МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГЭУ»)

К Г Э У

УТВЕРЖДАЮ
Директор института цифровых технологий и экономики
_____Ю.В. Торкунова
«24» ноября 2021 г.

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 04637A9600B7AE93974C7182805C6B90EF
Владелец: Торкунова Юлия Владимировна
Действителен: с 17.06.2022 до 17.06.2023

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА

Направление подготовки	09.04.01 Информатика и вычислительная техника
Направленность (профиль)	Инженерия искусственного интеллекта
Квалификация	Магистр
Форма обучения	Очная

Перечень сведений о рабочей программе	Учетные данные
Образовательная программа Инженерия искусственного интеллекта	**Код ОП** 09.04.01
Направление подготовки Информатика и вычислительная техника	**Код направления и уровня подготовки** 09.04.01

Программа составлена автором:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

Программа оформлена в соответствие с ПОЛОЖЕНИЕМ О ПОРЯДКЕ РАЗРАБОТКИ И УТВЕРЖДЕНИЯ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ – ПРОГРАММ БАКАЛАВРИАТА, ПРОГРАММ СПЕЦИАЛИТЕТА И ПРОГРАММ МАГИСТРАТУРЫ В КГЭУ

Рекомендовано учебно-методическим советом Института цифровых технологий и экономики ФГБОУ ВО «КГЭУ»
Протокол № 4 от 24.11.2021 г.

1. Цель, задачи и планируемые результаты обучения по дисциплине

Целью освоения дисциплины является изучение математического аппарата, используемого в основе методов обработки естественных языков, программных инструментов для обработки естественных языков и приобретение практических навыков в профессиональной деятельности.

Задачами дисциплины являются:

– получение теоретических знаний и практических навыков обработки естественно-языковых текстов;

– выявление сложностей, связанных с применением существующих методов обработки естественно-языковых текстов;

– использование полученных знаний по разработке, адаптации и использованию новейших средств информатики для обработки текстов на естественных языках.

Компетенции, формируемые у обучающихся, запланированные результаты обучения по дисциплине, соотнесенные с индикаторами достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине (знать, уметь, владеть)
Профессиональные компетенции (ПК)		
ПК-7 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях	ПК-7.1 Руководит проектами в области сквозной цифровой субтехнологии «Компьютерное зрение»	*Знать:* принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение» *Уметь:* руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение»
	ПК-7.2 Руководит проектами в области сквозной цифровой субтехнологии «Обработка естественного языка»	*Знать:* принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» *Уметь:* руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»
	ПК-7.3 Исследует и анализирует развитие новых направлений и	*Знать:* современное состояние и перспективы развития новых направлений, методов и технологий в

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине (знать, уметь, владеть)
Профессиональные компетенции (ПК)		
	перспективных методов и технологий в области искусственного интеллекта, участвует в исследовательских проектах по развитию перспективных направлений в области искусственного интеллекта (алгоритмическая имитация биологических систем принятия решений, автономное самообучение и развитие адаптивности алгоритмов к новым задачам, автономная декомпозиция сложных задач, поиск и синтез решений)	области искусственного интеллекта *Уметь:* проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения
ОПК-9 Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта	ОПК-9.1 Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта	*Знать:* инструментальные среды, программно-технические платформы для решения профессиональных задач *Уметь:* применять инструментальные среды, программно-технические платформы для решения профессиональных задач
	ОПК-9.2 Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	*Знать:* принципы разработки оригинальных программных средств для решения профессиональных задач *Уметь:* разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта

2. Место дисциплины в структуре ОПОП

Дисциплина *Обработка естественного языка* относится к части, формируемой участниками образовательных отношений учебного плана по направлению подготовки 09.04.01 Информатика и вычислительная техника.

Код	Предшествующие дисциплины	Последующие дисциплины

компетенции	(модули), практики, НИР, др.	(модули), практики, НИР, др.
ПК-7	Операционная система Linux Глубокие нейронные сети на Python Компьютерное зрение	Проектный практикум 3 Учебная практика (проектная практика) Производственная практика (преддипломная практика) Выполнение и защита выпускной квалификационной работы
ОПК-9	Программирование на Python Машинное обучение Глубокие нейронные сети на Python Компьютерное зрение	Выполнение и защита выпускной квалификационной работы

Для освоения дисциплины обучающийся должен:

Знать:

– основные методы анализа данных;

– основные методы принятия решений.

Уметь:

– выбирать подходящие методы анализа данных;

– выбирать подходящие методы принятия решений.

Владеть:

– методами анализа данных;

– методами анализа данных.

3. Структура и содержание дисциплины

3.1. Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетных единицы (ЗЕ), всего 108 часов, из которых 26 часов составляет контактная работа обучающегося с преподавателем (занятия лекционного типа 8 часов, занятия семинарского типа (практические, семинарские занятия, лабораторные работы и т.п.) 16 часов, групповые и индивидуальные консультации 0 часов, прием экзамена (КПА), экзамен - 0 часов, самостоятельная работа обучающегося 82 часа, контроль самостоятельной работы (КСР) - 2 часа.

Вид учебной работы	Всего часов	Семестр
		3
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	108	108
КОНТАКТНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ С ПРЕПОДАВАТЕЛЕМ, в том числе:	26	26
Лекции (Лек)	8	8
Практические (семинарские) занятия (Пр)	16	16
Консультации	0	0
Контроль самостоятельной работы и иная контактная работа (КСР)	2	2
Контактные часы во время аттестации (КПА)	0	0

САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ (СРС), в том числе:	82	82
Подготовка к промежуточной аттестации в форме: *зачёта*	0	0
ФОРМА ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ (За – зачет, ЗО – зачет с оценкой, Э – экзамен)	За	За

3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Разделы дисциплины	**Семестр**	**Распределение трудоемкости (в часах) по видам учебной работы, включая СРС**								**Итого**	**Формируемые результаты обучения (знания, умения, навыки)**	**Литература**	**Формы текущего контроля**	**Формы промежуточной аттестации**	**Максимальное количество баллов по балльно - рейтинговой системе**
		Занятия лекционного типа	Занятия практического / семинарского типа	Лабораторные работы	Групповые консультации	Самостоятельная работа студента, в т.ч.	*Контроль самостоятельной работы (КСР)*	*подготовка к промежуточной*	Сдача зачета / экзамена						
Раздел 1. Теоретические аспекты обработки естественного языка.	3	1	2			12				13	ПК-7.1-З1, ПК-7.1-У1, ПК-7.2-З1, ПК-7.2-У1	Л1.1, Л1.2, Л2.1	ПЗ		13
Раздел 2. Предварительная обработка текста.	3	1	2			10				13	ПК-7.2-З1, ПК-7.2-У1	Л1.1, Л1.2, Л2.1	ПЗ		12
Раздел 3. Векторизация текста.	3	1	2			10				13	ПК-7.3-З1, ПК-7.3-У1	Л1.1, Л1.2, Л2.1	ПЗ		13
Раздел 4. Машинное	3	1	2			10				13	ОПК-9.1-	Л1.1,	ПЗ		12

обучение для обработки текстов.											З1, ОПК-9.1-У1	Л1.2, Л2.1			
Раздел 5. Нейронные сети в решении задач текстовой обработки.	3	1	2			10				13	ОПК-9.1-З1, ОПК-9.1-У1	Л1.1, Л1.2, Л2.1	ПЗ		13
Раздел 6. Языковая модель.	3	1	2			10				13	ОПК-9.2-З1, ОПК-9.2-У1	Л1.1, Л1.2, Л2.1	ПЗ		12
Раздел 7. Поиск именованных сущностей.	3	1	2			10				13	ОПК-9.2-З1, ОПК-9.2-У1	Л1.1, Л1.2, Л2.1	ПЗ		13
Раздел 8. Механизм внимания. Трансформер.	3	1	2			10				13	ОПК-9.2-З1, ОПК-9.2-У1	Л1.1, Л1.2, Л2.1	ПЗ		12
ИТОГО		8	16			82	2		1	108					100

3.3. Тематический план лекционных занятий

№ п/п	Темы лекционных занятий	Трудоемкость, час.
1	Синтаксический, морфологический, семантический и графематический анализ, омонимия, задачи лингвистического анализа	1
2	Очистка текста, токенизация, стемминг, лемматизация, удаление стоп-слов, фильтрация наиболее частотных и наименее частотных слов.	1
3	Построение словаря, мешок слов, TF-IDF, word2vec, fasttext, LDA, LSI, GloVe.	1
4	Решение задач классификации и определения тональности методами классического машинного обучения на основе векторных моделей.	1
5	Архитектуры нейронных сетей для обработки текстов: рекуррентные (LSTM, GRU), одномерные сверточные. Применение нейронных сетей для обработки тектов.	1
6	Языковая модель и дистрибутивная семантика. Обучение векторной модели. Задача генерации текста. Различные подходы к генерации текста.	1
7	Задача поиска именованных сущностей в тексте. Применение нейронных сетей для поиска именованных сущностей.	1
8	Механизм внимания в нейронных сетях. Применение механизма внимания для обработки текста. Нейронные сети с архитектурой	1

	Transformer. Нейронные сети BERT, GPT. Перенос обучения.	
	Всего	**8**

3.4. Тематический план практических занятий

№ п/п	Темы практических работ	Трудоемкость, час.
1	Предварительная обработка текста для анализа.	2
2	Векторизация текста.	2
3	Классификация текста с использованием классических методов машинного обучения.	2
4	Классификация текста с использованием глубоких нейронных сетей.	2
5	Языковая модель. Обучение языковой модели.	2
6	Автоматическая генерация текста.	2
7	Поиск именованных сущностей в тексте.	2
8	Механизм внимания в нейронных сетях. Сети с архитектурой Transformer.	1
9	Перенос обучения в задачах обработки текстов.	1
	Всего	**16**

3.5. Тематический план лабораторных работ

Данный вид работы не предусмотрен учебным планом.

3.6. Самостоятельная работа студента

Номер раздела дисциплины	Вид СРС	Содержание СРС	Трудоемкость, час.
1	Изучение теоретического материала, выполнение домашних заданий	Изучение теоретического материала, подготовка к тестированию	12
2	Изучение теоретического материала, выполнение домашних заданий	Изучение теоретического материала, подготовка к тестированию	10
3	Изучение теоретического материала, выполнение домашних заданий	Изучение теоретического материала, подготовка к тестированию	10
4	Изучение теоретического материала, выполнение домашних заданий	Изучение теоретического материала, подготовка к тестированию	10
5	Изучение теоретического	Изучение теоретического материала, подготовка к	10

	материала, выполнение домашних заданий	тестированию	
6	Изучение теоретического материала, выполнение домашних заданий	Изучение теоретического материала, подготовка к тестированию	10
7	Изучение теоретического материала, выполнение домашних заданий	Изучение теоретического материала, подготовка к тестированию	10
7	Изучение теоретического материала, выполнение домашних заданий	Изучение теоретического материала, подготовка к тестированию	10
		Всего	82

4. Образовательные технологии

В процессе обучения используются:

- дистанционные курсы, размещенные на площадке LMS Moodle, URL: http://lms.kgeu.ru/;

- электронные образовательные ресурсы (ЭОР), размещенные в личных кабинетах студентов Электронного университета КГЭУ, URL: http://e.kgeu.ru/

5. Оценивание результатов обучения

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля успеваемости, проводимого по балльно-рейтинговой системе (БРС), и промежуточной аттестации.

Текущий контроль успеваемости осуществляется в течение семестра, включает выполнение практических заданий.

Итоговой оценкой результатов освоения дисциплины является оценка, выставленная во время промежуточной аттестации обучающегося (*зачёт*) с учетом результатов текущего контроля успеваемости. Результат (зачтено / не зачтено) промежуточной аттестации в форме зачёта определяется по совокупности результатов текущего контроля успеваемости по дисциплине.

Обобщенные критерии и шкала оценивания уровня сформированности компетенции (индикатора достижения компетенции) по итогам освоения дисциплины:

Планируемые результ-	Обобщенные критерии и шкала оценивания результатов обучения			
	неудовлетво-	удовлетворительно	хорошо	отлично

таты обучения	рительно			
	не зачтено	зачтено		
Полнота знаний	*Уровень знаний ниже минимальных требований, имеют место грубые ошибки*	*Минимально допустимый уровень знаний, имеет место много негрубых ошибок*	*Уровень знаний в объеме, соответствующем программе, имеет место несколько негрубых ошибок*	*Уровень знаний в объеме, соответствующем программе подготовки, без ошибок*
Наличие умений	*При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки*	*Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме*	*Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами*	*Продемонстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме*
Наличие навыков (владение опытом)	*При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки*	*Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами*	*Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами*	*Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов*
Характеристика сформированности компетенции (индикатора достижения компетенции)	*Компетенция в полной мере не сформирована. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач*	*Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач, но требуется дополнительная практика по большинству практических задач*	*Сформированность компетенции в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных практических (профессиональных) задач*	*Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных практических (профессиональных) задач*
Уровень сформированности компетенции (индикатора достижения компетенции)	Низкий	Ниже среднего	Средний	Высокий

Шкала оценки результатов обучения по дисциплине:

Код компе-	Код индикатора	Заплани-рованные	Уровень сформированности компетенции (индикатора достижения компетенции)

тенции	достижения компетенции	результаты обучения по дисциплине	Высокий	Средний	Ниже среднего	Низкий
			Шкала оценивания			
			отлично	хорошо	удовлет-ворительно	неудов-летвори-тельно
			зачтено			не зачтено
ПК-7	ПК-7.1	знать:				
		принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение»	Знает все основные принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение», не допускает ошибок	Знает многие основные принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение», может допустить несколько негрубых ошибок	Знает некоторые основные принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение», допускает много негрубых ошибок	Уровень знаний ниже минимального требования, допускает грубые ошибки
		уметь:				
		руководить проектами по созданию, внедрению и поддержке систем искусственного	Демонстрирует умение руководить проектами по	Демонстрирует умение руководить проектами по	Частично демонстрирует умение руководить проектам	Не сформировано умение руководить проектам

		интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение»	созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение», не допускает ошибок	созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение», может допустить несколько негрубых ошибок	и по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение», допускает много негрубых ошибок	и по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение», допускает грубые ошибки
	ПК-7.2	знать:				
		принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»	Знает все основные принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнол	Знает многие основные принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой	Знает некоторые основные принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной	Уровень знаний ниже минимального требования, допускает грубые ошибки

			огии «Обработка естественного языка», не допускает ошибок	субтехнологии «Обработка естественного языка», может допустить несколько негрубых ошибок	цифровой субтехнологии «Обработка естественного языка», допускает много негрубых ошибок	
		уметь:				
		руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»	Демонстрирует умение руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка», не допускает ошибок	Демонстрирует умение руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка», может допустить несколько негрубых ошибок	Частично демонстрирует умение руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка», допускает много негрубых ошибок	Не сформировано умение руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка», допускает грубые ошибки
	ПК-7.3	знать:				
		современное состояние и перспективы развития новых направлений, методов и	Знает все основные современное состояние и	Знает многие основные современное состояние	Знает некоторые основные современное	Уровень знаний ниже минимального требован

		технологий в области искусственного интеллекта	перспективы развития новых направлений, методов и технологий в области искусственного интеллекта, не допускает ошибок	и перспективы развития новых направлений, методов и технологий в области искусственного интеллекта, может допустить несколько негрубых ошибок	состояние и перспективы развития новых направлений, методов и технологий в области искусственного интеллекта, допускает много негрубых ошибок	ия, допускает грубые ошибки
		уметь:				
		проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения	Демонстрирует умение проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения, не допускает ошибок	Демонстрирует умение проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения, может допустить несколько негрубых ошибок	Частично демонстрирует умение проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения, допускает много негрубых ошибок	Не сформировано умение проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения, допускает грубые

						ошибки
ОПК-9	ОПК-9.1	знать:				
		инструментальные среды, программно-технические платформы для решения профессиональных задач	Знает все основные инструментальные среды, программно-технические платформы для решения профессиональных задач, не допускает ошибок	Знает многие основные инструментальные среды, программно-технические платформы для решения профессиональных задач, может допустить несколько негрубых ошибок	Знает некоторые основные инструментальные среды, программно-технические платформы для решения профессиональных задач, допускает много негрубых ошибок	Уровень знаний ниже минимального требования, допускает грубые ошибки
		уметь:				
		применять инструментальные среды, программно-технические платформы для решения профессиональных задач	Демонстрирует умение применять инструментальные среды, программно-технические платформы для решения профессиональных задач, не допускает ошибок	Демонстрирует умение применять инструментальные среды, программно-технические платформы для решения профессиональных задач, может допустить несколько негрубых ошибок	Частично демонстрирует умение применять инструментальные среды, программно-технические платформы для решения профессиональных задач, допускает много негрубых ошибок	Не сформировано умение применять инструментальные среды, программно-технические платформы для решения профессиональных задач, допускает грубые ошибки
	ОПК-9.2	знать:				
		принципы разработки оригинальных программных	Знает все основные принципы разработк	Знает многие основные принципы	Знает некоторые основные	Уровень знаний ниже минимал

		средств для решения профессиональных задач	и оригинальных программных средств для решения профессиональных задач, не допускает ошибок	разработки оригинальных программных средств для решения профессиональных задач, может допустить несколько негрубых ошибок	принципы разработки оригинальных программных средств для решения профессиональных задач, допускает много негрубых ошибок	ьного требования, допускает грубые ошибки
		уметь:				
		разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	Демонстрирует умение разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта, не допускает ошибок	Демонстрирует умение разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта, может допустить несколько негрубых ошибок	Частично демонстрирует умение разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта, допускает много негрубых ошибок	Не сформировано умение разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта, допускает грубые ошибки

Оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации приведены в Приложении к рабочей программе дисциплины. *Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре-разработчике в бумажном и электронном виде.*

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Учебно-методическое обеспечение

Основная литература

№ п/п	Автор(ы)	Наименование	Вид издания (учебник, учебное пособие, др.)	Место издания, издательство	Год издания	Адрес электронного ресурса	Кол-во экземпляров в библиотеке КГЭУ
1	Ганегедара Т.	Обработка естественного языка с Tensorflow	руководство	М.: ДМК Пресс	2020	https://e.lanbook.com/book/140584	1
2	Гольдберг Й.	Нейросетевые методы в обработке естественного языка	руководство	М.: ДМК Пресс	2019	https://e.lanbook.com/book/131704	1

Дополнительная литература

№ п/п	Автор(ы)	Наименование	Вид издания (учебник, учебное пособие, др.)	Место издания, издательство	Год издания	Адрес электронного ресурса	Кол-во экземпляров в библиотеке КГЭУ
1	Паттерсон Дж., Гибсон А.	Глубокое обучение с точки зрения практика	учебник	М.: ДМК Пресс	2018	https://e.lanbook.com/book/116122	1

6.2. Информационное обеспечение

6.2.1. Электронные и интернет-ресурсы

№ п/п	Наименование электронных и интернет-ресурсов	Ссылка
1	*Электронно-библиотечная система «Лань»*	https://e.lanbook.com/
2	*Электронно-библиотечная система «ibooks.ru»*	https://ibooks.ru/
3	*Электронно-библиотечная система «book.ru»*	https://www.book.ru/
4	*Энциклопедии, словари, справочники*	http://www.rubricon.com
5	*Портал "Открытое образование"*	http://npoed.ru
6	*Единое окно доступа к образовательным ресурсам*	http://window.edu.ru
7	Научная электронная библиотека	http://elibrary.ru
8	Портал искусственного интеллекта	http://www.aiportal.ru/
9	Портал изучения средств построения нечётких	http://matlab.exponenta.ru

	интеллектуальных систем	/fuzzylo gic/index.php
10	Интеллектуальные технологии идентификации	http://matlab.exponenta.ru /fuzzylo gic/book5/index.php

6.2.2. Профессиональные базы данных

№ п/п	Наименование профессиональных баз данных	Адрес	Режим доступа
1	*Официальный интернет-портал правовой информации*	http://pravo.gov.ru	http://pravo.gov.ru
2	*Справочная правовая система «Консультант Плюс»*	http://consultant.ru	http://consultant.ru
3	*Справочно-правовая система по законодательству РФ*	http://garant.ru	http://garant.ru

6.2.3. Информационно-справочные системы

№ п/п	Наименование информационно-справочных систем	Адрес	Режим доступа
1	*Научная электронная библиотека*	http://elibrary.ru	http://elibrary.ru
2	*Российская государственная библиотека*	http://www.rsl.ru	http://www.rsl.ru
3	*Международная реферативная база данных научных изданий zbMATH*	http://www.zbmath.org	http://www.zbmath.org
4	*Международная реферативная база данных научных изданий Springerlink*	http:// link.springer.com	http:// link.springer.com
5	*Образовательный портал*	http://www.ucheba.com	http://www.ucheba.com

6.2.4. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№ п/п	Наименование программного обеспечения	Способ распространения (лицензионное/свободно)	Реквизиты подтверждающих документов
1	Windows 7 Профессиональная (Pro)	Пользовательская операционная система	№2011.25486 от 28.11.2011
2	Visual Studio Express	Инструмент создания Web приложений	https://visualstudio.microsoft.com/ru/vs/express/
3	Браузер Chrome	Система поиска информации в сети интернет (включая	https://www.google.com /intl/ru/chrome/

		русскоязычный интернет).	
4	Браузер Firefox	Свободный веб-браузер	https://www.mozilla.org/ ru/firefox/new/
5	OpenOffice	Пакет офисных приложений.Одним из первых стал поддерживать новый открытый формат OpenDocument. Официально поддерживается на платформах Linux	https://www.openoffice. org/ru/download/index.h tml
6	Adobe Acrobat	Пакет программ	https://get.adobe.com/ru/ reader/
7	LMS Moodle	Это современное программное обеспечение	https://download.moodle .org/releases/latest/

7. Материально-техническое обеспечение дисциплины

№ п/п	Вид учебной работы	Наименование специальных помещений и помещений для СРС	Оснащенность специальных помещений и помещений для СРС
1	Лекционные занятия	Учебная аудитория для проведения занятий лекционного типа В-103	180 посадочных мест, доска аудиторная, акустическая система, проектор, усилитель-микшер для систем громкой связи, экран, микрофон, миникомпьютер, монитор, подключение к сети "Интернет", доступ в электронную информационно-образовательную среду
2	Лабораторные работы	Учебная лаборатория В-617	44 посадочных места (20 по центру - 24 по краю), доска ученическая, моноблок (10 шт.), подключение к сети «Интернет», доступ в электронную информационно-образовательную среду
		Лаборатория В-619	46 посадочных мест (24 по центру + 22 по краю), доска ученическая; моноблок (12 шт.), подключение к сети «Интернет», доступ в электронную информационно-образовательную среду
3	Практические занятия	Учебная лаборатория В-617	44 посадочных места (20 по центру - 24 по краю), доска ученическая, моноблок (10 шт.), подключение к сети «Интернет», доступ в электронную

			информационно-образовательную среду
		Лаборатория В-619	46 посадочных мест (24 по центру + 22 по краю), доска ученическая; моноблок (12 шт.), подключение к сети «Интернет», доступ в электронную информационно-образовательную среду
4	Самостоятельная работа обучающегося	Компьютерный класс с выходом в Интернет В-600а	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультимедийный проектор, компьютер (ноутбук), экран), видеокамеры, программное обеспечение
		Читальный зал библиотеки	Специализированная мебель, компьютерная техника с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, мультимедийный проектор, экран, программное обеспечение

8. Особенности организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов

Лица с ограниченными возможностями здоровья (ОВЗ) и инвалиды имеют возможность беспрепятственно перемещаться из одного учебно-лабораторного корпуса в другой, подняться на все этажи учебно-лабораторных корпусов, заниматься в учебных и иных помещениях с учетом особенностей психофизического развития и состояния здоровья.

Для обучения лиц с ОВЗ и инвалидов, имеющих нарушения опорно-двигательного аппарата, обеспечены условия беспрепятственного доступа во все учебные помещения. Информация о специальных условиях, созданных для обучающихся с ОВЗ и инвалидов, размещена на сайте университета www//kgeu.ru. Имеется возможность оказания технической помощи ассистентом, а также услуг сурдопереводчиков и тифлосурдопереводчиков.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушенным слухом справочного, учебного материала по дисциплине обеспечиваются следующие условия:

- для лучшей ориентации в аудитории, применяются сигналы оповещения о начале и конце занятия (слово «звонок» пишется на доске);

- внимание слабослышащего обучающегося привлекается педагогом жестом (на плечо кладется рука, осуществляется нерезкое похлопывание);

- разговаривая с обучающимся, педагогический работник смотрит на него, говорит ясно, короткими предложениями, обеспечивая возможность чтения по губам.

Компенсация затруднений речевого и интеллектуального развития слабослышащих обучающихся проводится путем:

- использования схем, диаграмм, рисунков, компьютерных презентаций с гиперссылками, комментирующими отдельные компоненты изображения;

- регулярного применения упражнений на графическое выделение существенных признаков предметов и явлений;

- обеспечения возможности для обучающегося получить адресную консультацию по электронной почте по мере необходимости.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушениями зрения справочного, учебного, просветительского материала, предусмотренного образовательной программой по выбранному направлению подготовки, обеспечиваются следующие условия:

- ведется адаптация официального сайта в сети Интернет с учетом особых потребностей инвалидов по зрению, обеспечивается наличие крупношрифтовой справочной информации о расписании учебных занятий;

- педагогический работник, его собеседник (при необходимости), присутствующие на занятии, представляются обучающимся, при этом каждый раз называется тот, к кому педагогический работник обращается;

- действия, жесты, перемещения педагогического работника коротко и ясно комментируются;

- печатная информация предоставляется крупным шрифтом (от 18 пунктов), тотально озвучивается;

- обеспечивается необходимый уровень освещенности помещений;

- предоставляется возможность использовать компьютеры во время занятий и право записи объяснений на диктофон (по желанию обучающихся).

Форма проведения текущей и промежуточной аттестации для обучающихся с ОВЗ и инвалидов определяется педагогическим работником в соответствии с учебным планом. При необходимости обучающемуся с ОВЗ, инвалиду с учетом их индивидуальных психофизических особенностей дается возможность пройти промежуточную аттестацию устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п., либо предоставляется дополнительное время для подготовки ответа.

9. Оценочные материалы

Задания по контрольно-оценочным мероприятиям в рамках текущей и промежуточной аттестации должны обеспечивать освоение и достижение результатов обучения (индикаторов) и предметного содержания дисциплины на соответствующем уровне.

9.1 Контрольная работа

Примерная тематика контрольных работ***:***
Проектирование пайплайна для задач обработки естественного языка.

Примерные задания в составе контрольных работ***:***
Спроектировать последовательность действий для решения задачи анализа текста с помощью машинного обучения. Пайплайн должен включать:

- Метод подготовки текста для обработки.
- Подход к токенизации текста.
- Подход к векторизации текста.
- Используемую модель машинного обучения.
- Метод обучения модели.
- Метод оценки качества модели.
- Использование обученной модели для решения задачи анализа текста.
- Другие шаги, которые могут понадобиться при решении задачи.

Примеры задач обработки естественного языка, для которых нужно составлять пайплайны:

- Классификация текста.
- Определение эмоциональной окраски текста.
- Автоматическая генерация текста.
- Поиск именованных сущностей в тексте.

9.2 Домашняя работа

Примерная тематика домашних работ***:***
Домашняя работа №1:
Обучение языковой модели для текстов на русском языке.

Домашняя работа №2:
Дообучение предварительно обученной сети BERT

Примерные задания в составе домашних работ***:***

1. Обучите языковую модель для русского языка и используйте ее для генерации текста. Для этого:
    - Подготовьте набор данных с текстами на русском языке. Можно использовать готовые наборы данных или создать собственный.
    - Обучите языковую модель на подготовленном наборе данных.
    - Используя обученную языковую модель сгенерируйте пять примеров текстов на русском языке.
    - Выложите набор данных, код и обученную модель в открытый доступ на GitHub.
    - Оформите презентацию или технологическую статью о ходе работы, обосновании принятых решений и результатах работы.
    - (Не обязательное задание). Запишите видео с демонстрацией работы созданного решения.

2. Дообучите предварительно обученную сеть с архитектурой Transformer для классификации текстов на русском языке. Для этого:
    - Подготовьте набор данных с текстами на русском языке для классификации. Можно использовать готовые наборы данных или создать собственный.
    - Выберите предварительно обученную нейронную сеть с архитектурой Transformer, подходящую для задачи классификации текстов на русском языке.
    - Выполните дообучение выбранной нейронной сети на подготовленном наборе данных.
    - Проведите тестирование классификации текстов с помощью дообученной нейронной сети и оцените качество работы сети.
    - Выложите набор данных, код и дообученную модель в открытый доступ на GitHub.
    - Оформите презентацию или технологическую статью о ходе работы, обосновании принятых решений и результатах работы.
    - (Не обязательное задание). Запишите видео с демонстрацией работы созданного решения.

    Пример дообучения нейронной сети BERT в TensorFlow – https://www.tensorflow.org/text/tutorials/fine_tune_bert
    Ноутбук с примером кода решения – https://colab.research.google.com/github/tensorflow/text/blob/master/docs/tutorials/fine_tune_bert.ipynb
    Пример дообучения нейронных сетей с архитектурой Transformer в Hugging Face – https://huggingface.co/transformers/training.html

9.3 Зачет(устные /письменные ответы на вопросы)

Список примерных вопросов для зачета:

1. Теоретические аспекты обработки естественного языка.
2. Особенности обработки текста на английском языке.
3. Особенности обработки текста на русском языке.
4. Предварительная обработка текста. Очистка текста. Удаление стоп-слов/наиболее и наименее частотных слов.
5. Токенизация, стемминг, лемматизация текста.
6. Методы векторизации текста: построение словаря, мешок слов.
7. Методы векторизации текста: TF-IDF.
8. Методы векторизации текста: word2vec.
9. Методы векторизации текста: fasttext
10. Методы векторизации текста: GloVe.
11. Классические методы машинного обучения для решения задач классификации текста.
12. Классические методы машинного обучения для решения определения тональности текста.
13. Архитектуры нейронных сетей для обработки текстов: LSTM.
14. Архитектуры нейронных сетей для обработки текстов: GRU.
15. Архитектуры нейронных сетей для обработки текста: одномерные сверточные сети.
16. Классификация текста с помощью нейронных сетей.
17. Определение тональности текста с помощью нейронных сетей.
18. Языковая модель.
19. Обучение языковой модели.
20. Основные подходы к генерации текста.
21. Задача поиска именованных сущностей в тексте.
22. Применение нейронных сетей для поиска именованных сущностей.
23. Механизм внимания в нейронных сетях.
24. Применение механизма внимания для обработки текста.
25. Архитектура нейронных сетей Transformer.
26. Предварительно обученные нейронные сети для обработки текстов BERT.
27. Предварительно обученные нейронные сети для обработки текстов GPT.
28. Перенос обучения для задач обработки текстов.
29. Классификация текста с помощью сетей с архитектурой Transformer.
30. Генерация текста с помощью сетей с архитектурой Transformer.
31. Поиск именованных сущностей в тексте с помощью сетей с архитектурой Transformer.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
К Г Э У **«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»**
(ФГБОУ ВО «КГЭУ»)

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
для проведения текущего контроля успеваемости и промежуточной аттестации студентов по итогам освоения дисциплины

Обработка естественного языка

Направление подготовки 09.04.01 Информатика и вычислительная техника

Направленность (профиль) Инженерия искусственного интеллекта

Квалификация Магистр

Форма обучения Очная

Составлено авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

1. Цель и задачи текущего контроля и промежуточной аттестации студентов по дисциплине «Обработка естественного языка»

Цель текущего контроля - систематическая проверка степени освоения программы дисциплины «Обработка естественного языка», уровня сформированности знаний, умений, навыков, компетенций на текущих занятиях

Задачи текущего контроля:

1. определение индивидуального учебного рейтинга студентов;
2. своевременное выполнение корректирующих действий по содержанию и организации процесса обучения; обнаружение и устранение пробелов в усвоении учебной дисциплины;
3. подготовки к промежуточной аттестации.

В течение семестра при изучении дисциплины реализуется комплексная система поэтапного оценивания уровня освоения – балльно-рейтинговая система. За каждый вид учебных действий студенты получают определенное количество баллов. В течение семестра студент может набрать до 60-ти баллов.

Цель промежуточной аттестации - проверка степени усвоения студентами учебного материала за время изучения дисциплины, уровня сформированности компетенций после завершения изучения дисциплины. Аттестация проходит в форме зачета.

Задачи промежуточной аттестации:

1. определение уровня усвоения учебной дисциплины;
2. определение уровня сформированности компетенций.

2. Основное содержание текущего контроля и промежуточной аттестации студентов

В результате изучения дисциплины «Обработка естественного языка» формируются следующие компетенции или их составляющие:

2.1. Основное содержание текущего контроля

Коды компетенций	Совокупность ожидаемых результатов образования студентов в форме компетенций по завершении модуля / освоения дисциплины	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3
ПК-7	ПК-7.1 Руководит проектами в области сквозной цифровой субтехнологии «Компьютерное зрение» ПК-7.2 Руководит проектами в области сквозной цифровой субтехнологии «Обработка естественного языка» ПК-7.3 Исследует и анализирует развитие новых направлений и перспективных методов и технологий в области искусственного интеллекта, участвует в исследовательских проектах по развитию перспективных направлений в области искусственного интеллекта (алгоритмическая имитация биологических систем принятия решений, автономное самообучение и развитие адаптивности	Контрольная работа; домашняя работа; практическая работа; зачёт

	алгоритмов к новым задачам, автономная декомпозиция сложных задач, поиск и синтез решений)	
ОПК-9	ОПК-9.1 Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта ОПК-9.2 Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	Контрольная работа; домашняя работа; практическая работа; зачёт

2.2 Основное содержание промежуточной аттестации студентов

Код и наименование компетенций, формируемые с участием дисциплины	**Индикаторы достижения компетенции**	**Планируемые результаты обучения**	**Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине**
1	**2**	**3**	**4**
ПК-7	ПК-7.1 Руководит проектами в области сквозной цифровой субтехнологии «Компьютерное зрение» ПК-7.2 Руководит проектами в области сквозной цифровой субтехнологии «Обработка естественного языка» ПК-7.3 Исследует и анализирует развитие новых направлений и перспективных методов и технологий в области искусственного интеллекта, участвует в исследовательских проектах по развитию перспективных направлений в области искусственного интеллекта (алгоритмическая имитация биологических систем принятия решений, автономное	Знать: принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение» Уметь: руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение» Знать: принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» Уметь: руководить проектами по созданию, внедрению и	Контрольная работа; домашняя работа; практическая работа; зачёт

	самообучение и развитие адаптивности алгоритмов к новым задачам, автономная декомпозиция сложных задач, поиск и синтез решений)	поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» Знать: современное состояние и перспективы развития новых направлений, методов и технологий в области искусственного интеллекта Уметь: проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения	
ОПК-9	ОПК-9.1 Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта ОПК-9.2 Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	Знать: инструментальные среды, программно-технические платформы для решения профессиональных задач Уметь: применять инструментальные среды, программно-технические платформы для решения профессиональных задач Знать: принципы разработки оригинальных программных средств для решения профессиональных задач Уметь: разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	Контрольная работа; домашняя работа; практическая работа; зачёт

3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины

1.Лекции: **коэффициент значимости совокупных результатов лекционных занятий – 0.5**		
Текущая аттестация на лекциях	**Сроки – семестр, учебная неделя**	**Максимальная оценка в баллах**
Контрольная работа	3 сем., 13 нед.	80
Самостоятельное изучение материала	3 сем., 1-15 нед.	20
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – *Зачет* **Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5**		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5		
Текущая аттестация на практических/семинарских занятиях	**Сроки – семестр, учебная неделя**	**Максимальная оценка в баллах**
Выполнение и оформление практических работ	3 сем., 15 нед.	50
Домашняя работа №1	3 сем., 10 нед.	25
Домашняя работа №2	3 сем., 14 нед.	25
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1		
Промежуточная аттестация по практическим/семинарским занятиям– не предусмотрена **Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – 0**		
3. Лабораторные занятия: Не предусмотрены **коэффициент значимости совокупных результатов лабораторных занятий – 0**		

4. Практические занятия

№ п/п	Примерный перечень тем практических работ
1	Предварительная обработка текста для анализа.
2	Векторизация текста.
3	Классификация текста с использованием классических методов машинного обучения.
4	Классификация текста с использованием глубоких нейронных сетей.
5	Языковая модель. Обучение языковой модели.
6	Автоматическая генерация текста.
7	Поиск именованных сущностей в тексте.
8	Механизм внимания в нейронных сетях. Сети с архитектурой Transformer.
9	Перенос обучения в задачах обработки текстов.

Примерная тематика контрольных работ***:***
Проектирование пайплайна для задач обработки естественного языка.

Примерные задания в составе контрольных работ***:***

Спроектировать последовательность действий для решения задачи анализа текста с помощью машинного обучения. Пайплайн должен включать:

1. Метод подготовки текста для обработки.
2. Подход к токенизации текста.
3. Подход к векторизации текста.
4. Используемую модель машинного обучения.
5. Метод обучения модели.
6. Метод оценки качества модели.
7. Использование обученной модели для решения задачи анализа текста.
8. Другие шаги, которые могут понадобится при решении задачи.

Примеры задач обработки естественного языка, для которых нужно составлять пайплайны:

- Классификация текста.
- Определение эмоциональной окраски текста.
- Автоматическая генерация текста.
- Поиск именованных сущностей в тексте.

Примерная тематика домашних работ***:***
Домашняя работа №1:
Обучение языковой модели для текстов на русском языке.

Домашняя работа №2:
Дообучение предварительно обученной сети BERT

Примерные задания в составе домашних работ***:***

1. Обучите языковую модель для русского языка и используйте ее для генерации текста. Для этого:
    - Подготовьте набор данных с текстами на русском языке. Можно использовать готовые наборы данных или создать собственный.
    - Обучите языковую модель на подготовленном наборе данных.
    - Используя обученную языковую модель сгенерируйте пять примеров текстов на русском языке.
    - Выложите набор данных, код и обученную модель в открытый доступ на GitHub.
    - Оформите презентацию или технологическую статью о ходе работы, обосновании принятых решений и результатах работы.
    - (Не обязательное задание). Запишите видео с демонстрацией работы созданного решения.

2. Дообучите предварительно обученную сеть с архитектурой Transformer для классификации текстов на русском языке. Для этого:
    - Подготовьте набор данных с текстами на русском языке для классификации. Можно использовать готовые наборы данных или создать собственный.
    - Выберите предварительно обученную нейронную сеть с архитектурой Transformer, подходящую для задачи классификации текстов на русском языке.
    - Выполните дообучение выбранной нейронной сети на подготовленном наборе данных.
    - Проведите тестирование классификации текстов с помощью дообученной нейронной сети и оцените качество работы сети.
    - Выложите набор данных, код и дообученную модель в открытый доступ на GitHub.
    - Оформите презентацию или технологическую статью о ходе работы, обосновании принятых решений и результатах работы.
    - (Не обязательное задание). Запишите видео с демонстрацией работы созданного решения.

    Пример дообучения нейронной сети BERT в TensorFlow – https://www.tensorflow.org/text/tutorials/fine_tune_bert
    Ноутбук с примером кода решения – https://colab.research.google.com/github/tensorflow/text/blob/master/docs/tutorials/fine_tune_bert.ipynb
    Пример дообучения нейронных сетей с архитектурой Transformer в Hugging Face – https://huggingface.co/transformers/training.html

Список примерных вопросов для зачета:

1. Теоретические аспекты обработки естественного языка.
2. Особенности обработки текста на английском языке.
3. Особенности обработки текста на русском языке.
4. Предварительная обработка текста. Очистка текста. Удаление стоп-слов/наиболее и наименее частотных слов.
5. Токенизация, стемминг, лемматизация текста.
6. Методы векторизации текста: построение словаря, мешок слов.

7. Методы векторизации текста: TF-IDF.
8. Методы векторизации текста: word2vec.
9. Методы векторизации текста: fasttext
10. Методы векторизации текста: GloVe.
11. Классические методы машинного обучения для решения задач классификации текста.
12. Классические методы машинного обучения для решения определения тональности текста.
13. Архитектуры нейронных сетей для обработки текстов: LSTM.
14. Архитектуры нейронных сетей для обработки текстов: GRU.
15. Архитектуры нейронных сетей для обработки текста: одномерные сверточные сети.
16. Классификация текста с помощью нейронных сетей.
17. Определение тональности текста с помощью нейронных сетей.
18. Языковая модель.
19. Обучение языковой модели.
20. Основные подходы к генерации текста.
21. Задача поиска именованных сущностей в тексте.
22. Применение нейронных сетей для поиска именованных сущностей.
23. Механизм внимания в нейронных сетях.
24. Применение механизма внимания для обработки текста.
25. Архитектура нейронных сетей Transformer.
26. Предварительно обученные нейронные сети для обработки текстов BERT.
27. Предварительно обученные нейронные сети для обработки текстов GPT.
28. Перенос обучения для задач обработки текстов.
29. Классификация текста с помощью сетей с архитектурой Transformer.
30. Генерация текста с помощью сетей с архитектурой Transformer.
31. Поиск именованных сущностей в тексте с помощью сетей с архитектурой Transformer.