

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет компьютерных технологий и прикладной математики

УТВЕРЖДАЮ:

Проректор по учебной работе,
качеству образования – первый
проректор

_____ Хагуров Т.А.

подпись
« 29 » августа 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
Б1. В.09 Технологии обработки языка, звуковых данных, включая
распознавание и синтез речи

Направление подготовки 02.03.01 Математика и компьютерные науки

Профиль Искусственный интеллект и аналитика данных

Форма обучения очная

Квалификация бакалавр

Краснодар 2025

Рабочая программа дисциплины Технологии обработки языка, звуковых данных, включая распознавание и синтез речи составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 02.03.01 Математика и компьютерные науки

Программу составил(и):

Г.В. Калайдина, доцент, канд. физ.-мат. наук



Д.Ю. Романюта, ст. преподаватель

И.О. Фамилия, должность, ученая степень, ученое звание


подпись

Рабочая программа дисциплины утверждена на заседании центра
искусственного интеллекта
протокол № 01 «28» августа 2025 г.

Руководитель центра ИИ Коваленко А.В.


подпись

Утверждена на заседании учебно-методической комиссии факультета
компьютерных технологий и прикладной математики
протокол № 01 «28» августа 2025 г.

Председатель УМК факультета Коваленко А.В.


подпись

Рецензенты:

Мостовой Евгений Викторович, генеральный директор ООО «Портал-Юг»,
e-mail: mostovoy@portal-yug.ru

Луценко Евгений Вениаминович, доктор экономических наук, кандидат технических наук, профессор кафедры компьютерных технологий и систем Федерального государственного бюджетного образовательного учреждения высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина», e-mail: prof.lutsenko@gmail.com

1 Цели и задачи изучения дисциплины (модуля)

1.1 Цель освоения 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.

1.2 Задачи дисциплины

1. 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).
2. Изучить методы представления, анализа и предобработки звуковых сигналов и текстовых данных, включая извлечение признаков (MFCC, spectrogram, embeddings) и нормализацию текстов.
3. Научить использовать современные инструменты и библиотеки (SpeechRecognition, torchaudio, librosa, Hugging Face Transformers, FastSpeech и др.) для решения прикладных задач распознавания и генерации речи.
4. Развить навыки построения и интеграции пайплайнов для обработки речевых и текстовых данных – от сбора и подготовки данных до обучения, оценки качества и внедрения моделей.
5. Подготовить студентов к проектированию и анализу систем речевого взаимодействия (чат-боты, голосовые ассистенты, системы субтитрования и озвучивания).
6. Показать профессиональные кейсы применения технологий распознавания и синтеза речи в индустрии (автоматизация колл-центров, доступность контента, медиасервисы, аналитика звонков и др.).

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи» относится к «Части, формируемая участниками образовательных отношений» Блока 1 «Дисциплины (модули)» учебного плана.

1.4 Профессиональные роли в структуре образовательной программы

*Роль 1: NLP Engineer / Speech AI Engineer (Инженер по обработке языка и речевых данных)
Разработка, обучение и внедрение моделей обработки естественного языка и звуковых сигналов.*

Задачи:

1. *Проектирование и реализация систем автоматического распознавания и синтеза речи.*
2. *Использование библиотек и фреймворков (PyTorch, Hugging Face, torchaudio, Whisper, Tacotron и др.) для создания речевых и языковых моделей.*
3. *Оптимизация и внедрение моделей в производственные пайплайны.*

Роль 2: Data Scientist (Специалист по анализу данных и машинному обучению)

Анализ текстовых и аудиоданных, построение интеллектуальных моделей и исследование закономерностей в данных.

Задачи:

1. *Предобработка и анализ звуковых и текстовых данных.*

2. Разработка и оценка моделей классификации, кластеризации, распознавания и генерации речи.

3. Интерпретация и визуализация результатов, оценка метрик качества распознавания и синтеза.

Роль 3: AI Product / Project Manager (Менеджер проектов в области ИИ и речевых технологий)

Управление процессами проектирования и внедрения решений, использующих технологии обработки речи и языка.

Задачи:

1. Постановка задач и определение требований к системам речевого взаимодействия.

2. Планирование и координация проектов, связанных с распознаванием и синтезом речи.

3. Оценка эффективности ИИ-решений, основанных на речевых технологиях, и их влияния на бизнес-процессы.

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

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

Код, уровень и формулировка компетенции	Индикаторы	Уровни освоения индикаторов компетенции
DL-4 П Способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка	DL-4.1 Применяет (проводя выбор и эксперименты) известные алгоритмы и библиотеки для обработки естественного языка, предобученные глубокие нейросетевые модели для прикладных задач анализа текстов, при необходимости дообучая и валидируя на собственных наборах данных	Выбирает и обосновывает использование современных моделей обработки текста (BERT, RoBERTa, GPT и др.) под конкретную задачу; проводит дообучение и оценку качества; анализирует результаты с точки зрения архитектуры, набора данных и применимости модели в реальных условиях.
FC-2 Б Способен проводить фронтирные исследования в области фундаментальных и генеративных моделей	FC-2.1 Исследует и разрабатывает большие языковые модели (LLM) и другие модели для символьных данных	Проводит эксперименты с моделями, выдвигает гипотезы
LLM-1 Б Способен применять и (или) разрабатывать генеративные модели и БЯМ	LLM-1.1 Знает архитектуры генеративных моделей	Описывает принципы и отличия архитектур современных генеративных моделей (Transformer, seq2seq, Diffusion, GAN, VAE, Tacotron, Whisper и др.). Сопоставляет их применимость для задач синтеза и распознавания речи, генерации текста, мультимодальных сценариев; обосновывает выбор архитектуры под конкретную задачу.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 4 зач. ед. (144 часов), их распределение по видам работ представлено в таблице

Вид учебной работы	Всего часов	Семестры (часы)					
		7					
Контактная работа, в том числе:	72,3	72,3					
Аудиторные занятия (всего):	68	68					
Занятия лекционного типа	34	34					
Лабораторные занятия	34	34					
Занятия семинарского типа (семинары, практические занятия)							
Иная контактная работа:	4,3	4,3					
Контроль самостоятельной работы (КСР)	4	4					
Промежуточная аттестация (ИКР)	0,3	0,3					
Самостоятельная работа, в том числе:	36	36					
Курсовая работа							
Проработка учебного (теоретического) материала	10	10					
Выполнение индивидуальных заданий (подготовка сообщений, презентаций)	16	16					
Реферат							
Подготовка к текущему контролю	10	10					
Контроль:	35,7	35,7					
Подготовка к экзамену							
Общая трудоёмкость	час.	144	144				
	в том числе контактная работа	72,3	72,3				
	зач. ед	4	4				

2.2 Структура дисциплины

Распределение видов учебной работы и их трудоёмкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 1 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			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				
		Все го	Аудиторная работа			Вне уди то рная рабо та
			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 текущему контролю		35,7				
Общая трудоемкость по дисциплине		144				

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1.	Введение в технологии обработки языка и звуковых данных	Понятие обработки естественного языка (NLP) и звуковых данных. История развития речевых технологий. Основные задачи: распознавание речи (ASR), синтез речи (TTS), анализ текстов, мультимодальные модели. Обзор современных направлений: Speech AI, Multimodal AI, Voice User Interfaces (VUI). Архитектуры и стандарты представления аудио- и текстовых данных. Современные библиотеки и экосистемы: Hugging Face, torchaudio, Whisper, SpeechBrain, Mozilla TTS. Роль речевых технологий в ИИ и индустрии.	<i>Вопросы экзамена 1-5</i>
2.	Основы цифровой обработки звуковых сигналов	Физические основы звука: частота, амплитуда, тембр. Дискретизация и квантование звукового сигнала. Спектральное и временное представление звука. Преобразование Фурье, мел-шкала, фильтрационные банки. Извлечение признаков: MFCC, Mel-spectrogram, chroma features, zero-crossing rate. Работа с шумом и реверберацией. Методы визуализации и анализа аудиосигналов с использованием библиотек librosa, torchaudio, matplotlib.	<i>Вопросы экзамена 6-20</i>
3.	Предобработка и анализ речевых и текстовых данных	Этапы предобработки речевых данных: удаление шумов, нормализация громкости, разметка и сегментация аудио. Предобработка текстовых данных: токенизация, стемминг, лемматизация, удаление стоп-слов. Представление текстов в виде признаков: bag-of-words, TF-IDF, word2vec,	<i>Вопросы экзамена 21-23</i>

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		embeddings. Использование современных моделей представления текста (BERT, RoBERTa). Анализ взаимосвязей между аудио- и текстовыми данными. Формирование обучающих выборок и балансировка данных	
4.	Архитектуры и принципы работы моделей распознавания речи (ASR)	Обзор классических моделей ASR: HMM, GMM, n-граммные языковые модели. Современные нейросетевые архитектуры: RNN, LSTM, GRU, attention-механизмы. Принципы CTC-обучения. Архитектура трансформеров в задачах распознавания речи (Wav2Vec2, Whisper). Работа с предварительно обученными моделями распознавания речи. Настройка и дообучение моделей под конкретные задачи. Методы повышения точности: language model fusion, data augmentation, fine-tuning.	<i>Вопросы экзамена 24-28</i>
5.	Модели синтеза речи (TTS) и генерация звуковых сигналов	История и подходы к синтезу речи: формантный, конкатенативный, параметрический, нейросетевой. Архитектуры Tacotron, Tacotron2, FastSpeech, Glow-TTS, VITS, HiFi-GAN. Особенности генерации спектрограмм и преобразования в аудиосигнал. Методы управления параметрами синтезированной речи (эмоции, темп, акцент). Обзор мультязычных и мультимодальных систем синтеза. Практическая работа с библиотеками TTS, SpeechT5, pyttsx3, gTTS.	<i>Вопросы экзамена 29-33</i>
6.	Языковые модели и интеграция с речевыми системами	Связь NLP и Speech AI. Архитектуры больших языковых моделей (LLM) и их применение в диалоговых системах. Интеграция распознавания речи с генерацией ответов (speech-to-text-to-speech pipeline). Примеры комплексных решений: Whisper + GPT, SpeechT5, SeamlessM4T, Gemini. Обработка контекста и диалога. Режимы реального времени и стриминг речи. Использование API и фреймворков для построения голосовых ассистентов.	<i>Вопросы экзамена 34-37</i>
7.	Методы оценки качества распознавания и синтеза речи	Понятие точности и естественности речи. Метрики качества распознавания: WER (Word Error Rate), CER (Character Error Rate). Метрики синтеза речи: MOS (Mean Opinion Score), STOI, PESQ. Субъективная и автоматическая оценка качества. Анализ ошибок распознавания: подстановки, пропуски, вставки. Методы повышения устойчивости моделей к шуму и акцентам. Интерпретация и валидация результатов экспериментов.	<i>Вопросы экзамена 38-41</i>
8.	Интеграция речевых технологий в прикладные системы	Архитектура приложений с распознаванием и синтезом речи. Пайплайны обработки аудио и текста. Реализация голосовых интерфейсов и чат-ботов с речевым вводом и выводом. Автоматическая генерация субтитров и озвучивание видео. Применение речевых технологий в колл-центрах, образовательных системах, медиасервисах. Использование API (OpenAI, Google Speech, Azure Speech, Yandex SpeechKit). Разработка мини-проекта по интеграции ASR и TTS.	<i>Вопросы экзамена 42-46</i>

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
9.	Этические и социальные аспекты применения речевых технологий	Этические вопросы сбора и использования речевых данных. Конфиденциальность и защита персональной информации. Проблемы искажения и предвзятости в данных. Влияние речевых технологий на общество, образование, доступность. Противодействие дипфейкам и манипуляциям аудиоконтентом. Стандарты качества и нормативные аспекты применения Speech AI. Принципы ответственного ИИ.	Вопросы экзамена 47-48

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.3 Лабораторные занятия

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Введение в технологии обработки языка и звуковых данных	Обзор основных библиотек для работы с аудио- и текстовыми данными (librosa, torchaudio, nltk, transformers). Обзор и тестирование современных открытых моделей для распознавания и синтеза речи (Whisper, SpeechT5, FastSpeech, Tacotron).	ЛР
2.	Основы цифровой обработки звуковых сигналов	Анализ и визуализация звуковых сигналов: временная и частотная области, построение спектрограмм. Извлечение признаков из аудиосигналов: вычисление MFCC, chroma features, mel-spectrogram; интерпретация признаков.	ЛР
3.	Предобработка и анализ речевых и текстовых данных	Предобработка аудиоданных: шумоподавление, нормализация уровня громкости, сегментация речевых фрагментов. Предобработка текстовых данных: токенизация, лемматизация, стемминг, удаление стоп-слов, построение словарей и частотных распределений. Представление текстов с помощью векторных моделей (TF-IDF, Word2Vec, BERT embeddings) и визуализация семантического пространства.	ЛР
4.	Архитектуры и принципы работы моделей распознавания речи (ASR)	Использование предобученных моделей распознавания речи (SpeechRecognition, Whisper API) для преобразования аудио в текст. Дообучение модели распознавания речи на пользовательском наборе данных и анализ результатов.	ЛР

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
		Оценка качества распознавания речи с использованием метрик WER и CER.	
5.	Модели синтеза речи (TTS) и генерация звуковых сигналов	Синтез речи с использованием нейросетевых моделей Tacotron2, FastSpeech, VITS или библиотек TTS. Настройка параметров синтезированной речи: выбор голоса, скорость, эмоции, интонация.	ТЛР
6.	Языковые модели и интеграция с речевыми системами	Интеграция распознавания речи с языковой моделью (Whisper + GPT): построение пайплайна Speech-to-Text-to-Response. Реализация двунаправленного голосового взаимодействия (Speech-to-Text-to-Speech) с помощью LLM и TTS.	ЛР
7.	Методы оценки качества распознавания и синтеза речи	Сравнительный анализ качества распознавания и синтеза речи с использованием объективных (WER, MOS) и субъективных оценок; визуализация ошибок.	ЛР
8.	Интеграция речевых технологий в прикладные системы	Разработка мини-приложения с речевым вводом и выводом (голосовой чат-бот, автоматическая озвучка, субтитрирование).	ЛР
9.	Этические и социальные аспекты применения речевых технологий	Анализ рисков и ограничений речевых технологий: работа с искажёнными данными, шумами, акцентами, оценка конфиденциальности и безопасности.	ЛР

Примечание: ЛР – отчет/защита лабораторной работы, КП - выполнение курсового проекта, КР - курсовой работы, РГЗ - расчетно-графического задания, Р - написание реферата, Э - эссе, К - коллоквиум, Т – тестирование, РЗ – решение задач.

2.3.3. Примерная тематика курсовых работ (проектов)

Не предусмотрены

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Проработка и повторение лекционного материала, материала учебной и научной литературы, подготовка к семинарским занятиям	Методические указания для подготовки к лекционным и семинарским занятиям, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г
2	Подготовка к текущему контролю	Методические указания для подготовки к лекционным и семинарским занятиям, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме на языке Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии, применяемые при освоении дисциплины (модуля)

В соответствии с требованиями ФГОС в программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: чтение лекций с использованием мультимедийных технологий; метод малых групп, разбор практических задач и кейсов.

При обучении используются следующие образовательные технологии:

- Технология коммуникативного обучения – направлена на формирование коммуникативной компетентности студентов, которая является базовой, необходимой для адаптации к современным условиям межкультурной коммуникации.

- Технология разноуровневого (дифференцированного) обучения – предполагает осуществление познавательной деятельности студентов с учётом их индивидуальных способностей, возможностей и интересов, поощряя их реализовывать свой творческий потенциал. Создание и использование диагностических тестов является неотъемлемой частью данной технологии.

- Технология модульного обучения – предусматривает деление содержания дисциплины на достаточно автономные разделы (модули), интегрированные в общий курс.

- Информационно-коммуникационные технологии (ИКТ) - расширяют рамки образовательного процесса, повышая его практическую направленность, способствуют интенсификации самостоятельной работы учащихся и повышению познавательной активности. В рамках ИКТ выделяются 2 вида технологий:

- Технология использования компьютерных программ – позволяет эффективно дополнить процесс обучения языку на всех уровнях.

- Интернет-технологии – предоставляют широкие возможности для поиска информации, разработки научных проектов, ведения научных исследований.

- Технология индивидуализации обучения – помогает реализовывать личностно-ориентированный подход, учитывая индивидуальные особенности и потребности учащихся.

- Проектная технология – ориентирована на моделирование социального взаимодействия учащихся с целью решения задачи, которая определяется в рамках профессиональной подготовки, выделяя ту или иную предметную область.

- Технология обучения в сотрудничестве – реализует идею взаимного обучения, осуществляя как индивидуальную, так и коллективную ответственность за решение учебных задач.

– Игровая технология – позволяет развивать навыки рассмотрения ряда возможных способов решения проблем, активизируя мышление студентов и раскрывая личностный потенциал каждого учащегося.

– Технология развития критического мышления – способствует формированию разносторонней личности, способной критически относиться к информации, умению отбирать информацию для решения поставленной задачи.

– Комплексное использование в учебном процессе всех вышеназванных технологий стимулируют личностную, интеллектуальную активность, развивают познавательные процессы, способствуют формированию компетенций, которыми должен обладать будущий специалист.

Основные виды интерактивных образовательных технологий включают в себя:

– работа в малых группах (команде) - совместная деятельность студентов в группе под руководством лидера, направленная на решение общей задачи путём творческого сложения результатов индивидуальной работы членов команды с делением полномочий и ответственности;

– проектная технология - индивидуальная или коллективная деятельность по отбору, распределению и систематизации материала по определенной теме, в результате которой составляется проект;

– анализ конкретных ситуаций - анализ реальных проблемных ситуаций, имевших место в соответствующей области профессиональной деятельности, и поиск вариантов лучших решений;

– развитие критического мышления – образовательная деятельность, направленная на развитие у студентов разумного, рефлексивного мышления, способного выдвинуть новые идеи и увидеть новые возможности.

Подход разбора конкретных задач и ситуаций широко используется как преподавателем, так и студентами во время лекций, лабораторных занятий и анализа результатов самостоятельной работы. Это обусловлено тем, что при исследовании и решении каждой конкретной задачи имеется, как правило, несколько методов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

Семестр	Вид занятия	Используемые интерактивные образовательные технологии	количество интерактивных часов
1	ЛР	Практические занятия в режимах взаимодействия «преподаватель – студент» и «студент – студент»	12
Итого			12

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Темы, задания и вопросы для самостоятельной работы призваны сформировать навыки поиска информации, умения самостоятельно расширять и углублять знания, полученные в ходе лекционных и практических занятий.

Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4. Оценочные и методические материалы

4.1 Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Промпт инжиниринг в профессиональной деятельности».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме тестовых заданий, кейсов и **промежуточной аттестации** в форме вопросов и заданий к **зачету**.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на зачете;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом,

– в форме электронного документа.

Для лиц с нарушениями слуха:

– в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация

1	Введение в технологии обработки языка и звуковых данных	DL-4	Лабораторная работа №1	Вопросы экзамена 1-5
2	Основы цифровой обработки звуковых сигналов	DL-4	Лабораторная работа №2	Вопросы экзамена 6-20
3	Предобработка и анализ речевых и текстовых данных	DL-4	Лабораторная работа №3	Вопросы экзамена 21-23
4	Архитектуры и принципы работы моделей распознавания речи (ASR)	DL-4.1, FC-2.1	Лабораторная работа №4	Вопросы экзамена 24-28
5	Модели синтеза речи (TTS) и генерация звуковых сигналов	DL-4.1, FC-2.1	Лабораторная работа №5	Вопросы экзамена 29-33
6	Языковые модели и интеграция с речевыми системами	DL-4.1, FC-2.1, LLM-1.1	Лабораторная работа №6	Вопросы экзамена 34-37
7	Методы оценки качества распознавания и синтеза речи	DL-4.1, FC-2.1, LLM-1.1	Лабораторная работа №7	Вопросы экзамена 38-41
8	Интеграция речевых технологий в прикладные системы	DL-4.1, FC-2.1, LLM-1.1	Лабораторная работа №8	Вопросы экзамена 42-46
9	Этические и социальные аспекты применения речевых технологий	LLM-1.1	Лабораторная работа №9	Вопросы экзамена 47-48

Показатели, критерии и шкала оценки сформированных компетенций

Соответствие пороговому уровню освоения компетенций планируемым результатам обучения и критериям их оценивания (оценка: **зачтено**):

DL-1 П Способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка.

Знает основные этапы обработки текстовых и звуковых сигналов (предобработка, извлечение признаков, векторизация, классификация). Применяет известные алгоритмы и библиотеки (например, NLTK, spaCy, Transformers, Librosa, SpeechBrain, torchaudio) для анализа текстов и речи. Использует предобученные модели для задач распознавания речи (ASR), синтеза речи (TTS) и анализа текстов (sentiment, NER, summarization). Проводит настройку и валидацию моделей на собственных наборах данных, оценивает метрики качества (WER, BLEU, F1, MOS и др.). Анализирует результаты экспериментов, выбирает оптимальные подходы и архитектуры для конкретной задачи.

FC-2 Б Способен проводить фронтальные исследования в области фундаментальных и генеративных моделей

Понимает принципы работы и архитектуры современных генеративных и самообучающихся моделей (GPT, BERT, T5, Whisper, VALL-E и др.). Проводит исследовательские эксперименты с моделями для распознавания, синтеза и генерации речевых и текстовых данных. Анализирует публикации и тренды (self-supervised learning, diffusion models для TTS, multimodal LLM).

Сравнивает результаты экспериментов с существующими решениями, оценивает новизну и перспективность подходов.

LLM-1 **Способен применять и (или) разрабатывать генеративные модели и БЯМ**
Б

Знает архитектуры LLM и мультимодальных моделей (GPT-4, Gemini, Whisper, AudioPaLM, VALL-E и др.).

Понимает принципы мультимодальности — объединение текстовых и звуковых данных.

Применяет LLM для анализа, преобразования и генерации текстов и речи (например, диалоговые системы, audiobook TTS).

Настраивает пайплайны взаимодействия LLM с моделями распознавания и синтеза речи.

Анализирует ограничения, этические и правовые аспекты использования генеративных технологий.

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Практические кейсы по тематике лабораторных работ

Лабораторная работа №1. Введение в обработку языка и речи

Кейс: «Голосовой помощник для студента»

Задача: Ознакомиться с основами обработки естественного языка и аудиосигналов, научиться работать с библиотеками NLTK, spaCy, SpeechRecognition.

Что нужно сделать:

Загрузить и проанализировать аудиофайлы с речью (SpeechRecognition);

Выполнить распознавание речи и получить текст;

Провести базовую предобработку текста (токенизация, лемматизация);

Сравнить точность распознавания при разных настройках.

Инструменты: Python, SpeechRecognition, spaCy, NLTK, Jupyter Notebook

Лабораторная работа №2. Анализ и визуализация речевых данных

Кейс: «Определение эмоционального окраса голоса»

Задача: Провести спектральный анализ речи и выделить признаки, связанные с интонацией, тоном и энергией сигнала.

Что нужно сделать:

Загрузить несколько аудиозаписей;

Построить спектрограммы, определить частотные диапазоны;

Выделить характеристики (pitch, energy, tempo);

Сравнить показатели для разных эмоциональных состояний.

Инструменты: Librosa, matplotlib, pandas, Jupyter Notebook

Лабораторная работа №3. Архитектуры моделей распознавания и синтеза речи

Кейс: «Сравнение Whisper, Wav2Vec2 и Silero»

Задача: Изучить архитектуры современных моделей ASR (Automatic Speech Recognition) и провести их сравнительный анализ.

Что нужно сделать:

Подготовить небольшой датасет аудиозаписей;

Провести распознавание с помощью моделей Whisper, Wav2Vec2, Silero;

Сравнить результаты по точности (WER, CER) и скорости обработки.

Инструменты: Hugging Face, OpenAI Whisper, TorchAudio, Jupyter Notebook

Лабораторная работа №4. Текстовая предобработка и векторизация

Кейс: «Подготовка корпуса для голосового чат-бота»

Задача: Выполнить предобработку корпуса текстов, включая очистку, нормализацию и векторизацию.

Что нужно сделать:

Очистить текст от шумов и пунктуации;

Выполнить токенизацию и лемматизацию;

Сравнить разные методы векторизации: TF-IDF, Word2Vec, BERT;

Подготовить данные для обучения модели распознавания смысла.

Инструменты: spaCy, gensim, sklearn, Hugging Face, Jupyter Notebook

Лабораторная работа №5. Распознавание речи в реальном времени

Кейс: «Голосовой стенограф»

Задача: Реализовать приложение, распознающее речь с микрофона и сохраняющее текст в реальном времени.

Что нужно сделать:

Настроить захват звука через микрофон;

Реализовать распознавание речи в потоковом режиме;

Добавить фильтрацию стоп-слов и пунктуацию;

Сохранить результаты в файл.

Инструменты: Python, SpeechRecognition, PyAudio, Whisper API, FastAPI

Лабораторная работа №6. Синтез речи

Кейс: «Голосовой ответ для умного ассистента»

Задача: Научиться использовать TTS (Text-to-Speech) системы для генерации естественного звучания речи.

Что нужно сделать:

Сравнить синтезаторы gTTS, VITS, Bark, Silero;

Сгенерировать речь на русском и английском языках;

Оценить качество синтеза по естественности и разборчивости;

Настроить параметры голоса (скорость, тон, тембр).

Инструменты: gTTS, Bark, VITS, Silero TTS, Jupyter Notebook

Лабораторная работа №7. Создание интерактивного голосового ассистента

Кейс: «Мини-чат-бот с голосовым вводом и ответом»

Задача: Объединить распознавание и синтез речи в едином пайплайне взаимодействия с пользователем.

Что нужно сделать:

Реализовать цепочку: микрофон → распознавание → генерация ответа → синтез речи;

Добавить простой интерфейс (CLI или веб);

Проверить устойчивость работы и задержку;

Подготовить демонстрацию работы.

Инструменты: Python, FastAPI/Streamlit, Whisper, gTTS, OpenAI API

Лабораторная работа №8. Метрики и тестирование качества моделей

Кейс: «Сравнение систем распознавания речи»

Задача: Оценить качество моделей ASR по метрикам Word Error Rate (WER) и Character Error Rate (CER).

Что нужно сделать:

Подготовить эталонные расшифровки;

Провести распознавание речи с помощью разных моделей;

Рассчитать метрики WER и CER;

Сформировать аналитический отчет.

Инструменты: jiwer, pandas, matplotlib, Jupyter Notebook

Лабораторная работа №9. Итоговый проект: Мультиязычный ассистент

Кейс: «Интеллектуальный ассистент для онлайн-обучения»

Задача: Создать полноценное приложение, использующее язык, звук и (при желании) изображение для взаимодействия с пользователем.

Что нужно сделать:

Реализовать сбор речи и преобразование в текст;

Обработать текст (анализ смысла, генерация ответа);

Синтезировать ответ голосом;

Подготовить отчет и защитить проект.

Инструменты: Python, FastAPI/Gradio/Streamlit, Whisper, GPT API, gTTS/Bark, Hugging Face

Пример лабораторной работы

Лабораторная работа №3: Архитектуры моделей распознавания и синтеза речи

Название: Распознавание речи с помощью моделей Whisper и Silero

Цель:

Научиться использовать современные модели автоматического распознавания речи (ASR – Automatic Speech Recognition) и анализировать их точность на русскоязычных аудиозаписях.

Задачи:

1. Освоить базовые принципы работы моделей распознавания речи.
2. Провести распознавание русской речи с помощью моделей Whisper и Silero.
3. Сравнить качество результатов и оценить метрику WER (Word Error Rate).
4. Сделать выводы о применимости моделей для различных условий записи.

Ожидаемые результаты:

- Умение использовать готовые модели ASR (Whisper, Silero).
- Понимание метрик качества распознавания речи.
- Навык сравнения результатов при разных условиях и шуме записи.

Инструменты и библиотеки:

- Python 3.10+
- OpenAI Whisper / Hugging Face Transformers
- Silero Speech-To-Text
- Librosa, jiwer, torch
- Jupyter Notebook / Google Colab

Исходные данные:

3 короткие аудиозаписи (.wav) с русской речью: запись в тишине, запись с шумом, запись с разным темпом речи.

Ход работы:

1. Загрузите аудиофайлы и установите необходимые библиотеки.
2. Проведите распознавание речи.
3. Используйте Silero STT для того же файла.
4. Сравните результаты распознавания с эталонным текстом.
5. Рассчитайте WER.
6. Сделайте выводы:
 - Какая модель лучше справляется с шумом?
 - Какая модель лучше подходит для встраиваемых решений?

Пример кода:

```
import whisper
model = whisper.load_model("small")
result = model.transcribe("speech_clean.wav")
print(result["text"])
import torch
model, decoder, utils = torch.hub.load(repo_or_dir='snakers4/silero-models',
                                       model='silero_stt',
                                       language='ru')

sample = torch.load("speech_clean.wav")
text = model(sample)
print(text)
from jiwer import wer
reference = "Добрый день, сегодня мы изучаем технологии распознавания речи"
hypothesis = result["text"]
print("WER:", wer(reference, hypothesis))
```

Требования к отчёту:

- Скриншоты кода и результатов распознавания
- Таблица WER для обеих моделей
- Вывод о качестве и применимости моделей

Критерии оценки (зачтено/незачтено):

Зачтено: модели установлены, распознавание выполнено, таблица и выводы есть.

Незачтено: отсутствуют результаты одной из моделей, нет анализа WER, нет отчёта.

Лабораторная работа № 6: Синтез речи

Название: Синтез речи с использованием моделей gTTS и Bark

Цель:

Освоить технологии синтеза речи (Text-to-Speech, TTS), сравнить качество звучания, выразительность и естественность разных моделей..

Задачи:

1. Сгенерировать русскую речь из текста с помощью gTTS и Bark.
2. Настроить параметры синтеза (скорость, тон, паузы).
3. Сравнить естественность и выразительность речи.
4. Сделать выводы о применимости моделей для интерактивных голосовых ассистентов.

Ожидаемые результаты:

- Умение использовать TTS-модели для генерации речи.
- Понимание параметров, влияющих на естественность звучания
- Навык сравнительного анализа результатов синтеза.

Инструменты и библиотеки:

- Python 3.10+
- gTTS (Google Text-to-Speech)
- Suno Bark / Coqui TTS

- IPython.display.Audio
- Jupyter Notebook / Google Colab

Исходные данные:

- 3 текста (по выбору студента):
- Приветствие (10–15 слов).
- Эмоциональная фраза (восклицание, обращение).
- Нейтральная справочная фраза.

Ход работы:

1. Установите библиотеки::

```
pip install gtts bark
```

2. Синтезируйте речь с помощью gTTS:

```
from gtts import gTTS
tts = gTTS("Добрый день! Сегодня мы изучаем синтез речи.", lang='ru')
tts.save("gtts_output.mp3")
```

3. Синтезируйте тот же текст с помощью Bark:

```
from bark import generate_audio
audio_array = generate_audio("Добрый день! Сегодня мы изучаем синтез речи.",
voice_preset="v2/ru_speaker_6")
```

4. Прослушайте и сравните звучание::

```
from IPython.display import Audio
Audio("gtts_output.mp3")
Audio(audio_array, rate=22050)
```

5. Оцените качество по параметрам: Естественность, Разборчивость, Эмоциональность, Время генерации

6. Сделайте выводы.

Требования к отчёту: Скриншоты кода, Таблица оценки качества, Аудиофайлы с результатами, Вывод о применимости моделей

Критерии оценки (зачтено/незачтено):

Зачтено: обе модели применены, результаты оформлены, анализ проведён.

Незачтено: нет одного из синтезов, отсутствует сравнение или вывод.

Проверяемые компетенции комплексом практических заданий: DL-4.1; FC-2.1; LLM-1.1

Зачетно-экзаменационные материалы для промежуточной аттестации (зачет)

Вопросы для подготовки к экзамену

1. Что включает в себя понятие «обработка естественного языка» (NLP)?
2. Какие основные задачи решаются методами NLP?
3. Чем различаются уровни лингвистического анализа: морфологический, синтаксический, семантический?
4. Что такое токенизация и зачем она нужна?
5. Как работает стемминг и лемматизация?
6. Объясните понятие векторного представления слов (word embeddings).
7. В чём отличие моделей Word2Vec, GloVe и FastText?
8. Что такое контекстуальные эмбединги и чем они лучше статических?
9. Объясните архитектуру трансформеров и их применение в NLP.
10. Чем различаются модели BERT и GPT по назначению и обучению?
11. Что такое attention-механизм и как он работает?
12. Какие этапы включает препроцессинг текстовых данных?
13. Что такое языковая модель и как она обучается?
14. В чём отличие между autoregressive и autoencoding моделями?
15. Что представляет собой архитектура encoder-decoder?
16. Объясните понятие fine-tuning и его роль в адаптации моделей.
17. Какие метрики применяются для оценки качества текстовой генерации?

18. Что такое perplexity и как она интерпретируется?
19. В чём разница между задачами классификации текста и генерации текста?
20. Что такое sequence-to-sequence модели? Приведите примеры.
21. Опишите этапы обработки звукового сигнала перед распознаванием речи.
22. Что такое MFCC и для чего они используются?
23. В чём заключается принцип работы алгоритма распознавания речи?
24. Что представляет собой акустическая модель в ASR (Automatic Speech Recognition)?
25. Какова роль языковой модели в системах распознавания речи?
26. Чем отличается распознавание речи в реальном времени от офлайн-режима?
27. Что такое end-to-end модели для распознавания речи? Примеры.
28. Объясните архитектуру моделей DeepSpeech и Whisper.
29. В чём состоит задача синтеза речи (Text-to-Speech)?
30. Что такое вокодер и какую роль он играет в TTS-системах?
31. Объясните разницу между параметрическими и нейронными методами синтеза речи.
32. Какие современные модели используются для синтеза речи? (например, Tacotron, FastSpeech)
33. Что такое prosody и как она моделируется в TTS?
34. Какие подходы используются для мультязычного распознавания и синтеза речи?
35. Что такое генеративные модели в контексте обработки звука и языка?
36. Объясните применение диффузионных моделей к синтезу звука и речи.
37. Какие вызовы существуют при объединении текстовой и аудиальной информации (мультимодальные модели)?
38. Как обеспечивается интерпретируемость и устойчивость моделей обработки речи и языка?
39. Что означает понятие точности и естественности речи в системах ASR и TTS?
40. Какие типы ошибок выделяются при анализе распознавания речи (подстановки, пропуски, вставки)?
41. Какие методы используются для повышения устойчивости моделей к шуму и акцентам?
42. Как устроена архитектура приложений с распознаванием и синтезом речи?
43. Что такое пайплайн обработки аудио и текста и из каких этапов он состоит?
44. Как применяются речевые технологии в колл-центрах, образовательных системах и медиасервисах?
45. Какие возможности предоставляют API для распознавания и синтеза речи (OpenAI, Google Speech, Azure Speech, Yandex SpeechKit)?
46. Как реализуются голосовые интерфейсы и чат-боты с речевым вводом и выводом?
47. Какие этические и юридические аспекты необходимо учитывать при работе с речевыми данными?
48. Какие критерии используются для оценки курсовой работы по дисциплине «Технологии обработки языка и звуковых данных»?

Перечень компетенций (части компетенции), проверяемых оценочным средством DL-4.1; FC-2.1; LLM-1.1

Практические задания к экзамену

1. Токенизация и лемматизация текста

Задание: реализуйте на Python конвейер для обработки текста — токенизация, удаление стоп-слов, стемминг и лемматизация. Протестируйте на корпусе новостных заголовков.

2. Векторизация слов и предложений
Задание: сравните качество представлений текста при использовании CountVectorizer, TF-IDF и Word2Vec. Проведите косинусное сравнение смысловой близости между фразами.
3. Анализ тональности (Sentiment Analysis)
Задание: обучите простую модель классификации (Logistic Regression / BERT) для определения тональности отзывов. Оцените точность, полноту и F1-меру.
4. Распознавание речи (ASR) с использованием готовой модели
Задание: реализуйте скрипт, который получает аудиофайл и преобразует его в текст, используя модель (например, openai/whisper-base). Сравните результаты на 3 записях.
5. Синтез речи (Text-to-Speech)
Задание: напишите скрипт, который получает текст и синтезирует речь с помощью модели (например, gTTS или Tacotron 2). Сравните качество синтеза при разных параметрах.
6. Извлечение признаков звука (Feature Extraction)
Задание: для набора аудиозаписей вычислите спектрограммы и MFCC. Визуализируйте различия между мужским и женским голосом.
7. Классификация звуков по спектрограмме
Задание: обучите простую CNN-модель для классификации коротких аудиофрагментов (например, речь/музыка/шум). Постройте графики точности и потерь обучения.
8. Обработка команд голосового ассистента
Задание: реализуйте пайплайн: распознавание речи → классификация команды → синтез ответа. Тест: 5 голосовых команд («включи свет», «открой браузер» и т. п.).
9. Сравнение архитектур для обработки текста и звука
Задание: выберите две архитектуры (например, BERT и Whisper). Подготовьте сравнительный отчёт: тип входных данных, структура, задачи, метрики.
10. Мультимодальная генерация: текст + аудио
Задание: реализуйте простую систему, которая получает текст и создаёт краткий аудиоотчёт (синтез речи). Дополнительно: создайте подзаголовок (резюме) текста с помощью LLM.
11. Нормализация и очистка речевых данных
Задание: разработайте функцию для очистки и сегментации аудиоданных (шумоподавление, обрезка тишины). Продемонстрируйте результат до и после обработки.
12. Fine-tuning модели BERT для задачи NER (распознавание сущностей)
Задание: обучите модель на подмножестве данных с аннотацией именованных сущностей. Оцените качество распознавания.
13. TTS с управлением эмоцией или тембром
Задание: создайте несколько аудиоответов с разными стилями (спокойный, радостный, нейтральный). Опишите, какие параметры влияют на тембр и выразительность.
14. Распознавание языка по звуковому сигналу
Задание: реализуйте модель или правило, различающее язык речи по аудиозаписи (например, русский / английский). Используйте признаки MFCC или мел-спектрограммы.
15. Сравнение моделей распознавания речи (Whisper, Vosk, Wav2Vec2)
Задание: протестируйте одну и ту же запись на трёх моделях. Сравните скорость работы и точность распознавания.
16. Создание корпуса данных для речевой задачи
Задание: соберите мини-корпус (текст + аудио), очистите и аннотируйте данные для обучения TTS или ASR. Оформите структуру в формате JSON.
17. Интерпретация предсказаний модели ASR

Задание: реализуйте визуализацию внимания (attention map) для аудио-фрагмента.

Объясните, какие участки звука наиболее значимы для распознавания.

18. Диалоговая система с распознаванием и синтезом речи

Задание: создайте простую систему «вопрос — ответ», где пользователь говорит, а система отвечает голосом. Используйте Whisper + TTS.

19. Сравнение текстовых и аудиальных признаков в задаче классификации эмоций

Задание: соберите данные (текст + голос), выделите признаки и сравните точность модели по тексту и по аудио отдельно.

20. Этические аспекты обработки речевых данных

Задание: разработайте рекомендации по защите персональных данных при работе с аудиозаписями (анонимизация, удаление биометрических меток).

Перечень компетенций (части компетенции), проверяемых оценочным средством DL-4.1; FC-2.1; LLM-1.1

4.2 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Методические рекомендации, определяющие процедуры оценивания лабораторной работы.

Текущая аттестация проводится по лабораторным работам, и может принести в копилку максимум **40 баллов**. В соответствии с критериями оценки выполнения лабораторных работ Пороговый уровень (удовлетворительно) 10 баллов, (хорошо) базовый 20 баллов, продвинутый уровень (отлично) 40 баллов. Текущий бал определяется усреднением баллов по всем (9) лабораторным работам и как результат будет принадлежать отрезку от 0 до 40 баллов.

Промежуточной аттестацией по дисциплине является экзамен. Максимальная оценка, которую можно получить в качестве оценки экзамена 60 баллов: Пороговый уровень 40 баллов, (хорошо) базовый 50 баллов, продвинутый уровень (отлично) 60 баллов.

В стандартной форме экзаменационная оценка определяется следующим соответствием суммарных баллов и оценки:

0 – 49 баллов «неудовлетворительно»;

50 – 70 баллов «удовлетворительно»;

71 – 85 баллов «хорошо»;

86 – 100 баллов «отлично».

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом,

– в форме электронного документа.

Для лиц с нарушениями слуха:

– в печатной форме,

– в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме,

– в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4.3. Методические указания по организации вычислительной инфраструктуры

Требования к аппаратному и программному обеспечению рабочих мест

1. Общие положения

Лабораторные работы предназначены для формирования практических навыков применения современных технологий обработки естественного языка и звуковых данных, включая распознавание и синтез речи. Работы построены по принципу "от простого к сложному" и охватывают полный цикл разработки речевых систем — от обработки аудиосигналов до построения комплексных приложений с использованием речевых интерфейсов.

2. Аппаратное обеспечение

Компьютер: Стационарный компьютер или ноутбук (мобильные устройства не подходят). Рекомендуется многоядерный процессор (Intel Core i3/i5/i7 или AMD Ryzen) и не менее 4 ГБ ОЗУ (лучше 8 ГБ и выше).

Интернет: Стабильное широкополосное подключение со скоростью не менее 5–10 Мбит/с.

3. Программное обеспечение

Обязательное: Современная ОС (Windows 10/11, macOS, GNU/Linux), актуальный веб-браузер (Chrome, Firefox и т.д.).

Дополнительное: Для выполнения заданий требуется установить:

Git для контроля версий.

IDE/редактор (VS Code, PyCharm).

Python 3.10+ и менеджер пакетов (pip/conda).

FFmpeg и Audacity для работы со звуком.

Docker (обязателен для преподавателя, рекомендуется для студентов) для контейнеризации.

Необходимо убедиться, что на всех машинах есть права для установки ПО, а брандмауэр не блокирует исходящие подключения к внешним сервисам.

4. Архитектура вычислительной инфраструктуры

Общая инфраструктура курса построена на принципах автоматизации и включает четыре ключевых компонента:

4.1 Система контроля версий (GitLab)

Все работы хранятся в приватных Git-репозиториях на платформе GitLab.

Преподаватель создает шаблонные репозитории с готовой структурой проекта, заданиями и конфигурацией для автоматической проверки.

Каждый студент работает в своей личной репозитории, созданном из шаблона.

4.2 CI/CD конвейер (GitLab Runner и Docker)

Для автоматической проверки решений используется конвейер непрерывной интеграции (CI/CD), который описывается в файле `.gitlab-ci.yml`.

Задачи CI (автотесты) выполняются на GitLab Runner в изолированных Docker-контейнерах. Это гарантирует единообразие среды выполнения для всех студентов и исключает проблему "работает на моей машине".

Runner может быть облачным (от GitLab.com) или развернутым преподавателем на собственном сервере.

4.3 Автоматическое тестирование (Автотесты)

В шаблоны репозитория встроены наборы автотестов (на Python/PyTest и др.). При каждом коммите студента конвейер автоматически запускает эти тесты в контейнере.

Результат (успех или failure) сразу отображается в интерфейсе GitLab, позволяя студенту оперативно исправлять ошибки.

4.4 Шаблоны репозитории и примеры

Шаблоны не только задают структуру, но и содержат примеры кода и конфигураций.

Это помогает студентам сразу начать работу над задачей, минимизируя время на настройку, и обеспечивает единообразие при проверке.

5. Чек-лист корректности реализации

Для проверки работоспособности инфраструктуры убедитесь в следующем:

Рабочие места: Всё необходимое ПО установлено и работает, интернет-соединение стабильно.

GitLab и репозитории: Создана группа курса, шаблонные репозитории подготовлены, у каждого студента есть личный репозиторий с правильными правами доступа.

CI/CD конвейер: При пуше кода в репозиторий пайплайн автоматически запускается и успешно выполняет все стадии в Docker-контейнере.

Автотесты и отчеты: Автотесты выполняются корректно, а результаты их работы (отчеты в формате JUnit/HTML) сохраняются как артефакты и отображаются в интерфейсе GitLab.

План-проспект методических указаний по организации лабораторных работ по дисциплине "Технологии обработки языка, звуковых данных, включая распознавание и синтез речи"

1. Общие сведения

Образовательная программа: Искусственный интеллект и аналитика данных

Дисциплина: " Технологии обработки языка, звуковых данных, включая распознавание и синтез речи "

Вид обеспечения: Проведение лабораторных работ

Условия применения:

- Курс рассчитан на студентов 4-го года обучения.
- Наличие доступа к вычислительным ресурсам (GPU/CPU) для работы с нейросетевыми моделями обработки текста и аудио.
- Использование открытых датасетов и библиотек, доступ к предобученным моделям распознавания и синтеза речи.

2. Цели, задачи и ожидаемые результаты

Цели:

- Закрепление теоретических знаний на практике.
- Развитие навыков работы с алгоритмами обработки текста и звука, распознавания и синтеза речи.
- Подготовка к решению реальных задач в индустрии речевых и текстовых технологий.

Задачи:

- Обеспечить студентов структурированными лабораторными работами.
- Предоставить доступ к необходимым вычислительным ресурсам.
- Организовать проверку и обратную связь по выполненным работам.

Ожидаемые результаты:

- Умение применять алгоритмы и библиотеки для обработки естественного языка и аудиосигналов.
- Навыки работы с предобученными моделями для распознавания и синтеза речи.
- Опыт создания собственных приложений и анализа их эффективности.

3. Порядок реализации

3.1. Задача №1: Подготовка лабораторных работ

1) Определение тем:

- Введение в технологии обработки языка и аудиоданных
- Архитектуры моделей распознавания речи и синтеза голоса
- Предобученные модели для обработки текста и аудио
- Распознавание речи с использованием API и локальных моделей
- Синтез речи и голосовые интерфейсы
- Обработка текстовых данных: токенизация, нормализация, семантический анализ
- Интеграция распознавания и синтеза в пайплайны
- Визуализация и анализ результатов обработки аудио и текста

2) Разработка заданий:

- Пошаговые инструкции.
- Примеры кода.
- Контрольные вопросы.

Разработка заданий для лабораторной работы «Распознавание речи с использованием предобученных моделей».

Название: Распознавание речи с использованием Whisper и Vosk

Цель: Научиться применять предобученные модели распознавания речи, оценивать качество распознавания и анализировать ошибки.

Задачи:

1. Провести распознавание аудиозаписей с различным качеством и шумами.
2. Сравнить результаты нескольких моделей (Whisper, Vosk, Google Speech-to-Text).
3. Проанализировать ошибки распознавания и предложить методы их снижения.

Ожидаемые результаты:

- Умение подключать предобученные модели распознавания речи.
- Навык оценки качества распознавания (WER, CER).
- Опыт анализа влияния качества аудиозаписи на результаты распознавания.

Инструменты и библиотеки:

- Python 3.10+
- Jupyter Notebook / Google Colab
- PyTorch, TensorFlow
- Hugging Face Transformers, Whisper, Vosk
- Matplotlib / Seaborn

Исходные данные: Аудиофайлы в формате WAV/MP3 с различной длительностью и шумами.

Ход работы:

1. Установите и импортируйте необходимые библиотеки
2. Загрузите аудиофайлы для распознавания.
3. Проведите распознавание речи с помощью Whisper и Vosk.
4. Сравните результаты и вычислите WER/CER.
5. Составьте таблицу с результатами и сделайте выводы.

Пример кода (опционально):

```
import whisper
model = whisper.load_model("base")
result = model.transcribe("sample_audio.wav")
print(result["text"])
```

Анализ результатов:

- Модели справляются с чистым аудио лучше, чем с шумным.
- Whisper показывает более точное распознавание длинных фраз.
- Vosk быстрее обрабатывает короткие аудиофрагменты, но с меньшей точностью.

Требования к отчёту:

- Таблица с аудиофайлами и результатами распознавания
- Примеры кода и выводы
- Аналитическая часть с предложениями по улучшению качества распознавания
- Скриншоты или вывод кода

Критерии оценки (зачтено/незачтено):

Зачтено: все этапы работы выполнены, результаты распознавания представлены, анализ проведен.

Незачтено: пропущены этапы работы, код неработоспособен, отсутствует анализ.

Формируемые компетенции:

- DL-4 П Способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка.
- FC-2 Б Способен проводить фронтирные исследования в области фундаментальных и генеративных моделей.
- LLM-1 Б Способен применять и (или) разрабатывать генеративные модели и БЯМ.

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

5.1 Учебная литература:

1. Чернышев, С. А. Основы программирования на Python : учебное пособие для вузов / С. А. Чернышев. — Москва : Издательство Юрайт, 2022. — 286 с. — (Высшее образование). — ISBN 978-5-534-14350-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/496893>. (дата обращения: 19.07.2025).
2. Постолиит, А.В. Основы искусственного интеллекта в примерах на Python : самоучитель / Анатолий Постолиит. – Санкт-Петербург : БХВ –Петербург, 2024 – 446 с. – ISBN 978-5-9775-1818-5 – Текст : электронный // Электронная библиотека КубГУ [сайт]. – URL: http://212.192.134.46/MegaPro/UserEntry?Action=Link_FindDoc&id=276189&idb=0 (дата обращения: 20.07.2025).
3. Галушкин, А. И. Нейронные сети: основы теории / А. И. Галушкин. - Москва : Горячая линия-Телеком, 2024. - 496 с. - URL: <https://e.lanbook.com/book/448412> (дата обращения: 14.01.2025). - Режим доступа: для авториз. пользователей. - ISBN 978-5-9912-0082-0. - Текст : электронный.

4. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. - Москва : Юрайт, 2024. - 85 с. - URL: <https://urait.ru/bcode/544780> (дата обращения: 01.07.2024). - Режим доступа: для авториз. пользователей. - ISBN 978-5-534-15561-7. - Текст: электронный. URL: http://megapro.kubsu.ru/MegaPro/UserEntry?Action=Link_FindDoc&id=276597&idb=0
5. Гудфеллоу, Ян. Глубокое обучение = Deep Learning / Ян Гудфеллоу, Йошуа Бенджио, Аарон Курвилль ; перевод с английского А. А. Слинкина. - 2-е цветное изд., испр. - Москва : ДМК Пресс, 2018. - 651 с. : ил. - Библиогр.: с. 604-645. - ISBN 978-1-491-93799-0. - ISBN 978-5-97060-618-6 : 2848 р 50 к. - Текст : непосредственный. URL: http://megapro.kubsu.ru/MegaPro/UserEntry?Action=Link_FindDoc&id=224359&idb=0

5.2. Периодическая литература

1. Sun, X., Li, J., Kovalenko, A.V., Feng, W., Ou, Y. Integrating Reinforcement Learning and Learning From Demonstrations to Learn Nonprehensile Manipulation //IEEE Transactions on Automation Science and Engineering, 2023, 20(3), 1735–1744, DOI: 10.1109/TASE.2022.3185071, Q1
2. Petukhova, A.V.; Kovalenko, A.V.; Ovsyannikova, A.V. Algorithm for Optimization of Inverse Problem Modeling in Fuzzy Cognitive Maps. Mathematics 2022, 10, 3452. DOI: 10.3390/math10193452, Q1
3. Kirillova, E.; Kovalenko, A.; Urtenov, M. Study of the Current–Voltage Characteristics of Membrane Systems Using Neural Networks. AppliedMath 2025, 5, 10. <https://doi.org/10.3390/appliedmath5010010>
4. Kadurin, Artur, et al. "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." Oncotarget 8.7 (2016): 10883.
5. Kadurin, Artur, et al. "druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico." Molecular pharmaceutics 14.9 (2017): 3098-3104.
6. Polykovskiy, Daniil, et al. "Molecular sets (MOSES): a benchmarking platform for molecular generation models." Frontiers in pharmacology 11 (2020): 565644.
7. Khrabrov, Kuzma, et al. " ∇^2 DFT: A Universal Quantum Chemistry Dataset of Drug-Like Molecules and a Benchmark for Neural Network Potentials." Advances in Neural Information Processing Systems 37 (2024): 36869-36889.
8. Polykovskiy, Daniil, et al. "Entangled conditional adversarial autoencoder for de novo drug discovery." Molecular pharmaceutics 15.10 (2018): 4398-4405.
9. Николенко, Сергей, Кадури, Артур и Архангельская Екатерина. Глубокое обучение. Издательский дом "Питер", 2017.
10. Фёдоров, Е. А., Булыгин, А. В. «Искусственный интеллект в страховании: от теории к практике» // *Вопросы экономики и управления*, 2023, № 2.
11. Сабитов К.Т., Исаев Д.В. ВОЗМОЖНОСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СТРАХОВАНИИ// Научный сетевой журнал «Столыпинский вестник» №9/2024. <https://cyberleninka.ru/article/n/vozmozhnosti-iskusstvennogo-intellekta-v-strahovanii?ysclid=mh5wh3euk7572963311>
12. Ассоциация страховых организаций России (АСО) — *Отчёты по цифровизации страхового рынка РФ* (ежегодно, 2020–2024).
13. Базы данных компании «Ист Вью» <http://dlib.eastview.com>
14. Электронная библиотека GREBENNIKON.RU <https://grebennikon.ru/>

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» www.biblioclub.ru
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Базы данных и аналитические платформы

1. Google BigQuery – облачная аналитика больших данных.
2. Apache Hadoop & Spark – официальная документация и ресурсы.
3. Kaggle – датасеты, соревнования и учебные материалы.
4. Cloudera – платформа для работы с Big Data.
5. Databricks – решения на основе Apache Spark.

Справочные системы и блоги

1. Towards Data Science (Medium) – статьи по Data Science и Big Data.
2. KDnuggets – новости, обучающие материалы и обзоры инструментов.
3. O'Reilly Data & AI – книги и статьи по Big Data и машинному обучению.
4. IBM Big Data Hub – кейсы и руководства по Big Data.

Ресурсы свободного доступа

1. [Apache Spark Documentation](#)
2. [Yandex Cloud Big Data](#)
3. [Kaggle Datasets](#)

6. Методические указания для обучающихся по освоению дисциплины (модуля)

По дисциплине «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи» предусмотрено проведение лекционных занятий, направленных на формирование систематизированного представления о современных методах и технологиях обработки естественного языка и аудиосигналов. В ходе лекций студенты знакомятся с базовыми концепциями Natural Language Processing (NLP) и Speech Processing, архитектурами нейронных моделей для распознавания и синтеза речи (RNN, CNN, Transformer, wav2vec2.0, Whisper, Tacotron, FastSpeech), а также с принципами предобработки, анализа и интерпретации текстовых и звуковых данных. Особое внимание уделяется изучению алгоритмов преобразования звука в текст (ASR) и текста в речь (TTS), методам извлечения признаков (MFCC, спектрограммы, мел-частотный анализ), а также современным подходам к обучению и адаптации предобученных моделей под конкретные задачи и языки. Рассматриваются практические примеры применения технологий распознавания и синтеза речи в диалоговых системах, голосовых интерфейсах, переводчиках, системах безопасности и анализа данных клиентов. Обсуждаются ключевые проблемы и риски речевых технологий: влияние шума и акцента на качество распознавания, этические и юридические аспекты обработки голосовых данных, защита персональной информации и соблюдение конфиденциальности. После каждой лекции студентам рекомендуется выполнять практические и лабораторные задания, включающие работу с аудио- и текстовыми данными, использование библиотек SpeechRecognition, Whisper, Vosk, Hugging Face Transformers, а также эксперименты с параметрами синтеза и распознавания речи для анализа точности и естественности получаемых результатов.

Лабораторные занятия ориентированы на формирование практических навыков обработки языка и звуковых данных. Студенты последовательно осваивают методы предобработки текстов и аудио, применение предобученных нейросетевых моделей для распознавания речи, синтеза речи и анализа текста. Выполняются упражнения по настройке

параметров моделей, таких как скорость синтеза, размер окна анализа спектра, чувствительность распознавания, а также по подготовке собственных датасетов и проведению их валидации. Особое внимание уделяется интеграции технологий в приложения и сервисы: создание простых веб-интерфейсов и API для распознавания речи и генерации аудио с помощью FastAPI, Gradio или Streamlit. В рамках лабораторных работ студенты изучают методы тестирования и оценки качества моделей, проводят сравнение различных архитектур и настроек, измеряют точность распознавания, естественность синтеза и полноту анализа текста. После каждого занятия выдаются задания для самостоятельной работы, включающие настройку собственных моделей и параметров, реализацию цепочек обработки аудио и текста, а также анализ и интерпретацию полученных результатов с целью закрепления практических навыков.

Самостоятельная работа обучающихся направлена на углублённое изучение методологических и прикладных аспектов обработки языка и звуковых данных, включая распознавание и синтез речи. Рекомендуется регулярно обращаться к учебникам, документации библиотек и API (SpeechRecognition, Whisper, Vosk, Hugging Face Transformers), а также к научным публикациям по теме (arXiv, статьи OpenAI, Google DeepMind, публикации в области речевых технологий). Студенты должны научиться формулировать прикладные задачи, такие как автоматическое распознавание речи, генерация аудиоконтента, создание голосовых ассистентов, анализ и синтез текстов, разрабатывать пайплайны обработки аудио и текста, настраивать параметры распознавания и синтеза, а также оценивать точность, естественность и полноту получаемых результатов. Особое внимание уделяется разработке комплексных решений, включающих подготовку и обработку данных, проектирование моделей и промптов для генерации речи и текста, а также интеграцию результатов в web-интерфейсы, десктопные или CLI-инструменты. Навыки Python-программирования и работы с базами данных активно применяются при обработке входных и выходных данных, построении пользовательских интерфейсов и автоматизации экспериментов.

Для студентов с ограниченными возможностями здоровья предусмотрены индивидуальные консультации и адаптированные материалы. Преподаватель помогает осваивать интерфейсы взаимодействия с ИИ, объясняет ключевые понятия в доступной форме, предоставляет инструкции с альтернативным форматированием. При необходимости используются голосовые интерфейсы, увеличенный масштаб экрана, сопровождение при выполнении заданий. Индивидуальный подход обеспечивает равные условия участия в образовательном процессе и достижения запланированных результатов обучения.

Рассмотрим примеры кейсов.

Подход, определяющий установление соответствия кейсов ИП и УГТ (5), позволяет четко соотносить этапы развития технологии с вовлеченностью партнера и снижать риски при переходе от лабораторных испытаний к промышленному внедрению.

УГТ5 - Проверка данного уровня проводится в средах имитационного моделирования эмулирующих условия приближенные к реальности. Таким образом реализуется основная цель, продемонстрировать уровень готовности технологии на модельной среде максимально приближенной к реальности, а также проверить соответствие технологии требованиям к производительности (провести профилирование на реальных объемах и убедиться в эффективности процедуры масштабирования).

Кейс 1. Голосовой ассистент для строительных и девелоперских проектов

Описание: Менеджеры и инженеры AVA Group ежедневно обрабатывают большое количество текстовых отчетов и голосовых совещаний по строительству объектов. Необходимо создать голосового ассистента, который распознаёт устные запросы и текстовые данные, автоматически генерирует структурированные отчеты о текущем статусе объектов и уведомляет сотрудников голосом или текстом.

Цель: Автоматизировать сбор и анализ информации о проектах, ускорить коммуникацию между подразделениями и снизить нагрузку на менеджеров.

Ожидаемый результат: Прототип, который:

Распознаёт голосовые и текстовые запросы;

Формирует сводки с ключевыми параметрами (статус, проблемы, рекомендации);

Озвучивает отчёты голосом и/или выводит их текстом;

Демонстрирует примеры промптов для генерации точных и структурированных ответов.

Кейс 2. Анализ клиентских обращений и отзывов для гостиниц и медицинских объектов

Описание: В отелях и медицинских объектах AVA Group поступает множество текстовых сообщений и голосовых звонков клиентов. Задача – создать NLP-модель, которая: классифицирует обращения, определяет их срочность, выявляет эмоциональный тон и ключевые темы.

Цель: Повысить качество обслуживания клиентов, ускорить обработку жалоб и запросов, выявлять типовые проблемы.

Ожидаемый результат: Система, способная:

Автоматически извлекать ключевые проблемы и пожелания клиентов;

Определять тональность сообщений (негативная/нейтральная/позитивная);

Генерировать краткие рекомендации для персонала;

Представить анализ нескольких обращений с настройкой промптов для повышения точности и релевантности.

Кейс 3. Генерация маркетингового и информационного контента для логистики и строительства

Описание: AVA Group регулярно нуждается в текстовом и аудиоконтенте для внутренних отчётов, презентаций, маркетинговых материалов и обучающих видео для сотрудников и клиентов. Необходимо использовать LLM и TTS для генерации связных текстов и аудиокomentarиев на основе исходных данных о проектах, логистике и объектах.

Цель: Снизить трудозатраты на подготовку материалов и ускорить внутреннюю и внешнюю коммуникацию.

Ожидаемый результат: Прототип, который:

Генерирует тексты и аудио по заданной тематике (логистика, строительство, объекты);

Настраивает стиль и тональность контента под целевую аудиторию;

Демонстрирует влияние промптов на точность и креативность вывода;

Позволяет создавать мультимодальные отчёты (текст + аудио).

Кейс 4. Автоматическая транскрипция и аннотация лекций

Описание: Университет записывает лекции в аудиоформате. Студенты и преподаватели хотят быстро получать текстовые конспекты и аннотации с ключевыми моментами. Задача – использовать модели распознавания речи (ASR) и генеративные языковые модели для автоматической транскрипции аудио и выделения важных тезисов.

Цель: Разработать инструмент, который по аудиофайлу формирует текстовую расшифровку с краткой аннотацией основных идей. Промпт-инжиниринг используется для формирования структурированного конспекта.

Ожидаемый результат: Прототип, который получает аудиофайл лекции и возвращает: 1) полный текст лекции, 2) краткую аннотацию, 3) ключевые термины и понятия. Студент демонстрирует несколько примеров с разных лекций и объясняет, как промпты помогают структурировать текст.

Кейс 5. Голосовой помощник для поддержки пользователей

Описание: Компания внедряет голосового ассистента для ответов на вопросы клиентов. Необходимо создать систему, которая распознаёт голос, извлекает суть запроса и формирует корректный текстовый или голосовой ответ.

Цель: Настроить мультимодальный агент с использованием ASR и TTS (Text-to-Speech), способный точно понимать запросы и выдавать осмысленные ответы. Промпт-инжиниринг используется для генерации формулировок ответов и контроля стиля речи.

Ожидаемый результат: Голосовой прототип, который может принимать вопросы клиентов (по голосу), формулировать ответы и воспроизводить их голосом. Студент демонстрирует работу на нескольких сценариях (типовые вопросы, сложные вопросы) и показывает, как промпт влияет на точность и естественность ответа.

Кейс 6. Автоматическая генерация сценариев диалогов для обучения операторов

Описание: В контакт-центрах обучение операторов требует множества примеров диалогов с клиентами. Задача – автоматизировать генерацию реалистичных диалогов с разными типами клиентов и ситуаций, включая эмоциональные нюансы.

Цель: Использовать LLM для генерации текстовых диалогов на основе промптов с указанием роли, ситуации и эмоций клиента. Промпт-инжиниринг позволяет создавать сценарии с контролируемой сложностью и эмоциональной окраской.

Ожидаемый результат: Набор диалогов с различными ролями и ситуациями, готовых для обучения операторов. Студент показывает примеры сгенерированных диалогов и объясняет, как структура промпта влияет на естественность и вариативность текста.

Кейс 7. Анализ эмоциональной окраски речи

Описание: Компания собирает аудиозаписи звонков с клиентами. Необходимо определить эмоциональное состояние говорящего (радость, гнев, нейтральность, тревога) и степень удовлетворённости.

Цель: Построить модель для анализа эмоций в аудио и генерации краткого отчёта по эмоциональному состоянию. Промпт-инжиниринг используется для превращения распознанного текста и аудио-фич в осмысленные описания.

Ожидаемый результат: Прототип, который получает аудиофайл и возвращает: текстовую расшифровку и эмоциональный анализ с оценкой интенсивности эмоций. Студент демонстрирует несколько примеров звонков и показывает, как промпт помогает создавать читаемые и понятные отчёты.

Кейс 8. Генерация подкастов и аудиоконтента

Описание: Медиа-компания хочет автоматизировать создание подкастов на основе текстовых сценариев и новостных заметок. Задача – сгенерировать аудиоконтент с естественной интонацией, темпом и эмоциональной окраской.

Цель: Настроить TTS-модель с промптами для управления стилем речи, паузами и интонацией, чтобы получить реалистичный подкаст.

Ожидаемый результат: Несколько аудиофайлов с разными стилями (новостной, образовательный, развлекательный). Студент демонстрирует примеры подкастов и объясняет, как промпты и настройки TTS влияют на выразительность и естественность речи.

Кейс 9. Мультимодальный перевод с аудио на текст и другой язык

Описание: Компания хочет автоматически переводить устные выступления с одного языка на другой с генерацией текста и аудио на целевом языке. Необходимо объединить ASR, LLM и TTS для создания мультимодального пайплайна.

Цель: Создать прототип, который распознаёт речь, переводит её и синтезирует аудиоверсию на новом языке. Промпт-инжиниринг используется для управления стилем перевода и форматом текста.

Ожидаемый результат: Пример перевода аудиозаписи с последующим синтезом на другом языке. Студент демонстрирует работу модели и объясняет, как промпты обеспечивают точность перевода и естественность синтеза речи.

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю)

7.1 Перечень информационно-коммуникационных технологий

Инструменты и библиотеки: Для выполнения практических и лабораторных заданий используются современные платформы и модели обработки текста и речи. Для работы с аудио-данными и синтезом речи применяются библиотеки: SpeechRecognition, Whisper, Vosk, gTTS и другие TTS/STT-решения. Для NLP-анализа и оценки качества генерации текста используются nltk, spaCy, rouge-score и аналогичные инструменты. Веб-разработка и интеграция систем выполняются на Python (версии 3.7+) с использованием FastAPI, Uvicorn, requests и dotenv. Для экспериментов и интерактивной работы с кодом и моделями применяются Google Colab и Jupyter Notebook. Оформление отчетов и презентация результатов может осуществляться в MS Word, Markdown-редакторе или других средствах текстового оформления.

Исходные данные: для экспериментов и лабораторных заданий используются учебные текстовые и аудио-материалы. Это могут быть: примеры устной и письменной речи, диалоги, технические тексты, инструкции, маркетинговые описания, звонки клиентов или аудиозаписи лекций. Для задач распознавания речи используются аудиофайлы разной длительности и качества, для синтеза речи — тексты, которые необходимо озвучить. Для мультимодальных экспериментов могут использоваться сопроводительные изображения или видеозаписи. Все данные выбираются или подготавливаются преподавателем заранее, могут быть как открытыми, так и сгенерированными специально для учебных целей.

Программное обеспечение и ИКТ: необходим доступ к интернету. Рекомендуется использовать среды разработки Jupyter Notebook или Google Colab, а также удобный код-редактор (Visual Studio Code). Для хранения и совместной работы со студентами можно задействовать учебную информационную систему (LMS). Кроме того, для анализа результатов и визуализации могут понадобиться Excel или Google Sheets.

1. Облачные платформы и сервисы
cloud.ru, YandexCloud, AWS/GCP/Azure – облачные вычисления
2. Системы управления версиями и коллаборации
Git/GitHub/GitLab – контроль версий кода и совместная разработка
4. Система управления обучением
5. Библиотеки и инструменты

Python (3.7+), библиотеки для NLP и аудио: SpeechRecognition, Whisper, Vosk, Hugging Face Transformers, librosa, numpy, matplotlib — для обработки текста и звуковых данных, анализа результатов и визуализации.

Moodle – сдача работ

7.2 Перечень лицензионного и свободно распространяемого программного обеспечения

1. Лицензионное ПО
VSCode – IDE для Python (свободнораспространяемое)
LibreOffice – оформление отчетов (свободнораспространяемое)

2. Свободное ПО (Open Source)
Hugging Face Transformers – предобученные модели (BERT, GPT)
Gensim – тематическое моделирование и word2vec
GitLab, GIT, MLFlow, Docker, Kubernetes, Terraform. Фреймворки для ML:
PyTorch/TensorFlow – разработка нейросетей
scikit-learn – классические алгоритмы ML

Инструменты для визуализации:

Streamlit/Gradio – создание веб-интерфейсов для моделей
Matplotlib/Seaborn – графики и анализ данных

СУБД:

SQLite/PostgreSQL – хранение структурированных данных
FAISS/Annoy – векторный поиск

8. Материально-техническое обеспечение по дисциплине (модулю)

Виртуальные машины, кластер Managed Kubernetes и ресурсы GPU в облаке предоставляется индустриальным партнером ПАО «Сбербанк»:

№	Продукт	Параметры продукта	Кол-во	Кол-во конфигураций	Ед. изм.
1	Виртуальная машина	Виртуальная машина 10% vCPU 2 vCPU 4 RAM	1	60	Шт
		ОС Ubuntu 22.04	1		Шт
		Системный диск SSD	1		Шт
			10		Гб
		Аренда публичного IP	1		Шт
2	Виртуальная машина с GPU	Виртуальная машина с GPU NVIDIA® Tesla® V100 2 GPU 8 vCPU 128 Гб RAM	1	1	Шт
		ОС Ubuntu_24.04	1		Шт
		Системный диск SSD	1		Шт
			2000		Гб
		Диск SSD	1		Шт
			4096		Гб
		Диск SSD	1		Шт
	4096		Гб		
3	K8S	Аренда публичного IP	1		Шт
		Master node 8 vCPU 16 RAM	1	1	Шт
		Worker node 10% доля 4 vCPU 32 RAM	5		Шт
		Worker node SSD-NVME	64		Гб
	Аренда публичного IP	1		Шт	

4	ML Inference Instance Type GPU	Время работы в месяц	40	1	Ч
		Инстанс 8 x NVIDIA® H100 NVLink PCIe 160 vCPU 1520 GB RAM	1		Шт
		Количество запросов к ML-моделям	1		Млн. Шт
		Кэш ML-моделей	160		Гб
5	LLM	Токены GigaChat 2 Max	50		Млн. Шт
		Токены Embeddings	400		Млн. Шт

Дополнительные облачные ресурсы предоставляются технологическим партнером Yandex Cloud.

№	Вид работ	Наименование учебной аудитории, ее оснащенность оборудованием и техническими средствами обучения
1.	Лекционные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
2.	Лабораторные занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, проектором, программным обеспечением
3.	Практические занятия	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения
4.	Групповые (индивидуальные) консультации	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
5.	Текущий контроль, промежуточная аттестация	Аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, компьютерами, программным обеспечением
6.	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета.

Примечание: Конкретизация аудиторий и их оснащение определяется ОПОП.