

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 02.06.2025 12:22:08

Уникальный программный ключ:

sa953a01204891083f939673078ef1a989dae18a

**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 направления подготовки/специальности:**

**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)

**2025 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 направлению 02.03.02 «Фундаментальная информатика и информационные технологии» и изучается в 5 семестре 3 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 36 тем и направлена на изучение формирования у студентов современных знаний и практических умений по разработке и применению методов искусственного интеллекта для анализа и обработки текстовой информации. Курс позволяет освоить основные этапы создания intelligent-NLP систем: от предобработки текстовых данных и извлечения признаков до построения и оценки моделей машинного обучения и глубинного обучения для автоматизации понимания, поиска, генерации и классификации текстов.

Целью освоения дисциплины является научить студентов строить и внедрять прикладные решения в области обработки естественного языка, используя методы ИИ, машинного и глубинного обучения; формировать практические навыки работы с современными NLP-библиотеками, проведения анализа и построения пайплайнов задач классификации, генерации, поиска и семантической обработки 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-1	Способен создавать и оценивать различные модели машинного обучения, архитектуру нейронных сетей и алгоритмы искусственного интеллекта с целью выбора наиболее эффективных решений для конкретных профессиональных задач	ПК-1.1 Может выбирать подходящий алгоритм машинного обучения и архитектуру нейронных сетей для конкретной задачи, учитывая особенности данных и требования к решению; ПК-1.2 Демонстрирует навыки обработки, представления и анализа данных для построения моделей машинного обучения; ПК-1.3 Владеет методами создания и обучения моделей с использованием различных алгоритмов и архитектур; ПК-1.4 Умеет оценивать соблюдение методологии разработки различных моделей машинного обучения, архитектур нейронных сетей и алгоритмов, анализировать качество моделей и разрабатывать стратегии для улучшения качества моделей;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	ПК-2.2 Демонстрирует навыки анализа данных с использованием статистических методов и инструментов; ПК-2.3 Владеет методами работы с различными алгоритмами машинного обучения и глубокого обучения для решения различных задач;
ПК-3	Способен разрабатывать и оптимизировать системы обучения с подкреплением и автоматизированного принятия	ПК-3.2 Демонстрирует умение разрабатывать и реализовывать алгоритмы обучения с подкреплением; ПК-3.3 Умеет интегрировать и оптимизировать системы автоматизированного принятия решений в информационных

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	решений в информационных системах, обеспечивая их эффективное функционирование и адаптацию для различных приложений	системах;

### 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-1	Способен создавать и оценивать различные модели машинного обучения, архитектуру нейронных сетей и алгоритмы искусственного интеллекта с целью выбора наиболее эффективных решений для конкретных профессиональных задач		Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная); Нейронные сети; Прикладные задачи машинного обучения; Методы машинного обучения; Основы глубокого обучения; Оптимизация моделей машинного обучения; Практикум по обработке естественного языка (NLP); Массово-параллельные вычисления в машинном обучении (GPU); Проектирование и разработка систем компьютерного зрения; <i>Анализ временных рядов**</i> ; <i>Информационный поиск**</i> ; <i>Генеративные модели**</i> ; <i>Обработка сигналов**</i> ;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью	Статистические методы и первичный анализ данных; Введение в базы данных; Программирование на языке Python; Лингвистические основы анализа естественного языка;	<i>Информационный поиск**</i> ; <i>Анализ временных рядов**</i> ; Нейронные сети; Оптимизация моделей машинного обучения; Практикум по обработке естественного языка (NLP);

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	извлечения полезной информации для обучения моделей искусственного интеллекта	Введение в компьютерное зрение; Программирование на языке C++; <i>Программирование на языке NodeJS**</i> ; <i>Программирование на языке Go**</i> ; Эксплуатационная практика (учебная); Технологическая (проектно-технологическая) практика (учебная);	Основы глубокого обучения; Проектирование и разработка систем компьютерного зрения; Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная);
ПК-3	Способен разрабатывать и оптимизировать системы обучения с подкреплением и автоматизированного принятия решений в информационных системах, обеспечивая их эффективное функционирование и адаптацию для различных приложений		<i>Преддипломная практика</i> ; <i>Технологическая (проектно-технологическая) практика (производственная)</i> ; <i>Эксплуатационная практика (производственная)</i> ; <i>Рекомендательные системы**</i> ; <i>Нейронные сети</i> ; <i>Прикладные задачи машинного обучения</i> ;

\* - 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>	72		72
Лекции (ЛК)	18		18
Лабораторные работы (ЛР)	36		36
Практические/семинарские занятия (СЗ)	18		18
<i>Самостоятельная работа обучающихся, ак.ч.</i>	45		45
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
<b>Общая трудоемкость дисциплины</b>	<b>ак.ч.</b>	<b>144</b>	<b>144</b>
	<b>зач.ед.</b>	<b>4</b>	<b>4</b>

## 5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

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

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Предобработка и представление текстовых данных	1.1	Введение в NLP и задачи анализа естественного языка	ЛК
		1.2	Основные этапы предварительной обработки текста: токенизация, нормализация, удаление стоп-слов	ЛК
		1.3	Методы представления текста: bag-of-words, one-hot encoding, TF-IDF, word embeddings	ЛК
		1.4	Импорт, очистка и первичный разбор жанровых текстов	ЛР
		1.5	Токенизация и разметка текстов с помощью NLTK/Spacy	ЛР
		1.6	Лемматизация, стемминг, выделение стоп-слов	ЛР
		1.7	Векторизация текста: CountVectorizer, TF-IDF	ЛР
		1.8	Построение и визуализация распределения лексических признаков	ЛР
		1.9	Работа с моделями word2vec и fastText для получения эмбеддингов	ЛР
		1.10	Анализ типовых ошибок предобработки: влияние на качество моделей	СЗ
		1.11	Сравнение моделей представления текста (различие в задачах)	СЗ
		1.12	Разбор примеров постановки задачи анализа текста: поиск релевантного подхода	СЗ
Раздел 2	Классификация и извлечение информации из текста с применением ML/AI	2.1	Классификация текста: базовые алгоритмы — наивный Байес, логистическая регрессия, SVM	ЛК
		2.2	Задачи извлечения информации: NER, морфо-синтаксическая разметка, выделение тематик	ЛК
		2.3	Методы оценки качества моделей на текстовых задачах (accuracy, F1, ROC-AUC)	ЛК
		2.4	Построение простого классификатора (например, определения тональности)	ЛР
		2.5	Настройка и выбор параметров модели, кросс-валидация	ЛР
		2.6	Практика извлечения именованных сущностей (NER) с использованием spaCy/HuggingFace	ЛР
		2.7	Морфолого-синтаксическая разметка с помощью открытых пакетов	ЛР
		2.8	Тематическое моделирование LDA и визуализация тематик	ЛР
		2.9	Оценка качества текстовых моделей, построение метрик	ЛР
		2.10	Групповой разбор кейсов классификации отзывов и новостей	СЗ
		2.11	Анализ ошибок: интерпретация результатов моделей	СЗ
		2.12	Практика постановки BERT/finetuning для продвинутых задач (обзор)	СЗ
Раздел 3	Генерация, семантика и современные решения глубокого обучения в NLP	3.1	Введение в рекуррентные (RNN) и трансформерные архитектуры; языковые модели	ЛК
		3.2	Генерация текстов: seq2seq, автокомплит, машинный перевод	ЛК
		3.3	Современные фреймворки и открытые языковые модели (BERT, GPT, T5): применение, перспективы	ЛК
		3.4	Применение Seq2Seq для генерации коротких	ЛР

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
			текстов/аннотаций	
		3.5	Автоматический перевод предложений с помощью трансформерных моделей	ЛР
		3.6	Применение BERT или аналогичных моделей для семантического поиска	ЛР
		3.7	Finetuning готовой модели под конкретный корпус	ЛР
		3.8	Генерация текстов с использованием GPT/BLOOM	ЛР
		3.9	Итоговый мини-проект: пайплайн анализа и генерации текстов (выбор задачи: чат-бот, тематический FAQ, семантическое индексирование и пр.)	ЛР
		3.10	Дискуссия: этика, подводные камни, фейки и ответственность разработчика	СЗ
		3.11	Обзор реальных NLP-проектов: разбор архитектур и стеков решений	СЗ
		3.12	Итоговое обсуждение: перспективы применения современных языковых моделей в разных отраслях	СЗ

\* - 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)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 25 шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	
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.	
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.	

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

## 7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

*Основная 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

*Дополнительная литература:*

1. Лейн Хобсон, Хапке Ханнес, Ховард Коул. Обработка естественного языка в действии. — СПб.: Питер, 2021. ISBN 978-5-4461-1371-2

*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](http://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/>

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

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

1. Курс лекций по дисциплине «Анализ естественного языка с помощью методов искусственного интеллекта».

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

**РАЗРАБОТЧИК:**

Заведующий кафедрой  
прикладного искусственного  
интеллекта

---

*Должность, БУП*

---

*Подпись*

Подолько Павел  
Михайлович

---

*Фамилия И.О.*

**РУКОВОДИТЕЛЬ БУП:**

Заведующий кафедрой  
прикладного искусственного  
интеллекта

---

*Должность БУП*

---

*Подпись*

Подолько Павел  
Михайлович

---

*Фамилия И.О.*

**РУКОВОДИТЕЛЬ ОП ВО:**

Заведующий кафедрой  
прикладного искусственного  
интеллекта

---

*Должность, БУП*

---

*Подпись*

Подолько Павел  
Михайлович

---

*Фамилия И.О.*