

АННОТАЦИЯ рабочей программы дисциплины
Б1.В.09 Технологии обработки языка, звуковых данных, включая распознавание и синтез речи

Объем трудоемкости: 4 з.е.

Цели дисциплины:

Формирование у студентов системных знаний, практических навыков и компетенций в области обработки естественного языка и звуковых данных, включая распознавание и синтез речи – ключевых направлений современного искусственного интеллекта. Освоение дисциплины направлено на подготовку специалистов, способных разрабатывать, обучать и интегрировать модели обработки речевых и текстовых сигналов в прикладные интеллектуальные системы, что необходимо для профессиональной деятельности аналитика данных, AI-инженера, разработчика машинного обучения и инженера по речевым технологиям.

Задачи дисциплины: в соответствии с поставленной целью состоят в следующем:

- Ознакомить студентов с архитектурой и принципами работы современных систем обработки естественного языка и звуковых данных, включая модели автоматического распознавания речи (ASR), синтеза речи (TTS), и языковые модели (LLM).
- Изучить методы представления, анализа и предобработки звуковых сигналов и текстовых данных, включая извлечение признаков (MFCC, spectrogram, embeddings) и нормализацию текстов.
- Научить использовать современные инструменты и библиотеки (SpeechRecognition, torchaudio, librosa, Hugging Face Transformers, FastSpeech и др.) для решения прикладных задач распознавания и генерации речи.
- Развить навыки построения и интеграции пайплайнов для обработки речевых и текстовых данных – от сбора и подготовки данных до обучения, оценки качества и внедрения моделей.
- Подготовить студентов к проектированию и анализу систем речевого взаимодействия (чат-боты, голосовые ассистенты, системы субтитрирования и озвучивания).
- Показать профессиональные кейсы применения технологий распознавания и синтеза речи в индустрии (автоматизация колл-центров, доступность контента, медиасервисы, аналитика звонков и др.).

Место дисциплины в структуре ООП ВО

Дисциплина «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи» относится к части, формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)» учебного плана. Для успешного освоения данной дисциплины необходимы знания следующих дисциплин: Математический анализ, Векторная алгебра, Основы программирования, Обработка данных на Python, Математические модели нейронных сетей, Нейросетевые технологии, Аналитика данных, A/B-тестирование и Uplift-моделирование.

Требования к уровню освоения дисциплины

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Индикаторы	Уровни освоения индикаторов компетенции
DL-4 Способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка	
DL-4.1 Применяет (проводя выбор и эксперименты) известные алгоритмы и библиотеки для обработки	Знать Основные алгоритмы NLP, библиотеки (Hugging Face,

естественного языка, предобученные глубокие нейросетевые модели для прикладных задач анализа текстов, при необходимости дообучая и валидируя на собственных наборах данных	sраCy, NLTK), архитектуры предобученных моделей (BERT, GPT), принципы дообучения и валидации. Уметь Выбирать и экспериментировать с моделями, дообучать на собственных данных, валидировать на текстовых задачах (классификация, NER, суммаризация). Владеть Применением продвинутых подходов (каузальный вывод, байесовская оптимизация, аплифт-моделирование); внедрением моделей в продакшн с мониторингом дрейфа и алертами.
DL-4.2. Определяет стек технологий, методов и алгоритмов для построения продуктов с обработкой естественного языка (диалоговые системы, вопросно-ответные системы, рекомендательные системы и т.д.)	Знать Стек технологий для диалоговых, вопросно-ответных, рекомендательных систем; методы векторизации текста (Word2Vec, SBERT, TF-IDF). Уметь Определять стек под задачу, подбирать алгоритмы и методы, анализировать векторные представления на семантическую близость. Владеть Самостоятельной разработкой или дообучением моделей векторизации под бизнес-задачу; анализом свойств векторных представлений.
DL-4.3 Имплементирует известные алгоритмы, архитектуры и модели обработки естественного языка на реальных данных, строит пайплайны обучения моделей и развертывания NLP-сервисов в продуктивной среде	Знать Архитектуры Transformer, CRF, пайплайны обучения и развертывания, принципы высоконагруженных NLP-систем (кеширование, батчинг, асинхронность). Уметь Имплементировать известные архитектуры, строить пайплайны обучения, развертывать NLP-сервисы в продуктивной среде. Владеть Разработкой кастомных гибридных моделей (Transformer+CRF); созданием высоконагруженных систем; автоматизацией цикла (A/B-тесты, активное обучение, авторетейн); оптимизацией моделей под устройства (квантизация, дистилляция, прунинг).
LLM-1 Способен применять и (или) разрабатывать генеративные модели и БЯМ	
LLM-1.1. Знает архитектуры генеративных моделей	Знать Архитектуры генеративных моделей (GPT, LLaMA, Mistral, T5, мультимодальные CLIP, Flamingo). Уметь Оценивать и сравнивать архитектуры, выбирать под задачу, применять генеративные модели для текста и речи. Владеть Проектированием новых архитектурных решений, включая мультимодальные модели.

LLM-1.2. Оценивает производительность генеративных моделей	Знать Метрики оценки генеративных моделей (BLEU, ROUGE, BERTScore, Perplexity, здравый смысл, токсичность). Уметь Оценивать производительность моделей, проводить эксперименты, интерпретировать метрики. Владеть Оптимизацией моделей под метрики и задачи; разработкой кастомных метрик.
LLM-1.4. Понимает принципы генерации в мультимодальных моделях	Знать Принципы генерации в мультимодальных моделях (текст→изображение, аудио→текст, видео→текст), архитектуры (Flamingo, GATO, ImageBind). Уметь Применять мультимодальные модели для задач распознавания и синтеза речи, связывания текста и звука. Владеть Разработкой собственных мультимодальных генеративных решений (например, звук→текст→команда).
FC-2 Способен проводить передовые исследования в области фундаментальных и генеративных моделей	
FC-2.1 Исследует и разрабатывает большие языковые модели (LLM) и другие модели для символьных данных	Знать Архитектуры LLM, методы масштабирования, нейросимволические подходы, современные исследования в области символьных данных. Уметь Исследовать и разрабатывать LLM, проводить эксперименты с символьными данными, анализировать научную литературу. Владеть Самостоятельной разработкой новых парадигм нейросимволической генерации для задач обработки языка и речи.

Содержание дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа СРС
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Введение в технологии обработки языка и звуковых данных	8	2		2	4
2.	Основы цифровой обработки звуковых сигналов	12	4		4	4
3.	Предобработка и анализ речевых и текстовых данных	12	4		4	4
4.	Архитектуры и принципы работы моделей распознавания речи (ASR)	12	4		4	4
5.	Модели синтеза речи (TTS) и генерация звуковых сигналов	12	4		4	4

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	СРС
1	2	3	4	5	6	7
6.	Языковые модели и интеграция с речевыми системами	12	4		4	4
7.	Методы оценки качества распознавания и синтеза речи	12	4		4	4
8.	Интеграция речевых технологий в прикладные системы	12	4		4	4
9.	Этические и социальные аспекты применения речевых технологий	12	4		4	4
ИТОГО по разделам дисциплины		104	34		34	36
Контроль самостоятельной работы (КСР)		4				
Промежуточная аттестация (ИКР)		0,3				
Подготовка к экзамену		35,7				
Общая трудоемкость по дисциплине		144				

Курсовые работы: не предусмотрены

Форма проведения аттестации по дисциплине: экзамен.

Автор: А.Д. Колотий, доцент КПМ, к.ф.-м.н., доцент
Г.В. Калайдина, доцент, к.ф.-м.н.