

Документ подписан простой электронной подписью

Информация о владельце:

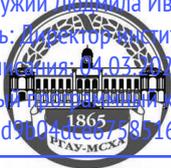
ФИО: Хоружий Людмила Ивановна

Должность: Директор института экономики и управления АПК

Дата подписания: 04.03.2025 16:54:58

Уникальный идентификатор документа:

1e90b132d910a4cc67581160b015dddf2cb1e6a9



МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ –
МСХА имени К.А. ТИМИРЯЗЕВА»
(ФГБОУ ВО РГАУ - МСХА имени К.А. Тимирязева)

Институт экономики и управления АПК
Кафедра Прикладной информатики

УТВЕРЖДАЮ:
Директор института
экономики и управления АПК
Л.И. Хоружий
« 28 » 08 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.05 «Обработка естественного языка»

для подготовки магистров

ФГОС ВО

Направление: 09.04.03 «Прикладная информатика»

Направленность: «Архитектура систем искусственного интеллекта»

Курс 2

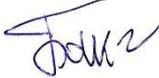
Семестр 4

Форма обучения: очная

Год начала подготовки: 2025

Москва, 2025

Разработчик: Бакнин М.Д., к.т.н.
(ФИО, ученая степень, ученое звание)


«28» августа 2025г.

Разработчик: Греченева А.В., к.т.н.
(ФИО, ученая степень, ученое звание)


«28» августа 2025г.

Рецензент: Ашмарина Т.И. к.э.н.
(ФИО, ученая степень, ученое звание)


«28» августа 2025г.

Программа составлена в соответствии с требованиями ФГОС ВО, профессионального стандарта и учебного плана по направлению подготовки 09.04.03 «Прикладная информатика».

Программа обсуждена на заседании кафедры прикладной информатики протокол № 1 от «28» августа 2025г.

Зав. кафедрой прикладной информатики
д.э.н., профессор Худякова Е.В.

(подпись)



«28» августа 2025г.

Согласовано:

Председатель учебно-методической
комиссии института экономики и управления АПК
к.э.н., доцент Гупалова Т.Н.

(подпись)



«28» августа 2025г.

Заведующий выпускающей кафедрой
прикладной информатики
д.э.н., профессор Худякова Е.В.

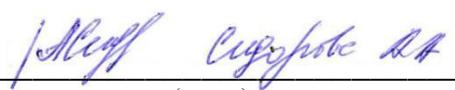
(подпись)



«28» августа 2025г.

Заведующий отделом комплектования ЦНБ

(подпись)



СОДЕРЖАНИЕ

<u>АННОТАЦИЯ</u>	4
<u>1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ</u>	4
<u>2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ</u>	5
<u>3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ</u>	5
<u>4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ</u>	5
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ ПО СЕМЕСТРАМ	5
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	7
4.3 ЛЕКЦИИ/ ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	8
<u>5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ</u>	9
<u>6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ</u>	10
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ	10
6.2. ОПИСАНИЕ ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ КОНТРОЛЯ УСПЕВАЕМОСТИ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ	11
<u>7.</u>	Ошибка! Закладка не определена.
7.1. ОСНОВНАЯ ЛИТЕРАТУРА	13
7.2. ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	13
7.3. НОРМАТИВНЫЕ ПРАВОВЫЕ АКТЫ	14
<u>8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ</u>	14
<u>9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ</u>	15
<u>10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ</u>	15
<u>11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ОБУЧАЮЩИМСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ</u>	16
<u>12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ</u>	17

АННОТАЦИЯ

рабочей программы учебной дисциплины Б1.В.05 «Обработка естественного языка»

для подготовки магистра по направлению 09.04.03 «Прикладная информатика», направленности «Архитектура систем искусственного интеллекта»

Цель освоения дисциплины: формирование у магистрантов компетенций в области разработки и внедрения систем обработки естественного языка (NLP), включая овладение современными методами анализа текстовых данных, архитектурами нейронных сетей и практическими навыками работы с предобученными языковыми моделями для решения прикладных задач искусственного интеллекта.

Задачи дисциплины:

1) Сформировать глубокое понимание принципов работы современных архитектур обработки естественного языка (RNN, Transformer, предобученные модели).

2) Научить магистрантов разрабатывать и адаптировать языковые модели для решения конкретных прикладных задач с использованием современных фреймворков.

3) Обучить методам предобработки текстовых данных, построению качественных датасетов и их аннотирования.

4) Развить навыки оценки качества NLP-моделей, проведения экспериментов и анализа результатов с применением строгих метрик.

5) Подготовить к внедрению NLP-систем в production, включая тестирование, валидацию, мониторинг и управление рисками.

6) Привить практические умения применения обработки естественного языка к реальным бизнес-задачам в контексте АПК.

Место дисциплины в учебном плане: учебная дисциплина Б1.В.05 «Обработка естественного языка» включена в часть, формируемую участниками образовательных отношений учебного плана по направлению подготовки 09.04.03 «Прикладная информатика», дисциплина осваивается на 2 курсе в 4 семестре.

Требования к результатам освоения дисциплины: в результате освоения дисциплины «Обработка естественного языка» формируются следующие компетенции (индикаторы) их достижения: ПКос-12.1; ПКос-12.2; ПКос-12.3; ПКос-4.1; ПКос-4.2; ПКос-4.3; ПКос-10.1; ПКос-10.2; ПКос-10.3; ПКос-11.1; ПКос-11.2; ПКос-11.3

Краткое содержание дисциплины: дисциплина «Обработка естественного языка» охватывает современные подходы обработки текстовых данных и развития языковых моделей на основе нейросетевых архитектур. Изучаются

основы NLP: токенизация, лемматизация, морфологический анализ, синтаксический разбор, особенности русского языка. Рассматриваются классические методы векторизации (Bag-of-Words, TF-IDF, Word2Vec, GloVe) и современные подходы получения контекстных эмбеддингов (ELMo, BERT, GPT).

Основной акцент сделан на архитектурах нейронных сетей для NLP: RNN/LSTM/GRU, механизме внимания (Attention) и Transformer-архитектуре. Подробно изучаются предобученные языковые модели (BERT, GPT, T5, BART) и русскоязычные варианты (ruBERT, ruGPT-3, YandexGPT). Рассматриваются стратегии fine-tuning, prompt engineering, параметрически-эффективные методы (LoRA, QLoRA).

Практический блок включает основные задачи NLP: Named Entity Recognition, machine translation, text summarization, question answering, seq2seq модели и RAG подходы. Завершающий раздел посвящён тестированию, валидации, обнаружению bias, интерпретируемости моделей (LIME, SHAP), мониторингу в production и развёртыванию NLP-систем через FastAPI, TorchServe и облачные сервисы.

Общая трудоемкость дисциплины: 144/4 (часы/зач. ед.), в том числе 4 часа практической подготовки.

Промежуточный контроль: экзамен.

1. Цель освоения дисциплины

Целью освоения дисциплины «Обработка естественного языка» магистрами направления 09.04.03 «Прикладная информатика» по направленностям «Архитектура систем искусственного интеллекта» является формирование у обучающихся теоретических, методических и технологических основ разработки и применения современных методов обработки естественного языка, включая проектирование и внедрение NLP-систем, ориентированных на решение задач анализа текстовых данных, извлечения информации, генерации текстов, машинного перевода и построения интеллектуальных информационных систем в прикладных областях.

Особое внимание уделяется изучению архитектур нейронных сетей для NLP (RNN, LSTM, GRU, Transformer), механизмов внимания, предобученных языковых моделей (BERT, GPT, T5, BART) и методов их адаптации, а также практическому использованию современных фреймворков (Hugging Face Transformers, spaCy, PyTorch, TensorFlow) и инструментов для разработки, тестирования и развёртывания NLP-компонентов в production-среде.

Значимость формирования компетенций в области обработки естественного языка в процессе профессиональной подготовки специалистов в области ИТ обусловлена требованиями цифровой трансформации экономики, возрастающей роли больших языковых моделей в автоматизации интеллектуальных задач, а также приоритетами государственной политики Российской Федерации в области внедрения искусственного интеллекта и интеллектуальных технологий. Освоение дисциплины направлено на обеспечение готовности выпускников к

практико-ориентированной деятельности в условиях цифрового общества, включая разработку и управление NLP-системами в сельском хозяйстве, пищевой промышленности и других отраслях АПК.

2. Место дисциплины в учебном процессе

Дисциплина «Обработка естественного языка» относится к части, формируемой участниками образовательных отношений учебного плана по направлению подготовки 09.04.03 «Прикладная информатика». Дисциплина реализуется в соответствии с требованиями ФГОС ВО, основной профессиональной образовательной программы (ОПОП), профессиональных стандартов в области информационных технологий и искусственного интеллекта, осваивается в 4 семестре магистратуры.

Предшествующими дисциплинами, на которых непосредственно базируется курс «Обработка естественного языка», являются: «Архитектура и организация ЭВМ», «Информационные системы и технологии», «Основы программирования», «Теоретические основы информатики». Данные курсы обеспечивают необходимую подготовку в области программирования, математических основ машинного обучения и принципов проектирования информационных систем.

Дисциплина «Обработка естественного языка» является основополагающей для изучения последующих дисциплин: «Интеллектуальный анализ данных и Business Intelligence», «Проектирование и управление информационными системами в АПК», «Системы принятия решений на основе ИИ», «Big Data и облачные вычисления». Полученные в ходе изучения курса знания и умения являются необходимыми при выполнении выпускной квалификационной работы (магистерской диссертации) и практической деятельности в области разработки интеллектуальных информационных систем.

Рабочая программа дисциплины «Обработка естественного языка» для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учётом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся в соответствии с требованиями законодательства Российской Федерации об образовании и локальными нормативными актами университета.

3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Образовательные результаты освоения дисциплины обучающимся, представлены в таблице 1.

4. Структура и содержание дисциплины

4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам

Общая трудоёмкость дисциплины составляет 4 зач. единиц (144 часов), их распределение по видам работ представлено в табл. 2.

Таблица 1

Требования к результатам освоения учебной дисциплины

№ п/п	Индекс компетенции	Содержание компетенции (или её части)	Индикатор компетенций	В результате выполнения курсового проекта по учебной дисциплине обучающиеся должны:		
				знать	уметь	владеть
	ПКос-12	Способность использовать и развивать методы научных исследований и инструментария в области проектирования и управления информационными системами в прикладных областях	ПКос-12.1 Знает алгоритмы методов научного исследования проектирования и управления информационными системами в экономике АПК	<ul style="list-style-type: none"> - алгоритмы методов научного исследования в области обработки естественного языка и управления NLP-системами в экономике АПК; - современные архитектуры языковых моделей (BERT, GPT, T5, BART) и методы их применения; - принципы работы трансформеров и механизмов внимания (Attention, Self-Attention, Cross-Attention); - методологию экспериментального исследования NLP-моделей; - метрики качества NLP-систем (BLEU, ROUGE, METEOR, BERTScore, perplexity) - подходы к fine-tuning и transfer learning в обработке естественного языка. 	-	-

			<p>ПКос-12.2 Применяет методы научного исследования проектирования и управления информационными системами в экономике АПК</p>	-	<ul style="list-style-type: none"> - применять методы научного исследования для разработки и оценки NLP-систем в экономике АПК; - проектировать архитектуру NLP-решений с использованием фреймворков Hugging Face Transformers, spaCy, NLTK; - проводить сравнительный анализ языковых моделей на базе PyTorch и TensorFlow; - адаптировать предобученные модели (ruBERT, ruGPT, mBART) для специфических задач; - разрабатывать методики оценки качества текстовых данных и результатов обработки; - использовать инструменты для аннотирования данных (Label Studio, Prodigy, Doccano). 	-
--	--	--	--	---	---	---

			<p>ПКос-12.3 Владеет инструментарием научного исследования проектирования и управления информационными системами в экономике АПК</p>	-	-	<ul style="list-style-type: none"> - инструментарием научного исследования NLP-систем (Weights & Biases, MLflow, TensorBoard); - навыками работы с фреймворками Hugging Face Transformers, spaCy, AllenNLP; - технологиями векторного представления текстов (Word2Vec, FastText, GloVe, sentence-transformers); - методами построения и анализа корпусов текстов с использованием NLTK, TextBlob; - инструментами для работы с большими языковыми моделями (LangChain, LlamaIndex); - практикой использования API открытых языковых моделей (OpenAI API, Anthropic Claude API, YandexGPT API).
--	--	--	--	---	---	--

	ПКос-4	Способность принимать эффективные проектные решения в условиях неопределенности и риска	ПКос-4.1 Знает как принимать эффективные проектные решения в условиях неопределенности и риска	<ul style="list-style-type: none"> - методы принятия проектных решений при разработке NLP-систем в условиях зашумленных и неразмеченных данных; - подходы к оценке качества и репрезентативности текстовых корпусов; - стратегии работы с несбалансированными датасетами в задачах классификации текстов; - методы количественной оценки неопределенности предсказаний языковых моделей (Uncertainty Estimation); - риски применения NLP-моделей (bias, hallucinations, adversarial attacks); - техники обеспечения надежности и безопасности NLP-систем. 	-	-
--	--------	---	--	---	---	---

			<p>ПКос-4.2 Умеет принимать эффективные проектные решения в условиях неопределенности и риска</p>	-	<ul style="list-style-type: none"> - принимать обоснованные решения по выбору архитектуры NLP-моделей в условиях ограниченных данных; - применять техники data augmentation для текстов (back-translation, paraphrasing, EDA); - использовать методы active learning и weak supervision для снижения затрат на разметку; - работать с инструментами синтетической генерации данных (ChatGPT, GPT-4, Anthropic Claude); - оценивать trade-off между качеством модели, вычислительными затратами и скоростью inference; - применять техники квантизации и pruning моделей (ONNX, TensorRT, Optimum); - использовать A/B-тестирование для оценки эффективности NLP-решений. 	-
--	--	--	---	---	---	---

			<p>ПКос-4.3 Владеет инструментами обоснования эффективных проектных решений в условиях неопределенности и риска</p>			<ul style="list-style-type: none"> - инструментами мониторинга и диагностики NLP-систем в production (Evidently AI, WhyLabs, Fiddler); - методами интерпретации и объяснения решений моделей (LIME, SHAP, Captum, InterpretML); - техниками обнаружения и митигации bias в языковых моделях; - навыками работы с фреймворками для управления рисками ML-систем (Fairlearn, AI Fairness 360); - практикой версионирования моделей и экспериментов (DVC, Neptune.ai); - методами оценки экономической эффективности внедрения NLP-решений.
--	--	--	---	--	--	--

	ПКос-10	Способность выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования	ПКос-10.1 Знает методы и инструменты для тестирования и валидации программных компонентов систем искусственного интеллекта	<ul style="list-style-type: none"> - методы тестирования и валидации NLP-компонентов (unit testing, integration testing, regression testing); - метрики качества для различных NLP-задач: классификации (accuracy, F1, precision, recall), генерации (BLEU, ROUGE, METEOR), информационного поиска (MRR, NDCG, MAP); - принципы оценки качества языковых моделей через perplexity и cross-entropy; - стандарты и бенчмарки для оценки NLP-систем (GLUE, SuperGLUE, XNLI, Russian SuperGLUE); - методы оценки производительности и latency моделей в inference; - подходы к стресс-тестированию NLP-систем (adversarial testing, robustness testing). 	-	-
--	---------	--	--	--	---	---

			<p>ПКос-10.2 Умеет выбирает подходящие методики для экспериментальной проверки работоспособности программных компонентов в соответствии с заданными критериями эффективности и качества</p>		<ul style="list-style-type: none"> - разрабатывать комплексные планы тестирования NLP-компонентов с использованием pytest, unittest; - проводить функциональное тестирование text processing pipelines в spaCy и Hugging Face; - применять инструменты для автоматического тестирования NLP-моделей (TextAttack, Checklist, robustness-gym); - выполнять профилирование производительности моделей с помощью cProfile, line_profiler, memory_profiler; - использовать фреймворки для нагрузочного тестирования API (Locust, Apache JMeter); - проводить error analysis и создавать confusion matrices для анализа ошибок классификации; - применять cross-validation и stratified sampling для надежной оценки качества моделей 	
--	--	--	---	--	--	--

			<p>ПКос-10.3 Владеет навыками разработки тестовых сценариев и проведения экспериментальных исследований для оценки производительности и надежности программных компонентов ИИ</p>	-	-	<ul style="list-style-type: none"> - навыками разработки тестовых датасетов и benchmark-наборов для NLP-задач; - инструментами continuous testing в CI/CD pipeline (GitHub Actions, GitLab CI, Jenkins); - методами автоматизированного тестирования с использованием Great Expectations для валидации данных; - техниками canary testing и blue-green deployment для безопасного развертывания NLP-моделей; - навыками работы с фреймворками для мониторинга качества моделей (Prometheus, Grafana, ELK stack); - практикой проведения A/B-экспериментов и multivariate testing; - инструментами для отладки трансформеров (bertviz, exBERT, Captum).
--	--	--	---	---	---	---

	ПКос-11	Способность разрабатывать и применять методы и алгоритмы машинного обучения для решения задач	ПКос-11.1 Знает парадигмы, методы и алгоритмы машинного обучения, включая обучение с учителем, без учителя, подкрепление и глубокое обучение	<ul style="list-style-type: none"> - архитектуры нейронных сетей для NLP: RNN, LSTM, GRU, CNN для текстов, Transformer, BERT, GPT; - методы обучения языковых моделей: supervised learning, self-supervised learning (MLM, CLM, NSP), few-shot и zero-shot learning; - алгоритмы для основных NLP-задач: токенизация (BPE, WordPiece, SentencePiece), NER, POS-tagging, dependency parsing, sentiment analysis, text generation, machine translation; - принципы работы энкодеров (BERT, RoBERTa, ELECTRA) и декодеров (GPT, GPT-2, GPT-3, GPT-4); - архитектуры encoder-decoder моделей (T5, BART, mT5, mBART); - методы prompt engineering и instruction tuning для больших языковых моделей; - техники LoRA, QLoRA для эффективного fine-tuning крупных моделей. 	-	-
--	---------	---	--	--	---	---

			<p>ПКос-11.2 Умеет применять алгоритмы машинного обучения для анализа данных, распознавания образов, предсказания и оптимизации в соответствии с конкретными задачами и требованиями</p>		<ul style="list-style-type: none"> - применять предобученные модели из Hugging Face Hub для задач классификации, NER, question answering; - выполнять fine-tuning BERT-подобных моделей с использованием библиотеки transformers; - реализовывать custom tokenizers с помощью tokenizers library (BPE, WordPiece, Unigram); - разрабатывать text classification pipelines с использованием sklearn, PyTorch, TensorFlow; - применять seq2seq модели для задач машинного перевода и суммаризации с помощью fairseq, OpenNMT; - использовать техники prompt engineering для работы с LLM (GPT-4, Claude, YandexGPT); - применять RAG (Retrieval-Augmented Generation) подходы с использованием LangChain, LlamaIndex. работать с векторными базами данных для semantic search (Pinecone, Weaviate, Milvus, ChromaDB, FAISS). 	
--	--	--	--	--	--	--

			<p>ПКос-11.3 Владеет инструментами и библиотеками для реализации методов машинного обучения, такими как TensorFlow, PyTorch, Scikit-learn, в контексте разработки программного обеспечения систем искусственного интеллекта</p>			<ul style="list-style-type: none"> - навыками разработки NLP-приложений с использованием PyTorch, TensorFlow, JAX/Flax; - фреймворками для быстрой разработки NLP-решений: Hugging Face Transformers, spaCy, AllenNLP, Flair; - инструментами для работы с русскоязычными моделями: DeepPavlov, ruBERT, ruGPT-3, YandexGPT; - библиотеками для предобработки текста: NLTK, TextBlob, pymorphy2, razdel, natasha; - техниками работы с эмбедингами: gensim (Word2Vec, FastText), sentence-transformers; - инструментами для создания chatbots и conversational AI: Rasa, Botpress, DialogFlow; - практикой развертывания NLP-моделей: FastAPI, Flask, TorchServe, TensorFlow Serving, ONNX Runtime; - методами оптимизации inference: mixed precision, quantization (int8, int4), distillation; - навыками работы с облачными NLP-сервисами: AWS Comprehend, Google Cloud
--	--	--	---	--	--	---

						NLP, Azure Cognitive Services, Yandex Cloud.
--	--	--	--	--	--	--

Таблица 2

Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость	
	час. всего/*	в т.ч. по семестрам
		№4
Общая трудоёмкость дисциплины по учебному плану	144/4	144/4
1. Контактная работа:		
Аудиторная работа:	46,4/4	46,4/4
<i>в том числе:</i>		
<i>лекции (Л)</i>	18	18
<i>практические занятия (ПЗ)</i>	26/4	26/4
<i>курсовой проект (консультация, защита)</i>		
<i>консультация перед экзаменом</i>	2	2
<i>контактная работа на промежуточном контроле (КРА)</i>	0,4	0,4
2. Самостоятельная работа (СРС)	73	73
<i>в том числе:</i>		
<i>курсовой проект (подготовка)</i>		
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к практическим занятиям, и т.д.)</i>	57	57
<i>подготовка к экзамену</i>	24,6	24,6
Вид промежуточного контроля:	Экзамен	

* в том числе практическая подготовка

4.2 Содержание дисциплины

Таблица 3

Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	ПЗ/С всего/*	ПКР	
Тема 1. Введение в обработку естественного языка и лингвистические основы	16	2	4	-	10
Тема 2. Векторные представления текстов и эмбединги	16	2	4	-	10
Тема 3. Архитектуры нейронных сетей для NLP: от RNN до Transformer	20/2	4	6/2	-	10
Тема 4. Предобученные языковые модели и механизмы внимания	18	4	4	-	10

Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа			Внеаудиторная работа СР
		Л	ПЗ/С всего/*	ПКР	
Тема 5. Прикладные задачи NLP: классификация, NER, генерация текстов	18	4	4/2	-	10
Тема 6. Тестирование, валидация и развертывание NLP-систем	29	2	4	-	23
Подготовка к экзамену	24,6	-	-	-	-
Консультация перед экзаменом	2	-	-	-	-
Контактная работа на промежуточном контроле (КРА)	0,4	-	-	0,4	-
Всего за 4 семестр	144/4	18	26/4	0,4	73
Итого по дисциплине	144/4	18	26/4	0,4	73

* в том числе практическая подготовка

Тема 1 Введение в обработку естественного языка и лингвистические основы

Основы обработки естественного языка: парадигмы, методы и архитектура NLP-систем. Эволюция методов NLP: от rule-based к статистическим и нейросетевым подходам. Основные задачи NLP: токенизация, лемматизация, POS-tagging, синтаксический анализ. Лингвистические основы: морфология, синтаксис, семантика русского языка. Проблемы обработки естественного языка: многозначность, контекст, морфологическая сложность. Обзор современных архитектур NLP-систем в экономике АПК. Инструментарий: NLTK, spaCy, pymorphy2, natasha.

Тема 2 Векторные представления текстов и эмбединги

Векторные представления слов и документов: от Bag-of-Words до трансформеров. Классические методы: Bag-of-Words, TF-IDF, n-граммы. Дистрибутивные модели: Word2Vec (CBOW, Skip-gram), GloVe. Контекстные эмбединги: ELMo, contextualized representations. Векторные представления предложений и документов: Doc2Vec, sentence-transformers. FastText для морфологически богатых языков (русский). Метрики схожести: косинусное расстояние, евклидова метрика. Визуализация эмбедингов: PCA, t-SNE, UMAP

Тема 3 Архитектуры нейронных сетей для NLP: от RNN до Transformer

Рекуррентные нейронные сети и их модификации для обработки последовательностей. Принципы работы RNN: обработка последовательностей, backpropagation through time. Проблемы vanishing/exploding gradients. LSTM (Long Short-Term Memory): архитектура, gates mechanism. GRU (Gated Recurrent Unit): упрощённая альтернатива LSTM. Bidirectional RNN для учёта контекста. CNN для обработки текстов: 1D convolutional layers. Применение в задачах sequence labeling, text classification. Механизм внимания и архитектура Transformer. Механизм внимания (Attention): query, key, value. Self-Attention и Multi-Head Attention. Архитектура Transformer: encoder-decoder structure. Positional encoding для учёта позиции токенов. Scaled Dot-Product Attention.

Преимущества Transformer перед RNN: параллелизация, long-range dependencies. Feed-Forward Networks, Layer Normalization, Residual Connections

Тема 4 Предобученные языковые модели и механизмы внимания

Архитектуры предобученных языковых моделей: BERT, GPT, T5. Концепция transfer learning в NLP. Encoder-only модели: BERT, RoBERTa, ELECTRA, DeBERTa. Masked Language Modeling (MLM) и Next Sentence Prediction (NSP). Decoder-only модели: GPT, GPT-2, GPT-3, GPT-4. Causal Language Modeling (CLM), autoregressive generation. Encoder-decoder модели: T5, BART, mT5, mBART. Русскоязычные модели: ruBERT, ruGPT-3, ruT5, YandexGPT. Особенности multilingual models (mBERT, XLM-RoBERTa). Fine-tuning и адаптация предобученных моделей. Стратегии fine-tuning: full fine-tuning, feature extraction, adapter layers. Parameter-Efficient Fine-Tuning (PEFT): LoRA, QLoRA, Prefix Tuning. Prompt engineering и prompt tuning. Few-shot и zero-shot learning. Instruction tuning для LLM. Domain adaptation для специализированных задач (АПК). Distillation для сжатия моделей. Практика работы с Hugging Face Transformers

Тема 5 Прикладные задачи NLP: классификация, NER, генерация текстов

Задачи sequence labeling: NER, POS-tagging, dependency parsing. Named Entity Recognition (NER): извлечение сущностей (организации, персоны, локации). Part-of-Speech tagging и dependency parsing. Архитектуры для sequence labeling: BiLSTM-CRF, Transformer-based NER. Метрики качества: precision, recall, F1 на уровне токенов и сущностей. Инструменты: spaCy, DeepPavlov, AllenNLP. Domain-specific NER для текстов АПК (культуры, регионы, предприятия). Методы аннотирования данных: Label Studio, Prodigy, Doccano. Генерация текстов и seq2seq задачи: машинный перевод, суммаризация, QA. Seq2seq модели с attention mechanism. Задачи генерации: machine translation, text summarization, question answering. Beam search и другие стратегии декодирования (greedy, top-k, nucleus sampling). Метрики качества генерации: BLEU, ROUGE, METEOR, BERTScore. Extractive vs. abstractive summarization. Retrieval-Augmented Generation (RAG) для QA систем. Работа с фреймворками: LangChain, LlamaIndex. Применение в задачах АПК: генерация отчётов, автоматическое реферирование.

Тема 6 Тестирование, валидация и развертывание NLP-систем

Оценка качества, тестирование и мониторинг NLP-моделей. Метрики качества для различных задач NLP (классификация, NER, генерация). Бенчмарки: GLUE, SuperGLUE, Russian SuperGLUE. Тестирование robustness: adversarial examples, stress testing. Инструменты: TextAttack, Checklist, robustness-gym. Обнаружение и митигация bias в языковых моделях. Interpretability: LIME, SHAP, Captum, attention visualization. Мониторинг моделей в production: Evidently AI, WhyLabs. A/B-тестирование NLP-систем. Continuous evaluation и модельный дрейф.

4.3 Лекции/ практические занятия

Таблица 4

Содержание лекций/ практических занятий и контрольные мероприятия

№ п/п	Название раздела, темы	№ и название лекций / практических занятий	Формируемые компетенции (индикаторы)	Вид контрольного мероприятия	Кол-во Часов / из них практическая подготовка
1.	Тема 1. Введение в обработку естественного языка и лингвистические основы	Лекция №1. Основы обработки естественного языка: парадигмы, методы и архитектура NLP-систем.	ПКос-12.1, ПКос-4.1	—	2
		Практическая работа №1. Предобработка текстов: токенизация, нормализация, морфологический анализ.	ПКос-12.2, ПКос-10.2	Защита работы	2
		Практическая работа №2. Построение корпуса текстов и извлечение лингвистических признаков.	ПКос-12.2, ПКос-12.3	Защита работы	2
2.	Тема 2. Векторные представления текстов и эмбединги	Лекция №2. Векторные представления слов и документов: от Bag-of-Words до трансформеров.	ПКос-11.1, ПКос-12.1	—	2
		Практическая работа №3. Обучение Word2Vec и FastText на корпусе текстов АПК.	ПКос-11.2, ПКос-12.2	Защита работы	2
		Практическая работа №4. Sentence embeddings и семантический поиск с использованием sentence-transformers.	ПКос-11.2, ПКос-11.3	Защита работы	2
3.	Тема 3. Архитектуры нейронных сетей для NLP: от RNN до Transformer	Лекция №3. Рекуррентные нейронные сети и их модификации для обработки последовательностей.	ПКос-11.1, ПКос-12.1	—	2
		Лекция №4. Механизм внимания и архитектура Transformer. Механизм внимания (Attention): query, key, value. Self-Attention и Multi-Head Attention.	ПКос-11.1, ПКос-12.1	—	2
		Практическая работа №5. Реализация LSTM для классификации текстов на PyTorch.	ПКос-11.2, ПКос-11.3	Защита работы	2
		Практическая работа №6. Кейс Россельхозбанка: классификация кредитных заявок с использованием RNN.	ПКос-11.2, ПКос-4.2, ПКос-12.2	Защита работы	2 / 2

		Практическая работа №7. Реализация механизма Self-Attention и визуализация attention weights.	ПКос-11.2, ПКос-12.2, ПКос-10.2	Защита работы	2
		Коллоквиум (после Темы 3)	ПКос-11.1, ПКос-12.1, ПКос-4.1	Устный опрос	
4.	Тема 4. Предобученные языковые модели и механизмы внимания	Лекция №5. Архитектуры предобученных языковых моделей: BERT, GPT, T5.	ПКос-11.1, ПКос-12.1	—	2
		Лекция №6. Fine-tuning и адаптация предобученных моделей.	ПКос-11.1, ПКос-4.1, ПКос-12.1	—	2
		Практическая работа №8. Fine-tuning BERT для классификации текстов с использованием Hugging Face.	ПКос-11.2, ПКос-11.3, ПКос-12.2	Защита работы	2
		Практическая работа №9. Prompt engineering для GPT-моделей и работа с LLM API.	ПКос-11.2, ПКос-4.2, ПКос-12.2	Защита работы	2
5.	Тема 5. Прикладные задачи NLP: классификация, NER, генерация текстов	Лекция №7. Задачи sequence labeling: NER, POS-tagging, dependency parsing. Named Entity Recognition (NER)	ПКос-11.1, ПКос-10.1, ПКос-12.1	—	2
		Лекция №8. Генерация текстов и seq2seq задачи: машинный перевод, суммаризация, QA. Seq2seq модели с attention mechanism.	ПКос-11.1, ПКос-12.1	—	2
		Практическая работа №10. Кейс Россельхозбанка: извлечение именованных сущностей из кредитных досье.	ПКос-11.2, ПКос-11.3, ПКос-4.2, ПКос-12.2	Защита работы	2 / 2
		Практическая работа №11. Построение системы автоматического реферирования с использованием T5.	ПКос-11.2, ПКос-11.3, ПКос-12.2	Защита работы	2
6.	Тема 6. Тестирование, валидация и развертывание NLP-систем	Лекция №9. Оценка качества, тестирование и мониторинг NLP-моделей.	ПКос-10.1, ПКос-10.2, ПКос-4.1, ПКос-12.1	—	2
		Практическая работа №12. Unit-тестирование и профилирование NLP пайплайнов.	ПКос-10.2, ПКос-10.3, ПКос-4.3	Защита работы	2
		Практическая работа №13. Развертывание NLP-модели с использованием FastAPI и Docker.	ПКос-10.2, ПКос-10.3, ПКос-11.3, ПКос-4.2	Защита работы	2

Таблица 5

Перечень вопросов для самостоятельного изучения дисциплины

№ п/п	Название раздела, темы	Перечень рассматриваемых вопросов для самостоятельного изучения
1.	Тема 1. Введение в обработку естественного языка и лингвистические основы	Сравнительный анализ корпусной лингвистики и дистрибутивной семантики для моделирования специализированной лексики АПК. Методы автоматического построения онтологий для аграрной предметной области. Особенности морфологического анализа агротехнических терминов и номенклатуры растений в русском языке. Применение тегирования частей речи (POS-tagging) для извлечения агрономических знаний из научных публикаций. Проблемы disambiguation (разрешения многозначности) в текстах о технологиях растениеводства и животноводства. ПКос-12.1, ПКос-12.2, ПКос-4.1
2.	Тема 2. Векторные представления текстов и эмбединги	Обучение domain-specific эмбедингов на корпусах аграрных текстов: выбор гиперпараметров и архитектуры для Word2Vec и FastText. Сравнительный анализ качества эмбедингов для русскоязычных терминов АПК при использовании различных корпусов (научные статьи, нормативные документы, новостные тексты). Применение Cross-lingual embeddings для работы с многоязычными агротехническими документами. Методы оценки качества эмбедингов через intrinsic evaluation (word similarity, analogy tasks) на примерах из сельского хозяйства. Построение тематических моделей (LDA, NMF) для анализа научно-исследовательской литературы по агрономии. ПКос-11.1, ПКос-12.1, ПКос-12.2
3.	Тема 3. Архитектуры нейронных сетей для NLP: от RNN до Transformer	Адаптация архитектур Encoder-Decoder для задач автоматического перевода агротехнической документации. Применение Attention mechanism для интерпретации решений модели при анализе кредитных заявок агропредприятий. Сравнительный анализ вычислительной эффективности RNN, LSTM, GRU и Transformer для обработки длинных документов (годовых отчетов, технико-экономических обоснований). Использование Hierarchical Attention Networks для классификации многостраничных документов в агробизнесе. Проблемы gradient clipping и learning rate scheduling при обучении моделей на небольших отраслевых датасетах АПК. ПКос-11.1, ПКос-11.2, ПКос-4.1, ПКос-4.2
4.	Тема 4. Предобученные языковые модели и механизмы внимания	Сравнительный анализ производительности русскоязычных моделей (ruBERT vs ruGPT-3 vs YandexGPT) на задачах классификации документов аграрного сектора. Методы Domain Adaptation и Continual Learning для адаптации LLM к специфике терминологии АПК без катастрофического забывания (catastrophic forgetting). Применение Low-Rank Adaptation (LoRA) и Quantized LoRA (QLoRA) для эффективного fine-tuning на ограниченных вычислительных ресурсах агропредприятий. Разработка prompt templates для извлечения структурированной информации из неструктурированных агротехнических отчетов. Chain-of-Thought prompting для улучшения рассуждений LLM при анализе сложных сценариев развития агробизнеса. Использование Retrieval-Augmented Generation (RAG) для создания вопросно-ответных систем по нормативно-правовой базе АПК. ПКос-11.1, ПКос-11.2, ПКос-11.3, ПКос-4.1, ПКос-4.2, ПКос-12.1
5.	Тема 5. Прикладные задачи NLP: классификация, NER, генерация текстов	Разработка custom NER моделей для извлечения специфических сущностей АПК: названий культур, сортов, агротехнологий, удобрений, средств защиты растений. Применение Conditional Random Fields (CRF) и BiLSTM-CRF для sequence labeling в задачах аннотирования агрономических протоколов. Методы автоматической генерации аналитических отчетов по результатам сельскохозяйственного сезона с использованием Data-to-Text подходов. Extractive vs Abstractive суммаризация научных публикаций по агрономии: выбор подхода и метрики оценки. Построение Question Answering

№ п/п	Название раздела, темы	Перечень рассматриваемых вопросов для самостоятельного изучения
		систем на базе корпуса нормативных документов Минсельхоза РФ. Применение Named Entity Linking для связывания упоминаний предприятий АПК с базами данных СПАРК и Контур.Фокус. Методы Relation Extraction для автоматического построения графов знаний в агробизнесе. ПКос-11.1, ПКос-11.2, ПКос-11.3, ПКос-10.1, ПКос-12.2, ПКос-4.2
6.	Тема 6. Тестирование, валидация и развертывание NLP-систем	Методы обнаружения и количественной оценки bias в моделях, обученных на исторических данных кредитования АПК (по регионам, размеру хозяйств, типу собственности). Применение Fairness Metrics (demographic parity, equalized odds) для оценки справедливости скоринговых NLP-моделей в агрокредитовании. Стратегии Data Drift Detection и Model Monitoring для NLP-систем, работающих с сезонными изменениями в агробизнесе. Техники Model Compression (pruning, quantization, knowledge distillation) для развертывания NLP-моделей на edge-устройствах в удаленных сельских районах. Организация Human-in-the-Loop пайплайнов для валидации результатов NER и классификации в критичных бизнес-процессах АПК. Методы Shadow Mode Deployment и Canary Release для безопасного внедрения обновленных версий NLP-моделей в production-системы банков и агрохолдингов. Разработка Service Level Agreements (SLA) и ключевых показателей производительности (KPI) для NLP-сервисов в инфраструктуре АПК. ПКос-10.1, ПКос-10.2, ПКос-10.3, ПКос-4.1, ПКос-4.2, ПКос-4.3, ПКос-12.3

5. Образовательные технологии

Таблица 6

Применение активных и интерактивных образовательных технологий

№ п/п	Тема и форма занятия	Наименование используемых активных и интерактивных образовательных технологий	
1	Тема 1. Введение в обработку естественного языка и лингвистические основы	Л	Лекция-визуализация
		ПЗ	Решение задач профессиональной направленности, проблемно-поисковое занятие, групповое обсуждение
2	Тема 2. Векторные представления текстов и эмбединги	Л	Лекция-визуализация
		ПЗ	Решение задач профессиональной направленности, проблемно-поисковое занятие, групповое обсуждение
3	Тема 3. Архитектуры нейронных сетей для NLP: от RNN до Transformer	Л	Лекция-визуализация
		ПЗ	Решение задач профессиональной направленности, проблемно-поисковое занятие, групповое обсуждение
4	Тема 4. Предобученные языковые модели и механизмы внимания	Л	Лекция-визуализация
		ПЗ	Решение задач профессиональной направленности, проблемно-поисковое занятие, групповое обсуждение
5	Тема 5. Прикладные задачи NLP: классификация, NER, генерация текстов	Л	Лекция-визуализация
		ПЗ	Решение задач профессиональной направленности, проблемно-поисковое занятие, групповое обсуждение

№ п/п	Тема и форма занятия	Наименование используемых активных и интерактивных образовательных технологий	
6	Тема 6. Тестирование, валидация и развертывание NLP-систем	Л	Лекция-визуализация
		ПЗ	Решение задач профессиональной направленности, проблемно-поисковое занятие, групповое обсуждение

6. Текущий контроль успеваемости и промежуточная аттестация по итогам освоения дисциплины

6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

1) Примеры заданий практических работ

Практическая работа №1

Тема: Предобработка текстов: токенизация, нормализация, морфологический анализ

Задание: Разработать Python-скрипт для комплексной предобработки корпуса текстов в области агробизнеса, включающий следующие операции:

1) Загрузка и парсинг текстовых документов (новостные статьи, технологические регламенты).

2) Токенизация и нормализация текста (приведение к нижнему регистру, удаление пунктуации, специальных символов).

3) Морфологический анализ с использованием `py morphology2` и `razdel`: определение леммы, части речи, граммематических признаков.

4) Удаление стоп-слов (использование расширенного словаря для терминов АПК).

5) Сохранение результатов в формате CSV с метаданями (исходный текст, токены, леммы, POS-теги).

Цель: Овладение навыками подготовки текстовых данных для анализа и машинного обучения, понимание особенностей морфологии русского языка в контексте аграрной терминологии.

Компетенции: ПКос-12.2, ПКос-10.2

Практическая работа №3

Тема: Обучение Word2Vec и FastText на корпусе текстов АПК

Задание: На основе собранного корпуса статей о растениеводстве и животноводстве реализовать пайплайн обучения векторных представлений слов:

1) Загрузить и предобработать корпус текстов (минимум 1000 документов)

2) Обучить модель Word2Vec с параметрами: размерность вектора 300, окно контекста 5, минимальная частота слова 3

3) Обучить модель FastText для обработки редких и составных терминов (минимум 3, максимум 6 символов)

4) Сравнить качество эмбедингов: найти аналогичные слова (word similarity), выполнить аналогические задачи (word analogy) для пар типа "пшеница-зерно" = "кукуруза-?"

5) Визуализировать пространство эмбедингов с помощью t-SNE, выделив кластеры по тематикам (культуры, вредители, удобрения)

Цель: Овладение методами получения контекстных векторных представлений текстов, практический опыт работы с инструментами gensim и fasttext для русскоязычных корпусов специализированной направленности.

Компетенции: ПКос-11.2, ПКос-12.2

Практическая работа №4

Тема: Sentence embeddings и семантический поиск с использованием sentence-transformers

Задание: Разработать приложение для семантического поиска документов в базе знаний по агрономии:

1) Загрузить предобученную модель sentence-transformers (например, multilingual-MiniLM-L12-v2 или русскоязычный аналог)

2) Создать корпус из 500+ документов (технологические карты выращивания культур, инструкции по применению удобрений)

3) Преобразовать все документы в эмбединги и сохранить в векторной базе данных (FAISS или Pinecone)

4) Реализовать функцию семантического поиска: для заданного пользовательского запроса найти топ-10 релевантных документов по косинусному расстоянию

5) Провести оценку качества поиска: сравнить с BM25 (лексическим поиском), рассчитать precision@10, recall@10

6) Визуализировать результаты поиска, включая значения сходства между запросом и найденными документами

Цель: Усвоение методов получения эмбедингов предложений и документов, практическое применение семантического поиска для информационно-поисковых систем в аграрной сфере.

Компетенции: ПКос-11.2, ПКос-11.3

Практическая работа №5

Тема: Реализация LSTM для классификации текстов на PyTorch

Задание: Построить нейросетевую модель на основе LSTM для классификации текстовых документов по категориям:

1) Подготовить датасет: собрать или использовать готовый набор текстов из 3-5 категорий (например: новости по растениеводству, животноводству, экономике АПК, технологии, государственная поддержка)

2) Реализовать пайплайн обработки данных: токенизация, преобразование в индексы, паддинг последовательностей

3) Определить архитектуру модели на PyTorch:

- Embedding layer (размерность векторов: 100-300);

- LSTM layer (скрытая размерность: 128-256);

- Dropout слой для регуляризации;
- Fully Connected слой для классификации;
- Softmax для вероятностного распределения по классам.
- 4) Обучить модель (epochs=20-50, batch_size=32, learning_rate=0.001)
- 5) Вычислить метрики качества на тестовой выборке: accuracy, precision, recall, F1-score, построить confusion matrix
- 6) Реализовать функцию для предсказания класса нового текста и вывода top-3 вероятностей

Цель: Освоение архитектур рекуррентных нейронных сетей и их применения для обработки последовательностей текстов, получение практического опыта работы с PyTorch.

Компетенции: ПКос-11.2, ПКос-11.3

Практическая работа №9

Тема: Prompt engineering для GPT-моделей и работа с LLM API

Задание: Разработать систему для генерации аналитических отчётов по результатам сельскохозяйственного сезона с использованием больших языковых моделей:

- 1) Получить доступ к API одной из LLM (OpenAI GPT-4, Anthropic Claude или YandexGPT)
- 2) Спроектировать систему prompt templates для различных типов анализа:
 - Синтез текстовых данных (агрегация информации из нескольких источников);
 - Генерация выводов и рекомендаций на основе структурированных показателей;
 - Создание краткого резюме для разных аудиторий (для фермера, для инвестора, для государственных органов);
- 3) Реализовать Chain-of-Thought prompting: разбить сложную задачу анализа на промежуточные шаги
- 4) Разработать функции для:
 - Отправки запросов к API с параметризацией (temperature, max_tokens, top_p);
 - Обработки ответов и их постобработки;
 - Управления токен-бюджетом и контролем затрат.
- 5) Протестировать качество генерируемых отчётов на примерах реальных или синтетических данных
- 6) Провести сравнение результатов между разными моделями и настройками промптов
- 7) Документировать best practices для написания эффективных промптов в контексте АПК

Цель: Приобретение навыков работы с API больших языковых моделей, овладение техниками prompt engineering для решения прикладных задач аналитики и генерации текстов.

Компетенции: ПКос-11.2, ПКос-4.2, ПКос-12.2.

2) Примерный перечень вопросов, выносимых на промежуточную аттестацию (экзамен на 2 курсе 4 семестре)

Раздел 1: Основы NLP и предобработка текстов

1. Дайте определение обработки естественного языка (NLP) и перечислите основные задачи, решаемые в этой области. Какова роль NLP-систем в управлении информационными системами в агропромышленном комплексе?

2. Описанные основные этапы предобработки текстов: токенизация, нормализация, морфологический анализ. Какие специфические особенности русского языка необходимо учитывать при предобработке аграрных текстов (например, при обработке названий культур, агротехнических терминов)?

3. Что такое стоп-слова и зачем их удаляют при предобработке текстов? Почему простой список стоп-слов может быть неадекватен для специализированных корпусов (например, в сфере АПК)?

Раздел 2: Векторные представления текстов

4. Объясните принцип работы методов Bag-of-Words (BoW) и TF-IDF. Каковы преимущества и ограничения этих подходов для представления семантики текстов?

5. Опишите архитектуру и принцип обучения Word2Vec с методами CBOW и Skip-gram. Почему Word2Vec более эффективен для захвата семантических отношений между словами по сравнению с TF-IDF?

6. Что представляет собой метод FastText и в чём его преимущество перед Word2Vec при работе с морфологически богатыми языками, такими как русский? Приведите примеры из аграрной терминологии.

7. Объясните концепцию контекстных эмбеддингов на примере модели ELMo. Почему контекстные представления слов превосходят статические эмбеддинги при решении задач полисемии?

Раздел 3: Архитектуры нейронных сетей для NLP

8. Описанные принцип работы рекуррентных нейронных сетей (RNN) и проблемы vanishing/exploding gradients. Как LSTM и GRU решают эти проблемы благодаря механизмам gate?

9. Объясните механизм внимания (Attention) в нейронных сетях: что такое query, key и value, и как они используются для взвешивания релевантности входных элементов?

10. Дайте подробное описание архитектуры Transformer, включая: многоголовое самовнимание (Multi-Head Self-Attention), позиционное кодирование, кроссовое внимание (Cross-Attention). Почему Transformer вытеснил RNN в большинстве современных NLP-приложений?

Раздел 4: Предобученные языковые модели и трансфер-обучение

11. Опишите архитектуру и методы предобучения моделей типа BERT (Masked Language Modeling, Next Sentence Prediction). Какие русскоязычные BERT-подобные модели существуют и где они применяются в системах управления АПК?

12. Объясните различие между decoder-only моделями (семейство GPT) и encoder-only моделями (семейство BERT). Почему GPT-модели лучше подходят для генерации текстов, а BERT — для задач классификации и извлечения информации?

13. Что такое fine-tuning и transfer learning? Описанные основные стратегии адаптации предобученных моделей: full fine-tuning, feature extraction, adapter layers, LoRA. В каких случаях параметрически-эффективные методы (LoRA, QLoRA) предпочтительны для практических приложений?

Раздел 5: Практические приложения NLP

14. Опишите задачу Named Entity Recognition (NER) для текстов аграрной направленности. Какие подходы используются для её решения (BiLSTM-CRF, Transformer-based)? Приведите примеры сущностей в контексте АПК (культуры, вредители, удобрения, регионы).

15. Объясните архитектуру seq2seq моделей и их применение для задач машинного перевода, суммаризации текстов и вопросно-ответных систем. Что такое beam search и nucleus sampling, и как они используются при декодировании выходной последовательности? Какие метрики качества (BLEU, ROUGE, BERTScore) применяются для оценки генерируемых текстов?

6.2. Описание показателей и критериев контроля успеваемости, описание шкал оценивания

Оценочные средства текущего контроля успеваемости и сформированности компетенций основана на подсчете баллов, «заработанных» студентом в течение семестра.

Успеваемость студента по дисциплине оценивается в баллах от 0 до 100.

Оценка знаний проводится по следующим критериям:

- посещение занятий – 20 баллов;
- выполнение практических заданий – 30 баллов;
- качество коллоквиума – 10 баллов;
- промежуточный контроль (экзамен) – 40 баллов.

Соответствие балльной оценки общепринятой 4-х балльной шкале оценок приведено в таблице 7.

Соответствие балльных оценок по 4-х балльной шкале

Балльная оценка	Оценка по 4хбалльной шкале	Оценка по шкале «Зачтено» / «Не зачтено»
0-59	Неудовлетворительно - 2	Не зачтено
60-69	Удовлетворительно - 3	Зачтено
70-89	Хорошо – 4	Зачтено
90-100	Отлично - 5	Зачтено

Критерии оценивания результатов обучения показаны в таблице 8

Критерии оценивания результатов обучения (экзамен)

Оценка	Критерии оценивания
Высокий уровень «5» (отлично)	оценку « отлично » заслуживает студент, освоивший знания, умения, компетенции и теоретический материал без пробелов; выполнивший все задания, предусмотренные учебным планом на высоком качественном уровне; практические навыки профессионального применения освоенных знаний сформированы. Компетенции , закреплённые за дисциплиной, сформированы на уровне – высокий.
Средний уровень «4» (хорошо)	оценку « хорошо » заслуживает студент, практически полностью освоивший знания, умения, компетенции и теоретический материал, учебные задания не оценены максимальным числом баллов, в основном сформировал практические навыки. Компетенции , закреплённые за дисциплиной, сформированы на уровне – хороший (средний).
Пороговый уровень «3» (удовлетворительно)	оценку « удовлетворительно » заслуживает студент, частично с пробелами освоивший знания, умения, компетенции и теоретический материал, многие учебные задания либо не выполнил, либо они оценены числом баллов близким к минимальному, некоторые практические навыки не сформированы. Компетенции , закреплённые за дисциплиной, сформированы на уровне – достаточный.
Минимальный уровень «2» (неудовлетворительно)	оценку « неудовлетворительно » заслуживает студент, не освоивший знания, умения, компетенции и теоретический материал, учебные задания не выполнил, практические навыки не сформированы. Компетенции , закреплённые за дисциплиной, не сформированы.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература

1. Риз Р. Обработка естественного языка на Java. – Москва: ДМК Пресс, 2016. – 264 с. – URL: <https://e.lanbook.com/book/93272>. – ISBN 978-5-97060-331-4.
2. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение: научное издание. – 2-е изд. – Москва: ДМК Пресс, 2018. – 652 с. – URL: <https://e.lanbook.com/book/107901>. – ISBN 978-5-97060-618-6.

7.2. Дополнительная литература

1. Ёылдырым С., Асгари-Ченаглу М. Осваиваем архитектуру Transformer. – Москва: ДМК Пресс, 2022. – 320 с. – URL: <https://e.lanbook.com/book/241148>. – ISBN 978-5-93700-106-1.
2. Гольдберг Й. Нейросетевые методы в обработке естественного языка: руководство. – Москва: ДМК Пресс, 2019. – 282 с. – URL: <https://e.lanbook.com/book/131704>. – ISBN 978-5-97060-754-1.
3. Ганегедара Т. Обработка естественного языка с TensorFlow: руководство. – Москва: ДМК Пресс, 2020. – 382 с. – URL: <https://e.lanbook.com/book/140584>. – ISBN 978-5-97060-756-5.

7.3. Нормативные правовые акты

1. Гост 19.001-77. Единая система программной документации: Общие положения. – М.: Изд.-во стандартов, 1994.
2. Гост 19.101-77. Единая система программной документации: Виды программ и программных документов. – М.: Изд.-во стандартов, 1994.
3. Гост 19.102-77. Единая система программной документации: Стадии разработки. – М.: Изд.-во стандартов, 1994.
4. Гост 19.105-78. Единая система программной документации: Общие требования к программным документам. – М.: Изд.-во стандартов, 1994.
5. Гост 19.201-78. Единая система программной документации: Техническое задание. Требования к содержанию и оформлению. – М.: Изд.-во стандартов, 1994.
6. Гост 19.202-78. Единая система программной документации: Спецификация. Требования к содержанию и оформлению. – М.: Изд.-во стандартов, 1994.
7. Гост 19.502-78. Единая система программной документации: Описание применения. Требования к содержанию и оформлению. – М.: Изд.-во стандартов, 1994.
8. Гост 19.404-79. Единая система программной документации: Пояснительная записка. Требования к содержанию и оформлению. – М.: Изд.-во стандартов, 1994.
9. Гост 3.11.09-82. Система технологической документации: Термины и определения основных понятий. – М.: Изд.-во стандартов, 1994.

10. Гост 34.201-89. Виды, комплектность и обозначение документов при создании автоматизированных систем. – М.: Изд.-во стандартов, 1991.
11. ГОСТ 34.601-90. Автоматизированные Системы Стадии создания. Комплекс стандартов на автоматизированные системы. - М.: Изд-во стандартов, 1997
12. ISO/IEC 12207:1995

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Hugging Face Documentation (Transformers, Datasets, Tokenizers) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://huggingface.co/docs>
2. Hugging Face Course (практический курс по современному NLP) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://huggingface.co/learn>
3. PyTorch Documentation (нейросетевые модели, обучение, инференс) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://pytorch.org/docs/stable/>
4. TensorFlow Documentation (глубокое обучение для NLP) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://www.tensorflow.org/>
5. Keras Documentation (высокоуровневый API для нейросетевых моделей) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://keras.io/>
6. scikit-learn Documentation (базовые ML-модели и метрики, baseline-подходы) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://scikit-learn.org/stable/>
7. spaCy Documentation (пайплайны NLP, NER, токенизация, синтаксис) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://spacy.io/usage>
8. NLTK Documentation (классические методы NLP, корпуса, прототипирование) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://www.nltk.org/>
9. Gensim Documentation (векторные представления, topic modeling) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://radimrehurek.com/gensim/>
10. fastText (модели subword-эмбеддингов и классификация текстов) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://fasttext.cc/>
11. Stanford CS224N (материалы курса по NLP и DL) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://web.stanford.edu/class/cs224n/>
12. Speech and Language Processing (Jurafsky & Martin, онлайн-версия) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://web.stanford.edu/~jurafsky/slp3/>

13. arXiv (архив научных публикаций по NLP/LLM) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://arxiv.org/>
14. Papers with Code (лидерборды, датасеты, SOTA-результаты по задачам NLP) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://paperswithcode.com/>
15. Universal Dependencies (UD) (деревья зависимостей, форматы разметки) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://universaldependencies.org/>
16. GLUE Benchmark (описание задач и протоколы оценки) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://gluebenchmark.com/>
17. SQuAD (датасет для QA, описание и материалы) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://rajpurkar.github.io/SQuAD-explorer/>
18. DeepPavlov Documentation (NLP для русского языка, пайплайны и модели) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://docs.deeppavlov.ai/>
19. Natasha (инструменты NLP для русского языка) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://github.com/natasha/natasha>
20. OpenCorpora (русскоязычные корпуса и морфологическая разметка) [Электронный ресурс]. — Режим доступа: открытый доступ. — <https://opencorpora.org/>
21. Kaggle Datasets (учебные датасеты по обработке текстов) [Электронный ресурс]. — Режим доступа: открытый доступ (в отдельных случаях требуется регистрация). — <https://www.kaggle.com/datasets>

9. Перечень программного обеспечения и информационных справочных систем

9.1. Программное обеспечение

1. **Python (3.x)** — среда разработки и выполнения программ для обработки текста и обучения моделей. <https://www.python.org/>
2. **Anaconda / Miniconda** — управление окружениями и пакетами Python. <https://www.anaconda.com/> / <https://docs.conda.io/en/latest/miniconda.html>
3. **Jupyter Notebook / JupyterLab** — интерактивные вычисления, отчётность по экспериментам. <https://jupyter.org/>
4. **Visual Studio Code** — редактор/IDE для разработки. <https://code.visualstudio.com/>
5. **PyCharm (Community Edition)** — IDE для Python (альтернатива). <https://www.jetbrains.com/pycharm/>
6. **Git** — система контроля версий (ведение проектов, воспроизводимость). <https://git-scm.com/>

7. **Docker** — контейнеризация и воспроизводимость окружений (по необходимости). <https://www.docker.com/>
8. **PyTorch** — обучение нейросетевых моделей NLP. <https://pytorch.org/>
9. **TensorFlow / Keras** — альтернативный стек DL для NLP. <https://www.tensorflow.org/> / <https://keras.io/>
10. **Hugging Face Transformers** — трансформерные модели и пайплайны NLP. <https://huggingface.co/docs/transformers/>
11. **Hugging Face Datasets** — датасеты и загрузчики для экспериментов. <https://huggingface.co/docs/datasets/>
12. **spaCy** — промышленный NLP-пайплайн (токенизация, NER, синтаксис). <https://spacy.io/usage>
13. **NLTK** — классические методы NLP и прототипирование. <https://www.nltk.org/>
14. **Gensim** — тематическое моделирование, эмбединги, similarity. <https://radimrehurek.com/gensim/>
15. **fastText** — subword-эмбединги и классификация текстов. <https://fasttext.cc/>
16. **scikit-learn** — baseline-модели, метрики, подготовка признаков. <https://scikit-learn.org/stable/>
17. **Label Studio / doccano** (по необходимости) — разметка текстовых данных (NER, классификация). <https://labelstud.io/> / <https://github.com/doccano/doccano>

9.2. Информационные справочные системы и интернет-ресурсы

1. **Hugging Face Hub** — каталоги моделей/датасетов, model cards. <https://huggingface.co/>
2. **Papers with Code** — лидеры (SOTA), датасеты, коды, метрики. <https://paperswithcode.com/>
3. **arXiv** — научные публикации по NLP/LLM. <https://arxiv.org/>
4. **OpenAlex** — метаданные научных публикаций и авторов (поиск литературы). <https://docs.openalex.org/>
5. **Crossref** — метаданные DOI и публикаций. <https://www.crossref.org/>
6. **Universal Dependencies** — стандарты синтаксической разметки и корпуса. <https://universaldependencies.org/>
7. **OpenCorpora** — русскоязычные корпуса и морфологические ресурсы. <https://opencorpora.org/>
8. **RuWordNet / WordNet** — лексико-семантические сети (синонимия/гиперонимия). <http://ruwordnet.ru/> / <https://wordnet.princeton.edu/>
9. **Wiktionary** — справочные лексические данные (ограниченно, с критической оценкой качества). <https://www.wiktionary.org/>
10. **Грамота.ру** — орфография/пунктуация/нормы русского языка (как справочная поддержка). <https://gramota.ru/>

11. **CyberLeninka / eLIBRARY** — поиск русскоязычной научной литературы (при наличии доступа). <https://cyberleninka.ru/> / <https://www.elibrary.ru/>

10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Таблица 9в та

Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	Тема 1–6	Anaconda	Платформа для разработки и анализа данных	Anaconda Inc.	2019
2	Тема 1–6	Jupyter Notebook / JupyterLab	Интерактивная среда для Python	Project Jupyter	2015
3	Тема 1–6	Visual Studio Code	Редактор исходного кода	Microsoft	2016
4	Тема 1–6	Git / GitHub	Система контроля версий	Linus Torvalds / GitHub Inc.	2005
5	Тема 1–2	NLTK, pymorphy2, razdel, natasha	Библиотеки для предобработки текстов и лингвистического анализа	Сообщество Python	2004–2023
6	Тема 1–2	spaCy, TextBlob	Фреймворки для обработки естественного языка	ExplosionAI, TextBlob community	2015–2023
7	Тема 3	gensim, FastText	Библиотеки для обучения Word2Vec и получения эмбедингов	Radim Řehůřek, Facebook AI	2009–2023
8	Тема 3	sentence-transformers	Библиотека для получения sentence embeddings	UKPLab	2019
9	Тема 3	scikit-learn	Библиотека машинного обучения	Scikit-learn developers	2010
10	Тема 3–4	PyTorch	Фреймворк для глубокого обучения	Meta (Facebook)	2016
11	Тема 3–4	TensorFlow, Keras	Фреймворк для глубокого обучения	Google, François Chollet	2015
12	Тема 4–5	Hugging Face Transformers	Библиотека предобученных языковых моделей	Hugging Face	2018
13	Тема 4–5	DeepPavlov	Фреймворк для разработки NLP-систем	Sber AI	2017
14	Тема 4–5	Transformers Hub (ruBERT, ruGPT-3, mBART, mT5)	Предобученные русскоязычные и многоязычные модели	Hugging Face, Sber AI, Meta	2018–2023

15	Тема 2–4	pandas, NumPy, matplotlib, seaborn	Библиотеки Python для анализа и визуализации данных	Сообщество Python	2008–2023
16	Тема 2–3	Plotly, Streamlit	Библиотеки для интерактивной визуализации и веб-интерфейсов	Plotly Inc., Streamlit Inc.	2012–2019
17	Тема 1–6	pytest, unittest	Инструменты модульного тестирования	Сообщество Python	2010–2023
18	Тема 5	LangChain	Фреймворк для разработки приложений с LLM	LangChain developers	2022
19	Тема 5	LlamaIndex (GPT Index)	Инструмент индексирования данных для RAG систем	Jerry Liu	2022
20	Тема 5–6	FastAPI, Flask	Фреймворки для разработки REST API	Sebastián Ramírez, Armin Ronacher	2018–2010
21	Тема 5–6	Docker	Платформа контейнеризации приложений	Docker Inc.	2013
22	Тема 6	TorchServe, TensorFlow Serving	Инструменты для развёртывания моделей	Meta, Google	2019–2018
23	Тема 6	ONNX Runtime	Среда выполнения для оптимизированных моделей	Microsoft	2019
24	Тема 1–6	Rasa	Платформа для разработки чат-ботов и диалоговых систем	Rasa Inc.	2016
25	Тема 5–6	Label Studio, Prodigy, Doccano	Инструменты для аннотирования и разметки данных	Heartex, Explosion AI, doccano community	2018–2021
26	Тема 6	Evidently AI, WhyLabs	Платформы мониторинга качества моделей	Evidently AI, WhyLabs Inc.	2020–2021
27	Тема 6	Weights & Biases, MLflow	Платформы для отслеживания экспериментов и управления моделями	Weights & Biases, Databricks	2017–2020
28	Тема 3–6	FAISS, Pinecone, Weaviate, Milvus, ChromaDB	Векторные базы данных для semantic search и RAG	Meta, Pinecone, Weaviate, Zilliz, Chroma	2017–2023
29	Тема 5–6	TextAttack, Checklist, robustness-gym	Инструменты для тестирования robustness NLP-моделей	Carnegie Mellon, Marco Túlio Ribeiro, Robustness community	2020–2023
30	Тема 6	LIME, SHAP, Captum	Библиотеки для интерпретируемости моделей	Marco Túlio Ribeiro, Scott Lundberg, Facebook	2016–2019
31	Тема 1–6	Jupyter Hub, Google Colab	Облачные платформы для интерактивных вычислений	Project Jupyter, Google	2018–2017
32	Тема 1–6	Mentimeter, Moodle	Платформы для интерактивного	Mentimeter AB, Moodle Pty Ltd	2002–2016

			обучения и управления курсами		
--	--	--	-------------------------------	--	--

Сведения об обеспеченности специализированными аудиториями, кабинетами, лабораториями

Наименование специальных* помещений и помещений для самостоятельной работы (№ учебного корпуса, № аудитории)	Оснащенность специальных помещений и помещений для самостоятельной работы**
1	2
Корпус 1, Аудитория 201 Количество рабочих мест: 24	Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE.
Корпус 1, Аудитория 203 Количество рабочих мест: 18	Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE. Структурное подразделение: Кафедра Цифровая кафедра
Корпус 1, Аудитория 206 Количество рабочих мест: 24	Встроенные сетевые адаптеры (Intel I219-V или Realtek RTL8111H), интерфейс RJ-45, скорость 10/100/1000 Мбит/с. Точки доступа: Ubiquiti UniFi AP AC Pro, стандарты IEEE 802.11a/b/g/n/ac, частоты 2.4 ГГц (450 Мбит/с) и 5 ГГц (1300 Мбит/с), поддержка MU-MIMO, питание PoE.
Центральная научная библиотека имени Н.И. Железнова	Читальные залы библиотеки
Студенческое общежитие	Комната для самоподготовки

11. Методические рекомендации обучающимся по освоению дисциплины

Освоение дисциплины «Обработка естественного языка» требует активной вовлечённости обучающихся в процесс решения прикладных задач, связанных с предобработкой текстовых данных, разработкой и тестированием NLP-систем, а также применением современных архитектур нейронных сетей и предобученных языковых моделей. Дисциплина ориентирована на развитие проектного и исследовательского подхода с использованием кейсов, включая практические задачи в области сельскохозяйственного производства, агрономии и управления информационными системами в АПК.

Лекционные занятия направлены на формирование системного понимания основ обработки естественного языка, архитектур нейронных сетей для NLP (RNN, LSTM, Transformer) и принципов применения предобученных языковых моделей (BERT, GPT, T5). Лекции сопровождаются презентациями, примерами кода, визуализацией механизмов внимания и архитектур моделей, проходят с использованием мультимедийного оборудования и интерактивных платформ (JupyterHub, Google Colab, Mentimeter, Moodle). Студентам рекомендуется изучить документацию фреймворков Hugging Face Transformers, PyTorch и TensorFlow параллельно с лекциями.

Студенты обязаны вести конспект лекций, дополняя его материалами из рекомендованной и дополнительной литературы, включая статьи из

конференций ACL, EMNLP, NAACL, официальную документацию Python-библиотек (NLTK, spaCy, transformers, PyTorch), а также отраслевые источники по применению NLP в агробизнесе. Перед каждой новой темой рекомендуется повторение ключевых концепций (например, архитектур encoder-decoder для seq2seq задач) и самостоятельное выполнение мини-заданий на закрепление (реализация простого tokenizer, обучение Word2Vec на малом корпусе).

Практические занятия проводятся в компьютерных классах, оборудованных современным ПО (Python 3.9+, Jupyter, PyTorch, TensorFlow, Hugging Face), и строятся по принципу: изучение сценария – разбор готового примера – адаптация под индивидуальное задание – рефлексия результатов. Каждое практическое задание направлено на развитие конкретных навыков, в том числе:

- 1) предобработка и нормализация текстов (токенизация, морфологический анализ, удаление стоп-слов) с использованием NLTK, spaCy, rymorphy2, razdel;
- 2) обучение и применение векторных представлений текстов (Word2Vec, FastText, sentence-transformers);
- 3) реализация архитектур нейронных сетей (RNN, LSTM, CNN) на PyTorch и TensorFlow;
- 4) использование и fine-tuning предобученных моделей (BERT, GPT, T5, ruBERT, ruGPT-3) через Hugging Face Transformers;
- 5) разработка NLP-пайплайнов для задач классификации, NER, генерации текстов, machine translation;
- 6) тестирование, валидация и оценка качества моделей (BLEU, ROUGE, METEOR, BERTScore, confusion matrix);
- 7) развёртывание NLP-систем через FastAPI, Docker и облачные сервисы.

Результаты работы оформляются в виде Jupyter-ноутбуков с документированным кодом, визуализациями, метриками качества и выводами, а также загружаются в систему контроля версий (GitHub) и LMS. При необходимости требуется подготовка краткого текстового отчёта с описанием использованных методов и результатов.

Наиболее трудоёмкие и критически важные темы дисциплины:

- 1) **Тема 3.** Архитектуры нейронных сетей для NLP (RNN, LSTM, GRU, CNN, Transformer, механизм внимания) - требует глубокого понимания математических основ и практической реализации сложных архитектур; этап, где большинство студентов испытывают сложности с пониманием backpropagation through time и механизмов gate в LSTM/GRU.
- 2) **Тема 4.** Предобученные языковые модели и трансфер-обучение (BERT, GPT, T5, fine-tuning, LoRA, prompt engineering) - охватывает наиболее современные и требуемые индустрией подходы; требует работы с большими объёмами данных, параметров моделей и вычислительных ресурсов; критична для развития компетенций в области больших языковых моделей.
- 3) **Тема 5.** Практические приложения NLP (seq2seq, NER, машинный перевод, суммаризация, RAG, LLM API) - объединяет все предыдущие

знания и требует интеграции множества компонентов в единое решение; потребует максимума времени на отладку и оптимизацию.

Рекомендуется отвести дополнительное время для самостоятельного изучения этих тем, использования дополнительных обучающих ресурсов (YouTube-каналы по PyTorch и Transformers, курсы на Coursera/Fast.ai, официальные tutorиалы Hugging Face).

Самостоятельная работа студентов включает:

- 1) изучение справочной и профессиональной документации по NLTK, spaCy, PyTorch, TensorFlow, Hugging Face Transformers;
- 2) выполнение расширенных задач по предобработке текстов, обучению моделей, проведению экспериментов;
- 3) подготовку мини-проектов (например, разработка NER-системы для аграрной терминологии, обучение классификатора новостей АПК, реализация semantic search приложения);
- 4) написание рефлексивных отчётов с анализом результатов, выводами о качестве моделей и рекомендациями для улучшения;
- 5) участие в обсуждении на форумах (Stack Overflow, GitHub Issues) при возникновении проблем.

Промежуточная и итоговая аттестация промежуточная аттестация проводится по результатам выполнения практических работ и их защиты. Итоговая форма контроля – экзамен.

Виды и формы отработки пропущенных занятий

Студент, пропустивший занятия обязан отработать:

Пропущенные лекции – предоставив преподавателю конспект лекции, ответив на вопросы устно, пройдя собеседование по пропущенной теме, пройти тестирование.

Пропущенные практические занятия – в форме выполненных заданий, устного опроса, посещения дополнительных занятий.

Защита индивидуальных заданий проводятся в часы в дни и часы, устанавливаемые преподавателем.

Пропуск занятия по документально подтвержденной дирекцией уважительной причине не является основанием для снижения оценки выполненной практической работы.

Методические рекомендации преподавателям по организации обучения по дисциплине

Преподавание курса «Обработка естественного языка» должно носить контекстный характер. В процессе обучения должна четко прослеживаться целевая установка на развитие личности; интеграционное единство форм, методов и средств обучения; взаимодействие обучаемых и педагогов; индивидуальный стиль педагогической деятельности.

Реализация технологий контекстного обучения в профессионально-образовательном процессе обеспечивается соблюдением следующих условий:

– мотивационное обеспечение субъектов педагогической деятельности и

учение, основанное на реализации их личностных функций в этом процессе;

- наличие четкой и диагностически заданной цели образования, т.е. измеримого представления об ожидаемом результате;
- представление учебного материала в виде системы познавательных и практических задач, ситуаций, заданий, проектов, упражнений и др.;
- указание способов взаимодействия субъектов профессионально-образовательного процесса;
- обозначение границ правилосообразной (алгоритмической) и творческой деятельности педагогов, допустимого отклонения от правил;
- обеспечение открытости обучения профессиональному будущему, направленность на его предвосхищение.

В результате изучения дисциплины студенты получают знания о распространении программного обеспечения, а также методологии и стандартах на основе лицензии и договоров, а также применять достижения отечественной и зарубежной науки и практики.

Методика преподавания дисциплины строится на сочетании лекций с практическими занятиями; групповыми и индивидуальными консультациями по отдельным разделам программы; внеаудиторной самостоятельной работой студентов (работа с учебниками, учебными пособиями, методическими указаниями, заданиями, специальной литературой, поиск необходимой информации в сети Интернет).

Лекционный курс, как одна из составляющей дисциплины «Обработка естественного языка», должен быть логическим и последовательным. Каждая лекция должна, согласно правилам дидактики, начинаться с актуализации знаний. Чтение лекций должно происходить на основе проблемного метода обучения, что будет стимулировать деятельность студентов к самостоятельному поиску знаний. Интерес к изучению материала преподаватель должен стимулировать, используя наглядные методы обучения (мультимедийные презентации, иллюстрации, стенды и т.д.). Помимо традиционной лекции необходимо использовать проблемные лекции, лекции-визуализации, бинарные лекции, дискуссии и т.д.

В начале каждой лекции следует четко формулировать цель, которую необходимо достигнуть посредством решения ряда задач. При этом сами задачи должны быть четко оговорены. Важная роль на лекции должна быть отведена дискуссии. Преподаватель заранее должен продумать траекторию изучения материала с вовлечением студентов в дискуссии. Это позволит на смену авторитарному методу обучению, укоренившемуся в современной системе образования, быть студентам собеседниками преподавателя. Эта особенность лекции важна для более глубокого понимания изучаемого материала.

Как и любое занятие, лекция должна заканчиваться подведением итогов и формулировкой выводов.

Что касается практических занятий, то для них должны соблюдаться такая же структура, как и для лекционных занятий: актуализация знаний, постановка цели и задач и т.д. Практическая работа также должна соответствовать принципам контекстного подхода, с использованием решения

исследовательских задач профессиональной направленности. На практических занятиях должны быть использованы технологии дифференцированного обучения студентов, уделяя большее внимание «слабым» студентам.

Практические занятия проводятся под руководством преподавателя. В рамках этих занятий производится анализ типовых ошибок, допущенных при выполнении заданий, рассматриваются наиболее удачные варианты. Студенты привлекаются к разбору и сравнительному анализу предлагаемых вариантов решений. Происходит коллективное обсуждение, в результате которого приобретаются навыки ведения дискуссии по обсуждаемым вопросам.

Успех закрепления знаний и умений определяется стройной системой подобранных вопросов для текущего контроля.

В процессе самостоятельной работы студенты отрабатывают теоретические положения, изложенные на лекциях, и изучают примеры, рассмотренные на практических занятиях.

Конкретная тема обсуждается с каждым студентом и учитывает направление научных интересов студента или тему выпускной квалификационной работы.

Большое значение в ходе самостоятельной работы студентов имеет работа над литературой и другими источниками информации (периодические издания, Интернет и т.д.).

Особенности методики преподавания данной дисциплины состоят в интенсификации теоретической, практической и самостоятельной работы студентов и широком применении активных и интерактивных форм и методов обучения.

Программу разработал:

Бакнин М.Д. к.т.н., доцент

Греченева А.В. к.т.н., доцент



РЕЦЕНЗИЯ

на рабочую программу дисциплины Б1.В.05 «Обработка естественного языка» ОПОП ВО по направлению подготовки 09.04.03 «Прикладная информатика», направленности «Архитектура систем искусственного интеллекта» (квалификация выпускника – магистр)

Ашмарина Татьяной Игоревной, кандидат экономических наук, доцентом кафедры экономики и организации производства ФГБОУ ВО «Российский государственный аграрный университет – МСХА им. К.А. Тимирязева» (далее по тексту рецензент), проведено рецензирование рабочей программы дисциплины Б1.В.05 «Обработка естественного языка» ОПОП ВО по направлению подготовки 09.04.03 «Прикладная информатика», направленности «Архитектура систем искусственного интеллекта» (магистр) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре прикладной информатики (разработчик – Бакнин Максим Дмитриевич, доцент кафедры прикладной информатики, кандидат технических наук, Греченева Анастасия Владимировна, доцент кафедры прикладной информатики, кандидат технических наук).

Рассмотрев представленные на рецензирование материалы, рецензент пришел к следующим выводам:

1. Предъявленная рабочая программа дисциплины «Обработка естественного языка» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению подготовки 09.04.03 «Прикладная информатика». Программа содержит все основные разделы, соответствует требованиям к нормативно-методическим документам.

2. Представленная в Программе актуальность учебной дисциплины в рамках реализации ОПОП ВО не подлежит сомнению – дисциплина относится к основной части, формируемой участниками образовательных отношений, учебного цикла – Б1.В.

3. Представленные в Программе цели дисциплины соответствуют требованиям ФГОС ВО направления подготовки 09.04.03 «Прикладная информатика».

4. В соответствии с Программой за дисциплиной «Обработка естественного языка» закреплена 4 компетенция (12 индикаторов). Дисциплина «Обработка естественного языка» и представленная Программа способна реализовать ее в объявленных требованиях. Результаты обучения, представленные в Программе в категориях знать, уметь, владеть соответствуют специфике и содержанию дисциплины и демонстрируют возможность получения заявленных результатов.

5. Общая трудоёмкость дисциплины «Обработка естественного языка» составляет 4 зачётных единицы (144/4 часов).

6. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Обработка естественного языка» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 09.04.03 «Прикладная информатика».

7. Представленная Программа предполагает использование современных образовательных технологий, используемые при реализации различных видов учебной работы. Формы образовательных технологий соответствуют специфике дисциплины.

8. Программа дисциплины «Обработка естественного языка» предполагает проведение занятий в интерактивной форме.

9. Виды, содержание и трудоёмкость самостоятельной работы студентов, представленные в Программе, соответствуют требованиям к подготовке выпускников, содержащимся во ФГОС ВО направления 09.04.03 «Прикладная информатика».

10. Представленные и описанные в Программе формы текущей оценки знаний (защита практических работ, групповое обсуждение) соответствуют специфике дисциплины и требованиям к выпускникам. Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме зачета во 2 семестре, что соответствует статусу дисциплины, как дисциплины, включенной в часть, формируемую

участниками образовательных отношений учебного цикла – Б1.В.05 ФГОС ВО направления 09.04.03 «Прикладная информатика».

11. Формы оценки знаний, представленные в Программе, *соответствуют* специфике дисциплины и требованиям к выпускникам.

12. Учебно-методическое обеспечение дисциплины представлено: основной литературой – 2 источника, дополнительной литературой – 3 наименования, Интернет-ресурсы – 6 источников и *соответствует* требованиям ФГОС ВО направления 09.04.03 «Прикладная информатика».

13. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Обработка естественного языка» и обеспечивает использование современных образовательных, в том числе интерактивных методов обучения.

14. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «Обработка естественного языка».

ОБЩИЕ ВЫВОДЫ

На основании проведенного рецензирования можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Обработка естественного языка» ОПОП ВО по направлению 09.04.03 «Прикладная информатика», направленности «Архитектура систем искусственного интеллекта» (квалификация выпускника – магистр), разработанная Бакниным М.Д., Греченовой А.В., соответствует требованиям ФГОС ВО, современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

Рецензент:

Ашмарина Т.И., кандидат экономических наук, доцент, доцент кафедры экономики и организации производства ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева



«28» августа 2025г.