ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«07» марта 2024 г.
Протокол №1

Рабочая программа дисциплины (модуля)
«Natural Language Processing (Обработка естественного языка)»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Математика и искусственный интеллект*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2024*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2024

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 5
4. **Содержание дисциплины (модуля)** 7
5. **Учебно-методическое обеспечение** 8
6. **Материально-техническое обеспечение** 8
7. **Методические и оценочные материалы** 10

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Natural Language Processing (Обработка естественного языка)» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 Математика и компьютерные науки, профиль Математика и искусственный интеллект, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Natural Language Processing (Обработка естественного языка)» играет ключевую роль в создании систем, которые могут понимать и генерировать человеческий язык, что открывает новые возможности для взаимодействия между людьми и машинами. Эта область обеспечивает развитие технологий, таких как чат-боты, автоматический перевод и анализ настроений, которые становятся все более важными в различных сферах, включая бизнес, медицину и образование.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 Математика и компьютерные науки, профиль Математика и искусственный интеллект и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) изучается на 4 курсе в 7 семестре.

Цель изучения дисциплины (модуля): приобретение знаний и навыков, необходимых для разработки и применения технологий, позволяющих компьютерам анализировать, понимать и взаимодействовать с человеческим языком в различных контекстах.

Задачи изучения дисциплины (модуля):

— формирование знаний и понимания по темам: основы предобработки текста и выделения признаков, принципы языкового моделирования и ключевые архитектуры нейронных сетей, применяемых в области NLP (RNN, Трансформер), подходы к машинному переводу, генерации текста и суммаризации, современные методы обучения и оптимизации языковых моделей (LLM), основы работы диалоговых систем и извлечения информации, применение мультимодальных моделей и технологий обработки речи;

— освоение умений: проводить предобработку и векторизацию текста, применять нейронные сети и LLM для задач NLP, обучать и оптимизировать языковые модели, разрабатывать и интегрировать NLP-приложения, оценивать и улучшать качество моделей;

— формирование навыков владения инструментами Python для NLP (HuggingFace, nltk, pytorch), методами обучения и оптимизации языковых моделей, навыками работы с мультимодальными системами и Text-to-speech.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— основные способы предобработки и векторизации текста;

— методы решения ключевых задач обработки естественного языка: классификация текстов, извлечение информации, тегирование последовательностей, генерация текста, суммаризация;

— принципы построения и методы дообучения предобученных языковых моделей (BERT, GPT и др.) для задач NLP;

— базовые концепции оценки качества и интерпретации работы языковых моделей в различных задачах обработки естественного языка;

— подходы к интеграции внешних знаний в LLM с использованием RAG (Retrieval Augmented Generation);

— методы построения агентных систем с применением LLM;

— принципы работы мультимодальных языковых моделей, способных обрабатывать и интегрировать различные модальности, такие как текст и звук;

уметь:

— конструировать и реализовывать пайплайны обработки текста для решения типовых и прикладных задач NLP;

— применять и дообучать предобученные языковые модели (например, BERT, GPT и др.) на собственные датасеты для решения задач обработки естественного языка;

— оптимизировать производительность и ускорять работу NLP-систем, используя методы сокращения размера моделей, квантизацию, дистилляцию и внедрение эффективных вычислительных решений;

— осуществлять комплексную оценку качества разработанных моделей с использованием метрик (accuracy, F1-score, BLEU, ROUGE, LLM-as-a-judge и др.), а также применять методы повышения качества (тюннинг гиперпараметров, аугментация данных и др.);

владеть:

— основными инструментами и библиотеками для NLP (NLTK, spaCy, HuggingFace Transformers,HuggingFace Datasets, vllm, PEFT, SentenceTransformers и др.);

— приемами работы с предобученными языковыми моделями, включая их применение, дообучение и оптимизацию;

— умением самостоятельно разрабатывать, оптимизировать и оценивать NLP-решения для прикладных задач, а также интерпретировать их результаты.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и	ОПК-1.1.	Знает основные концепции и теории в области математического анализа и смежных дисциплин; методы и подходы, используемые в различных областях математики
		ОПК-1.2.	Умеет применять математические методы для решения профессиональных задач
		ОПК-1.3.	Имеет практический опыт разработки и реализации математических моделей в профессиональной деятельности

	математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности		
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов
		ПК-1.2.	Умеет корректно ставить и формулировать математические задачи, применять строгие методы доказательства и анализировать полученные результаты
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач
ПК-2.	Способен решать типовые задачи профессиональной деятельности в области искусственного интеллекта, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области разработки, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Контактная работа		Контроль	Самостоятельная работа	
		Лекции	Семинары			
1	Encoder-based модели	10	10		42	Домашние задания, Подготовка к семинару
2	Большие языковые модели (LLM)	10	10		42	Подготовка к семинару
3	Современные возможности и расширения больших языковых моделей (LLM)	10	10		42	Домашние Подготовка к семинару
	Зачет с оценкой			4		
	Итого:	***30***	***30***	***4***	***164***	
Объем дисциплины (модуля) (в ак. ч.)		***228***				
Объем дисциплины (модуля) (в зач. ед.)		***6***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Encoder-based модели	Векторизация и классификация текстов. Предобученные языковые модели на основе архитектуры Кодировщик Трансформера (BERT и др.). Ранжирование и методы информационного поиска. Мультилингвальность и доменная адаптация предобученных языковых моделей
2	Большие языковые модели (LLM)	Языковое моделирование и обучение больших языковых моделей (LLM). Дообучение и fine-tuning LLM. Efficient Transformers, PEFT и оценка эффективности LLM. Сжатие и оптимизация LLM. Mixture-of-Experts и параллельные вычисления
3	Современные возможности и расширения больших языковых моделей (LLM)	Retrieval Augmented Generation (RAG). LLM-агенты. Мультимодальные языковые модели. Кодовые LLM. Альтернативные архитектуры LLM. Обработка речи и аудиосигналов

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Хобсон Л. Обработка естественного языка в действии : практическое руководство / Л. Хобсон, Х. Ханнес, Х. Коул. - Санкт-Петербург : Питер, 2020. - 576 с. - (Серия «Для профессионалов»). - ISBN 978-5-4461-1371-2.

2. Риз, Р. Обработка естественного языка на Java : практическое руководство / Р. Риз ; пер. с англ. А. В. Снастина. - 2-е изд. - Москва : ДМК Пресс, 2023. - 266 с. - ISBN 978-5-89818-333-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2102620.

3. Хобсон, Л. Обработка естественного языка в действии : практическое руководство / Л. Хобсон, Х. Ханнес, Х. Коул. - Санкт-Петербург : Питер, 2020. - 576 с. - (Серия «Для профессионалов»). - ISBN 978-5-4461-1371-2. - Текст : электронный. - URL: https://znanium.com/catalog/product/1733506.

4. Васильев, Ю. Обработка естественного языка. Python и spaCy на практике / Ю. Васильев. - Санкт-Петербург : Питер, 2021. - 256 с. - ISBN 978-5-4461-1506-8. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2141633.

5. Бенгфорт, Б. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка : практическое руководство / Б. Бенгфорт, Р. Билбро, Т. Охеда. - Санкт-Петербург : Питер, 2020. - 368 с. - (Серия «Бестселлеры O'Reilly»). - ISBN 978-5-4461-1153-4. - Текст : электронный. - URL: https://znanium.com/catalog/product/1733693.

Дополнительная литература:

1. Бенгфорт, Б. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка : практическое руководство / Б. Бенгфорт, Р. Билбро, Т. Охеда. - Санкт-Петербург : Питер, 2020. - 368 с. - (Серия «Бестселлеры O'Reilly»). - ISBN 978-5-4461-1153-4.

2. Николенко, С. И. Глубокое обучение : практическое руководство / С. И. Николенко, А. Кадурин, Е. Архангельская. - Санкт-Петербург : Питер, 2019. - 480 с. - (Серия «Библиотека программиста»). - ISBN 978-5-496-02536-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/1760785.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также

помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

— столами и стульями;
— компьютерной техникой;
— механическими калькуляторами;
— специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование"	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое

Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Natural Language Processing (Обработка естественного языка)» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары (аудиторная работа), домашние задания, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удается, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Участие в семинаре (практическом занятии) – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным,

опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины.

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Natural Language Processing (Обработка естественного языка)»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме ***зачета с оценкой,*** при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	**Пятибалльная оценка**	**Общая характеристика результата обучения по дисциплине (модулю)**
10	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	
8	Отлично	
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он
6	Хорошо	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Natural Language Processing (Обработка естественного языка)» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	60%	2	Набор задач по темам недели
Аудиторная работа	10%	7	Активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии
Зачет с оценкой	30%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Natural Language Processing (Обработка естественного языка)»: «(0,6 × среднее за домашние задания + 0,1 × среднее за аудиторную работу + 0,3 × зачет с оценкой».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1.

Задание 1.

1. Выберите текстовый корпус (например, набор новостных статей или отзывов о продуктах).
2. Выполните предобработку текста, включая:
   - Удаление пунктуации и специальных символов.
   - Приведение текста к нижнему регистру.

- Стемминг или лемматизацию.
- Удаление стоп-слов.

3. Реализуйте метод "мешок слов" (Bag of Words) для выделения признаков из вашего корпуса.
4. Создайте таблицу, отображающую частоту слов в вашем наборе данных.

Ожидаемый результат: Отчет, содержащий результаты предобработки, таблицу частоты слов и краткий анализ полученных данных.

Задание 2.

1. Используя библиотеку Gensim, обучите модель Word2Vec на выбранном текстовом корпусе (можно использовать корпус из первого задания).
2. Проведите анализ векторных представлений, выбрав несколько слов и найдите их ближайшие "соседи" в векторном пространстве.
3. Объясните, как векторные представления слов помогают в решении задач NLP.
4. Создайте простую языковую модель на основе N-грамм, используя ваш текстовый корпус, и оцените ее производительность на тестовом наборе данных.

Ожидаемый результат: Отчет с результатами обучения модели Word2Vec, списком ближайших соседей для выбранных слов и кратким анализом языковой модели.

Задание 3.

1. Реализуйте простую рекуррентную нейронную сеть (RNN) для задачи классификации текста (например, классификация отзывов как положительных или отрицательных). Используйте фреймворк TensorFlow или PyTorch.
2. Обучите модель на размеченном наборе данных (например, IMDB для отзывов о фильмах).
3. Сравните производительность RNN с более сложной архитектурой, такой как LSTM или GRU, и проанализируйте результаты.
4. Реализуйте механизм внимания для улучшения производительности вашей модели. Опишите, как механизм внимания помогает в контексте вашей задачи.

Ожидаемый результат: Отчет, содержащий результаты обучения RNN и LSTM/GRU, графики производительности и анализ влияния механизма внимания на результаты.

Домашнее задание 2.

Задание 1.

1. Изучите архитектуру Трансформер, включая механизмы внимания и позиционное кодирование. Создайте визуальную схему, иллюстрирующую основные компоненты модели.
2. Реализуйте простую языковую модель на основе кодировщика Трансформера (например, BERT) с использованием библиотеки Hugging Face Transformers. Обучите модель на задаче классификации текстов (например, классификация отзывов).
3. Проведите оценку производительности модели на тестовом наборе данных и проанализируйте, какие факторы могут влиять на результаты.

Ожидаемый результат: Отчет с визуальной схемой архитектуры, кодом реализации модели, результатами оценки производительности и анализом факторов.

Задание 2.

1. Выберите генеративную языковую модель на основе декодера Трансформера (например, GPT-2 или GPT-3). Изучите, как работает prompt tuning и реализуйте его для вашей модели.

2. Проведите эксперимент, сравнив производительность модели с и без prompt tuning на задаче генерации текста (например, продолжение предложений).

3. Исследуйте методы оптимизации, такие как P-tuning и LoRA. Реализуйте один из этих методов для улучшения производительности вашей модели и оцените результаты.

Ожидаемый результат: Отчет, включающий результаты экспериментов с prompt tuning, сравнение производительности, реализацию метода оптимизации и анализ полученных данных.

Задание 3.

1. Реализуйте RAG (Retrieval-Augmented Generation) систему, используя предобученные модели и набор данных для извлечения информации (например, Wikipedia). Объясните, как система использует механизмы извлечения и генерации.

2. Создайте простую диалоговую систему, реализовав задачи intent detection и slot filling. Используйте библиотеку Rasa или аналогичную для реализации.

3. Проведите тестирование вашей диалоговой системы на наборе пользовательских запросов и проанализируйте результаты, включая точность определения намерений и заполнения слотов.

Ожидаемый результат: Отчет, содержащий описание RAG системы, код реализации диалоговой системы, результаты тестирования и анализ производительности.

Примерные вопросы для подготовки к семинарам

Encoder-based модели

1. Что такое векторизация текста и какие основные методы используются для преобразования текста в векторы в контексте encoder-based моделей?
2. Как работает процесс классификации текстов с использованием BERT-моделей? Опишите ключевые шаги.
3. Какие преимущества имеет архитектура кодировщика трансформера по сравнению с традиционными методами обработки текста?
4. Назовите основные компоненты модели BERT и их роли в обработке текста.
5. Как происходит предобучение BERT-модели и какие задачи используются для этого?
6. Что такое маскированное языковое моделирование (MLM) в BERT и почему оно важно?
7. Как применяется BERT для задач ранжирования в информационном поиске?
8. Опишите механизм attention в трансформерах и его роль в encoder-based моделях.
9. Какие метрики используются для оценки качества ранжирования в методах информационного поиска?
10. Что такое мультилингвальность в предобученных моделях и как она реализуется в моделях вроде mBERT?
11. Как происходит доменная адаптация предобученных языковых моделей для специфических задач?
12. Приведите примеры задач, где encoder-based модели превосходят другие подходы в NLP.
13. Какие ограничения имеют encoder-based модели при работе с длинными текстами?

14. Как интегрировать encoder-based модели в пайплайн обработки данных для классификации текстов?

15. Обсудите этические аспекты использования предобученных моделей, таких как BERT, в реальных приложениях.

Большие языковые модели (LLM)

1. Что такое языковое моделирование и как оно применяется в обучении больших языковых моделей (LLM)?
2. Опишите процесс обучения LLM с использованием трансформерной архитектуры.
3. Какие данные используются для предобучения LLM и почему они важны?
4. Что такое дообучение (fine-tuning) LLM и в чем его отличие от предобучения?
5. Как работает метод PEFT (Parameter-Efficient Fine-Tuning) и какие преимущества он дает?
6. Назовите ключевые метрики для оценки эффективности LLM в задачах генерации текста.
7. Что такое Efficient Transformers и как они оптимизируют вычисления в LLM?
8. Опишите методы сжатия LLM, такие как квантизация и pruning, и их влияние на производительность.
9. Как работает Mixture-of-Experts (MoE) в архитектуре LLM и почему это полезно для масштабирования?
10. Какие параллельные вычисления используются при обучении больших LLM, например, data parallelism или model parallelism?
11. Как оценивать эффективность LLM с точки зрения вычислительных ресурсов и качества генерации?
12. Приведите примеры задач, где fine-tuning LLM дает значительные улучшения.
13. Какие вызовы возникают при оптимизации LLM для развертывания на мобильных устройствах?
14. Обсудите роль параллельных вычислений в снижении времени обучения LLM.
15. Как балансировать между размером модели и ее эффективностью в практических приложениях LLM?

Современные возможности и расширения больших языковых моделей (LLM)

1. Что такое Retrieval Augmented Generation (RAG) и как оно улучшает генерацию текста в LLM?
2. Опишите процесс работы RAG: как происходит извлечение информации и ее интеграция в ответы модели?
3. Какие преимущества имеет RAG по сравнению с чистой генерацией текста в LLM?
4. Что такое LLM-агенты и как они используются для автономного выполнения задач?
5. Приведите примеры приложений LLM-агентов в области автоматизации и принятия решений.
6. Как работают мультимодальные языковые модели, такие как GPT-4V, и какие модальности они поддерживают?
7. Назовите ключевые вызовы в интеграции текста, изображений и других модальностей в мультимодальных LLM.
8. Что такое кодовые LLM и как они применяются для генерации и анализа программного кода?
9. Опишите процесс fine-tuning LLM для задач программирования, например, в моделях вроде CodeLlama.

10. Какие альтернативные архитектуры LLM существуют помимо трансформеров, например, на основе RNN или state-space моделей?

11. Как обрабатываются речевые сигналы в LLM и какие модели используются для распознавания речи?

12. Что такое text-to-speech (TTS) и speech-to-text (STT) в контексте LLM, и как они интегрируются?

13. Обсудите роль LLM в обработке аудиосигналов, включая шумоподавление и генерацию музыки.

14. Какие этические вопросы возникают при использовании мультимодальных и кодовых LLM в реальных приложениях?

15. Как оценивать эффективность RAG и LLM-агентов с точки зрения точности, скорости и надежности?

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Назовите основные компоненты в архитектуре модели BERT.	энкодера	УК-1
2	Назовите метод предобучения, используемый в BERT для маскированного языкового моделирования.	MLM	УК-1
3	Укажите задачи, используемые для предобучения BERT.	задачи	УК-1
4	Определите основное преимущество архитектуры кодировщика трансформера по сравнению с RNN.	параллелизм	УК-1
5	Назовите механизм, отвечающий за внимание в трансформерах.	attention	УК-2
6	Укажите основные метрики для оценки качества ранжирования в информационном поиске.	метрики	УК-2
7	Определите способ доменной адаптации предобученных моделей для специфических задач.	fine-tuning	УК-2
8	Назовите ограничение encoder-based моделей при работе с длинными текстами.	длина	УК-2
9	Укажите основные шаги в процессе классификации текстов с использованием BERT.	шага	ОПК-1
10	Назовите метод векторизации текста, используемый в encoder-based моделях.	embedding	ОПК-1
11	Определите количество параметров в больших языковых моделях, где применяется Mixture-of-Experts.	миллиарды	ОПК-1
12	Укажите тип параллельных вычислений, используемых при обучении LLM.	data	ОПК-1
13	Назовите метод оптимизации LLM для снижения размера модели.	квантизация	ПК-1
14	Определите основные задачи, где fine-tuning LLM дает улучшения.	генерации	ПК-1

15	Укажите механизм работы Efficient Transformers для оптимизации вычислений.	sparsity	ПК-1
16	Назовите преимущество метода PEFT в дообучении LLM.	эффективность	ПК-1
17	Определите модальности, поддерживаемые мультимодальными LLM.	текста	ПК-2
18	Укажите тип генерации, используемой в Retrieval Augmented Generation.	augmented	ПК-2
19	Назовите применение кодовых LLM в программировании.	генерация	ПК-2
20	Определите альтернативные архитектуры LLM помимо трансформеров.	RNN	ПК-2