

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Декан ФИиВТ

УТВЕРЖДАЮ /А.А. Кречетов/
(Ф.И.О. декана (директора института))

19.02.2026 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б.1.2.16 Большие языковые модели

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

09.03.04 Программная инженерия

Квалификация выпускника

Бакалавр

(бакалавр/магистр/специалист)

Направленность

Разработка программных систем и систем искусственного
интеллекта

Курс 4
Семестр 7

Распределение учебного времени

Трудоемкость по учебному плану	<u>144 / 4</u>	часов/зачетных единиц
Лекции	<u>32</u>	часов
Лабораторные работы	<u>32</u>	часов
Практические занятия	<u>-</u>	часов
Иная контактная работа	<u>-</u>	часов
Всего контактной работы (без учета экз.)	<u>64</u>	часов
Контактная работа по экзамену	<u>-</u>	часов
Курсовой проект (работа)	<u>-</u>	семестр
Самостоятельная работа обучающихся (без учета экз.)	<u>80</u>	часов
Самостоятельная работа по подготовке к экзамену	<u>-</u>	часов
Экзамен	<u>-</u>	семестр
Зачет	<u>-</u>	семестр
БРК, ДЗ	<u>7</u>	семестр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 09.03.04 Программная инженерия

Программу составили:

заведующий кафедрой	ИиСП	СОГЛАСОВАНО	А.В. Бородин
(должность)	(кафедра)		(И.О. Фамилия)
доцент	ИиСП	СОГЛАСОВАНО	Д.С. Шумков
(должность)	(кафедра)		(И.О. Фамилия)

РАССМОТРЕНА и ОДОБРЕНА на заседании кафедры, за которой закреплена дисциплина
Кафедра информатики и системного программирования

(наименование кафедры)			
09.02.2026	протокол №	6	
(дата)			
Заведующий кафедрой	СОГЛАСОВАНО	А.В. Бородин	
		(И.О. Фамилия)	

Рабочая программа СОГЛАСОВАНА с факультетом (институтом), выпускающей(ими) кафедрой(ами).

СООТВЕТСТВУЕТ действующей ОП.

Заведующий кафедрой	СОГЛАСОВАНО	А.В. Бородин
		(И.О. Фамилия)

Председатель методической комиссии факультета (института), в который входит выпускающая кафедра

СОГЛАСОВАНО	А.А. Кречетов
	(И.О. Фамилия)

Эксперт(ы): Галочкин Александр Владимирович, генеральный директор ООО «Трэвэл Лайн»

Рабочая программа проверена и зарегистрирована в УМЦ 19.02.2026 г.

Специалист учебно-методического центра СОГЛАСОВАНО /Т.А. Смирнова/

Раздел 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является достижение планируемых результатов обучения, соответствующих установленным в ОПОП индикаторам достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения
1. ML-1 Способен применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ	ML-1.1 Позиционирует собственную задачу в заданной области знания с точки зрения трендов современного искусственного интеллекта	знания: (Б) Знает базовые формулировки задач в заданной предметной области, соотносит их с базовыми направлениями ИИ, знает общие актуальные технологии. умения: (П) Умеет анализировать и сопоставлять задачу с современными трендами, выделять специфику задачи в контексте последних достижений ИИ навыки: (Э) Владеет аргументацией выбора постановки задачи с учётом cutting-edge технологий, владеет инструментарием выявления потенциала интеграции новых подходов на стыке дисциплин.
	ML-1.2 Определяет тенденции развития, оценивает новизну и практическую значимость своих решений с точки зрения современного искусственного интеллекта	знания: (Б) Знает основные тенденции развития ИИ, оценивает новизну решения на уровне известных практик. Знает минусы традиционных подходов в сравнении с современными принципами, заложенными в технологии ИИ умения: (П) Объясняет причины появления концепции больших данных (БД), разницу определений. Выявляет различные категории проблем больших данных с примерами. Умеет анализировать динамику появления новых технологий, сопоставлять собственные решения с современными исследованиями и индустриальными стандартами. навыки: (Э) Владеет методологией прогноза перспективных направлений развития ИИ, оценки новизны и значимости решений с опорой на state-of-the-art публикации.

	ML-1.3 Оценивает конкурирующие решения и разработки с точки зрения трендов современного искусственного интеллекта	знания: (Б) Знаком с конкурирующими решениями по базовым характеристикам (точность, скорость, применимость), ориентируясь на общепринятые подходы (например, сравнение различных ML-решений). умения: (П) Умеет оценивать конкурентные решения с учётом современных трендов (например, использование современных архитектур, подходов к интерпретируемости, устойчивости, энергоэффективности), анализирует преимущества и ограничения. навыки: (Э) Владеет инструментарием бенчмаркинга конкурирующих решений на основе актуальных метрик и датасетов, учитывает аспекты explainable AI, robustness, fairness, compliance, формулирует рекомендации по выбору и развитию наиболее перспективных
2. ML-2 Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками	ML-2.1 Различает основные типы задач машинного обучения и применяет на практике принципы их решения	знания: (Б) Знает основные типы задач машинного обучения (обучением с учителем, без учителя и с подкреплением). Применяет типовые подходы к решению базовых задач с использованием готовых инструментов и библиотек (ScikitLearn). умения: (П) Умеет выбирать и обосновывать методы решения задач машинного обучения с учётом характеристик данных и бизнес-контекста, настраивать базовые модели и проводить их оценку. навыки: (П) Владеет базовым инструментарием решения задач машинного обучения с учётом характеристик данных и бизнес-контекста.
	ML-2.2 Применяет методы предварительной обработки данных и работы с признаками	знания: (Б) Знаком с методами разведочного анализа данных, работы с пропущенными значениями и выбросами. умения: (П) Умеет проводить разведочный анализ данных, работать с пропущенными значениями и выбросами. навыки: (П) Владеет методами feature engineering: отбора создания и преобразования признаков.
	ML-2.3 Решает проблемы несбалансированных данных и оценивает качество моделей	знания: (Б) Знает основные метрики оценки качества для задач классификации и регрессии. умения: (П) Применяет различные типы кросс-валидации. Оценивает качество моделей с учетом bias-variance trade-off. навыки: (П) Владеет базовым инструментарием оценки качества для задач классификации и регрессии.

3. ML-3 Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения	ML-3.1 Обосновывает способы и варианты применения классических методов и моделей машинного обучения в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи.	знания: (Б) Знает базовые модели МО с учителем (метод ближайших соседей, деревья принятия решений, линейная регрессия, логистическая регрессия, метод опорных векторов) с пониманием их математической сущности. умения: (П) Умеет обосновать выбор конкретных алгоритмов и их параметров в зависимости от задачи и данных. навыки: (Э) Владеет методами разработки и адаптации собственных алгоритмических решений на основе классических методов. Обосновывает математически сложные решения.
	ML-3.2 Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ	знания: (Б) Знает постановки задач в терминах обучения на частично размеченных данных; знает стандартные алгоритмы для работы с данными и обучения моделей с использованием неполной разметки. Знает постановки задач в терминах задачи ранжирования; знает простейшие методы ранжирования (pointwise подход). умения: (П) Умеет применять методы байесовской классификации и ансамблевых методов МО (бэггинг, бустинг, стэкинг моделей), а также производных от них (случайные леса, градиентный бустинг на деревьях). Применяет классические методы МО для временных рядов (ARIMA, экспоненциальное сглаживание, линейная регрессия с лагами). навыки: (П) Владеет инструментами оценки качества моделей ранжирования и сравнения ранжирующих моделей между собой. Владеет методами обучения типа pairwise и listwise. Знает и применяет на практике различные архитектуры ранжированного поиска (одно-, двух-, трех-стадийное ранжирование).
	ML-3.3 Оценивает результативность применения классических методов и моделей машинного обучения в задачах ИИ на основе сопоставления с аналогами	знания: (Б) Знает основные метрики и подходы к сравнительному анализу нескольких моделей с учётом бизнес-требований. умения: (П) Умеет учитывать основные ограничения классического МО и области применения. Умеет разрабатывать рекомендации по использованию или отказу от классических моделей в пользу более сложных методов (ensemble, deep learning) на основе обоснованного анализа. навыки: (Э) Владеет методами оценки теоретических ограничений алгоритмов и способен находить баланс между различными подходами. Владеет методами системного анализа эффективности моделей на уровне бизнес-эффекта, затрат и рисков. Может объяснить результаты моделей заказчику.

4. ПК-1 Владение навыками использования различных технологий разработки программного обеспечения	ПК-1.1 Знает современные технологии разработки ПО (структурное, объектно-ориентированное)	знания: Знает современные технологии разработки ПО (структурное, объектно-ориентированное) умения: навыки:
	ПК-1.2 Умеет использовать современные технологии разработки ПО	знания: умения: Умеет использовать современные технологии разработки ПО навыки:
	ПК-1.3 Имеет навыки использования современных технологий разработки ПО	знания: умения: навыки: Имеет навыки использования современных технологий разработки ПО

Раздел 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к элективным дисциплинам (модулям) ОПОП.

Дисциплина является элективной

Для продолжения формирования заявленных компетенций необходимы знания предшествующих дисциплин: Основы технологий искусственного интеллекта (ML-1), Основы генеративных систем искусственного интеллекта (ML-1), Основы технологий искусственного интеллекта (ML-2), Анализ данных (ML-2), Основы машинного обучения (ML-2), Основы генеративных систем искусственного интеллекта (ML-2), Основы технологий искусственного интеллекта (ML-3), Основы машинного обучения (ML-3), Основы генеративных систем искусственного интеллекта (ML-3), Параллельное программирование (ПК-1)

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих государственной итоговой аттестации в форме: Подготовка к процедуре защиты и защита выпускной квалификационной работы (ML-1), Подготовка к процедуре защиты и защита выпускной квалификационной работы (ML-2), Подготовка к процедуре защиты и защита выпускной квалификационной работы (ML-3), Подготовка к процедуре защиты и защита выпускной квалификационной работы (ПК-1)

Раздел 3. ОПИСАНИЕ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

Для формирования заявленных компетенций используются методологические технологии, реализующие деятельностный, личностно-ориентированный, практико-ориентированный подходы.

Основными стратегическими технологиями являются: дискуссионные, исследовательские, лекционные занятия, практические и лабораторные занятия, процедуры самообучения, доступ обеспечивается с помощью электронного курса и сервисов образовательного портала пгту на платформе lms moodle (elearning.volgatech.net), интегрированного с федеральной платформой гис сдос online.edu.ru открытый образовательный портал (mooped.net).

На достижение конкретных целей обучения направлены применяемые тактические технологии: задания, информационные, классическая лекция, проблемная лекция, лекция с обсуждением новых научных достижений в сфере ии по результатам конференции и публикаций и др.), семинары с ит-практиками, конкурсы, хакатоны, кейс-чемпионаты, мастер-классы, командные проектные задания и др.

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

7 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Введение в тематику LLM	144	ML-1, ML-2, ML-3, ПК-1
Лекция. Лекция №1. Введение в Большие Языковые Модели (LLM) и эволюция NLP. Исторический контекст NLP. Появление механизма внимания. "Attention is All You Need" и архитектура трансформера. Влияние LLM на AI ландшафт.	4	
Лекция. Лекция №2. Архитектура трансформера: детали. Механизмы Self-Attention и Multi-Head Attention. Позиционное кодирование. Слои нормализации и остаточные соединения. Feed-Forward Network (FFN) в трансформере. Архитектуры: Encoder-Decoder против Decoder-Only (Causal LLM).	4	
Лекция. Лекция №3. Предварительное обучение (Pre-training) LLM. Цели и стратегии предварительного обучения. Датасеты для предварительного обучения. Токенизация. Масштабирование LLM.	2	
Лекция. Лекция №4. Донастройка (Fine-tuning). Назначение. Supervised Fine-tuning - SFT. Обучение с подкреплением на основе обратной связи от человека.	2	
Лекция. Лекция №5. Промпт-инжиниринг и In-Context Learning. Концепция промпт-инжиниринга. Парадигма "In-Context Learning". Типы промптинга. Продвинутое техники промптинга. Параметры генерации.	2	
Лекция. Лекция №6. Эффективная донстройка и развертывание LLM (PEFT). Проблемы полной донстройки LLM. Parameter-Efficient Fine-Tuning (PEFT) методы. Квантование моделей. Службы развертывания LLM.	4	
Лекция. Лекция №7. Оценка и бенчмаркинг LLM. Метрики для оценки LLM. Оценка человеком. Типовые бенчмарки и датасеты. Оценка LLM с помощью других LLM.	4	
Лекция. Лекция №8. Продвинутое приложения и мультимодальные LLM. Retrieval-Augmented Generation (RAG). LLM как агенты и использование инструментов. Мультимодальные LLM.	4	

Лекция. Лекция №9. Вызовы, ограничения и этические аспекты LLM. Основные ограничения LLM. Этическая сторона LLM. Open против Closed Source LLM: дискуссия о преимуществах и недостатках.	4
Лекция. Лекция №10. Будущее LLM и направления исследований. Текущие исследовательские тренды. LLM в науке. LLM на производстве. Регулирование сферы применения LLM. Философские вопросы применения LLM.	2
Лабораторная работа. Лабораторная работа №1. Знакомство с Hugging Face Transformers и генерация текста.	4
Лабораторная работа. Лабораторная работа №2. Основы промпт-инжиниринга и In-Context Learning.	4
Лабораторная работа. Лабораторная работа №3. Донастройка (Fine-tuning) LLM для конкретной задачи.	2
Лабораторная работа. Лабораторная работа №4. Эффективная донастройка (PEFT) с LoRA.	2
Лабораторная работа. Лабораторная работа №5. Использование LLM для вопросов и ответов (QA) и RAG.	4
Лабораторная работа. Лабораторная работа №6. LLM как агенты и использование инструментов.	2
Лабораторная работа. Лабораторная работа №7. Оценка LLM.	4
Лабораторная работа. Лабораторная работа №8. Исследовательская задача	10

Задания для самостоятельной работы, в том числе выполнение Проработка лекций.		
Подготовка к лабораторным работам.		
Самостоятельное изучение материалов: Foster D. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. - O'Reilly Media, 2019. - 330 p. Huyen C. AI Engineering: Building Applications with Foundation Models. - O'Reilly Media, 2025. - 532 p. Halloway E. THE AI ENGINE ACCELERATOR FOR BEGINNERS : Your Practical Road Map to Harness the Power of Artificial Intelligence, Craft Powerful Prompts, and Improve Task Automation with Ease. - O'Reilly Media, 2025. - 223 p. Natural Language Processing with Transformers: Fundamentals and Core Applications: A Practical Guide. From Beginner to Intermediate in Building Intelligent Language Applications / Quantum Technologies. - O'Reilly Media, 2024. - 1063 p. Data Engineering Foundations: Core Techniques for Data Analysis with Pandas, NumPy, and Scikit-Learn (Advanced Data Analysis Series Book 1) / Quantum Technologies. - O'Reilly Media, 2024. - 962 p. Caldwell P. The AI Prompt Engineering Principles: The Timeless Hidden Solution For Beginners To Stop Wasting Hours On Trial-And-Error And Get Exact, High-Quality AI Results. - O'Reilly Media, 2025. - 227 p.	80	
Иная контактная работа:	0	

Раздел 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Изучение дисциплины рекомендуется начать с ознакомления с рабочей программой, ее структурой и содержанием разделов. Учебный материал структурирован, изучение дисциплины осуществляется в тематической последовательности.

Занятия лекционного типа дают систематизированные знания по дисциплине, концентрируют внимание на наиболее сложных и важных вопросах. Во время лекционных занятий рекомендуется вести конспектирование учебного материала; обращать внимание на формулировки и категории, раскрывающие суть проблемы, явления или процесса; зафиксировать выводы и практические рекомендации.

Подготовка к **занятиям семинарского типа** включает ознакомление с планом лабораторного занятия; работу с конспектом лекций, выполнение домашнего задания, работу с учебной и учебно-методической литературой, научными изданиями и электронными образовательными ресурсами, рекомендованными рабочей программой дисциплины.

Содержание **самостоятельной работы** определяется рабочей программой дисциплины, оценочными и методическими материалами, заданиями и указаниями преподавателя. Самостоятельная работа может осуществляться в аудиторной и внеаудиторной формах. Эффективным средством осуществления самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к образовательной программе, рабочей программе дисциплины, к электронным библиотечным системам, профессиональным базам данных и информационным справочным

Дополнительно содержание дисциплины и самостоятельная работа в рамках дисциплины включает изучение исследований и разработок, представленных в научных публикациях на конференциях уровня А*/А, статьях, опубликованных в журналах 1 уровня Белого списка научных журналов Минобрнауки России, в том числе:

- International Conference on Learning Representations (ICLR). Сайт: <https://iclr.cc/> Международная конференция по обучению представлением. Имеет открытый процесс рецензирования и специальные секции (workshop tracks), ориентированные на студентов и исследователей, представляющих прикладные решения и обучающие эксперименты с использованием библиотек Python.
- International Conference on Machine Learning (ICML). Сайт: <https://icml.cc/> Одна из ведущих конференций по машинному обучению. Включает постерные и демонстрационные треки (demo, education), где принимаются проекты по анализу данных, визуализации результатов и разработке обучающих приложений на Python.
- AAAI Conference on Artificial Intelligence. Сайт: <https://aaai.org/conference/> Конференция Ассоциации по развитию искусственного интеллекта. Содержит специальную программу Student Abstract and Poster Program, ориентированную на студенческие и начальные исследовательские работы по искусственному интеллекту, включая разработку алгоритмов и систем на Python.
- International Joint Conference on Artificial Intelligence (IJCAI). Сайт: <https://www.ijcai.org/> Международная объединённая конференция по искусственному интеллекту. Имеет секцию AI Education Track, где рассматриваются проекты, направленные на развитие методов обучения ИИ и использование Python в образовательных целях.

– рецензируемый журнал «Искусственный интеллект и принятие решений» (Белый список, K1). Публикует оригинальные научные статьи и обзоры по широкому кругу проблем в области искусственного интеллекта и интеллектуальных систем, анализа и принятия решений, систем поддержки принятия решений, их практического применения.

– международный, междисциплинарный рецензируемый журнал «Форсайт» (Белый список K1), публикует оригинальные научные статьи с результатами передовых теоретических и прикладных исследований в ключевых областях стратегического планирования, научно-технической и инновационной политики, информирующий читателей о возможных альтернативных сценариях будущего для своевременного принятия правильных решений.

– рецензируемый журнал «Программирование» (Белый список, K1).. Публикует статьи по всем проблемам, связанным с теоретическим и практическим программированием: операционные системы, технологии программирования, языки программирования и компиляторы, параллельное программирование, верификация и тестирование программ, машинная графика, компьютерная алгебра и т.п.

- –международный журнал "Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications" (Белый список, K2). Публикует оригинальные работы в области распознавания образов, анализа, понимания и обработки изображений, а также в смежных областях теоретической и прикладной информатики.
- –рецензируемый журнал «Т-Comm – Телекоммуникации и Транспорт» (Белый список, K2). Публикует достижения и проблемы российских инфокоммуникаций; развитие интеллектуальных транспортных систем.

Периодичность проведения, формы текущего контроля успеваемости, система оценивания

хода освоения дисциплин представлены в рабочей программе.

Формой промежуточной аттестации по дисциплине является **балльно-рейтинговый контроль**.

Список рекомендуемой литературы для самостоятельного изучения:

1. Foster D. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. - O'Reilly Media, 2019. - 330 p.
2. Huyen C. AI Engineering: Building Applications with Foundation Models. - O'Reilly Media, 2025. - 532 p.
Halloway E. THE AI ENGINE ACCELERATOR FOR BEGINNERS : Your Practical Road Map to Harness the Power of Artificial Intelligence, Craft Powerful Prompts, and Improve Task Automation with Ease. - O'Reilly Media, 2025. - 223 p.
3. Natural Language Processing with Transformers: Fundamentals and Core Applications: A Practical Guide. From Beginner to Intermediate in Building Intelligent Language Applications / Cuantum Technologies. - O'Reilly Media, 2024. - 1063 p.
4. Data Engineering Foundations: Core Techniques for Data Analysis with Pandas, NumPy, and Scikit-Learn (Advanced Data Analysis Series Book 1) / Cuantum Technologies. - O'Reilly Media, 2024. - 962 p.
5. Caldwell P. The AI Prompt Engineering Principles: The Timeless Hidden Solution For Beginners To Stop Wasting Hours On Trial-And-Error And Get Exact, High-Quality AI Results. - O'Reilly Media, 2025. - 227 p.

Раздел 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-методическое обеспечение

№№ п/п	Список используемой литературы	Количество экземпляров печатных изданий, имеющих в библиотеке, или электронный адрес издания (ресурса) в сети Интернет
УЧЕБНЫЕ, УЧЕБНО-МЕТОДИЧЕСКИЕ И НАУЧНЫЕ ИЗДАНИЯ		
1.	Кревецкий, Александр Владимирович. Основы технологий искусственного интеллекта [Электронный ресурс] : учебное пособие / А. В. Кревецкий, Ю. А. Ипатов, Н. И. Роженцова ; под общей редакцией А. В. Кревецкого; Министерство науки и высшего образования Российской Федерации, ФГБОУ ВО "Поволжский государственный технологический университет". Йошкар-Ола: ПГТУ, 2023. - 272 с. ISBN 978-5-8158-2358-	https://portal.volgatech.net/books/Krevetskiy_Osnovy_tekhnologii_iskusstvennogo_intellekta_2023.pdf

2.	Сидоркина, Ирина Геннадьевна. Системы искусственного интеллекта [Текст] : [учеб. пособие для вузов по направлению 230100 "Информатика и вычисл. техника"] / И. Г. Сидоркина. М.: Кнорус, 2011. - 245 с. ISBN 978-5-406-00449-4. Экземпляры: всего 88.	88
3.	Остроух, А. В. Системы искусственного интеллекта [Электронный ресурс] : монография / Остроух А. В., Суркова Н. Е.; Суркова Н. Е. 4-е изд., стер. Санкт-Петербург: Лань, 2024. - 228 с. ISBN 978-5-507-47478-3.	https://e.lanbook.com/book/379988
4.	Нейросетевые технологии и прикладные модели машинного обучения в инфокоммуникациях [Текст] : практикум для направлений подготовки 11.04.02, 11.03.02 "Инфокоммуникационные технологии и системы связи", 09.03.02 "Информационные системы и технологии" / В. В. Овчинников, А. А. Кислицын, А. А. Елсуков [и др.]; Министерство науки и высшего образования Российской Федерации, ФГБОУ ВО "Поволжский государственный технологический университет", 2024. - 82, [1] с. ISBN 978-5-8158-2421-8. Экземпляры: всего 2.	2 / https://portal.volgatech.net/books/Neyrosetevyye_tekhnologii_i_prikladnyye_modeli_mashinnogo_obucheniya_v_infokommunikatsiyakh_2025.pdf
ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ		
1.	Научная электронная библиотека eLIBRARY.RU	http://elibrary.ru
2.	Научная электронная библиотека «Киберленинка»	http://cyberleninka.ru
ПРОФЕССИОНАЛЬНЫЕ БАЗЫ ДАННЫХ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ		
1.	Справочно-правовая система Консультант+	http://www.consultant.ru
2.	Информационно-правовой портал Гарант	http://www.garant.ru

6.2. Материально-техническая база и программное обеспечение

№№ п/п	Аудитории для проведения учебных занятий, самостоятельной работы и проведения государственной итоговой аттестации	Перечень основного оборудования	Программное обеспечение
1.	425 (III)	Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач
2.	428 (III)	Ноутбук ASUS EeePC 1215N 12,1" (1), Ноутбук HP Compaq	Microsoft Windows Enterprise, Справочная

		6730sT5870 2.00 ГГц +сумка (1), Проектор Мультимедийный Hitachi CP-X2 (1), Проектор мультимедийный Hitachi CP-X444 (1), Экран на штативе 180*180 (1), Экран настенный 200*200см Braun Roll Vision (1), Комплект учебной мебели (1)	правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач
3.	521 (I)	Комплект учебной мебели (1)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач
4.	522 (I)	Анализатор спектра NS-30A (1), Антенна M102 в компл. с кабелем ВЧ TNCm-SMAm (1), Блок питания лаборат. НУ 3003 D-3 (1), Внешний HDD WD 2TB 3.0 , 3.5"USB (1), Внешний накопитель 1 Seagate Original USB 3.0 4 Tb (1), Внешний накопитель флешка USB TRANSCEND Jetflash 780 64 Gb (1), Гигабитный управляемый коммутатор на 16 портов (1), Измеритель CN -801 HP (1), Кондиционер AEG ACS-09HR (1), Многофункциональный измерительный прибор (1), Монитор 20 "Beng FP 202W (2), Монитор LCD Samsung 17" SM 713N (1), МФУ Canon i-SENSYS MF 4018 (1), МФУ 1 Лазерный Canon i-Sensys MF226 (1), Набор ВЧ переходников (1), Ноутбук Dell Latitude E6520 Intel Core I5 Processor 2520M 15,6" (2), Ноутбук TOSHIBA Satellite L655-1H2-RU	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач

	(1), Паяльная станция AOYUE 968 (1), Переключатель ZX80-DR230 (1), Персональный компьютер 3 Atlant A2X4/4G(3)/512Mb/монитор Pyama 2209/3Y (1), ПК RAMEC GALE LCD LG 23"/Intel i5 4590/MSI B85M- E45/2x4DDR3/GT740 2Gb/500Gb/клав (28), Преобразователь SP-200-24-AC-DC в кожухе 199x99x50мм (1), Приемо- передающая программно- конфигурируемая радиоплатформа G32 (1), Принтер Canon LBP 2900 лазерный с кабелем (1), Проектор мультимедийный Hitachi CP-EX250 (1), Проектор мультимедийный Hitachi CP-EX251N (1), Сист. блок Pen D 945 3.4 DDR 2 1024*2/FDD 3.5/250 Gb/DVD- RW/кл+мышь+коврик (1), Систем.блок CPU Intel Core i7- 6700/ASRod Z-170/32Gb/GTX 1070/200Gb/Wi-Fi +клав, (1), Станок сверлильный 350 вт (1), Универсальная приёмо-передающая платформа для проектирования СВЧ-систем компл.mg (1), Усилитель LZY-22 (1), Усилитель ZHL-3A-S (1), Комплект учебной мебели (1)	
--	---	--

Раздел 7. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ/ ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Критерии оценивания индикаторов достижения компетенций направлены на:

- усвоение теоретического материала (объем знаний, глубина усвоения), предусмотренного рабочей программой;
- умение излагать материал (четкость, грамотность изложения материала, точность и полнота воспроизведения учебного материала);
- умение применять теоретические знания при решении практических заданий.

Шкала оценивания представлена ниже.

Уровень сформированности элементов компетенции	Критерии оценивания	Шкала оценивания
Пороговый уровень	Обучающийся имеет знания основного материала, проявляет умение логично его излагать, но может допускать неточности в изложении материала, недостаточно правильные формулировки, испытывает затруднения в выполнении практических заданий.	удовлетворительно
Продвинутый	Обучающийся твердо знает программный материал,	хорошо

уровень	излагает его грамотно и по существу, не допускает существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения	
Высокий уровень	Обучающийся глубоко и прочно усвоил программный материал, грамотно и логически стройно его излагает, дает исчерпывающие ответы на поставленные вопросы. В ответе тесно увязывается теория с практикой, при этом обучающийся не затрудняется с ответом при видоизменении задания, свободно справляется с задачами, вопросами и другими видами применения знаний, показывает знакомство с монографической литературой, периодическими изданиями, правильно обосновывает принятые решения, свободно владеет разносторонними навыками, приемами выполнения практических работ	отлично

7.1. Текущий контроль успеваемости

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины (модуля) и производится с применением технологии рейтингового контроля в соответствии с технологической картой дисциплины. Порядок составления технологической карты и алгоритм проведения процедуры оценивания видов деятельности обучающихся, направленных на освоение знаний, умений, навыков и/или опыта деятельности, по накопительной системе в баллах устанавливается положением о системе РИТМ в ФГБОУ ВО «ПГТУ»

7.2. Промежуточная аттестация обучающихся

Промежуточная аттестация обучающихся направлена на оценивание результатов обучения по дисциплине (модулю) и проводится с использованием фондов оценочных средств.

Примеры типовых контрольных заданий из базы фонда оценочных средств по образовательной программе.

1. Что является ключевой архитектурной инновацией, которая легла в основу современных LLM?

- a) Рекуррентные нейронные сети (RNN).
- b) Сверточные нейронные сети (CNN).
- c) Механизм Self-Attention (само-внимания).
- d) Многослойные перцептроны (MLP).

2. Для чего используется позиционное кодирование в архитектуре Трансформера?

- a) Для ускорения вычислений в Multi-Head Attention.
- b) Для добавления информации о порядке токенов в последовательности.
- c) Для уменьшения размерности входных эмбеддингов.

d) Для маскирования токенов в процессе Masked Language Modeling.

3. Какая из следующих моделей является примером архитектуры "Decoder-Only" (Causal LLM)?

- a) BERT
- b) T5
- c) 2
- d) ELMo

4. Какой метод предварительного обучения используется в BERT для понимания двунаправленного контекста?

- a) Causal Language Modeling (CLM).
- b) Masked Language Modeling (MLM).
- c) Next Sentence Prediction (NSP).
- d) И b, и c.

5. Что такое "Scaling Laws" (законы масштабирования) в контексте LLM?

- a) Законы, описывающие, как производительность LLM улучшается с увеличением размера модели, данных и вычислительных ресурсов.
- b) Правила для масштабирования входных данных перед подачей в LLM.
- c) Методы для уменьшения размера LLM для более эффективного развертывания.
- d) Алгоритмы для обучения LLM на распределенных системах.

6. Какова основная цель Reinforcement Learning from Human Feedback (RLHF) при обучении LLM?

- a) Увеличить размер словаря токенизатора.
- b) Выровнять поведение модели с человеческими предпочтениями и инструкциями.
- c) Ускорить процесс предварительного обучения LLM.
- d) Уменьшить количество параметров в модели.

7. Что означает "Zero-shot prompting"?

- a) Предоставление модели нескольких примеров в промпте.

- b) Генерация ответа модели без каких-либо примеров в промпте.
- c) Изменение весов модели для адаптации к новой задаче.
- d) Использование очень короткого промпта.

8. Техника промптинга, которая включает в себя просьбу к LLM показать свои рассуждения шаг за шагом перед предоставлением окончательного ответа, называется:

- a) Few-shot prompting.
- b) Chain-of-Thought (CoT) prompting.
- c) Negative prompting.
- d) Role-playing prompting.

9. Какой параметр генерации контролирует "креативность" или "случайность" вывода LLM?

- a) `max_length`.
- b) `num_beams`.
- c) `temperature`.
- d) `no_repeat_ngram_size`.

10. Почему методы Parameter-Efficient Fine-Tuning (PEFT), такие как LoRA, стали популярными?

- a) Они значительно увеличивают точность LLM по сравнению с полной донастройкой.
- b) Они позволяют донастраивать LLM, используя значительно меньше вычислительных ресурсов и памяти.
- c) Они устраняют проблему галлюцинаций в LLM.
- d) Они полностью заменяют необходимость предварительного обучения.

11. В LoRA, что представляет собой "ранг" (rank) адаптации?

- a) Количество слоев в LLM, которые донастраиваются.
- b) Размерность низкоранговых матриц, используемых для аппроксимации обновления весов.
- c) Количество эпох обучения.

d) Позиция токена в последовательности.

12. Какую проблему в основном решают системы Retrieval-Augmented Generation (RAG)?

a) Недостаточная скорость генерации текста.

b) Высокие вычислительные затраты на предварительное обучение LLM.

c) Проблема "галлюцинаций" и ограниченность знаний LLM на момент обучения.

d) Сложность настройки параметров генерации.

13. Что такое "Function Calling" или "Tool Use" для LLM-агентов?

a) Способ, которым LLM вызывает внутренние функции для оптимизации своих ответов.

b) Возможность LLM взаимодействовать с внешними API, базами данных или функциями для выполнения задач.

c) Процесс,

при котором LLM генерирует код на Python.

d) Метод для отладки ошибок в LLM.

14. Что из перечисленного является примером мультимодальной LLM?

a) BERT

b) Llama 2

c) CLIP (Contrastive Language-Image Pre-training)

d) 2

15. Какая метрика чаще всего используется для оценки языковых моделей, измеряя, насколько хорошо модель предсказывает следующий токен в последовательности?

a) BLEU score.

b) ROUGE score.

c) Frechet Inception Distance (FID).

d) Perplexity.

16. Какой из следующих бенчмарков предназначен для оценки широкого спектра знаний и способности рассуждений LLM в различных предметных областях?

- a) SQuAD (Stanford Question Answering Dataset).
- b) MMLU (Massive Multitask Language Understanding).
- c) GLUE (General Language Understanding Evaluation).
- d) SST-2 (Stanford Sentiment Treebank).

17. Что такое "галлюцинации" в контексте LLM?

- a) Ошибки в коде, приводящие к сбою модели.
- b) Генерация моделью фактов или информации, которая не соответствует действительности или не подкреплена входными данными.
- c) Визуальные эффекты, генерируемые мультимодальными LLM.
- d) Проблема, когда модель забывает ранее изученную информацию.

18. Какое этическое соображение относится к предвзятости (bias) в LLM?

- a) Модель потребляет слишком много энергии.
- b) Модель воспроизводит и усиливает стереотипы или несправедливое отношение, присутствующие в обучающих данных.
- c) Модель генерирует слишком длинные ответы.
- d) Сложность объяснения решений модели.

19. Если вам нужно донастроить Llama 2 (7B) на небольшом пользовательском датасете, но у вас есть GPU с ограниченной памятью (например, 16GB), какой метод вы, скорее всего, выберете?

- a) Полная донастройка (full fine-tuning).
- b) Предварительное обучение (pre-training) с нуля.
- c) LoRA (Low-Rank Adaptation) или QLoRA.
- d) Удаление нескольких слоев из модели, чтобы она поместилась в память.

20. Для чего в Transformer-архитектуре используются "Residual Connections" (остаточные соединения) и "Layer Normalization"?

- a) Для увеличения скорости генерации токенов.
- b) Для стабилизации обучения глубоких сетей и предотвращения проблем с исчезающим/взрывающимся градиентом.
- c) Для шифрования информации, передаваемой между слоями.
- d) Для уменьшения количества параметров модели.

Перечень вопросов для проведения промежуточной аттестации

Вопросы для промежуточной аттестации:

- Опишите ключевые этапы развития обработки естественного языка (NLP), начиная от статистических методов и заканчивая современными LLM.
- Объясните, какие проблемы ранних нейронных сетей (RNN, LSTM) были решены с появлением архитектуры Трансформера.
- Объясните принцип работы механизма **Self-Attention** (само-внимания). Как он позволяет обрабатывать зависимости между словами в последовательности?
- Что такое **Queries, Keys, Values** в механизме внимания? Объясните роль масштабирования (Scaled Dot-Product Attention).
- Почему необходим **позиционный кодирование (Positional Encoding)** в Трансформерах, и как оно работает?
- Что такое **Multi-Head Attention**? Каковы его преимущества по сравнению с Single-Head Attention?
- Опишите роль **Encoder-Decoder** и **Decoder-Only** (Causal LLM) архитектур. Приведите примеры известных моделей для каждой категории и их типичные задачи.
- Объясните, зачем нужна токенизация в LLM.
- Опишите принцип работы одного из распространенных алгоритмов токенизации (например, Byte Pair Encoding - BPE или WordPiece).
- Объясните цель предварительного обучения LLM.
- Опишите два распространенных метода предварительного обучения: **Language Modeling (Causal LM)** и **Masked Language Modeling (MLM)**. В чем их ключевые отличия?
- Какие типы данных используются для предварительного обучения LLM, и почему их качество и объем так важны?
- Что такое "законы масштабирования" в контексте LLM? Как производительность модели связана с размером модели, объемом данных и вычислительными ресурсами?
- Обсудите implications (последствия) этих законов.

Донастройка (Fine-tuning) и Обучение следовать инструкциям:

- Почему необходима донастройка LLM после предварительного обучения?
- Объясните процесс **Supervised Fine-Tuning (SFT)**, включая подготовку данных (Instruction Tuning).
- Опишите основные этапы **Reinforcement Learning from Human Feedback (RLHF)**. Какова его цель?
- Какие альтернативы RLHF существуют (например, DPO)?
- Что такое промпт-инжиниринг? Какова его важность для эффективного использования LLM?
- Объясните и сравните следующие техники промптинга: **Zero-shot, One-shot, Few-shot prompting**.
- Опишите **Chain-of-Thought (CoT) prompting** и его преимущества для решения задач, требующих рассуждений.
- Что такое "In-Context Learning"? Как LLM могут "учиться" без изменения своих весов?
- Опишите роль и влияние следующих параметров генерации текста: **Temperature, Top-K Sampling, Top-P (Nucleus Sampling), Beam Search**. Когда какой параметр предпочтительнее использовать?
- Почему полная донастройка (full fine-tuning) больших LLM часто непрактична?
- Объясните принцип работы **LoRA (Low-Rank Adaptation)**. Каковы его основные преимущества?
- Что такое **QLoRA**?
- Какую проблему решают RAG-системы?
- Опишите основные компоненты RAG-системы (ретривер, генератор). Как они взаимодействуют?
- Что такое LLM-агент? Каковы его основные способности (планирование, выбор инструмента, исполнение)?
- Объясните концепцию "Function Calling" или "Tool Use" в LLM. Приведите пример сценария, где это может быть полезно.
- Что такое мультимодальные LLM? Приведите примеры таких моделей и их приложений (например, text-to-image, image-to-text).
- Какие метрики используются для оценки качества LLM? Обсудите ограничения традиционных NLP-метрик (Perplexity, BLEU) для генеративных задач.
- Что такое **бенчмарки LLM** (например, MMLU, HELM)? Какова их цель?

- Обсудите роль **человеческой оценки** при оценке LLM.
- Назовите и объясните основные ограничения современных LLM (например, галлюцинации, предвзятость, отсутствие "понимания" мира, проблемы с рассуждениями).
- Расскажите о ключевых этических проблемах, связанных с LLM: **предвзятость (bias), токсичность, безопасность (safety), конфиденциальность, авторское право, дезинформация.**