



Федеральное государственное образовательное бюджетное учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)

Кафедра искусственного интеллекта
Факультет информационных технологий и анализа больших данных

Документ подписан усиленной неквалифицированной электронной подписью.

Организация: «ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»

Сертификат: mwVHJB6aRrbt4dLjowNLf+NVDIY6Zy2

Утверждено: Проректор по учебной и методической работе, Е.А. Каменева

Дата: 29.05.2026 г.

Е.В. Романова , П.В. Никитин

Прикладные задачи машинного обучения

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки

09.03.04 - Программная инженерия,

Образовательная программа «Инженер-разработчик программного обеспечения»

Профиль:

«Инженер-разработчик программного обеспечения»

Рекомендовано

*Факультет информационных технологий и анализа больших данных
(протокол № 08 от 19.05.2026 г.)*

Одобрено

*Кафедра искусственного интеллекта
(протокол № 4 от 06.05.2026 г.)*



Содержание

1. Наименование дисциплины	3
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	3
3. Место дисциплины в структуре образовательной программы	7
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	7
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	8
5.1. Содержание дисциплины	8
5.2. Учебно – тематический план	11
5.3. Содержание семинаров, практических занятий	12
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	14
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	14
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю	19
7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	20
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	7
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	9
10. Методические указания для обучающихся по освоению дисциплины	36
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)	38
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	38



1. Наименование дисциплины

Прикладные задачи машинного обучения

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКП-1	Способность описывать, анализировать и проектировать интерфейс программных модулей с учетом требований к ним	Демонстрирует знания основных понятий интерфейсов программных модулей, понятие внешней и внутренней среды, читает и понимает готовую программную документацию в части описания интерфейсов	Знать: основные понятия интерфейсов программных модулей (API, SDK, ABI), различия между внешней средой (операционная система, драйверы, другие приложения, сетевые протоколы) и внутренней средой (ядро программы, библиотеки, внутренние структуры данных и функции) программной системы, а также структуру и назначение технической документации по описанию интерфейсов Уметь: читать и интерпретировать техническую документацию на программные библиотеки и API (включая описания функций, передаваемых и возвращаемых параметров, форматов данных, кодов ошибок и исключений), и применять полученную информацию для интеграции программных модулей
		Понимает достоинства и недостатки различных архитектурных решений в области проектирования интерфейсов программных модулей, мо-	Знать: достоинства и недостатки различных архитектурных решений при проектировании интерфейсов программных модулей, включая монолитный API, микросервисную архитектуру, REST, gRPC, асинхронные очереди



Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
		жет критически анализировать существующие решения	сообщений (например, RabbitMQ, Kafka) и интеграцию на уровне библиотек/пакетов Уметь: критически анализировать существующие интерфейсные решения в контексте конкретной прикладной задачи машинного обучения, выявлять их «узкие места» с точки зрения производительности, масштабируемости, безопасности и удобства сопровождения, и предлагать пути рефакторинга или замены архитектуры
		Описывает интерфейс программной системы в формализованном виде по определенным стандартам, демонстрирует знания общепринятых стандартов описания архитектуры программной системы	Знать: общепринятые стандарты и нотации для формализованного описания архитектуры программных систем и интерфейсов (UML, OpenAPI/Swagger для REST API, AsyncAPI для асинхронных интерфейсов, gRPC/proto3 для gRPC-сервисов, C4 model), а также требования к структуре и содержанию архитектурной документации Уметь: описывать интерфейс программной системы в формализованном виде с использованием выбранного стандарта (например, спецификация OpenAPI для REST API, диаграммы классов и последовательностей UML), фиксируя входные/выходные данные, форматы сообщений, протоколы взаимодействия и зависимости между компонентами
		Проектирует интерфейс программного модуля с учетом требований к программной системе в целом и с	Знать: принципы проектирования интерфейсов программных модулей с учётом системных требований (производительность, масштабируемость, безопасность, отказоустойчивость) и требований к интеграции с другими



Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
		учетом интеграции с другими программными модулями	модулями (совместимость форматов данных, версионирование API, обработка ошибок, асинхронное взаимодействие) Уметь: разрабатывать спецификацию интерфейса программного модуля, обеспечивающую его бесшовную интеграцию с существующими или проектируемыми компонентами программной системы, выбирать оптимальные протоколы обмена и форматы данных, а также предусматривать механизмы обработки исключительных ситуаций и обеспечения обратной совместимости
ОПК-3	Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	Проводит самостоятельный поиск информации в открытых источниках по определенной заданной тематике	Знать: современные открытые источники научно-технической информации в области анализа данных и машинного обучения (arXiv, Google Scholar, IEEE Xplore, официальные репозитории Kaggle и GitHub), методы и критерии оценки релевантности и достоверности найденных материалов Уметь: формулировать поисковые запросы для сбора информации по заданной теме, осуществлять поиск и критический анализ найденных источников, структурировать отобранные материалы для решения поставленной прикладной задачи
		Проводит систематический обзор источников информации, анализирует содержащиеся в них данные,	Знать: методологию проведения систематического обзора литературы (формулировка исследовательского вопроса, критерии включения/исключения источников, оценка качества и достоверности данных, методы синтеза инфор-



Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
		делает и обосновывает выводы на основе проведенного обзора	мации из различных источников) Уметь: систематизировать найденные источники по заданным критериям, сопоставлять и критически анализировать противоречивые данные из разных источников, выявлять общие закономерности и формулировать аргументированные выводы по результатам обзора
		Демонстрирует знания основных требований информационной безопасности, основных алгоритмов защиты информации, в том числе с использованием криптографических протоколов	Знать: основные требования информационной безопасности при обработке конфиденциальных данных в системах машинного обучения, базовые алгоритмы защиты информации (гомоморфное шифрование, дифференциальная приватность, безопасная агрегация) и основы применения криптографических протоколов (SSL/TLS, Private Set Intersection) для обеспечения конфиденциальности при обучении и инференсе моделей Уметь: выбирать и обосновывать методы защиты информации (федеративное обучение, гомоморфное шифрование, безопасные многосторонние вычисления) в зависимости от специфики решаемой задачи машинного обучения и требований к конфиденциальности данных



3. Место дисциплины в структуре образовательной программы

Дисциплина «Прикладные задачи машинного обучения» относится к «Модулю "Технологии машинного обучения"».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

очно-заочная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 8 (в часах)
Общая трудоемкость дисциплины	3/108	108
Контактная работа – Аудиторные занятия	24	24
<i>Лекции</i>	<i>8</i>	<i>8</i>
<i>Семинары, практические занятия</i>	<i>16</i>	<i>16</i>
<i>Самостоятельная работа</i>	<i>84</i>	<i>84</i>
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Зачет	Зачет



5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Методы оптимизации для больших данных и распределенных систем

Введение в распределенное обучение. Архитектуры распределенного обучения: Модели Master-Slave; Peer-to-Peer модели. Методы агрегации градиентов: синхронная vs асинхронная агрегация; преимущества и недостатки каждого подхода. Обработка больших данных: инструменты для работы с большими данными (Hadoop, Spark); влияние больших данных на выбор методов оптимизации. Обучение на графических процессорах (GPU): оптимизация алгоритмов для GPU; параллелизация вычислений. Методы распределенного SGD: алгоритмы для распределенного SGD; сравнение с традиционными методами. Масштабируемость методов оптимизации: проблемы масштабируемости и их решения; подходы к улучшению производительности на больших данных; применение распределенных методов в реальных задачах; Будущее распределенного обучения: новые тренды и направления исследований.

Тема 2. Методы распределенного и федеративного обучения

Введение в распределенное обучение: определение и основные концепции распределенного обучения; отличия между распределенным и централизованным обучением. Типы распределенного обучения: параллельное обучение: данные и вычисления; модельное параллельное обучение; данные параллельное обучение. Архитектуры распределенного обучения: Master-Slave архитектура; Peer-to-Peer архитектура; микросервисные архитектуры. Федеративное обучение: определение и принципы федеративного обучения; особенности федеративного обучения по сравнению с традиционным обучением. Применения федеративного обучения: примеры использования в мобильных устройствах; применение в здравоохранении и финансовом секторе. Алгоритмы федеративного обучения: основные алгоритмы, такие как Federated Averaging (FedAvg) и его варианты; обсуждение улучшений и оптимизаций алгоритмов. Безопасность и конфиденциальность: вопросы безопасности данных в распределенном и федеративном обучении; методы обеспечения конфиденциальности, такие как дифференциальная приватность. 7 Вызовы и ограничения: проблемы с коммуникацией и задержками; неоднородность данных и вычислительных мощностей; сложности с синхронизацией моделей. Оценка



производительности: метрики для оценки эффективности распределенного и федеративного обучения; сравнение с традиционными методами обучения. Будущее распределенного и федеративного обучения: тенденции и перспективы развития технологий; возможные направления исследований и применения.

Тема 3. Большие языковые модели (LLM) и их архитектуры

Архитектуры LLM: обзор архитектур трансформеров; ключевые компоненты, такие как механизмы внимания и слои. Обучение больших языковых моделей: методы обучения, включая обучение с учителем и без учителя; использование предобучения и дообучения. Предобучение и дообучение: стратегии предобучения на больших объемах данных; адаптация моделей к специфическим задачам через дообучение. Тонкая настройка (fine-tuning): подходы к тонкой настройке моделей для конкретных приложений; влияние объема данных на эффективность тонкой настройки. Оценка производительности LLM: метрики для оценки качества генерации текста и понимания языка; сравнение с другими моделями. Этика и ответственность в использовании LLM: проблемы предвзятости, конфиденциальности и безопасности; обсуждение последствий использования LLM в обществе. Применения больших языковых моделей: примеры использования в чат-ботах, системах автоматического перевода и генерации контента. Тенденции и будущее LLM: новые направления исследований и разработки; прогнозы по эволюции архитектур и методов обучения. Интеграция LLM в бизнес-процессы: возможности применения в различных отраслях; влияние на автоматизацию и оптимизацию рабочих процессов.

Тема 4. Мульти-модальные модели и их интеграция

Введение в мульти-модальные модели: определение и основные характеристики мульти-модальных подходов; их роль в интеграции различных типов данных (текст, изображения, звук и т.д.). 8 Архитектуры мульти-модальных моделей: обзор архитектур, поддерживающих работу с несколькими модальностями; ключевые компоненты, такие как механизмы внимания и совместное представление данных. Обучение мульти-модальных моделей: методы обучения, включая совместное обучение и предобучение на различных модальностях; использование больших объемов данных из разных источников. Предобучение и дообучение мульти-модальных моделей: стратегии предобучения на многомодальных данных; адаптация моделей к специфическим задачам через дообучение на целевых модальностях. Интеграция данных из разных модальностей: подходы к объединению информации из текста, изображений и звука; методы синхронизации и согласования данных. Оценка производительности мульти-модальных моделей: метрики для оценки качества генерации и понимания данных из различных модальностей; сравнение с



однородными моделями. Применения мульти-модальных моделей: примеры использования в системах распознавания объектов, автоматическом переводе, генерации контента и других областях. Этика и ответственность в использовании мульти-модальных моделей: проблемы предвзятости, конфиденциальности и безопасности; обсуждение последствий использования таких моделей в обществе. Тенденции и будущее мульти-модальных моделей: новые направления исследований и разработки; прогнозы по эволюции архитектур и методов обучения. Интеграция мульти-модальных моделей в бизнес-процессы: возможности применения в различных отраслях; влияние на автоматизацию, анализ данных и пользовательский опыт.

Тема 5. Современные исследования в машинном обучении

Обзор литературы по актуальным направлениям машинного обучения: анализ последних публикаций в ведущих научных журналах; обзор ключевых конференций и их материалов (NeurIPS, ICML, CVPR и др.); тенденции в исследованиях: от классических алгоритмов к глубокому обучению. Предпринты и открытые исследования: изучение платформ, таких как arXiv, для поиска новейших исследований; обсуждение значимых предстоящих работ и их потенциального влияния на область; критический анализ методологий и результатов из предпринтов. 9 Практические примеры применения машинного обучения: кейсы успешного применения машинного обучения в различных отраслях (медицина, финансы, маркетинг, производство); разбор реальных проектов и стартапов, использующих машинное обучение для решения конкретных задач; влияние машинного обучения на бизнес-процессы и принятие решений. Инновационные технологии и инструменты: обзор новых фреймворков и библиотек для разработки моделей машинного обучения; исследование новых подходов к интерпретации моделей; разработка и использование автоматизированных систем для машинного обучения (AutoML). Этика и социальные аспекты применения машинного обучения: обсуждение проблем предвзятости в алгоритмах и их влияние на результаты; вопросы конфиденциальности данных и соблюдения этических норм; роль машинного обучения в решении социальных проблем. Будущее машинного обучения: тренды и перспективы: прогнозы по развитию технологий и методов машинного обучения; обсуждение междисциплинарных исследований и их влияния на область; потенциал интеграции с другими технологиями (блокчейн, IoT, квантовые вычисления).



5.2. Учебно – тематический план

Очно-заочная форма обучения

п/ п	Наименование тем (разделов) дисциплины	Трудоемкость в часах				
		Всего	Контактная работа - Аудиторная работа			Самостоятельная работа
			Общая, в т.ч.:	Лекции	Семинары, практические занятия	
1	Методы оптимизации для больших данных и распределенных систем	21	6	2	4	15
2	Методы распределенного и федеративного обучения	23	6	2	4	17
3	Большие языковые модели (LLM) и их архитектуры	21	6	2	4	15
4	Мульти-модальные модели и их интеграция	20	3	1	2	17
5	Современные исследования в машинном обучении	23	3	1	2	20
В целом по дисциплине		108	24	8	16	84



5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Методы оптимизации для больших данных и распределенных систем	Методы оптимизации в машинном обучении, включая градиентный спуск и его стохастические варианты. Преимущества и недостатки различных подходов. Пример задания: оптимизация гиперпараметров моделей с использованием Bayesian Optimization и разработка адаптивных методов обучения для глубоких нейронных сетей.	Интерактивная форма, практикум в малых группах (2–4 студента), коллективное обсуждение решений. Семинары с ИТ-практиками, конкурсы, чемпионаты, мастер-классы, командные проектные задания.
Методы распределенного и федеративного обучения	Концепции распределенного и федеративного обучения, их преимущества и недостатки. Подходы к обеспечению безопасности и конфиденциальности данных. Пример задания: создание системы для обучения моделей на мобильных устройствах с использованием федеративного подхода и разработка распределенных систем для обработки больших объемов данных в облачных средах.	Интерактивная форма, практикум в малых группах (2–4 студента), коллективное обсуждение решений. Семинары с ИТ-практиками, конкурсы, чемпионаты, мастер-классы, командные проектные задания.
Большие языковые модели (LLM) и их архитектуры	Архитектуры больших языковых моделей, таких как Transformer, и их применение в обработке текста. Проблемы, связанные с обучением и использованием LLM, включая предвзятость и потребление ресурсов. Пример задания: разработка чат-ботов на основе LLM для поддержки клиентов и создание систем автоматического перевода текста с использованием больших языковых моделей.	Интерактивная форма, практикум в малых группах (2–4 студента), коллективное обсуждение решений. Семинары с ИТ-практиками, конкурсы, чемпионаты, мастер-классы, командные проектные задания.



Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Мульти-модальные модели и их интеграция	Интеграция различных типов данных, таких как текст, изображения и звук, для решения сложных задач. Вызовы, возникающие при обучении мультимодальных моделей, а также способы оценки их производительности. Пример задания: разработка системы распознавания объектов на изображениях с текстовыми описаниями и мультимедийные поисковые движки, которые комбинируют текстовые и визуальные данные.	Интерактивная форма, практикум в малых группах (2–4 студента), коллективное обсуждение решений. Семинары с ИТ-практиками, конкурсы, чемпионаты, мастер-классы, командные проектные задания.
Современные исследования в машинном обучении	Текущие тренды в исследованиях машинного обучения, проблемы предвзятости и интерпретируемости моделей. Новые методы и подходы, предложенные в недавних публикациях. Пример задания: исследование по улучшению интерпретируемости глубоких нейронных сетей и разработка новых алгоритмов машинного обучения для применения в области экономики и финансов.	Интерактивная форма, практикум в малых группах (2–4 студента), коллективное обсуждение решений. Семинары с ИТ-практиками, конкурсы, чемпионаты, мастер-классы, командные проектные задания.



6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Методы оптимизации для больших данных и распределенных систем	1. В чём заключается компромисс между синхронной и асинхронной агрегацией градиентов при распределённом обучении? Какие факторы влияют на выбор подхода? 2. Как устроен метод эластичного усреднения (Elastic Averaging SGD) и за счёт чего он позволяет улучшить сходимость по сравнению с классическим распределённым SGD? 3. Какие проблемы масштабируемости возникают при обучении моделей машинного обучения на данных объёмом более 1 ТБ, и какие существуют подходы к их решению? 4. Как влияние архитектуры GPU-памяти (Shared Memory vs. Global Memory) на эффективность параллелизации вычислений градиента?	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашнего задания.
Методы распределенного и федеративного обучения	1. В чём заключается различие между горизонтальным и вертикальным федеративным обучением? Приведите примеры задач, для которых подходит каждый из подходов. 2. Как работает алгоритм Federated Averaging (FedAvg) и почему усреднение весов модели на сервере	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашнего задания.



Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
	эквивалентно усреднению градиентов при независимых идентично распределённых данных? 3. Какие дополнительные механизмы безопасности (помимо дифференциальной приватности) применяются в федеративном обучении для защиты от атак восстановления данных (например, атак по градиентам)? 4. Как проблема неоднородности данных (non-IID данные) влияет на сходимость федеративного обучения и какие существуют алгоритмы для её смягчения (например, FedProx, SCAFFOLD)?	
Большие языковые модели (LLM) и их архитектуры	1. В чём отличие механизмов Grouped-Query Attention (GQA) и Multi-Query Attention (MQA) от стандартного Multi-Head Attention, и как они влияют на скорость инференса LLM? 2. Почему в больших языковых моделях возникает проблема «галлюцинаций» (hallucinations), и какие подходы к её решению существуют (например, RLHF, chain-of-thought, retrieval-augmented generation)? 3. Как работает механизм KV-кэширования при авторегрессивной генерации текста в трансформерных декодерах, и почему он критически важен для производительности LLM? 4. Каковы принципы работы методов квантизации (4-bit, 8-bit) для больших языковых моделей (например, GPTQ,	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашнего задания.



Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
	AWQ, bitsandbytes) и как они влияют на точность предсказаний?	
Мульти-модальные модели и их интеграция	1. Какие существуют стратегии объединения (fusion) информации из разных модальностей (раннее, позднее, гибридное), и каковы их преимущества и недостатки в задачах типа VQA (Visual Question Answering)? 2. Как работает механизм кросс-модального внимания в архитектурах типа Flamingo или LLaVA, и как он позволяет согласовывать текстовые и визуальные представления? 3. В чём заключается проблема выравнивания модальностей (modality alignment) при обучении мультимодальных моделей, и как она решается с использованием контрастивных методов (например, CLIP, ALIGN)? 4. Какие способы синхронизации потоков данных из разных модальностей (аудио, видео, текст) во времени существуют, и как обрабатываются ситуации асинхронного поступления данных в реальных системах?	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашнего задания.
Современные исследования в машинном обучении	1. Каковы основные методологические критерии критической оценки качества статей на ведущих конференциях (NeurIPS, ICML, ICLR) при проведении систематического обзора литературы?	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашнего задания.



Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
	2. В чём отличие препечатных платформ (arXiv, bioRxiv) от рецензируемых журналов с точки зрения скорости распространения знаний и контроля качества, и как правильно интерпретировать результаты из предпринтов? 3. Какие направления исследований в области машинного обучения сейчас являются наиболее активно развивающимися согласно анализу публикаций за последние 2–3 года (например, агентные системы, диффузионные модели, LLM-агенты)? 4. Как междисциплинарные исследования на стыке машинного обучения и квантовых вычислений могут повлиять на развитие оптимизационных методов в ближайшие 5–10 лет?	



6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные вопросы для подготовки к контрольной работе

1. В чём отличие синхронной и асинхронной агрегации градиентов? Назовите преимущества и недостатки каждого подхода.
2. Какие инструменты для работы с большими данными (Hadoop, Spark) используются при распределённом обучении?
3. Как оптимизируются алгоритмы для обучения на GPU? Что такое параллелизация вычислений?
4. В чём заключается проблема масштабируемости методов оптимизации и каковы подходы к её решению?
5. Что такое федеративное обучение? В чём его отличие от традиционного централизованного обучения?
6. Опишите алгоритм Federated Averaging (FedAvg). Как он работает?
7. Какие методы обеспечения конфиденциальности используются в федеративном обучении?
8. Назовите основные вызовы распределённого обучения: проблемы коммуникации, задержек и неоднородности данных.
9. Опишите архитектуру трансформера. Какие ключевые компоненты входят в его состав?
10. В чём разница между предобучением (pre-training) и дообучением (fine-tuning) LLM?
11. Какие метрики используются для оценки качества генерации текста?
12. Назовите основные этические проблемы при использовании LLM (предвзятость, конфиденциальность, безопасность).
13. Что такое мультимодальная модель? Приведите примеры типов данных, которые она может обрабатывать.
14. Как происходит объединение (fusion) информации из разных модальностей?
15. Назовите примеры применения мультимодальных моделей в реальных задачах.
16. В чём отличие мультимодального обучения от обучения на однородных данных?
17. Назовите ведущие конференции и журналы в области машинного обучения.
18. Что такое препринты? Какие платформы используются для их публикации?
19. Приведите примеры успешного применения машинного обучения в финансах или медицине.
20. Что такое AutoML? Для каких задач он применяется?



Примерные задания контрольной работы

1. Реализовать синхронную и асинхронную версии распределённого SGD для простой модели логистической регрессии и сравнить время сходимости.
2. Реализовать алгоритм FedAvg для задачи классификации на разделённых данных (эмуляция федеративного обучения с 3 клиентами).
3. Выполнить тонкую настройку (fine-tuning) модели BERT на датасете IMDB для классификации тональности и вычислить ассурасу.
4. Использовать мультимодальную модель CLIP для поиска изображений по текстовому запросу на датасете из 100 изображений.
5. Подготовить краткий обзор (2–3 страницы) трёх свежих статей с arXiv по одной из тем: распределённое обучение, LLM или мультимодальные модели.

Дополнительная информация

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры искусственного интеллекта Факультета информационных технологий и анализа больших данных.



7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
ПКП-1. Способность описывать, анализировать и проектировать интерфейс программных модулей с учетом требований к ним	Демонстрирует знания основных понятий интерфейсов программных модулей, понятие внешней и внутренней среды, читает и понимает готовую программную документацию в части описания интерфейсов	Знать: основные понятия интерфейсов программных модулей (API, SDK, ABI), различия между внешней средой (операционная система, драйверы, другие приложения, сетевые протоколы) и внутренней средой (ядро программы, библиотеки, внутренние	Изучить официальную документацию API библиотеки scikit-learn (интерфейсы классов BaseEstimator и TransformerMixin), объяснить назначение методов fit(), transform() и predict(), описать их входные и выходные параметры, структуру возвращаемых данных, а также написать код-пример, демонстрирующий корректную и некорректную (с генерацией исключений) передачу параметров при использовании этих интерфейсов.



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		структуры данных и функции) программной системы, а также структуру и назначение технической документации по описанию интерфейсов Уметь: читать и интерпретировать техническую документацию на программные библиотеки и API (включая описания функций, передаваемых и возвращаемых параметров, форматов данных, кодов ошибок и исключений), и применять полученную информацию для интегра-	



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		ции программных модулей	
	Понимает достоинства и недостатки различных архитектурных решений в области проектирования интерфейсов программных модулей, может критически анализировать существующие решения	Знать: достоинства и недостатки различных архитектурных решений при проектировании интерфейсов программных модулей, включая монолитный API, микросервисную архитектуру, REST, gRPC, асинхронные очереди сообщений (например, RabbitMQ, Kafka) и интеграцию на уровне библиотек/пакетов Уметь: критически анализировать существующие интер-	Изучить архитектуру системы конвейерной обработки данных для чат-бота, где модуль подготовки признаков (Python-скрипт) и модель классификации интенгов (развёрнутая через Flask REST API) взаимодействуют синхронно «запрос-ответ». Провести критический анализ этой архитектуры, выявив её достоинства (простота развёртывания) и недостатки (проблемы с масштабированием, возможные блокировки на этапе подготовки признаков), а затем предложить альтернативные архитектурные решения (например, перевести часть обработки в асинхронные очереди с помощью брокера сообщений, перейти на gRPC для уменьшения задержек) и обосновать применимость каждого из них.



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		фейсные решения в контексте конкретной прикладной задачи машинного обучения, выявлять их «узкие места» с точки зрения производительности, масштабируемости, безопасности и удобства сопровождения, и предлагать пути рефакторинга или замены архитектуры	
	Описывает интерфейс программной системы в формализованном виде по определенным стандартам, демонстрирует знания обще-	Знать: общепринятые стандарты и нотации для формализованного описания архитектуры программных систем и интерфейсов (UML,	Разработать и представить в формализованном виде описание интерфейса программной системы для сервиса прогнозирования налоговых рисков: используя стандарт OpenAPI 3.0, специфицировать REST-эндпоинты для отправки данных о контрагенте и получения прогноза (входные/выходные JSON-схемы, форматы ошибок, параметры аутентификации), а с помощью диаграммы компонентов UML описать архитектуру системы, включая модуль предобработки данных, модель машинного обучения (XGBoost) и модуль генерации отчётов.



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
	принятых стандартов описания архитектуры программной системы	OpenAPI/Swagger для REST API, AsyncAPI для асинхронных интерфейсов, gRPC/proto3 для gRPC-сервисов, C4 model), а также требования к структуре и содержанию архитектурной документации Уметь: описывать интерфейс программной системы в формализованном виде с использованием выбранного стандарта (например, спецификация OpenAPI для REST API, диаграммы классов и по-	



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		следовательностей UML), фиксируя входные/выходные данные, форматы сообщений, протоколы взаимодействия и зависимости между компонентами	
	Проектирует интерфейс программного модуля с учетом требований к программной системе в целом и с учетом интеграции с другими программными модулями	Знать: принципы проектирования интерфейсов программных модулей с учётом системных требований (производительность, масштабируемость, безопасность, отказоустойчивость) и требований к интеграции с другими модулями (совместимость фор-	Разработать проект интерфейса модуля предобработки данных для системы кредитного скоринга, интегрируемого между модулем сбора данных (выдаёт сырые данные клиентов в формате JSON) и модулем обучения модели (ожидает нормализованные и закодированные признаки в формате CSV). Спроектировать входной и выходной форматы интерфейса, предложить механизмы валидации входных данных и обработки ошибок (отсутствие обязательных полей, некорректные типы), а также специфицировать способ интеграции модуля (REST API или вызов библиотечной функции) с обоснованием выбранного решения с учётом требований к производительности и отказоустойчивости всей системы.



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		матов данных, версионирование API, обработка ошибок, асинхронное взаимодействие) Уметь: разрабатывать спецификацию интерфейса программного модуля, обеспечивающую его бесшовную интеграцию с существующими или проектируемыми компонентами программной системы, выбирать оптимальные протоколы обмена и форматы данных, а также предусматривать механизмы обработки	



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		исключительных ситуаций и обеспечения обратной совместимости	
ОПК-3. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с уче-	Проводит самостоятельный поиск информации в открытых источниках по определенной заданной тематике	Знать: современные открытые источники научно-технической информации в области анализа данных и машинного обучения (arXiv, Google Scholar, IEEE Xplore, официальные репозитории Kaggle и GitHub), методы и критерии оценки релевантности и достоверности найденных материалов Уметь: формулировать поисковые за-	Провести самостоятельный поиск и анализ не менее пяти актуальных научных статей или публикаций (за последние 3 года) по теме «Прогнозирование временных рядов в финансовой сфере с использованием методов машинного обучения», подготовить краткий обзор-реферат с указанием использованных в работах методов, датасетов, достигнутых метрик и выводов о применимости подходов для конкретной бизнес-задачи.



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
том основных требований информационной безопасности		просы для сбора информации по заданной теме, осуществлять поиск и критический анализ найденных источников, структурировать отобранные материалы для решения поставленной прикладной задачи	
	Проводит систематический обзор источников информации, анализирует содержащиеся в них данные, делает и обосновывает выводы на основе проведенного обзора	Знать: методологию проведения систематического обзора литературы (формулировка исследовательского вопроса, критерии включения/исключения источников, оценка качества и достоверности дан-	Провести систематический обзор не менее семи источников (научных статей, технических отчётов, кейсов) по теме «Методы выявления мошеннических операций с банковскими картами», сгруппировать их по используемым алгоритмам (логистическая регрессия, случайный лес, градиентный бустинг, нейронные сети), сравнить достигнутые показатели качества (AUC-ROC, F1-score) и сформулировать обоснованный вывод о наиболее эффективных подходах для внедрения в реальной системе фрод-мониторинга.



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		ных, методы синтеза информации из различных источников) Уметь: систематизировать найденные источники по заданным критериям, сопоставлять и критически анализировать противоречивые данные из разных источников, выявлять общие закономерности и формулировать аргументированные выводы по результатам обзора	
	Демонстрирует знания основных требований информационной безопасности, основных алгоритмов	Знать: основные требования информационной безопасности при обработке конфиденциальных данных	Разработать архитектуру системы конфиденциального машинного обучения для банковского скоринга с использованием федеративного подхода, при котором модель обучается на распределённых данных трёх банков без передачи персональных данных между ними. Обосновать выбор криптографических протоколов (гомоморфное шифро-



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
	защиты информации, в том числе с использованием криптографических протоколов	в системах машинного обучения, базовые алгоритмы защиты информации (гомоморфное шифрование, дифференциальная приватность, безопасная агрегация) и основы применения криптографических протоколов (SSL/TLS, Private Set Intersection) для обеспечения конфиденциальности при обучении и инференсе моделей Уметь: выбирать и обосновывать методы защиты информации	вание PSI для синхронизации клиентов, SSL для защиты каналов связи) и описать, как обеспечивается выполнение требований законодательства о персональных данных.



Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
		(федеративное обучение, гомоморфное шифрование, безопасные многосторонние вычисления) в зависимости от специфики решаемой задачи машинного обучения и требований к конфиденциальности данных	



Примеры практико-ориентированных заданий

1. Реализовать распределённое обучение модели линейной регрессии на данных, разделённых на 4 части, с использованием синхронного усреднения градиентов и сравнить время обучения с централизованным вариантом.
2. Эмулировать федеративное обучение на датасете MNIST, разделённом между 5 клиентами с неоднородным распределением классов (каждому клиенту достаются только 2 цифры), и оценить качество модели после каждого раунда.
3. Выполнить дообучение модели Llama 3 8B с использованием QLoRA на небольшом датасете инструкций (50 примеров) и оценить качество сгенерированных ответов с помощью BERTScore.
4. Использовать мультимодальную модель CLIP для построения системы поиска изображений по текстовому запросу на датасете из 500 изображений и вычислить precision@10 .
5. Провести критический анализ трёх недавних статей с arXiv по теме федеративного обучения, сравнив методологии, датасеты и полученные результаты.

Примеры вопросов для подготовки к промежуточной аттестации

Примерные вопросы для подготовки к зачету

1. В чём отличие синхронной и асинхронной агрегации градиентов в распределённом обучении?
2. Какие инструменты для работы с большими данными используются в распределённом машинном обучении?
3. Что такое федеративное обучение? Назовите его основные принципы.
4. Опишите алгоритм Federated Averaging (FedAvg). Как он работает?
5. Какие методы обеспечения конфиденциальности используются в федеративном обучении?
6. Назовите основные вызовы и ограничения распределённого и федеративного обучения.
7. Опишите архитектуру трансформера. Какие ключевые компоненты входят в её состав?
8. В чём разница между предобучением (pre-training) и дообучением (fine-tuning) LLM?
9. Что такое LoRA и QLoRA? Для чего они применяются при дообучении больших моделей?
10. Какие метрики используются для оценки качества генерации текста LLM?
11. Назовите этические проблемы при использовании больших языковых моделей.
12. Что такое мультимодальная модель? Приведите примеры архитектур.



13. Как происходит объединение (fusion) информации из разных модальностей?
14. Назовите примеры применения мультимодальных моделей в реальных задачах.
15. Что такое CLIP? Для решения каких задач используется эта модель?
16. Назовите ведущие конференции и журналы в области машинного обучения.
17. Что такое препринты? Какие платформы используются для их публикации?
18. Приведите примеры успешного применения машинного обучения в финансах или медицине.
19. Что такое AutoML? Для каких задач он применяется?
20. Каковы современные тренды и перспективы развития распределённого обучения и LLM?

**Промежуточная аттестация обучающихся проводится в формате тестирования*



8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Андриянов, Н. А. Построение и оценка моделей машинного обучения. 01.04.02 «Прикладная математика и информатика», всех профилей (программы подготовки магистров) [Электронный ресурс] / Андриянов Н. А., Никитин П. В. Москва : Финансовый университет, 2023 140 с. Книга из коллекции Финансовый университет - Информатика <https://e.lanbook.com/book/345374>. [БИК ID: RU-LAN-BOOK-345374]
2. Практикум по машинному обучению на Python / Коротеев М. В., Одинцова В. А., Плешакова Е. С. Ч. 2 : Практикум по машинному обучению на Python. Часть 2 : учебное пособие по дисциплине: «Машинное обучение» для студентов, обучающихся по направлениям: 01.03.02 «Прикладная математика и информатика», 09.03.03 «Прикладная информатика», всех профилей (программы подготовки бакалавров) . Ч. 2 / Коротеев М. В., Одинцова В. А., Плешакова Е. С. Москва : Финансовый университет, 2024 77 с. Книга из коллекции Финансовый университет - Информатика <https://e.lanbook.com/book/431105>. [БИК ID: RU-LAN-BOOK-431105]
3. Коротеев, М.В. Практикум по машинному обучению на Python : учебное пособие по дисциплине "Машинное обучение" для студентов, обучающихся по направлению: "Прикладная информатика", всех профилей (программы подготовки бакалавров), институт онлайн-образования / М.В. Коротеев, В.А. Одинцова, Е.С. Плешакова ; Финуниверситет, Департамент анализа данных и машинного обучения факультета информационных технологий и анализа больших данных Москва : Финуниверситет, 2023 1 файл (1,64 Мб) : ил. Доступ из локальной сети Финуниверситета (чтение) <http://elib.fa.ru/rbook/books137315.pdf> (дата обращения 16.10.2023) + Текст : электронный. [БИК ID: RU\FA\books\137315]

Дополнительная литература:

1. Никитин, П.В. Машинное обучение с подкреплением : Учебник / П.В. Никитин, С.А. Корчагин, Е.В. Романова Электрон. дан. Москва : КноРус, 2026 236 с. Режим доступа: book.ru Internet access <https://book.ru/book/959175> ISBN 978-5-406-15268-3. [БИК ID: RU\bookru\bibl\959175]

Нормативные правовые акты:

-



9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
2. Электронно-библиотечная система "Университетская библиотека ОНЛАЙН" <http://biblioclub.ru/>
3. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
4. Электронно-библиотечная система Znanium <http://www.znanium.com>
5. Электронно-библиотечная система издательства "Лань" <https://e.lanbook.com/>
6. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
7. • Электронные продукты издательства Elsevier <http://www.sciencedirect.com>
8. • Коллекция научных журналов Oxford University Press <https://academic.oup.com/journals/>
9. • Электронные коллекции книг и журналов издательства Springer: <http://link.springer.com/>
10. Электронная коллекция книг издательства Springer: Springer eBooks <http://link.springer.com/>
11. Applied Machine Learning / Microsoft. – <https://www.edx.org/course/applied-machine-learning-microsoft-dat203-3x-1>
12. Data Science Essentials / Microsoft. – <https://www.edx.org/course/data-science-essentials-microsoft-dat203-1x-3>
13. Профессиональный ресурс по машинному обучению. – <https://machinelearning.ru>
14. Специализация «Машинное обучение и анализ данных» / МФТИ и Яндекс. – <https://www.coursera.org/specializations/machine-learning-data-analysis>



10. Методические указания для обучающихся по освоению дисциплины

Лекционные занятия направлены на формирование теоретических основ прикладного анализа данных и машинного обучения, а также на развитие понимания их применения при решении практических задач. В процессе освоения лекционного материала обучающимся рекомендуется систематически посещать занятия и вести конспект, отражающий ключевые понятия, методы и подходы, рассматриваемые в рамках дисциплины. Особое внимание следует уделять взаимосвязи теоретических моделей с прикладными задачами, в том числе задачами, возникающими в бизнес-практике. Важным элементом освоения дисциплины является анализ примеров применения методов машинного обучения в различных предметных областях, что позволяет сформировать целостное представление о возможностях и ограничениях используемых подходов. Обучающимся также рекомендуется самостоятельно изучать дополнительные источники информации по рассматриваемым темам, а возникающие вопросы и затруднения выносить на обсуждение в рамках практических занятий. Особое внимание в ходе изучения лекционного материала следует уделять вопросам постановки задач машинного обучения, интерпретации результатов моделирования, а также анализу взаимосвязи между качеством исходных данных, выбором методов и итоговой эффективностью решений. Существенную роль играет понимание места машинного обучения в процессе поддержки принятия решений. Практические занятия по дисциплине организованы по проектному принципу и направлены на формирование устойчивых навыков решения прикладных задач машинного обучения. В ходе выполнения практических работ обучающиеся последовательно проходят основные этапы разработки решений, включая постановку задачи, подготовку и анализ данных, построение моделей, оценку их качества и интерпретацию полученных результатов. Особое значение придается активному участию обучающихся в анализе прикладных кейсов, что способствует развитию способности применять теоретические знания на практике. В процессе выполнения заданий рекомендуется использовать современные инструментальные средства обработки данных и реализации моделей машинного обучения, уделяя внимание корректности выполнения всех этапов работы, а также качеству получаемых результатов. Результаты выполнения практических заданий должны оформляться в виде отчетов, содержащих описание поставленной задачи, обоснование выбранных методов, анализ полученных результатов и выводы о возможности их практического применения. В рамках самостоятельной работы обучающимся рекомендуется дорабатывать выполненные задания, проводить дополнительный анализ данных и моделей, а также изучать альтернативные подходы к решению поставленных задач. Практическая



направленность дисциплины предполагает необходимость регулярной работы с данными и инструментами анализа, а также формирование навыков комплексного подхода к решению задач машинного обучения. Важным результатом освоения дисциплины является способность разрабатывать прикладные решения, ориентированные на достижение практического эффекта и использование в реальных условиях.



11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

Комплект лицензионного программного обеспечения:

1. Python 3.8
2. Jupyter Notebook (Windows 10)
3. LMS Moodle
4. Реляционная СУБД
5. Docker
6. Anaconda
7. библиотеки Keras
8. PyTorch
9. Libre Office
10. Yandex Cloud
11. Mail Cloud
12. Git (клиент)
13. Среда разработки (PyCharm, VS Code или др.)
14. Антивирус Kaspersky

Современные профессиональные базы данных и информационные справочные системы

1. Информационно-правовая система «Гарант»
2. Информационно-правовая система «Консультант Плюс»
3. Язык программирования Python 3. <https://pythonworld.ru/>
4. БД «СПАРК-ИНТЕРФАКС» URL: <https://vk.com/www.spark-interfax.ru>
5. Информационно-справочная система «СПАРК»
6. GitLab (<https://gitlab.com>) Альтернатива GitHub для CI/CD

Сертифицированные программные и аппаратные средства защиты информации

1. не предусмотрены

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

1. Учебная аудитория для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная



оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), набор демонстрационного оборудования (проектор, экран)