

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Структурное подразделение «Брикс кафедры»

УТВЕРЖДЕНА:
на заседании кафедры
Протокол №15 от 18 марта 2025 г.

Рабочая программа дисциплины

**«ВВЕДЕНИЕ В БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ / INTRODUCTION TO LARGE
LANGUAGE MODELS»**

Направление: 09.03.01 Информатика и вычислительная техника

Искусственный интеллект и компьютерные науки /Artificial Intelligence and Computer
Science

Квалификация: Бакалавр

Форма обучения: очная

Документ подписан простой
электронной подписью
Составитель программы:
Афанасьев Александр
Диомидович
Дата подписания: 18.06.2025

Документ подписан простой
электронной подписью
Утвердил: Киреев Анна
Павловна
Дата подписания: 19.06.2025

Документ подписан простой
электронной подписью
Согласовал: Афанасьев
Александр Диомидович
Дата подписания: 18.06.2025

Год набора – 2025

Иркутск, 2025 г.

1 Перечень планируемых результатов обучения по дисциплине, соотнесённых с планируемыми результатами освоения образовательной программы

1.1 Дисциплина «Введение в большие языковые модели / Introduction to Large Language Models» обеспечивает формирование следующих компетенций с учётом индикаторов их достижения

Код, наименование компетенции	Код индикатора компетенции
ДК-1 Способность осуществлять деятельность, находящуюся за пределами непосредственной профессиональной сферы	ДК-1.2

1.2 В результате освоения дисциплины у обучающихся должны быть сформированы

Код индикатора	Содержание индикатора	Результат обучения
ДК-1.2	Способен анализировать и применять принципы работы больших языковых моделей для решения практических задач в различных областях профессиональной деятельности	Знать Основные принципы архитектуры больших языковых моделей (LLM). Методы обучения и дообучения LLM, а также особенности генерации текста и обработки запросов. Области применения LLM в профессиональной деятельности и ограничения моделей (этические, технические). Современные инструменты и платформы для работы с LLM. Уметь Проводить анализ задач и выбирать подходящие методы применения LLM для их решения. Формулировать запросы (промнты) для эффективного взаимодействия с языковыми моделями. Использовать LLM для генерации, классификации, извлечения информации и автоматизации процессов. Оценивать качество и релевантность результатов, выявлять и корректировать ошибки модели. Владеть Навыками работы с API и библиотеками для интеграции LLM в программные решения. Методами тонкой настройки (fine-tuning) и адаптации моделей под конкретные задачи. Инструментами анализа и визуализации результатов работы LLM.

		Критическим мышлением для оценки возможностей и ограничений языковых моделей.
--	--	---

2 Место дисциплины в структуре ООП

Изучение дисциплины «Введение в большие языковые модели / Introduction to Large Language Models» базируется на результатах освоения следующих дисциплин/практик: «Введение в специальность / Introduction to the specialty», «Программирование / Programming», «Большие данные, методы обработки / Big Data, Processing Methods»

Дисциплина является предшествующей для дисциплин/практик: «Глубокое обучение в технологиях обработки естественного языка (NLP) / Deep Learning in Natural Language Processing Technologies (NLP)», «Производственная практика: технологическая (проектно-технологическая) практика / Manufacturing Practice: Technological Practice»

3 Объем дисциплины

Объем дисциплины составляет – 3 ЗЕТ

Вид учебной работы	Трудоемкость в академических часах (Один академический час соответствует 45 минутам астрономического часа)	
	Всего	Семестр № 6
Общая трудоемкость дисциплины	108	108
Аудиторные занятия, в том числе:	54	54
лекции	18	18
лабораторные работы	0	0
практические/семинарские занятия	36	36
Самостоятельная работа (в т.ч. курсовое проектирование)	54	54
Трудоемкость промежуточной аттестации	0	0
Вид промежуточной аттестации (итогового контроля по дисциплине)	Зачет	Зачет

4 Структура и содержание дисциплины

4.1 Сводные данные по содержанию дисциплины

Семестр № 6

№ п/п	Наименование раздела и темы дисциплины	Виды контактной работы						СРС		Форма текущего контроля
		Лекции		ЛР		ПЗ(СЕМ)				
		№	Кол. Час.	№	Кол. Час.	№	Кол. Час.	№	Кол. Час.	
1	2	3	4	5	6	7	8	9	10	11
1	Концепция больших языковых моделей, их роль в современном ИИ.	1	6					1, 2, 3	12	Тест

2	Обучение больших языковых моделей	2	4					2, 4	4	Тест
3	Создание базы знаний в контексте обработки естественного языка	3	8			1, 2	10	2, 4	4	Тест
4	Основы промт-инжиниринга.					3	4	2	2	Обзор статьи
5	Дообучение ChatGPT на основе базы знаний					4	6	1, 2, 4	14	Тест
6	Разработка нейросотрудников на GPT					5, 6	16	5	18	Творческое задание
	Промежуточная аттестация									Зачет
	Всего		18				36		54	

4.2 Краткое содержание разделов и тем занятий

Семестр № 6

№	Тема	Краткое содержание
1	Концепция больших языковых моделей, их роль в современном ИИ.	Обзор концепции больших языковых моделей, их роль в современном ИИ и обработке естественного языка. Эволюция языковых моделей. От статистических моделей и RNN к трансформерам и современным LLM. Сферы и способы применения: GPT-3, GPT-4, ChatGPT и др. Обзор архитектуры трансформеров: принципы работы трансформеров, механизм внимания, энкодер-декодер.
2	Обучение больших языковых моделей	Процессы предобучения, дообучения, сбор и подготовка данных. Методы генерации текста и ответы на запросы. Как LLM генерируют осмысленные тексты, работа с контекстом и последовательностями.
3	Создание базы знаний в контексте обработки естественного языка	Основные термины и инструменты при создании базы знаний в контексте обработки естественного языка: чанки; сплиттеры (CharacterTextSplitter, RecursiveCharacterTextSplitter, MarkdownHeaderTextSplitter); токены. Подходы к разделению текста: простой сплиттер, сплиттер предложений, регулярный сплиттер, сплиттер символов, на основе пунктуаций, NLTK сплиттер). Эмбединги (Embeddings): word2vec, GloVe, BERT. Платформа LangChain для разработки приложений на основе больших языковых моделей.
4	Основы промт-	Организация работы промтов (схема). Явные

	инжиниринга.	инструкции, скрытый контекст, системные настройки. Вредоносные действия со стороны пользователей. Примеры атак. Уязвимость промтов. Базовые принципы защиты.
5	Дообучение ChatGPT на основе базы знаний	Использование open source инструмента FAISS (Facebook AI Similarity Search) для организации запроса в пользовательском хранилище векторов. Сочетание FAISS с большой языковой моделью (LLM) для извлечения контекстно релевантной информации из домена базы знаний. Понятие Prompt. Роли ChatGPT: System (инструкция для ChatGPT = промт), User (запрос пользователя), Assistant (ответ ChatGPT).
6	Разработка нейро-сотрудников на GPT	Обзор нейро-сотрудников на GPT и особенности их функционирования. Создание нейро-сотрудника с нуля на python.

4.3 Перечень лабораторных работ

Лабораторных работ не предусмотрено

4.4 Перечень практических занятий

Семестр № 6

№	Темы практических (семинарских) занятий	Кол-во академических часов
1	Изучение инструментов для создания базы знаний в контексте обработки естественного языка	6
2	Создание базы знаний	4
3	Практика создания промтов.	4
4	Использование open source инструмента FAISS для организации запроса в пользовательском хранилище векторов.	6
5	Разработка нейросотрудников на GPT	10
6	Интеграция нейро-сотрудников в production - дополнительные возможности chatGPT	6

4.5 Самостоятельная работа

Семестр № 6

№	Вид СРС	Кол-во академических часов
1	Анализ научных публикаций	16
2	Ведение терминологического словаря	12
3	Выполнение тренировочных и обучающих тестов	2
4	Выполнение тренировочных и обучающих тестов в дистанционном режиме	6

5	Оформление отчетов по лабораторным и практическим работам	18
---	---	----

В ходе проведения занятий по дисциплине используются следующие интерактивные методы обучения: Дискуссия, кейс-технология, лекция с ошибками, мозговой штурм

5 Перечень учебно-методического обеспечения дисциплины

5.1 Методические указания для обучающихся по освоению дисциплины

5.1.1 Методические указания для обучающихся по практическим занятиям

Рекомендуется активно использовать современные инструменты и платформы для практического взаимодействия с LLM, экспериментировать с параметрами моделей и анализировать полученные результаты. Делать систематические записи полученных результатов.

5.1.2 Методические указания для обучающихся по самостоятельной работе:

Регулярно изучайте теоретический материал по основам больших языковых моделей, используя рекомендованные источники и гайды.

Практикуйтесь в работе с современными инструментами и платформами, выполняя задания по генерации, анализу и дообучению моделей.

Составляйте собственные конспекты и схемы, фиксируйте результаты экспериментов, анализируйте влияние параметров моделей на качество генерации.

Используйте разнообразные форматы самостоятельной работы: чтение, написание кода, подготовка отчетов и презентаций, что способствует развитию системного и критического мышления.

6 Фонд оценочных средств для контроля текущей успеваемости и проведения промежуточной аттестации по дисциплине

6.1 Оценочные средства для проведения текущего контроля

6.1.1 семестр 6 | Тест

Описание процедуры.

Тест проводится на платформе электронного образования ИРНИТУ.

Критерии оценивания.

Оценка "5" - от 95% выполненных заданий

Оценка "4" - от 80% до 94% выполненных заданий

Оценка "3" - от 60% до 79% выполненных заданий

Оценка "2" - менее 60%

6.1.2 семестр 6 | Обзор статьи

Описание процедуры.

Чтение статьи или ее фрагмента с последующим ее обсуждением или выполнение теста на понимание прочитанного материала.

Критерии оценивания.

Оценка результата понимания статьи на основе результата выполненного теста.

6.1.3 семестр 6 | Творческое задание

Описание процедуры.

Разработка нейро-сотрудника или ассистента (консультанта) на основе GPT по согласованию с преподавателем.

Критерии оценивания.

Рабочая модель нейро-сотрудника или ассистента (консультанта) на основе GPT

6.2 Оценочные средства для проведения промежуточной аттестации

6.2.1 Критерии и средства (методы) оценивания индикаторов достижения компетенции в рамках промежуточной аттестации

Индикатор достижения компетенции	Критерии оценивания	Средства (методы) оценивания промежуточной аттестации
ДК-1.2	Демонстрирует знания и навыки по применению LLM. Применяет LLM для решения практических задач, умеет анализировать результаты. Разрабатывает и оптимизирует решения с использованием LLM, критически оценивает их эффективность.	Тест, творческое задание

6.2.2 Типовые оценочные средства промежуточной аттестации

6.2.2.1 Семестр 6, Типовые оценочные средства для проведения зачета по дисциплине

6.2.2.1.1 Описание процедуры

Выполнение практических заданий и 4-х тестов в системе электронного образования. Представление отчета по творческому заданию.

6.2.2.1.2 Критерии оценивания

Зачтено	Не зачтено
Выполнено 60% практических заданий и 4-х тестов в системе электронного образования. Представлен отчет по творческому заданию.	Выполнено менее 60% практических заданий, не все тесты сделаны в системе электронного образования. Нет отчета по творческому заданию.

7 Основная учебная литература

1. Келен О. Разработка приложений на базе GPT-4 и ChatGPT / О. Келен, М.-А. Блете. — 2-е изд. — М.: Издательский дом «Питер», 2024. — 192 с. — (Бестселлеры O'Reilly). — ISBN 978-601-08-3731-7.

8 Дополнительная учебная литература и справочная

1. Ростовцев В. С. Искусственные нейронные сети : учебник для вузов / В. С. Ростовцев, 2023. - 216.

9 Ресурсы сети Интернет

1. <http://library.istu.edu/>
2. <https://e.lanbook.com/>
4. Подборка книг Awesome LLM Books <https://github.com/Jason2Brownlee/awesome-llm-books/tree/main?tab=readme-ov-file>
5. 8.6. Языковые модели от Яндекс Образование <https://education.yandex.ru/handbook/ml/article/yazykovye-modeli>
6. Создание персонального AI-ассистента с использованием RAG
Подробное руководство по построению ассистента на основе Retrieval-Augmented Generation с использованием ChromaDB, Llama и FastAPI.
Ссылка: <https://habr.com/ru/articles/873336/>
7. LLM + RAG: ингредиенты приготовления умного ИИ-ассистента (видео)
Обзор теории и практических кейсов применения RAG для повышения качества ответов LLM.
Ссылка: https://www.youtube.com/watch?v=_hYjp-2efzA
8. Интеграция LLM в корпоративные чат-боты: RAG-подход и примеры
Рассмотрение метода Retrieval Augmented Generation и примеров использования в бизнесе.
Ссылка: <https://habr.com/ru/companies/X5Tech/articles/834832/>
9. Retrieval-Augmented Generation (RAG): виды, типы и примеры
Обзор архитектур RAG, примеры кода с использованием Haystack и LangChain, пошаговая реализация.
Ссылка: <https://nerdit.ru/retrieval-augmented-generation-rag-vidy-tipy-i-primery/>
10. LangChain vs LlamaIndex: проектируем RAG и разбираемся, что лучше
Сравнение популярных фреймворков для реализации RAG-пайплайнов с примерами кода.
Ссылка: <https://www.braintools.ru/article/11446>

10 Профессиональные базы данных

1. <http://new.fips.ru/>
2. <http://www1.fips.ru/>

11 Перечень информационных технологий, лицензионных и свободно распространяемых специализированных программных средств, информационных справочных систем

1. Microsoft Windows (XP Prof + Vista Bussines) rus VLK поставка 08_2007
2. Excel Link concurrent AcademicEdition

12 Материально-техническое обеспечение дисциплины

1. Проектор Epson EB-W04LCD.WXGA 1280*800.3000:1.2800 ANSI Lumens

2. Компьютер Asustek P8H6-M/Intel Core i5
2400/4Gb/HDD2TB/DVD-RW/ATX550W/LCD22/ИБП1