ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

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

Рабочая программа дисциплины (модуля)
«Эксплуатация и внедрение моделей машинного обучения»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика»*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2024*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

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

Электронный документ

Москва
2024

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 7
4. **Содержание дисциплины (модуля)** 7
5. **Учебно-методическое обеспечение** 9
6. **Материально-техническое обеспечение** 9
7. **Методические и оценочные материалы** 11

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Эксплуатация и внедрение моделей машинного обучения» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Эксплуатация и внедрение моделей машинного обучения» позволяет студентам получить практические навыки, необходимые для успешной интеграции машинного обучения в бизнес-процессы, что критически важно для достижения конкурентных преимуществ. Кроме того, оно способствует пониманию вызовов и решений, связанных с масштабированием и поддержкой МО-моделей в условиях реального времени.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) изучается на 3 или 4 курсе в 5 или 7 семестре на выбор. Доступна к изучению после успешного освоения дисциплины (модуля) «Машинное обучение».

Цель изучения дисциплины (модуля): освоение студентами методов и практик внедрения, развертывания и поддержки моделей машинного обучения в реальных производственных системах.

Задачи изучения дисциплины (модуля):

— изучить основные этапы жизненного цикла моделей машинного обучения и организовать воспроизводимость экспериментов через трекинг кода, данных и метрик;

— освоить практики управления данными, включая версионирование, контроль качества и построение надёжных пайплайнов обработки;

— научиться проектировать и автоматизировать процессы обучения, тестирования и деплоя моделей с использованием инструментов mlops;

— развить навыки разработки и мониторинга масштабируемых решений для оффлайн и онлайн моделей в промышленных условиях;

— освоить методы интеграции и развёртывания больших языковых моделей (llm) в приложениях с применением llmops-подходов и индустриальных стандартов кодирования.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— как обеспечить воспроизводимость в машинном обучении;

— основные этапы жизненного цикла моделей;

— как устроен процесс управления данными для обучения, включая версионирование и обеспечение качества;

— в чём отличие между оффлайн и онлайн моделями и как планировать задачи для их обучения и внедрения;

— как подготовить модели машинного обучения к масштабированию;

— место llm в современной индустрии.

уметь:

— организовывать трекинг кода и результатов экспериментов;

— работать с пайплайнами обработки данных и обучения моделей;

— деплоить модели;

— автоматизировать процесс обучения и деплоя модели;

— разрабатывать тесты для проверки качества моделей и настраивать мониторинг для контроля их работы;

— развёртывать и применять llm инструменты.

владеть:

— навыком выстраивания mlops процесса (внедрение и поддержка работы моделей машинного обучения в продакшне);

— навыком работы с mlops и llmops инструментами;

— навыком масштабирования приложений;

— навыком использования и встраивания больших языковых моделей в приложения;

— навыком написания индустриального кода для сервисов машинного обучения.

2. Перечень планируемых результатов обучения

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

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности.
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области

		ОПК-1.2.	Умеет анализировать и формулировать математические задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов.
		ПК-1.2.	Умеет корректно ставить и формулировать математические задачи, применять строгие методы доказательства и анализировать полученные результаты.
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач
ПК-2.	Способен решать типовые задачи профессиональной деятельности в области искусственного интеллекта, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области искусственного интеллекта, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
		Лекции	Семинары			
1	Машинное обучение в индустрии. Базовые инструменты инженера машинного обучения	2	2		8	Домашнее задание
2	Работа с данными в машинном обучении и трекинг экспериментов машинного обучения	4	4		16	Домашнее задание
3	Подготовка моделей машинного обучения к деплою	6	6		26	Домашнее задание
4	Мониторинг	4	4	4	16	Домашнее задание
5	Автоматизация обучения и деплоя моделей машинного обучения	4	4		16	Домашнее задание
6	Масштабирование и отказоустойчивость приложений машинного обучения	2	2		8	Домашнее задание
7	LLMOps	8	8		32	Домашнее задание Творческое задание
	Зачёт с оценкой			4		Проект
	Итого:	***30***	***30***	***8***	***122***	
Объем дисциплины (модуля) (в ак. ч.)		***190***				
Объем дисциплины (модуля) (в зач. ед.)		***5***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Машинное обучение в индустрии. Базовые инструменты инженера машинного обучения	Введение. Инструменты разработки
2	Работа с данными в машинном обучении и трекинг экспериментов машинного обучения	Работа с данными в машинном обучении Трекинг экспериментов машинного обучения и воспроизводимость

Электронный документ

3	Подготовка моделей машинного обучения к деплою	Деплой моделей машинного обучения Индустриальный код
4	Мониторинг	Мониторинг и проект
5	Автоматизация обучения и деплоя моделей машинного обучения	Автоматизация пайплайна обучения моделей CI/CD для машинного обучения
6	Масштабирование и отказоустойчивость приложений машинного обучения	Масштабирование и отказоустойчивость приложений машинного обучения
7	LLMOps	Индустриальное использование больших языковых моделей ИИ Агенты в индустрии Мониторинг и контроль использования больших языковых моделей

5. Учебно-методическое обеспечение

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

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

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

Основная литература:

1. Вишневский, В. М. Теория очередей и машинное обучение : монография / В.М. Вишневский, Д.В. Ефросинин. — Москва : ИНФРА-М, 2025. — 370 с. : ил. — (Научная мысль). - ISBN 978-5-16-020572-4. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2204580.

2. Родзин, С. И. Машинное обучение: метаэвристики дифференциально-векторного движения : учебное пособие / С. И. Родзин, О. Н. Родзина. - Чебоксары : Среда, 2024. - 140 с. - ISBN 978-5-907830-17-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2171753.

Дополнительная литература:

1. Степашкина, А. С. Численные методы и машинное обучение в метрологии : учебное пособие / А. С. Степашкина. - Москва ; Вологда : Инфра-Инженерия, 2024. - 144 с. - ISBN 978-5-9729-1954-3. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2170894.

6. Материально-техническое обеспечение

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

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

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

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

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

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

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

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

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое

Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Эксплуатация и внедрение моделей машинного обучения» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, творческие задания, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

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

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

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Домашнее задание – набор задач по темам недели.

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

Творческое задание — это форма учебной деятельности, где студенты самостоятельно разрабатывают оригинальные проекты, интегрируя математические методы и компьютерные технологии для решения нестандартных задач и демонстрации инновационного мышления.

Для успешного выполнения творческого задания рекомендуется глубоко исследовать тему, экспериментировать с различными подходами и документировать процесс, чтобы не только решить задачу, но и представить результаты в виде отчета или презентации,

подчеркивающей креативность и практическую применимость.

Проект – исследовательская работа по дисциплине (модулю) и презентация результатов.

Для успешной подготовки к проекту рекомендуется: четко определить цели и задачи проекта; составить план работы, разбив проект на этапы с указанием сроков выполнения каждого из них; использовать разнообразные источники информации и инструменты для исследования темы; регулярно проверять прогресс и вносить коррективы в план, если это необходимо.

Бонусные баллы — это оценки, которые студенты могут получить за выполнение дополнительных заданий.

Формат бонусных баллов позволяет студентам улучшить общую оценку по дисциплине (модулю) и стимулирует углубленное изучение материала.

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

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

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Эксплуатация и внедрение моделей машинного обучения»

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

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

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

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

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

Дисциплина (модуль) «Эксплуатация и внедрение моделей машинного обучения» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	60%	Набор задач по темам недели
Творческое задание	10%	Форма учебной деятельности, где студентам предлагается разработать оригинальный проект или решить нестандартную задачу
Зачёт с оценкой	30%	Проект - Исследовательская работа по дисциплине (модулю) и презентация результатов

В рамках изучения дисциплины (модуля) возможно получение бонусных баллов.

Формула расчёта итоговой оценки по дисциплине (модулю) «Эксплуатация и внедрение моделей машинного обучения»: «0,6 × среднее за домашние задания + 0,1 × творческое задание + 0,3 × зачёт с оценкой».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1.

1. Опишите три основные области применения машинного обучения в современной индустрии и приведите по одному примеру для каждой.
2. Расскажите о ключевых ролях и обязанностях ML инженера в проекте по разработке ML-системы.
3. Составьте список из пяти популярных инструментов или библиотек, используемых ML инженерами, и кратко опишите назначение каждого.
4. Найдите и проанализируйте статью или кейс о внедрении ML в бизнес-процесс: какие задачи решались, и какую роль играл ML инженер?
5. Опишите, какие навыки и знания необходимы ML инженеру для эффективной работы с большими данными и развертыванием моделей.

Домашнее задание 2.

1. Соберите небольшой датасет (например, с открытых источников) и опишите процесс его очистки: какие шаги вы предприняли и почему.
2. Проведите базовый анализ данных: рассчитайте основные статистики (среднее, медиану, стандартное отклонение) для выбранных числовых признаков.
3. Постройте визуализации данных (гистограммы, scatter plot, boxplot) и сделайте выводы о распределении и взаимосвязях признаков.
4. Опишите методы обработки пропущенных значений и выбросов, которые вы применили к своему датасету.
5. Подготовьте данные для обучения модели: выполните нормализацию или стандартизацию признаков и объясните выбор метода.

Домашнее задание 3.

1. Сформулируйте гипотезу для ML эксперимента на основе выбранного датасета и опишите критерии успеха модели.
2. Опишите, как можно использовать систему трекинга экспериментов (например, MLflow или Weights & Biases) для управления экспериментами.
3. Проведите эксперимент по обучению модели с разными параметрами и зафиксируйте результаты (метрики качества). Сделайте выводы о влиянии параметров на качество.
4. Опишите основные шаги по оптимизации модели для продакшена (например, уменьшение размера, ускорение инференса).
5. Создайте простой REST API (например, с использованием Flask или FastAPI) для развертывания обученной модели и опишите процесс её вызова.

Примерное описание задания и критерии оценивания к творческому заданию

Нужно разработать и представить концепцию промышленного решения на основе модели машинного обучения (или большой языковой модели), включающую полный цикл её жизнедеятельности: от управления данными и трекинга экспериментов до деплоя, мониторинга, автоматизации и масштабирования.

Задание может быть выполнено в виде прототипа, архитектурной схемы, технического описания или демонстрационного проекта. Студенты выбирают реальную

или вымышленную бизнес-задачу (например, рекомендательная система, чат-бот на основе llm, детектор мошенничества) и проектируют mlops- или llmops-процессы для её решения.

Работа должна отражать понимание различий между оффлайн и онлайн сценариями, подходов к версионированию данных и моделей, организации пайплайнов, настройке мониторинга и обеспечению отказоустойчивости.

Критерии оценивания творческого задания:

— полнота охвата этапов жизненного цикла модели: управление данными, трекинг экспериментов, обучение, деплой, мониторинг и автоматизация;

— обоснованность выбора архитектуры и инструментов (например, mlflow, dagshub, kubeflow, tensorrt, langchain, vector db и др.) с учётом требований к масштабируемости и отказоустойчивости;

— наличие элементов автоматизации процессов обучения и обновления модели, включая ci/cd-подходы;

— реализация или описание механизма мониторинга: отслеживание дрейфа данных, качества предсказаний, задержек инференса и других ключевых метрик;

— интеграция больших языковых моделей (в случае llm-задачи): способ встраивания, управление контекстом, безопасность, cost control и оптимизация производительности.

— Примерное задание для проекта

Задание для проекта, кейс от ООО «Аренадата Софтвер»

Студентам предлагается разработать защищённую платформу для разработки и эксплуатации ИИ-сервисов:

1. Описание задания:

Организации, которые создают и эксплуатируют решения на основе данных и машинного обучения, сталкиваются с длительным развёртыванием рабочих сред, разрозненными учётными записями, неодинаковой конфигурацией серверов и отсутствием единой точки управления базами данных и виртуальными рабочими местами.

Студенческая команда должна спроектировать и развернуть в учебном изолированном контуре прототип централизованной платформы для команды аналитиков, разработчиков и инженеров эксплуатации.

2. Этапы работы:

— В VMmanager создать виртуальную инфраструктуру, шаблоны машин, сети, роли и ограничения ресурсов.

— На базе Astra Linux Special Edition подготовить серверные и пользовательские виртуальные машины, настроить службы, журналирование и базовые меры защиты.

— В ALD Pro организовать домен, группы пользователей и централизованное управление доступом.

— В Tantor развернуть базу данных для хранения метаданных наборов данных, экспериментов и версий моделей; настроить роли, резервное копирование и мониторинг.

— В Termidesk опубликовать стандартизированные виртуальные рабочие места и назначить их группам пользователей.

— Провести интеграционное тестирование, описать эксплуатационные процедуры и представить работающий стенд.

3. Используемые ресурсы:

— учебные лицензии и дистрибутивы VMmanager, Astra Linux Special Edition, ALD Pro, Tantor и Termidesk;

— синтетический набор данных и описание предметной области;

— перечень пользовательских ролей и сценариев доступа;

— официальная документация продуктов.

4. Проект считается успешно выполненным, если:

— все компоненты развёрнуты в согласованной архитектуре и доступны в учебном контуре;

— пользователь из назначенной группы может войти в виртуальное рабочее место через Termidesk;

— создание, блокирование и изменение прав пользователя централизованно отражаются на доступе;

— виртуальные машины создаются из подготовленного шаблона и имеют заданные ограничения ресурсов;

— приложение или тестовый клиент подключается к Tantor только с разрешёнными правами;

— выполняется резервное копирование и подтверждён тест восстановления учебной базы;

— события входа, отказов и основных операций фиксируются и доступны для анализа;

— комплект документации позволяет повторить развёртывание и провести проверку;

— команда демонстрирует решение по заранее подготовленному сценарию.

5. По итогам выполнения проекта студенты должны представить:

— работающий учебный стенд централизованной платформы;

— архитектурная схема с описанием назначения компонентов и сетевых связей;

— шаблоны виртуальных машин и виртуальных рабочих мест;

— настроенный домен, группы пользователей и правила доступа;

— экземпляр СУБД Tantor с учебной схемой данных, ролями и регламентом резервного копирования;

— комплект сценариев функционального и интеграционного тестирования;

— эксплуатационная инструкция администратора;

— краткая инструкция пользователя;

— итоговая демонстрация и защита решения (15-20 минут).

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Назовите основные компоненты архитектуры gpu, отвечающие за выполнение потоков и управление памятью	потоковые мультипроцессоры, иерархия памяти, контроллеры	УК-1
2	Объясните, как модель выполнения simt обеспечивает параллелизм в cuda-ядрах	за счёт одновременного выполнения одинаковых инструкций на нескольких потоках варпа	УК-1
3	Укажите, зачем используется индексация blockIdx.x * blockDim.x + threadIdx.x в cuda	для вычисления глобального индекса потока в одномерной сетке	УК-1
4	Назовите тип доступа к глобальной памяти, при котором достигается максимальная пропускная способность	слитный доступ / coalesced access	УК-1

5	Объясните, как разделяемая память используется для оптимизации ядра редукции суммы	для хранения промежуточных результатов и уменьшения обращений к глобальной памяти	УК-2
6	Назовите инструмент nvidia, позволяющий профилировать производительность cuda-ядер с детализацией по памяти и вычислениям	nvidia nsight compute	УК-2
7	Укажите, как примитив warp shuffle позволяет обмениваться данными между потоками варпа без использования разделяемой памяти	с помощью инструкции __shfl_sync и прямой передачи значений	УК-2
8	Объясните, в каких случаях атомарные операции необходимы при работе с глобальной памятью	при одновременной записи нескольких потоков в одну ячейку памяти	УК-2
9	Назовите способ интеграции пользовательского cuda-ядра в pytorch с поддержкой автоматического дифференцирования	реализация custom function с forward и backward через c++/cuda и pybind11	ОПК-1
10	Объясните, как pytorch сохраняет контекст для вычисления градиентов в кастомных операциях	с помощью ctx.save_for_backward и реализации backward-метода	ОПК-1
11	Назовите язык программирования gpu-ядер, позволяющий писать высокопроизводительный код без глубоких знаний cuda	triton	ОПК-1
12	Объясните, как разбиение на тайлы (tiling) в triton улучшает эффективность использования кэша	за счёт локальной обработки блоков данных и снижения обращений к глобальной памяти	ОПК-1
13	Назовите технику, позволяющую уменьшить пиковое потребление памяти при обучении за счёт перевычисления активаций	контрольное сохранение / gradient checkpointing	ПК-1
14	Укажите, какая стратегия распределённого обучения реплицирует модель на нескольких gpu с синхронизацией градиентов	distributeddataparallel / ddp	ПК-1
15	Назовите режим смешанной точности, сочетающий fp16 для вычислений и fp32 для накопления градиентов	автоматическая смешанная точность / amp	ПК-1
16	Назовите инструмент pytorch для анализа производительности модели с визуализацией времени выполнения операций	pytorch profiler	ПК-2
17	Объясните, как torch.compile ускоряет выполнение модели	за счёт слияния операторов, устранения накладных расходов и	ПК-2

		использования бэкендов (inductor)	
18	Назовите метод оптимизации инференса, при котором модель экспортируется в высокопроизводительный движок	ускорение вывода с помощью tensorrt	ПК-2
19	Укажите, какая техника используется в flashattention для снижения объёма обращений к памяти	tile-based обработка с кэшированием в разделяемой памяти	ПК-2
20	Назовите современный тренд, при котором вычисления выполняются только для активных частей модели	условные вычисления / conditional computation	ПК-2