МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Рязанский государственный радиотехнический университет»

Кафедра «Безопасность жизнедеятельности и экология»

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

по дисциплине

«Сопровождение моделей машинного обучения»

Направление

09.03.04 Программная инженерия

Профиль

Программная инженерия систем искусственного интеллекта
Программное обеспечение систем искусственного интеллекта

Квалификация выпускника – бакалавр

Форма обучения – очная

Рязань

1. ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов и процедур, предназначенных для оценки качества освоения обучающимися данной дисциплины как части основной образовательной программы.

Цель – оценить соответствие знаний, умений и уровня приобретённых компетенций обучающихся целям и требованиям основной образовательной программы в ходе текущего контроля и промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности компетенций и индикаторов их достижения, приобретаемых обучающимися в соответствии с требованиями ФГОС ВО.

Контроль знаний обучающихся проводится в форме текущего контроля (лабораторные работы) и промежуточной аттестации (зачёт).

2. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ ЭТАПОВ ИХ ФОРМИРОВАНИЯ

При освоении дисциплины формируются следующие компетенции:

- **ПК-19** – способен управлять процессом жизненного цикла ИИ-продукта (LC-4);
  - **ПК-19.1** – осуществляет запуск и ведение проекта в области ИИ, в том числе планирование и контроль задач, оценку ресурсов (LC-4.1);
  - **ПК-19.2** – контролирует реализацию проекта в соответствии с разработанной архитектурой проекта (LC-4.3).
- **ПК-20** – способен использовать различные технологии разработки программного обеспечения, включая современные;
  - **ПК-20.3** – организует внедрение и сопровождение разработанного программного обеспечения.
- **ПК-26** – способен проводить анализ бизнес-проблем с оценкой перспективности применения ИИ для их решения, осуществлять постановку задачи, формулировать требования к системе ИИ (LC-1);
  - **ПК-26.1** – формализует бизнес-цели и вырабатывает под них стратегии внедрения ИИ (LC-1.1);
  - **ПК-26.2** – выбирает оптимальные технологии под конкретные требования проекта внедрения ИИ (LC-1.2).
- **ПК-27** – способен проектировать и поддерживать архитектуру систем искусственного интеллекта (LC-3);
  - **ПК-27.1** – создаёт и развивает архитектуры системы ИИ на всех этапах жизненного цикла (LC-3.1).

Указанные компетенции формируются в соответствии со следующими этапами:

- формирование и развитие теоретических знаний (лекционные занятия, самостоятельная работа);
- приобретение и развитие практических умений (лабораторные работы, самостоятельная работа);
- закрепление знаний и навыков (выполнение и защита лабораторных работ, подготовка к зачёту).

3. ПОКАЗАТЕЛИ И КРИТЕРИИ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ

Сформированность каждой компетенции оценивается по трёхуровневой шкале:

- **пороговый уровень** – обязательный для всех обучающихся; соответствует освоению более 40% приведённых знаний, умений и навыков;
- **продвинутый уровень** – соответствует освоению более 60% знаний, умений и навыков;
- **эталонный уровень** – соответствует освоению более 80% знаний, умений и навыков.

При освоении менее 40% компетенция считается несформированной.

Таблица 1. Показатели достижения индикаторов компетенций

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
ПК-19.1	Знать принципы контролирования реализации системы ИИ в соответствии с её разработанной архитектурой. Уметь применять принципы контролирования реализации системы ИИ. Владеть навыками контролирования реализации системы ИИ.	7	Зачёт
ПК-19.2	Знать принципы запуска и ведения проекта по разработке системы ИИ. Уметь осуществлять планирование и контроль задач, оценку ресурсов при запуске и ведении проекта. Владеть навыками запуска и ведения проекта по разработке системы ИИ.	7	Зачёт

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
ПК-20.3	Знать методы и инструменты внедрения и сопровождения разработанного ПО в области ИИ. Уметь выполнять внедрение и сопровождение разработанного ПО в области ИИ. Владеть методами и инструментами внедрения и сопровождения ПО в области ИИ.	7	Зачёт
ПК-26.1	Знать принципы формализации бизнес-цели заказчика по внедрению ИИ. Уметь выявлять заинтересованные стороны во внедрении ИИ. Владеть принципами формулировки требований сторон, заинтересованных во внедрении ИИ (интервью, анкетирование, наблюдение).	7	Зачёт
ПК-26.2	Знать принципы анализа окружения места внедрения ИИ (пользовательское окружение, конфигурация оборудования, программное окружение, операционная среда). Уметь выявлять функциональные и нефункциональные требования. Владеть навыками анализа окружения и формулировки требований к системе ИИ.	7	Зачёт
ПК-27.1	Знать принципы и паттерны проектирования архитектуры систем ИИ. Уметь применять эффективные паттерны при проектировании архитектуры систем ИИ. Владеть навыками проектирования архитектуры систем ИИ.	7	Зачёт

4. ПАСПОРТ ОЦЕНОЧНЫХ МАТЕРИАЛОВ ПО ДИСЦИПЛИНЕ

Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или её части)	Наименование оценочного средства
Тема 1. Введение и базовое развертывание модели машинного обучения (от прототипа к продакшену, жизненный цикл ML-модели, развертывание как REST API, упаковка в Docker)	ПК-20.3, ПК-26.1, ПК-26.2	Зачёт
Тема 2. Инструменты и платформы для MLOps (MLflow, Kubeflow, AWS SageMaker, Azure ML, настройка логирования и мониторинга)	ПК-20.3, ПК-19.1, ПК-19.2	Зачёт
Тема 3. Мониторинг и диагностика моделей машинного обучения (детектирование дрейфа данных, сдвиг целевой переменной, деградация, системный мониторинг)	ПК-20.3, ПК-19.1, ПК-19.2	Зачёт
Тема 4. Автоматизация процессов (стратегии переобучения, версионирование данных и моделей, CI/CD пайплайны, Airflow/Prefect)	ПК-20.3, ПК-19.1, ПК-19.2, ПК-27.1	Зачёт
Тема 5. Качество и эксплуатация моделей машинного обучения (тестирование, интерпретируемость, безопасность, этика, вывод из эксплуатации)	ПК-20.3, ПК-19.1, ПК-19.2, ПК-27.1	Зачёт
Лабораторный практикум (12 лабораторных работ)	ПК-19.1, ПК-19.2, ПК-20.3, ПК-26.1, ПК-26.2, ПК-27.1	Защита лабораторных работ

5. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ

5.1. Промежуточная аттестация (зачёт)

Зачёт проводится в форме устного ответа по билетам, включающим теоретический вопрос и практическое задание (задачу). Также может использоваться тестирование.

5.1.1. Тестовые вопросы закрытого типа (20 вопросов)

1. Что такое MLOps?
а) Набор практик на стыке машинного обучения, DevOps и инженерии данных для управления жизненным циклом ML-моделей;
б) Язык программирования;
в) Тип базы данных;
г) Сетевой протокол.
Правильный ответ: а)

2. Какой этап не относится к жизненному циклу ML-модели?
а) Разработка;
б) Развертывание;
в) Мониторинг;
г) Ручной ввод данных без автоматизации.
Правильный ответ: г)

3. Что такое Docker?
а) Технология контейнеризации приложений;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

4. Что такое MLflow?
а) Платформа для управления жизненным циклом ML-моделей (логирование, версионирование, реестр моделей);
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

5. Что такое Data Drift?
а) Изменение распределения входных данных с течением времени;
б) Ошибка в коде;
в) Тип базы данных;
г) Метод оптимизации.
Правильный ответ: а)

6. Что такое Concept Drift?
а) Изменение зависимости между входными данными и целевой переменной;

б) Ошибка в коде;
в) Тип базы данных;
г) Метод оптимизации.
Правильный ответ: а)

7. Какая библиотека используется для детектирования и визуализации Data Drift?
а) Evidently AI;
б) NumPy;
в) Pandas;
г) Matplotlib.
Правильный ответ: а)

8. Что такое PSI (Population Stability Index)?
а) Метрика для оценки сдвига распределения данных;
б) Тип нейронной сети;
в) Метод оптимизации;
г) Алгоритм шифрования.
Правильный ответ: а)

9. Что такое JS-дивергенция?
а) Метрика для оценки различия между распределениями;
б) Тип базы данных;
в) Метод оптимизации;
г) Сетевой протокол.
Правильный ответ: а)

10. Что такое Prometheus?
а) Система мониторинга и сбора метрик;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

11. Что такое Grafana?
а) Платформа для визуализации метрик и создания дашбордов;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

12. Что такое стратегия переобучения по триггеру дрейфа?
а) Переобучение модели при обнаружении дрейфа данных;
б) Переобучение по расписанию;
в) Переобучение по запросу;
г) Отказ от переобучения.
Правильный ответ: а)

13. Что такое DVC?
а) Инструмент версионирования данных и моделей;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

14. Что такое CI/CD?
а) Непрерывная интеграция и непрерывная доставка;
б) Тип базы данных;
в) Метод оптимизации;
г) Сетевой протокол.
Правильный ответ: а)

15. Что такое Apache Airflow?
а) Инструмент для оркестрации конвейеров данных;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

16. Что такое pytest?
а) Фреймворк для автоматического тестирования на Python;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

17. Что такое SHAP?
а) Метод интерпретации предсказаний моделей машинного обучения;
б) Тип базы данных;
в) Метод оптимизации;
г) Сетевой протокол.
Правильный ответ: а)

18. Что такое LIME?
а) Метод интерпретации предсказаний моделей машинного обучения;
б) Тип базы данных;
в) Метод оптимизации;
г) Сетевой протокол.
Правильный ответ: а)

19. Что такое Adversarial Robustness Toolbox (ART)?
а) Библиотека для тестирования моделей на устойчивость к состязательным атакам;
б) Реляционная СУБД;
в) Операционная система;

г) Язык программирования.
Правильный ответ: а)

20. Что такое ONNX?
а) Открытый формат для представления моделей машинного обучения;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

5.1.2. Тестовые вопросы открытого типа (10 вопросов)

1. Опишите жизненный цикл ML-модели и основные этапы MLOps.

2. Объясните, как развернуть ML-модель в виде REST API-сервиса. Приведите примеры инструментов.

3. Что такое Data Drift и Concept Drift? Как их детектировать?

4. Опишите процесс настройки мониторинга ML-модели с помощью Prometheus и Grafana.

5. Что такое версионирование данных и моделей? Приведите примеры инструментов (DVC, MLflow).

6. Опишите стратегии переобучения моделей (по расписанию, по триггеру дрейфа).

7. Как организовать CI/CD пайплайн для ML-модели с использованием Apache Airflow?

8. Какие виды тестов используются для ML-моделей и систем? Приведите примеры.

9. Что такое XAI (Explainable AI)? Опишите методы SHAP и LIME.

10. Опишите план вывода модели из эксплуатации (graceful shutdown, архивирование, пост-мортем).

5.1.3. Типовые практические задания (задачи) к зачёту

1. Упаковать ML-модель в Docker-контейнер и развернуть REST API-сервис.
2. Настроить логирование и мониторинг с помощью MLflow.
3. Провести детектирование и визуализацию Data Drift с помощью Evidently AI.
4. Рассчитать и визуализировать метрики дрейфа (PSI, JS-дивергенция) для числовых и категориальных признаков.
5. Создать дашборд для мониторинга ML-системы (Prometheus/Grafana).
6. Реализовать стратегию переобучения по расписанию и по триггеру дрейфа.
7. Настроить версионирование данных (DVC) и моделей (MLflow).
8. Разработать и оркестрировать CI/CD/CT пайплайн с помощью Airflow.

9. Разработать комплекс автоматических тестов (pytest) для ML-пайплайна.
10. Интегрировать SHAP-объяснения в ML-сервис для объяснения индивидуальных прогнозов.

5.1.4. Вопросы к зачёту (30 вопросов)

1. От прототипа к продакшену. Жизненный цикл ML-модели.
2. Развертывание модели как REST API-сервиса.
3. Упаковка модели в Docker-контейнер.
4. MLOps как набор практик на стыке ML, DevOps и инженерии данных.
5. MLOps-паспорт жизненного цикла модели машинного обучения.
6. Инструменты и платформы для MLOps: обзор и сравнение (MLflow, Kubeflow, AWS SageMaker, Azure ML).
7. Настройка логирования и мониторинга с помощью MLflow.
8. Мониторинг данных. Детектирование дрейфа данных (Data Drift).
9. Концептуальный и популяционный сдвиг. Метрики дрейфа (PSI, JS-дивергенция).
10. Детектирование сдвига в прогнозах (Target/Prediction Drift).
11. Сдвиг целевой переменной и деградация модели.
12. Системный мониторинг ML-систем. Создание дашборда (Prometheus/Grafana).
13. Стратегии переобучения моделей (по расписанию, по триггеру дрейфа).
14. Версионирование данных и моделей с помощью DVC и MLflow.
15. CI/CD пайплайны для машинного обучения. Airflow/Prefect.
16. Разработка и оркестрация CI/CD/CT пайплайна.
17. Тестирование ML-моделей и систем: Data Tests, Model Tests, Drift Tests.
18. Разработка комплекса автоматических тестов (pytest).
19. Интерпретируемость и объяснимость моделей в продакшене. XAI.
20. Генерация объяснений для прогнозов модели с помощью SHAP/LIME.
21. Интеграция SHAP-объяснений в ML-сервис.
22. Безопасность и этика ML-систем.
23. Анализ устойчивости модели к Adversarial-атакам.
24. Аудит модели: проверка на устойчивость к атакам и смещения (Bias).
25. Управление производительностью и вывод из эксплуатации.
26. Пошаговый план вывода модели из эксплуатации (graceful shutdown).
27. Данные и артефакты для архивирования, срок архивирования.
28. Пост-мортем (post-mortem) анализ эффективности модели.
29. Оптимизация модели (ONNX) и разработка плана вывода из эксплуатации.
30. Перспективные направления развития MLOps.

6. КРИТЕРИИ ОЦЕНИВАНИЯ

6.1. Шкала оценивания тестирования

Шкала	Критерий
3 балла (эталонный уровень)	Процент верных ответов от 85 до 100%
2 балла (продвинутый уровень)	Процент верных ответов от 75 до 84%
1 балл (пороговый уровень)	Процент верных ответов от 60 до 74%
0 баллов	Процент верных ответов от 0 до 59%

6.2. Шкала оценивания теоретического вопроса

Шкала	Критерий
3 балла (эталонный уровень)	Дан полный ответ, показаны глубокие систематизированные знания, приведены примеры, даны ответы на дополнительные вопросы.
2 балла (продвинутый уровень)	Дан полный ответ, но на некоторые дополнительные вопросы ответы даны с помощью наводящих вопросов.
1 балл (пороговый уровень)	Дан неполный ответ, на дополнительные вопросы ответы даны только с помощью преподавателя.
0 баллов	Студент не смог ответить на вопрос.

6.3. Шкала оценивания практического задания

Шкала	Критерий
3 балла (эталонный уровень)	Задание решено верно, получены корректные результаты, сделаны выводы.
2 балла (продвинутый уровень)	Задание решено верно, но имеются незначительные технические неточности.
1 балл (пороговый уровень)	Задание решено верно с дополнительными наводящими вопросами преподавателя.

Шкала	Критерий
0 баллов	Задание не решено.

6.4. Итоговая оценка

На зачёт выносится тестовое задание, практическое задание и теоретический вопрос. Максимальный балл – 9.
Оценка «зачтено» выставляется при сумме баллов не менее 5 и выполнении всех предусмотренных лабораторных работ.
Оценка «не зачтено» – при сумме менее 5 баллов или наличии несданных лабораторных работ.

Оператор ЭДО ООО "Компания "Тензор"
ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
СОГЛАСОВАНО **ФГБОУ ВО "РГРТУ", РГРТУ,** Овечкин Геннадий Владимирович, Заведующий кафедрой ВПМ **05.11.26** 14:46 (MSK) Простая подпись