (Hugging Face). Практика: fine-tuning с LoRA vs. full fine-tuning: сравнение качества, памяти, скорости	СЗ	FC-2.1, DL-4.1
		2.8	Анализ ошибок и предвзятостей NLP-моделей	Типичные ошибки: OOV, длинные тексты (truncation), domain mismatch, языковые нюансы (ирония, сарказм, диалект). Предвзятости: гендерные, расовые, культурные стереотипы в эмбедингах и моделях. Методы обнаружения: probing, counterfactual testing. Практика: выявление предвзятостей в BERT, анализ ошибок fine-tuned модели	СЗ	SS-3.1, DL-4.2
		2.9	Мини-проект: NLP-система на трансформерах	Сквозная задача: выбор задачи (классификация, NER, QA, similarity) → выбор предобученной модели → fine-tuning	СЗ	DL-4.2, ПК-3.1,

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы *	Формируемые индикаторы
				(full или LoRA) → оценка по метрикам → анализ ошибок → сравнение с baseline из раздела 1 → критерии приёмки. Документирование: Model Card, описание экспериментов		FC-2.1
Раздел 3	Генеративные модели и продвинутые задачи NLP	3.1	Генеративные модели: GPT, T5, BART	GPT: causal language modeling, авторегрессивная генерация. T5: text-to-text формат, предобучение (span corruption). BART: denoising autoencoder. Стратегии декодирования: greedy, beam search, top-k sampling, top-p (nucleus), temperature. Практика: генерация текста с GPT-2, сравнение стратегий декодирования	СЗ	DL-4.1, FC-2.1
		3.2	Суммаризация текстов	Extractive summarization: TextRank, выбор предложений. Abstractive summarization: T5, BART, mBART для русского. Fine-tuning для суммаризации. Метрики: ROUGE-1, ROUGE-2, ROUGE-L, BERTScore. Проблемы: галлюцинации, factual consistency. Практика: суммаризация новостей, сравнение extractive vs. abstractive	СЗ	DL-4.2, DL-4.3
		3.3	Машинный перевод	Neural Machine Translation (NMT): seq2seq + attention, Transformer. MarianMT, mBART, NLLB. Fine-tuning на параллельном корпусе. Метрики: BLEU, chrF, COMET. Проблемы: low-resource languages, domain adaptation. Практика: fine-tuning модели перевода на специализированном корпусе	СЗ	DL-4.2, DL-4.1
		3.4	Генерация текста: контролируемая генерация	Conditional generation: управление стилем, темой, длиной. Prompt engineering для генеративных моделей. Constrained decoding (обзор). Обнаружение сгенерированного текста (DetectGPT, обзор). Практика: генерация текста с контролем параметров, промпт-инженерия для различных задач	СЗ	DL-4.3, FC-2.1
		3.5	Анализ тональности: продвинутые методы	Aspect-based sentiment analysis (ABSA): извлечение аспектов + определение тональности по аспекту. Emotion detection: многоклассовая тональность (радость, гнев, страх). Stance detection. Обработка сарказма и иронии. Практика: ABSA с помощью fine-tuned трансформера	СЗ	DL-4.2, DL-4.3
		3.6	Relation Extraction и Information Extraction	Задача Relation Extraction: определение типа связи между сущностями в тексте. Подходы: classification-based (BERT + entity markers), generative (T5). Open Information Extraction.	СЗ	DL-4.2, FC-2.1

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы *	Формируемые индикаторы
				Построение триплетов (subject, relation, object). Практика: извлечение отношений из текстов, формирование структурированных данных		
		3.7	Мультиязычные модели и кросс-языковой перенос	XLM-RoBERTa, mBERT: обучение на множестве языков. Zero-shot cross-lingual transfer: обучение на английском, применение к русскому. Translate-train vs. translate-test. Практика: кросс-языковая классификация (обучение на EN → тестирование на RU), оценка потерь качества	СЗ	DL-4.1, FC-2.1
		3.8	Мини-проект: генеративная NLP-система	Сквозная задача: выбор генеративной задачи (суммаризация, перевод, генерация ответов, парафраз) → выбор модели (T5/BART/mT5) → fine-tuning → оценка (ROUGE/BLEU/BERTScore) → анализ качества (галлюцинации, factual consistency) → критерии приёмки. Документирование	СЗ	DL-4.3, ПК-3.1, DL-4.2
Раздел 4	Развёртывание, оценка и прикладные NLP-системы	4.1	Развёртывание NLP-моделей	Экспорт: ONNX для трансформеров (optimum). Оптимизация: distillation (DistilBERT), quantization (dynamic, static), pruning (обзор). FastAPI сервис: tokenization → inference → postprocessing. Docker. Замеры: latency, throughput. Практика: развёртывание NLP-модели как REST API	СЗ	DL-4.3, DL-1.5
		4.2	Retrieval-Augmented Generation (RAG)	Проблема: LLM галлюцинируют, знания устаревают. RAG: retriever (semantic search) + generator (LLM). Архитектура: документы → эмбединги → vector store (FAISS/Chroma) → поиск релевантных → формирование промпта → генерация ответа. LangChain / LlamaIndex (обзор). Практика: построение RAG-системы для QA по корпусу документов	СЗ	FC-2.1, DL-4.3
		4.3	Чат-боты и диалоговые системы	Типы: задаче-ориентированные (task-oriented) и открытые (open-domain). Intent classification + slot filling. Dialogue state tracking (обзор). Генеративные чат-боты: fine-tuning на диалогах. Оценка: BLEU, human evaluation, safety. Практика: создание простого чат-бота на основе fine-tuned модели	СЗ	DL-4.3, DL-4.2
		4.4	Text-to-SQL и работа со структурированными данными	Задача: перевод вопроса на естественном языке в SQL-запрос. Подходы: seq2seq, LLM-based (промпт с описанием схемы). Бенчмарки: Spider, WikiSQL. Практика: реализация	СЗ	DL-4.3, FC-2.1

Номер раздела	Наименование раздела дисциплины	Наименование темы	Содержание темы	Вид учебной работы *	Формируемые индикаторы
			text-to-SQL с помощью LLM (промпт-инженерия) на учебной БД		
		4.5 Мультимодальные модели: текст + изображения	CLIP: contrastive learning текст-изображение. BLIP: image captioning, VQA. Практика: image captioning на пользовательских изображениях, visual QA, поиск изображений по текстовому запросу. Обсуждение: границы мультимодального понимания	СЗ	FC-2.1, DL-4.1
		4.6 Критическая оценка NLP-систем	Когда NLP работает хорошо, когда — плохо: зависимость от домена, языка, объёма данных. Проблемы: галлюцинации, confabulation, bias, toxicity. Оценка рисков: sensitive domains (медицина, право, финансы). Критерии приёмки NLP-системы: accuracy, coverage, safety, fairness. Практика: составление чек-листа критической оценки NLP-модели	СЗ	SS-3.1, ПК-3.1
		4.7 Формулирование ТЗ и критериев приёмки NLP-системы	Процесс: бизнес-задача → анализ осуществимости → требования к данным и качеству → выбор подхода → критерии приёмки (метрики, edge cases, safety, latency). Примеры из промышленности: автоклассификация обращений, автоматическая суммаризация, QA для поддержки. Практика: составление ТЗ для NLP-системы	СЗ	ПК-3.1, DL-4.3
		4.8 Итоговый проект: прикладная NLP-система	Финальная интеграция: выбор прикладной задачи (классификация обращений, NER для документов, QA по базе знаний, суммаризация, чат-бот, RAG) → подготовка данных → выбор и дообучение модели → оценка по метрикам → анализ ошибок и предвзятостей → развёртывание (API + Docker) → критерии приёмки. Документирование: Model Card, ТЗ, план тестирования. Презентация и peer review	СЗ	DL-4.2, DL-4.3, ПК-3.1, FC-2.1

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	Персональные компьютеры, необходимое ПО
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	Персональные компьютеры, необходимое ПО

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Язык и искусственный интеллект: проблемы, вызовы, решения [Электронный ресурс] : монография / Н. Ю. Бородулина, Е. Ю. Воякина, В. С. Григорьева и др. ; под общ. ред. Л. Г. Поповой. – Тамбов : Издательский центр ФГБОУ ВО «ТГТУ», 2025. ISBN 978-5-8265-2940-9

2. Лейн, Х. Обработка естественного языка в действии : практическое руководство для разработчиков / Х. Лейн, М. Дышель. — 2-е изд. — Санкт-Петербург : Питер, 2026. — 736 с. — (Для профессионалов). — ISBN 978-5-4461-4385-6.

Дополнительная литература:

1. Гольдберг Й. Нейросетевые методы в обработке естественного языка / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – 282 с.: ил. ISBN 978-5-97060-754-1

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН

<https://mega.rudn.ru/MegaPro/Web>

- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>

- ЭБС «Юрайт» <http://www.biblio-online.ru>

- ЭБС «Консультант студента» www.studentlibrary.ru

- ЭБС «Знаниум» <https://znanium.ru/>

2. Базы данных и поисковые системы

- Sage <https://journals.sagepub.com/>

- Springer Nature Link <https://link.springer.com/>

- Wiley Journal Database <https://onlinelibrary.wiley.com/>

- Наукометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**