МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Рабочая программа дисциплины (модуля)

ПРИКЛАДНЫЕ БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

Рабочая программа дисциплины (модуля) «Прикладные большие языковые модели» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика (утв. приказом Минобрнауки России от 19.09.2017г. №916) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Кригер А.Б.

Утверждена на заседании научно-образовательный центр "искусственный интеллект" от 27.05.2026 , протокол № 5

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кригер А.Б.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1582918206
Номер транзакции	0000000000F6CABD
Владелец	Кригер А.Б.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Целью освоения дисциплины является ознакомление слушателей с методами обработки текста на естественном языке, а также методами обработки слабоструктурированных данных и извлечения информации.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине	
			Код результата	Формулировка результата
09.04.03 «Прикладная информатика» (М-ПИ)				

В процессе освоения дисциплины решаются задачи воспитания гармонично развитой, патриотичной и социально ответственной личности на основе традиционных российских духовно-нравственных и культурно-исторических ценностей, представленные в таблице 1.2.

Таблица 1.2 – Целевые ориентиры воспитания

Воспитательные задачи	**Формирование ценностей**	**Целевые ориентиры**
Формирование научного мировоззрения и культуры мышления		
Формирование осознания ценности научного мировоззрения и критического мышления	Гуманизм	Системное мышление
Формирование коммуникативных навыков и культуры общения		
Формирование навыков публичного выступления и презентации своих идей	Взаимопомощь и взаимоуважение	Умение работать в команде и взаимопомощь

2 Место дисциплины (модуля) в структуре ОПОП

дисциплина относится к обязательной части Блока 1, Дисциплины (модули)

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обучения	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудоемкость	Объем контактной работы (час)						СРС	Форма аттестации
				(З.Е.)	Всего	Аудиторная			Внеаудиторная			
						лек.	прак.	лаб.	ПА	КСР		
09.04.03 Прикладная информатика	ОФО	М01.Б	4	4	33	8	24	0	1	0	111	Э

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы	Код результата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Основы языковых моделей	РД1	2	6	0	26	практическое задание
2	Большие языковые модели (LLM)	РД1, РД2, РД4, РД6	2	6	0	30	практическое задание
3	Применение языковых моделей	РД2, РД3, РД5, РД6	2	6	0	30	практическое задание
4	Будущее языковых моделей	РД4, РД6	2	6	0	25	практическое задание
Итого по таблице			**8**	**24**	**0**	**111**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Основы языковых моделей.

Содержание темы: Что такое языковые модели? • Определение языковой модели. • Основные принципы работы языковых моделей. • Различия между статистическими и нейронными языковыми моделями. Структура языковых моделей • Архитектура трансформеров. • Энкодеры и декодеры. • Механизм внимания. .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекции, практические занятия.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

Тема 2 Большие языковые модели (LLM).

Содержание темы: Особенности больших языковых моделей (GPT, BERT), процессы их обучения. Сравнение разных моделей, подготовка данных для обучения. Предварительное обучение, тонкая настройка, аугментация данных. .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекции, практические занятия.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

Тема 3 Применение языковых моделей.

Содержание темы: Применение в науке и медицине. Анализ медицинских текстов, поиск и классификация научных статей. Применение в маркетинге и IT. .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекции, практические занятия.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

Тема 4 Будущее языковых моделей.

Содержание темы: Перспективы развития, новые архитектуры и задачи. Исследование текущих трендов, прогнозирование будущего. Навыки и знания для работы с моделями, образовательные программы. .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекции, практические занятия.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков

и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Алетдинова, А. А. Интеллектуальный анализ больших данных : учебное пособие / А. А. Алетдинова, М. Ш. Муртазина. - Новосибирск : Изд-во НГТУ, 2023. - 66 с. - ISBN 978-5-7782-4899-1. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2246179 (дата обращения: 31.05.2026)

2. Бессмертный, И. А. Искусственный интеллект. Введение в многоагентные системы : учебник для вузов / И. А. Бессмертный. — Москва : Издательство Юрайт, 2026. — 148 с. — (Высшее образование). — ISBN 978-5-534-20348-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/589921 (дата обращения: 19.05.2026).

3. Принятие решений на основе анализа слабоструктурированных данных социальных сетей : монография / Н. Г. Ярушкина, И. А. Андреев, А. С. Желепов, В. С. Мошкин , научный редактор Н. Г. Ярушкина. — Ульяновск : УлГТУ, 2022. — 113 с. — ISBN 978-5-9795-2275-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/322841 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

7.2 Дополнительная литература

1. Загорулько, Ю. А. Искусственный интеллект. Инженерия знаний : учебное пособие для вузов / Ю. А. Загорулько, Г. Б. Загорулько. — Москва : Издательство Юрайт, 2024. — 93 с. — (Высшее образование). — ISBN 978-5-534-07198-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/540987 (дата обращения: 19.05.2026).

2. Терлецкий, А. С. Нейронные сети и искусственный интеллект: Основы нейронных сетей на языке Python : учебно-методическое пособие / А. С. Терлецкий, Е. С. Терлецкая. — Липецк : Липецкий ГПУ, 2023. — 76 с. — ISBN 978-5-907792-40-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/439343 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Образовательная платформа "ЮРАЙТ"
2. Электронно-библиотечная система "ZNANIUM.COM"
3. Электронно-библиотечная система "ЛАНЬ"
4. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/
5. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/
6. Информационно-справочная система "Консультант Плюс" - Режим доступа: http://www.consultant.ru/

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Коммутатор SuperStack 3 (16*10/100 19")
- Монитор облачный 23" LG23CAV42K/мышь Geniu
- Мультимедийный проектор №1 Casio XJ-V2
- Облачный монитор 23" LG CAV42K
- Облачный монитор LG Electronics черный +клавиатура+мышь
- П/К DNS Office T300, мышь Genius NetScroll 100, клавиатура Genius KB-06X, монитор AOC919 19"
- Проектор Casio XJ-V1
- Уст-во бесп.питания UPS-3000

Программное обеспечение:

- Microsoft Office Professional Plus 2016
- Python

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

ПРИКЛАДНЫЕ БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
09.04.03 «Прикладная информатика» (М-ПИ)		

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения		Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС	
			Текущий контроль	Промежуточная аттестация
Очная форма обучения				
РД1	Знание : базовые принципы построения языковых моделей, классы больших языковых моделей	1.1. Основы языковых моделей	Практическая работа	Тест
		1.2. Большие языковые модели (LLM)	Практическая работа	Тест
РД2	Умение : выбирать языковую модель для решения прикладной задачи	1.2. Большие языковые модели (LLM)	Практическая работа	Тест
		1.3. Применение языковых моделей	Практическая работа	Тест
РД3	Навык : применять большие языковые модели для решения практических бизнес задач	1.3. Применение языковых моделей	Практическая работа	Тест
РД4	Знание : архитектуры больших языковых моделей	1.2. Большие языковые модели (LLM)	Практическая работа	Тест
		1.4. Будущее языковых моделей	Практическая работа	Тест
РД5	Умение : обучать большие языковые модели для решения частных прикладных задач	1.3. Применение языковых моделей	Практическая работа	Тест

РД6	Навык : выбора инструментальных средств, построенных на больших языковых моделях, для решения исследовательских задач	1.2. Большие языковые модели (LLM)	Практическая работа	Тест
		1.3. Применение языковых моделей	Практическая работа	Тест
		1.4. Будущее языковых моделей	Практическая работа	Тест

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Краткие методические указания

При выполнении задания допускается объединение магистрантов в группы по 2-5 человек. Практические работы позволяет определить уровень усвоения материала. Перед выполнением работы необходимо ознакомиться с теоретическим материалом, проработать методы решения задач, рассмотренных в типовых примерах. После выполнения работы предусмотрен письменный аналитический отчет и публичная защита полученных результатов моделирования, сравнение и обсуждение результатов всех участников

Тест содержит тестовые задания с выбором одного или нескольких правильных ответов, ввод числа.

Шкала оценки

Оценка	Баллы	Описание
5	20	выставляется студенту, если он ответил на более, чем 90 % поставленных вопросов, учитывая множественный выбор ответов теста
4	11-15	выставляется студенту, если он ответил на 76-90%% поставленных вопросов, учитывая множественный выбор ответов теста
3	6-10	выставляется студенту, если он ответил на 61-75%% поставленных вопросов, учитывая множественный выбор ответов теста
2	0–5	выставляется студенту, если он ответил не более, чем на 60% поставленных вопросов, учитывая множественный выбор ответов теста

Шкала оценки

Оценка	Баллы	Описание
5	17–18	выставляется студенту, если его группа качественно собрала, обработала информацию и представила результаты по заданной теме, он лично отвечал на вопросы представителей других групп и преподавателя, четко представляя свою позицию и аргументируя точку зрения
4	14–16	выставляется студенту, если его группа качественно собрала, обработала информацию и представила результаты по заданной теме, но он лично, отвечая на вопросы представителей других групп и преподавателя, недостаточно четко представлял свою позицию и аргументировал точку зрения
3	9–13	баллов выставляется студентам групп, недостаточно качественно собравших, обработпвших информацию или представивших результаты по заданной теме, он лично, отвечая на вопросы представителей других групп и преподавателя, недостаточно четко предста влял свою позицию и аргументировал точку зрения
2	0–8	баллов выставляется студентам групп, недостаточно качественно собравших, обработавших информацию и представивших результаты по заданной теме, или он лично, не участвовал в обсуждении

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

5 Примерные оценочные средства

5.1 Примеры тестовых заданий

1. Что является основной задачей языковой модели?
A) Генерация случайных текстов
B) Понимание и генерация осмысленного текста
C) Классификация текстов по категориям
D) Распознавание сущностей в тексте

2. Какая архитектура лежит в основе современных больших языковых моделей, таких как BERT или GPT?
A) Рекуррентные нейронные сети (RNN)
B) Трансформеры
C) Нейронные сети прямого распространения
D) Вероятностные графовые модели

3. Какой элемент языковой модели отвечает за преобразование текста в вектор чисел, который максимально точно описывает исходные данные?
A) Декодер
B) Энкодер
C) Механизм внимания
D) Токенизатор

4. Как называется процесс присвоения уникальных идентификаторов каждому слову или знаку во входном тексте?
A) Токенизация
B) Эмбеддинг
C) Преобразование
D) Кодирование

5. Какие параметры используются в процессе эмбеддинга для представления слов в виде векторов?
A) Матрица запроса (Q), матрица ключевых значений (K), матрица значений (V)
B) Векторное представление слов
C) Позиционные эмбеддинги
D) Линейные преобразования

6. Что представляет собой механизм внимания в структуре трансформера?
A) Способность модели фокусироваться на определенных частях текста
B) Алгоритм оптимизации для уменьшения потери информации
C) Метод для учета контекста при переводе текста
D) Процесс преобразования текста в числовой формат

7. Какой этап следует после прямого прохождения данных по сети в процессе обучения языковой модели?
A) Вычисление ошибки
B) Обратное распространение ошибки
C) Корректировка весов
D) Предсказание результата

8. Что подразумевается под предварительным обучением языковой модели?
A) Обучение модели на общих данных перед специализированным дообучением
B) Начальное обучение на небольшом наборе данных
C) Обучение на синтетических данных
D) Параллельное обучение нескольких моделей

9. Каким образом обычно осуществляется подготовка данных для обучения языковой модели?
A) Автоматическая токенизация и эмбеддинг
B) Ручная разметка данных

C) Автоматическая очистка данных

D) Комбинированный метод

10. Какую роль играют позиционные эмбеддинги в структуре языковой модели?

A) Учет позиции слова в предложении

B) Определение семантической близости слов

C) Преобразование текста в числовой формат

D) Улучшение точности модели

Краткие методические указания

Каждый вариант теста содежит десять вопросов на тему идеи и структуры языковых моделей, с выбором одного правильного ответа.

Тест засчитыается если студент ответил правильно более чем на 7 вопросов

Шкала оценки

Шкала оценки

Оценка	Баллы	Описание
5	20	выставляется студенту, если он ответил на более, чем 90 % поставленных вопросов, уч итывая множественный выбор ответов теста
4	11-15	выставляется студенту, если он ответил на 76-90%% поставленных вопросов, учитыва я множественный выбор ответов теста
3	6-10	выставляется студенту, если он ответил на 61-75%% поставленных вопросов, учитыва я множественный выбор ответов теста
2	0–5	выставляется студенту, если он ответил не более, чем на 60% поставленных вопросов, учитывая множественный выбор ответов теста

5.2 Контрольный тест

Перечень контрольных вопросов:

1. Что такое языковая модель?
2. Какие виды языковых моделей существуют?
3. В чем заключается основная задача языковой модели?
4. Что такое трансформеры и какова их роль в современных языковых моделях?
5. Какие особенности отличают большие языковые модели (LLM)?
6. Какие методы обучения языковых моделей вы знаете?
7. В чем заключаются отличия между моделями BERT и GPT?
8. Каким образом осуществляется подготовка данных для обучения языковых моделей?
9. Как производится токенизация и эмбеддинг текста в процессе работы с языковыми моделями?
10. Какие области применения языковых моделей вам известны?
11. Какие перспективы развития языковых моделей вы видите в ближайшем будущем?
12. Какие библиотеки и инструменты используются для работы с языковыми моделями?
13. Какие практические задачи можно решить с помощью языковых моделей в медицине?
14. Какие сложности возникают при обучении и настройке языковых моделей?
15. Какие навыки и знания необходимы для успешной работы с языковыми моделями?

Краткие методические указания

Для подготовки к итоговому тестированию необходимо масостоятельно подготовиться в соответствии с перечнем вопросв

Шкала оценки

Шкала оценки

Оценка	Баллы	Описание
5	20	выставляется студенту, если он ответил на более, чем 90 % поставленных вопросов, уч итывая множественный выбор ответов теста

4	11-15	выставляется студенту, если он ответил на 76-90%% поставленных вопросов, учитыва я множественный выбор ответов теста
3	6-10	выставляется студенту, если он ответил на 61-75%% поставленных вопросов, учитыва я множественный выбор ответов теста
2	0–5	выставляется студенту, если он ответил не более, чем на 60% поставленных вопросов, учитывая множественный выбор ответов теста

5.3 Примеры заданий для выполнения практических работ

Сгенерировать текст доклада на выбранную тему. Выбрать инструмент для работы. Обучить Языковую модель если необходимо, составить промт.

Сгенерировать аналитический экономичесий отчет. Выбрать инструмент для работы. Обучить Языковую модель если необходимо, составить промт.

Сгенерировать иллюстании к реферату. Выбрать инструмент для работы. Обучить Языковую модель если необходимо, составить промт.

Сгенерировать иллюстании научному отчету. Выбрать инструмент для работы. Обучить Языковую модель если необходимо, составить промт.

Создантъ презентацию с помощью нейронной сети. Выбрать инструмент для работы. Обучить Языковую модель если необходимо, составить промт.

Краткие методические указания

Шкала оценки

Шкала оценки

Оценка	Баллы	Описание
5	10	выставляется студенту, если он ответил на более, чем 90 % поставленных вопросов, уч итывая множественный выбор ответов теста
4	8–9	выставляется студенту, если он ответил на 76-90%% поставленных вопросов, учитыва я множественный выбор ответов теста
3	6–7	выставляется студенту, если он ответил на 61-75%% поставленных вопросов, учитыва я множественный выбор ответов теста
2	0–5	выставляется студенту, если он ответил не более, чем на 60% поставленных вопросов, учитывая множественный выбор ответов теста

Ключи для ФОС
по дисциплине «Прикладные большие языковые модели»

Тест по языковым моделям

1. **Что является основной задачей языковой модели?**
   A) Генерация случайных текстов
   B) Понимание и генерация осмысленного текста
   C) Классификация текстов по категориям
   D) Распознавание сущностей в тексте

2. **Какая архитектура лежит в основе современных больших языковых моделей, таких как BERT или GPT?**
   A) Рекуррентные нейронные сети (RNN)
   B) Трансформеры
   C) Нейронные сети прямого распространения
   D) Вероятностные графовые модели

3. **Какой элемент языковой модели отвечает за преобразование текста в вектор чисел, который максимально точно описывает исходные данные?**
   A) Декодер
   B) Энкодер
   C) Механизм внимания
   D) Токенизатор

4. **Как называется процесс присвоения уникальных идентификаторов каждому слову или знаку во входном тексте?**
   A) Токенизация
   B) Эмбеддинг
   C) Преобразование
   D) Кодирование

5. **Какие параметры используются в процессе эмбеддинга для представления слов в виде векторов?**
   A) Матрица запроса (Q), матрица ключевых значений (K), матрица значений (V)
   B) Векторное представление слов
   C) Позиционные эмбеддинги
   D) Линейные преобразования

6. **Что представляет собой механизм внимания в структуре трансформера?**
   A) Способность модели фокусироваться на определенных частях текста
   B) Алгоритм оптимизации для уменьшения потери информации
   C) Метод для учета контекста при переводе текста
   D) Процесс преобразования текста в числовой формат

7. **Какой этап следует после прямого прохождения данных по сети в процессе обучения языковой модели?**
   A) Вычисление ошибки

B) Обратное распространение ошибки
C) Корректировка весов
D) Предсказание результата

8. **Что подразумевается под предварительным обучением языковой модели?**
A) Обучение модели на общих данных перед специализированным дообучением
B) Начальное обучение на небольшом наборе данных
C) Обучение на синтетических данных
D) Параллельное обучение нескольких моделей

9. **Каким образом обычно осуществляется подготовка данных для обучения языковой модели?**
A) Автоматическая токенизация и эмбеддинг
B) Ручная разметка данных
C) Автоматическая очистка данных
D) Комбинированный метод

10. **Какую роль играют позиционные эмбеддинги в структуре языковой модели?**
A) Учет позиции слова в предложении
B) Определение семантической близости слов
C) Преобразование текста в числовой формат
D) Улучшение точности модели

Дополнительные вопросы по языковым моделям и трансформерам

1. Что такое языковая модель в машинном обучении?

A) Алгоритм для распознавания речи
B) Модель, предсказывающая следующее слово в последовательности
C) Система для классификации изображений
D) Метод шифрования текста
Правильный ответ: B) Модель, предсказывающая следующее слово в последовательности

2. Какой тип архитектуры НЕ используется в современных языковых моделях?

A) Трансформеры
B) Рекуррентные нейронные сети (RNN)
C) Сверточные нейронные сети (CNN)
D) Деревья решений
Правильный ответ: D) Деревья решений

3. Какой алгоритм лежит в основе GPT (Generative Pre-trained Transformer)?

A) LSTM
B) Трансформер
C) Random Forest
D) SVM
Правильный ответ: B) Трансформер

4. Что делает механизм self-attention в трансформерах?

A) Сравнивает слова между собой внутри последовательности

B) Удаляет лишние слова из текста

C) Переводит текст на другой язык

D) Сжимает данные для экономии памяти

Правильный ответ: A) Сравнивает слова между собой внутри последовательности

5. Как называется процесс разбиения текста на слова или подстроки?

A) Лемматизация

B) Токенизация

C) Стемминг

D) Нормализация

Правильный ответ: B) Токенизация

6. Какой тип эмбеддингов учитывает порядок слов в предложении?

A) Word2Vec

B) Позиционные эмбеддинги

C) One-Hot Encoding

D) TF-IDF

Правильный ответ: B) Позиционные эмбеддинги

7. Какой компонент трансформера отвечает за генерацию текста?

A) Энкодер

B) Декодер

C) Pooling-слой

D) Dropout-слой

Правильный ответ: B) Декодер

8. Что такое BERT (Bidirectional Encoder Representations from Transformers)?

A) Модель, учитывающая контекст слева и справа от слова

B) Модель только для генерации текста

C) Алгоритм сжатия данных

D) Метод визуализации текста

Правильный ответ: A) Модель, учитывающая контекст слева и справа от слова

9. Какой метод используется для дообучения языковой модели под конкретную задачу?

A) Fine-tuning

B) Data Augmentation

C) Feature Engineering

D) Clustering

Правильный ответ: A) Fine-tuning

10. Какой параметр в трансформерах определяет, насколько сильно модель обращает внимание на разные части входа?

A) Learning Rate

B) Batch Size

C) Attention Weights

D) Dropout Rate

Правильный ответ: C) Attention Weights

11. Какой алгоритм НЕ является языковой моделью?

A) GPT-3

B) BERT

C) ResNet

D) T5

Правильный ответ: C) ResNet

12. Что такое "контекстное окно" в языковых моделях?

A) Максимальное количество слов, которые модель может обработать за раз

B) Размер шрифта в выводе модели

C) Количество слоев в нейросети

D) Метод визуализации внимания

Правильный ответ: A) Максимальное количество слов, которые модель может обработать за раз

13. Какой метод помогает языковым моделям работать с редкими словами?

A) Subword Tokenization (например, Byte Pair Encoding)

B) Удаление редких слов

C) Замена всех слов на синонимы

D) Случайная замена слов

Правильный ответ: A) Subword Tokenization (например, Byte Pair Encoding)

14. Какой loss function чаще всего используется при обучении языковых моделей?

A) Mean Squared Error (MSE)

B) Cross-Entropy Loss

C) Hinge Loss

D) Binary Cross-Entropy

Правильный ответ: B) Cross-Entropy Loss

15. Какая задача НЕ является типичной для языковых моделей?

A) Генерация текста

B) Классификация изображений

C) Машинный перевод

D) Ответы на вопросы (QA)

Правильный ответ: B) Классификация изображений

16. Какой подход использует GPT для генерации текста?

A) Авторегрессия (предсказание следующего слова на основе предыдущих)

B) Обработка текста целиком за один проход

C) Случайный подбор слов

D) Использование только энкодера

Правильный ответ: A) Авторегрессия (предсказание следующего слова на основе предыдущих)

17. Что такое "zero-shot learning" в контексте языковых моделей?

A) Обучение без данных

B) Выполнение задачи без явного обучения на ней

C) Обучение только на одном примере

D) Удаление всех параметров модели

Правильный ответ: B) Выполнение задачи без явного обучения на ней

18. Какой метод позволяет языковой модели учитывать многозначность слов?

A) Контекстные эмбеддинги (например, из BERT)

B) One-Hot Encoding

C) Фиксированные word vectors

D) Удаление многозначных слов

Правильный ответ: A) Контекстные эмбеддинги (например, из BERT)

19. Какой параметр в GPT-3 делает её такой мощной?

A) Огромное количество параметров (175 млрд)

B) Использование только RNN

C) Отсутствие механизма внимания

D) Обучение только на одном языке

Правильный ответ: A) Огромное количество параметров (175 млрд)

20. Какой алгоритм используется для оптимизации весов в языковых моделях?

A) Градиентный спуск (Gradient Descent)

B) Random Search

C) Genetic Algorithms

D) K-Means

Правильный ответ: A) Градиентный спуск (Gradient Descent)

21. Что такое "перплексия" (perplexity) в языковых моделях?

A) Мера неопределенности модели

B) Количество слоев в сети

C) Размер словаря

D) Скорость генерации текста

Правильный ответ: A) Мера неопределенности модели

22. Какой тип данных НЕ используется для обучения языковых моделей?

A) Тексты из интернета

B) Книги и статьи

C) Аудиозаписи

D) Код программ

Правильный ответ: C) Аудиозаписи

23. Какой компонент трансформера помогает избежать переобучения?

A) Dropout

B) Batch Normalization

C) Residual Connections

D) Все перечисленные

Правильный ответ: D) Все перечисленные

24. Какой метод позволяет модели "забывать" ненужную информацию?

A) Dropout

B) Attention

C) Layer Normalization

D) Gradient Clipping

Правильный ответ: A) Dropout

25. Какая архитектура лучше подходит для длинных текстов?

A) RNN

B) Трансформеры

C) CNN

D) Markov Chains

Правильный ответ: B) Трансформеры

26. Какой подход использует BERT для обучения?

A) Маскирование слов (Masked Language Modeling)

B) Авторегрессия

C) Только supervised learning

D) Обучение без учителя

Правильный ответ: A) Маскирование слов (Masked Language Modeling)

27. Как называется процесс уменьшения размерности эмбеддингов?

A) Pooling

B) PCA

C) Quantization

D) Все варианты верны

Правильный ответ: D) Все варианты верны

28. Какой метод помогает ускорить обучение больших моделей?

A) Mixed Precision Training

B) Уменьшение batch size

C) Отключение внимания

D) Обучение только на CPU

Правильный ответ: A) Mixed Precision Training

29. Какой алгоритм НЕ является трансформером?

A) GPT

B) BERT

C) ELMo

D) T5

Правильный ответ: C) ELMo

30. Какой параметр влияет на "креативность" текста в GPT?

A) Temperature

B) Batch Size

C) Learning Rate

D) Количество эпох

Правильный ответ: A) Temperature

Ключи к тестам

Тест по языковым моделям

1. B
2. B
3. B
4. A
5. B
6. A
7. A
8. A
9. D
10. A

Дополнительные вопросы по языковым моделям и трансформерам

1. B
2. D
3. B
4. A
5. B
6. B
7. B
8. A
9. A
10. C
11. C
12. A
13. A
14. B
15. B
16. A
17. B
18. A
19. A
20. A
21. A
22. C
23. D
24. A
25. B
26. A
27. D
28. A
29. C
30. A

Контрольный тест

1. **Что такое языковая модель?**
   - Ответ: Это алгоритм машинного обучения, который предсказывает вероятность последовательности слов, способный понимать, генерировать и обрабатывать естественный язык.
2. **Какие виды языковых моделей существуют?**
   - Ответ: Статистические (n-gram), нейронные (RNN, LSTM), трансформерные (GPT, BERT) и гибридные модели.
3. **В чем заключается основная задача языковой модели?**
   - Ответ: Понимание и генерация осмысленного текста, предсказание следующего слова в последовательности.
4. **Что такое трансформеры и какова их роль в современных языковых моделях?**
   - Ответ: Архитектура нейросетей с механизмом внимания, позволяющая эффективно обрабатывать длинные текстовые зависимости (основа для GPT, BERT).
5. **Какие особенности отличают большие языковые модели (LLM)?**
   - Ответ: Огромное количество параметров (миллиарды), обучение на разнообразных данных, способность к few-shot learning.
6. **Какие методы обучения языковых моделей вы знаете?**
   - Ответ: Самообучение (self-supervised), обучение с учителем, дообучение (fine-tuning), обучение с подкреплением.
7. **В чем заключаются отличия между моделями BERT и GPT?**
   - Ответ: BERT - двунаправленная модель для понимания текста, GPT - однонаправленная авторегрессионная модель для генерации.
8. **Каким образом осуществляется подготовка данных для обучения языковых моделей?**
   - Ответ: Очистка текста, токенизация, создание масок (для BERT), формирование контекстных пар (для GPT).

9. **Как производится токенизация и эмбеддинг текста в процессе работы с языковыми моделями?**
   - Ответ: Токенизация - разбиение на слова/подслова (BPE), эмбеддинг - преобразование токенов в векторные представления.
10. **Какие области применения языковых моделей вам известны?**
   - Ответ: Машинный перевод, чат-боты, генерация текста, классификация, суммаризация, QA-системы.
11. **Какие перспективы развития языковых моделей вы видите в ближайшем будущем?**
   - Ответ: Уменьшение размеров при сохранении эффективности, мультимодальность, улучшение логического мышления.
12. **Какие библиотеки и инструменты используются для работы с языковыми моделями?**
   - Ответ: Hugging Face Transformers, TensorFlow/PyTorch, spaCy, NLTK, FastAPI для развертывания.
13. **Какие практические задачи можно решить с помощью языковых моделей в медицине?**
   - Ответ: Анализ медицинских записей, генерация отчетов, помощь в диагностике, обработка научных статей.
14. **Какие сложности возникают при обучении и настройке языковых моделей?**
   - Ответ: Вычислительные ресурсы, переобучение, смещение в данных, интерпретируемость решений.
15. **Какие навыки и знания необходимы для успешной работы с языковыми моделями?**
   - Ответ: Машинное обучение, NLP, программирование (Python), работа с большими данными, понимание лингвистики.