Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Баламирзоев Назим Лиодинович
Должность: Ректор
Дата подписания: 24.02.2026 11:50:41
Уникальный программный ключ:
5cf0d6f89e80f49a334f6a4ba58e91f3326b9926

Министерство науки и высшего образования Российской Федерации

ФГБОУ ВО «Дагестанский государственный технический университет»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине «Анализ защищенности систем искусственного интеллекта»
(указывается индекс и наименование дисциплины)

Уровень образования — магистратура
(бакалавриат/магистратура/специалитет)

Направление подготовки — 10.04.01 Информационная безопасность
(код, наименование направления подготовки)

Направленность — Кибербразведка и противодействие угрозам с применением технологий искусственного интеллекта
(наименование)

Разработчик _____ Качаева Г.И., к.э.н.
(подпись) (ФИО, уч. степень, уч. звание)

Фонд оценочных средств обсужден на заседании кафедры ИБиПИ «05» февраля 2026 г., протокол № 6/1

Зав. выпускающей кафедрой _____ Качаева Г.И., к.э.н.
(подпись) (ФИО, уч. степень, уч. звание)

г. Махачкала 2026

СОДЕРЖАНИЕ

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств (далее - ФОС) является неотъемлемой частью рабочей программы дисциплины «Анализ защищенности систем искусственного интеллекта» и предназначен для контроля и оценки образовательных достижений обучающихся (в т.ч. самостоятельной работе обучающихся), освоивших программу данной дисциплины.

Целью разработки фонда оценочных средств является установление соответствия уровня подготовки обучающихся требованиям федерального государственного образовательного стандарта высшего образования (далее - ФГОС ВО) по направлению подготовки 10.04.01 Информационная безопасность.

Рабочей программой дисциплины «Анализ защищенности систем искусственного интеллекта» предусмотрено формирование следующих компетенций:

1) *ПК-2 Способен выполнять мониторинг и ситуационный анализ обстановки в сфере информационной безопасности;*
2) *ПК-6 Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования.*

Формой аттестации по дисциплине является экзамен.

2. РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ, ПОДЛЕЖАЩИЕ ПРОВЕРКЕ

В результате аттестации по дисциплине осуществляется комплексная проверка индикаторов достижения компетенций их формирования в процессе освоения ОПОП.

Таблица 1.

Результаты обучения: индикаторы достижения	**Формируемые компетенции**
ПК -2.2 Способен разрабатывать процедуры мониторинга обстановки в сфере информационной безопасности	ПК-2 Способен выполнять мониторинг и ситуационный анализ обстановки в сфере информационной безопасности
ПК-6.2 Проводит экспериментальную проверку работоспособности систем искусственного интеллекта	ПК-6 Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования

3. ОЦЕНКА ОСВОЕНИЯ ДИСЦИПЛИНЫ

3.1. Контроль и оценка освоения дисциплины по темам (разделам)

Предметом оценки служат индикаторы достижения компетенций, предусмотренные ОПОП, направленные на формирование профессиональных компетенций.

Таблица 2.

Элемент дисциплины	Формы и методы контроля			
	Текущий контроль		Промежуточная аттестация	
	Форма контроля	Проверяемые компетенции/ индикаторы достижения	Форма контроля	Проверяемые компетенции/ индикаторы достижения
Раздел 1. Методы сбора данных из открытых источников				
Тема1.1 Введение в угрозы систем искусственного интеллекта	Письменная работа №1 Устный опрос Лабораторная работа №1 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.2 Таксономия угроз ИИ	Письменная работа №2 Устный опрос Лабораторная работа №2 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.3 Индустриальная практика анализа защищённости систем искусственного интеллекта ч.1	Письменная работа №3 Устный опрос Лабораторная работа №3 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.4 Индустриальная практика анализа защищённости систем искусственного интеллекта ч.2	Письменная работа №4 Устный опрос Лабораторная работа №4 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.5 Эксплуатационные атаки	Письменная работа №5 Устный опрос Лабораторная работа №5 Самостоятельная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2

	Реферат			
Тема 1.6 Защита на этапе инференса.	Письменная работа №6 Устный опрос Лабораторная работа №6 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.7 Атаки на конфиденциальность моделей.	Письменная работа №7 Устный опрос Лабораторная работа №7 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.8 Методы обеспечения приватности моделей	Письменная работа №8 Устный опрос Лабораторная работа №8 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.9 Атаки на целостность и доступность моделей	Письменная работа №9 Устный опрос Лабораторная работа №9 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 1.10 Защита от экстракции и саботажа	Письменная работа №10 Устный опрос Лабораторная работа №10 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Раздел 2. Методы анализа данных				
Тема 2.1 Анализ уязвимостей в конвейерах MLOps и инфраструктуре	Письменная работа №11 Устный опрос Лабораторная работа №11 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 2.2 Безопасность supply chain в ML	Письменная работа №12 Устный опрос	ПК-2: ПК-2.2; ПК-6:	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6:

	Лабораторная работа №12 Самостоятельная работа Реферат	ПК-6.2		ПК-6.2
Тема 2.3 Анализ защищенности ИИ в предметных областях: киберразведка и промышленные системы	Письменная работа №13 Устный опрос Лабораторная работа №13 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 2.4 Анализ защищенности ИИ в предметных областях: здравоохранение и биометрия	Письменная работа №14 Устный опрос Лабораторная работа №14 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 2.5 Методологии аудита безопасности систем ИИ	Письменная работа №15 Устный опрос Лабораторная работа №15 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 2.6 Нормативно-правовое регулирование безопасности ИИ	Письменная работа №16 Устный опрос Лабораторная работа №16 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2
Тема 2.7 Формирование отчетов и управление рисками.	Письменная работа №17 Устный опрос Лабораторная работа №17 Самостоятельная работа Реферат	ПК-2: ПК-2.2; ПК-6: ПК-6.2	Экзаменационная работа	ПК-2: ПК-2.2; ПК-6: ПК-6.2

3.2. Перечень заданий для текущего контроля

Формируемая компетенция: ПК-2

Перечень заданий закрытого типа

Задание № 1. Какой из перечисленных типов мониторинга в первую очередь предназначен для отслеживания метрик производительности системы (latency, throughput) и утилизации аппаратных ресурсов (GPU, CPU) в рабочей среде ML-модели?
A) Мониторинг дрифта данных (Data Drift).
B) Мониторинг качества работы ML-модели.
C) Технический мониторинг (Technical Monitoring).
D) Мониторинг выбросов (OoD-сэмплов).

Задание № 2. На каком из перечисленных этапов разработки процедуры мониторинга безопасности ИИ-системы происходит определение конкретных индикаторов (KPI), таких как доля ложноположительных срабатываний на определенном типе данных или время реакции на инцидент?
A) Этап проведения аудита существующей инфраструктуры.
B) Этап разработки программы (плана) внедрения системы мониторинга.
C) Этап формирования рабочей группы.
D) Этап интеграции с бизнес-процессами.

Задание № 3. Установите соответствие между типом уязвимости/атаки на систему ИИ и основной целью мониторинга для её обнаружения.

Тип угрозы ИИ	**Цель соответствующего мониторинга**
1. Состязательные атаки (FGSM, PGD)	A) Отслеживание аномальной активности и паттернов запросов к ML-API, которые могут указывать на попытку извлечения или реконструкции модели.
2. Отравление данных (Data Poisoning)	B) Контроль распределения и статистик входных данных, детектирование аномальных выборок, которые могли быть внесены на этапе обучения.
3. Атаки на экстракцию модели (Model Stealing)	C) Анализ уверенности модели для конкретных предсказаний и поиск аномалий, указывающих на специально сконструированные вредоносные входные данные.
4. Утечка через членство (Membership Inference)	D) Мониторинг разницы в поведении модели на тренировочных данных и новых данных для выявления возможности определить, была ли конкретная запись в обучающем наборе.

Задание **№ 4** Установите соответствие между этапом разработки процедуры мониторинга и ключевым содержательным результатом этого этапа

Этап разработки процедуры	**Ключевой результат этапа**
1. Проведение аудита и анализ рисков	A) Формирование документа, описывающего инструменты, метрики, роли, расписание и шаблоны отчетности.
2. Определение индикаторов и источников данных	B) Создание модели угроз (Threat Model) для ИИ-системы, например, на основе таксономии MITRE ATLAS.
3. Выбор инструментов и разработка регламентов	C) Запуск регулярных проверок, настройка алертов и утверждение процесса реагирования на инциденты.
4. Внедрение и функционирование	D) Утверждение списка контролируемых KPI (например, accuracy drop, доля OoD-данных) и перечня логов/метрик для их расчета.

Задание № 5. Определите последовательность действий команды SOC при обработке инцидента, связанного с возможной состязательной атакой на систему распознавания образов.

1. Провести анализ логов модели на аномальные паттерны в поступающих данных.
2. Изолировать сегмент сети с атакованной системой для предотвращения эскалации.
3. Внести информацию об используемых состязательных паттернах в базу знаний для будущего обнаружения.
4. Подтвердить гипотезу атаки, сгенерировав состязательные примеры и проверив модель в тестовой среде.
5. Получить алерт о резком падении точности системы в реальном времени.

. Перечень заданий открытого типа

Задание № 1. Как называется тип атаки на машинное обучение, целью которой является создание специальных входных данных (шум, артефакты), заставляющих модель совершать целенаправленную ошибку на этапе её использования (инференса)?
Задание № 2. Как называется базовый метод защиты от состязательных атак, заключающийся в включении специально созданных вредоносных примеров в тренировочный набор данных?
Задание № 3. Какой метод обеспечения конфиденциальности, формально ограничивающий влияние отдельной записи из обучающей выборки на итоговую модель, часто применяется для защиты от атак типа Membership Inference?
Задание № 4. Дополните определение, вставляя пропущенное слово:
Стандартизированный перечень известных уязвимостей и угроз информационной безопасности, используемый и в контексте систем ИИ, называется _____.
Задание № 5. Дополните определение, вставляя пропущенное слово:
Фреймворк _____ предоставляет таксономию атак на системы машинного обучения, аналогичную MITRE ATT&CK для классических кибератак.
Правильный ответ: MITRE ATLAS.

Формируемая компетенция: ПК- 6

Перечень заданий закрытого типа

Задание № 1. Какой из перечисленных методов НЕ является методом экспериментальной проверки устойчивости (robustness) модели машинного обучения к состязательным атакам?
A) Измерение accuracy на чистом тестовом наборе данных.
B) Оценка accuracy после применения атаки PGD.
C) Тестирование модели на специально созданном наборе состязательных примеров.
D) Измерение метрики Attack Success Rate (ASR).

Задание № 2. Какой этап экспериментальной проверки безопасности системы ИИ следует непосредственно за этапом «Определение целей тестирования и критериев успеха»?
A) Разработка плана исправления обнаруженных уязвимостей.
B) Подготовка итогового отчета.
C) Разработка сценариев и выбор инструментов тестирования.
D) Внедрение защитных механизмов в production.

Задание № 3. Установите соответствие между типом экспериментальной проверки систем ИИ и его основной целью.

Тип проверки	Основная цель
1. Стресс-тестирование (Stress Testing)	A) Оценка способности системы обрабатывать входные данные, на которых она не обучалась (Out-of-Distribution).
2. Фаззинг-тестирование (Fuzzing Testing)	B) Проверка корректности работы системы под экстремально высокой нагрузкой или при частичных отказах.
3. Тестирование на OoD-данных	C) Определение минимальных изменений во входных данных, приводящих к кардинально разным

	предсказаниям.
4. Поиск границ принятия решений	D) Обнаружение неожиданных сбоев путем подачи на вход случайных, невалидных или искаженных данных.

Задание № 4. Установите соответствие между ключевым документом, формируемым в рамках проверки, и фазой жизненного цикла тестирования

Документ / Артефакт	**Фаза жизненного цикла тестирования**
1. План тестирования (Test Plan)	A) Фаза подготовки и планирования. Содержит цели, объем, критерии входа/выхода, подход и график.
2. Сценарии тестирования (Test Cases)	B) Фаза исполнения. Описывает последовательность шагов, входные данные и ожидаемые результаты для конкретной проверки.
3. Отчет об инциденте (Bug Report)	C) Фаза анализа результатов и отчетности. Фиксирует обнаруженные дефекты, их критичность и шаги для воспроизведения.
Документ / Артефакт	Фаза жизненного цикла тестирования

Задание № 5. Определите последовательность этапов эксперимента по проверке модели на устойчивость к атаке отравления данных Data Poisoning.

1. Подготовить контрольную "чистую" выборку данных для оценки эталонного качества модели.
2. Обучить новую версию модели на отравленном наборе данных.
3. Зафиксировать разницу в метриках качества и эффективности на целевых классах между двумя версиями модели.
4. Внедрить в часть обучающего набора скомпрометированные данные по заданному сценарию атаки.
5. Протестировать обе версии модели на контрольной и целевых выборках.

Перечень заданий закрытого типа

Задание № 1. Как называется основной количественный показатель успешности состязательной атаки, рассчитываемый как доля входных данных, на которых атака привела к целевому ошибочному предсказанию?

Задание № 2. Какой метод экспериментальной проверки предполагает подачу на вход модели заведомо некорректных, искаженных или случайных данных с целью вызова её сбоя или получения неожиданного поведения?

Задание № 3. Как называется практика проведения экспериментов по проверке безопасности на изолированной, идентичной production, копии системы?

Задание № 4. Дополните определение, вставляя пропущенное слово:

Процесс подтверждения путем объективного исследования, что программная система соответствует заданным спецификациям и требованиям, называется _____.

Задание № 4**.** Дополните определение, вставляя пропущенное слово:

Экспериментальная проверка, имитирующая поведение реального злоумышленника с целью найти и эксплуатации уязвимостей в работающей системе, называется тестированием на _____.

4. ПЕРЕЧЕНЬ ЗАДАНИЙ ДЛЯ ОЦЕНКИ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

Формируемая компетенция: ПК-2

Перечень заданий закрытого типа

Задание № 1. Какой из перечисленных типов мониторинга в первую очередь предназначен для отслеживания метрик производительности системы (latency, throughput) и утилизации аппаратных ресурсов (GPU, CPU) в рабочей среде ML-модели?
A) Мониторинг дрифта данных (Data Drift).
B) Мониторинг качества работы ML-модели.
C) Технический мониторинг (Technical Monitoring).
D) Мониторинг выбросов (OoD-сэмплов).

Задание № 2. На каком из перечисленных этапов разработки процедуры мониторинга безопасности ИИ-системы происходит определение конкретных индикаторов (KPI), таких как доля ложноположительных срабатываний на определенном типе данных или время реакции на инцидент?
A) Этап проведения аудита существующей инфраструктуры.
B) Этап разработки программы (плана) внедрения системы мониторинга.
C) Этап формирования рабочей группы.
D) Этап интеграции с бизнес-процессами.

Задание № 3. Для обнаружения входных данных, на которых система ИИ не обучалась (Out-of-Distribution, OoD), и которые могут привести к ошибочным предсказаниям, НЕ применяется метод:
A) Uncertainty estimation (оценка уверенности модели).
B) Логирование промежуточных значений (яркость, объем сегментов).
C) Детекция аномалий на основе статистических порогов.
D) Генерация состязательных примеров методом FGSM.

Задание № 4. Выявление факта того, что конкретная запись присутствовала в тренировочном наборе данных модели, является целью атаки на конфиденциальность, называемой:
A) Model Inversion.
B) Membership Inference.
C) Model Stealing.
D) Adversarial Evasion.

Задание № 5. Какой из инструментов НЕ является специализированным для анализа защищенности и тестирования на устойчивость моделей машинного обучения?
A) IBM Adversarial Robustness 360 Toolbox (ART).
B) Grafana (инструмент для визуализации и мониторинга метрик).
C) CleverHans.
D) Foolbox.

Задание № 6. При обнаружении значительного изменения распределения входных данных после подключения к системе нового источника (например, новой камеры видеонаблюдения) говорят о возникновении:
A) Дрифта данных (Data Drift / Concept Drift).
B) Отравления данных (Data Poisoning).
C) Состязательной атаки (Adversarial Attack).
D) Утечки модели (Model Leakage).

Задание № 7. Установите соответствие между типом уязвимости/атаки на систему ИИ и основной целью мониторинга для её обнаружения.

Тип угрозы ИИ	Цель соответствующего мониторинга
1. Состязательные атаки (FGSM, PGD)	A) Отслеживание аномальной активности и паттернов запросов к ML-API, которые могут указывать на попытку извлечения или реконструкции модели.
2. Отравление данных (Data Poisoning)	B) Контроль распределения и статистик входных данных, детектирование аномальных выборок, которые могли быть внесены на этапе обучения.
3. Атаки на экстракцию модели (Model Stealing)	C) Анализ уверенности модели для конкретных предсказаний и поиск аномалий, указывающих на специально сконструированные вредоносные входные данные.
4. Утечка через членство (Membership Inference)	D) Мониторинг разницы в поведении модели на тренировочных данных и новых данных для выявления возможности определить, была ли конкретная запись в обучающем наборе.

Задание **№ 8.** Установите соответствие между этапом разработки процедуры мониторинга и ключевым содержательным результатом этого этапа

Этап разработки процедуры	Ключевой результат этапа
1. Проведение аудита и анализ рисков	A) Формирование документа, описывающего инструменты, метрики, роли, расписание и шаблоны отчетности.
2. Определение индикаторов и источников данных	B) Создание модели угроз (Threat Model) для ИИ-системы, например, на основе таксономии MITRE ATLAS.
3. Выбор инструментов и разработка регламентов	C) Запуск регулярных проверок, настройка алертов и утверждение процесса реагирования на инциденты.
4. Внедрение и функционирование	D) Утверждение списка контролируемых KPI (например, accuracy drop, доля OoD-данных) и перечня логов/метрик для их расчета.

Задание № 9. Определите последовательность действий команды SOC при обработке инцидента, связанного с возможной состязательной атакой на систему распознавания образов.

1. Провести анализ логов модели на аномальные паттерны в поступающих данных.
2. Изолировать сегмент сети с атакованной системой для предотвращения эскалации.
3. Внести информацию об используемых состязательных паттернах в базу знаний для будущего обнаружения.
4. Подтвердить гипотезу атаки, сгенерировав состязательные примеры и проверив модель в тестовой среде.
5. Получить алерт о резком падении точности системы в реальном времени.

Задание № 10. Определите порядок разработки процедуры регулярного мониторинга уязвимостей в конвейере машинного обучения MLOps.

1. Интегрировать процедуру сканирования в CI/CD-конвейер для автоматического выполнения.
2. Составить регламент с указанием ответственных лиц, частоты проверок и шагов реагирования.
3. Протестировать процедуру на staging-среде и зафиксировать базовые показатели.
4. Выбрать и настроить инструменты для сканирования образов контейнеров, библиотек и конфигураций.
5. Определить критические компоненты конвейера, подлежащие проверке.

. Перечень заданий открытого типа

Задание № 1. Как называется тип атаки на машинное обучение, целью которой является создание специальных входных данных (шум, артефакты), заставляющих модель совершать целенаправленную ошибку на этапе её использования (инференса)?

Задание № 2. Как называется базовый метод защиты от состязательных атак, заключающийся в включении специально созданных вредоносных примеров в тренировочный набор данных?

Задание № 3. Какой метод обеспечения конфиденциальности, формально ограничивающий влияние отдельной записи из обучающей выборки на итоговую модель, часто применяется для защиты от атак типа Membership Inference?

Задание № 4. Как называется процесс создания копии (суррогатной модели) целевой ML-системы путем многократных запросов к её публичному API?

Задание № 5. Дополните определение, вставляя пропущенное слово:
Стандартизированный перечень известных уязвимостей и угроз информационной безопасности, используемый и в контексте систем ИИ, называется _____.

Задание № 6. Дополните определение, вставляя пропущенное слово:
Фреймворк _____ предоставляет таксономию атак на системы машинного обучения, аналогичную MITRE ATT&CK для классических кибератак.
Правильный ответ: MITRE ATLAS.

Формируемая компетенция: ПК- 6

Перечень заданий закрытого типа

Задание № 1. Какой из перечисленных методов НЕ является методом экспериментальной проверки устойчивости (robustness) модели машинного обучения к состязательным атакам?
A) Измерение accuracy на чистом тестовом наборе данных.
B) Оценка accuracy после применения атаки PGD.
C) Тестирование модели на специально созданном наборе состязательных примеров.
D) Измерение метрики Attack Success Rate (ASR).

Задание № 2. Какой этап экспериментальной проверки безопасности системы ИИ следует непосредственно за этапом «Определение целей тестирования и критериев успеха»?
A) Разработка плана исправления обнаруженных уязвимостей.
B) Подготовка итогового отчета.
C) Разработка сценариев и выбор инструментов тестирования.
D) Внедрение защитных механизмов в production.

Задание № 3. При проведении стресс-тестирования ML-модели на предмет её доступности (availability) наиболее релевантной метрикой будет:
A) Точность (Accuracy) на валидационной выборке.
B) Время обучения (Training Time).
C) Среднее время ответа (Latency) и пропускная способность (Throughput) под высокой нагрузкой.
D) Размер модели в мегабайтах.

Задание № 4. Какой из инструментов предоставляет готовую библиотеку для реализации, оценки и защиты от состязательных атак и является стандартом де-факто в исследовательском сообществе?
A) TensorFlow Extended (TFX).
B) MLflow.

C) IBM Adversarial Robustness 360 Toolbox (ART).
D) Apache Airflow.

Задание № 5. Для экспериментальной проверки модели на устойчивость к Membership Inference Attack необходимо:
A) Проверить, насколько сильно меняются предсказания модели при добавлении случайного шума к входным данным.
B) Обучить несколько «теневых» моделей и оценить, может ли атакующая модель отличить данные, участвовавшие в обучении, от новых.
C) Генерировать запросы к API модели для создания её суррогатной копии.
D) Измерить, сколько энергии потребляет GPU при инференсе модели.

Задание № 6. Какой тип проверки работоспособности системы ИИ фокусируется на поиске уязвимостей в программных зависимостях, конфигурации контейнеров и оркестраторов (Kubernetes)?
A) Функциональное тестирование (Functional Testing).
B) Тестирование безопасности инфраструктуры MLOps (MLOps Security Testing).
C) Fuzz-тестирование (Fuzzing) API модели.
D) Статический анализ кода модели (Static Code Analysis).

Задание № 7. Установите соответствие между типом экспериментальной проверки систем ИИ и его основной целью.
.

Тип проверки	Основная цель
1. Стресс-тестирование (Stress Testing)	A) Оценка способности системы обрабатывать входные данные, на которых она не обучалась (Out-of-Distribution).
2. Фаззинг-тестирование (Fuzzing Testing)	B) Проверка корректности работы системы под экстремально высокой нагрузкой или при частичных отказах.
3. Тестирование на OoD-данных	C) Определение минимальных изменений во входных данных, приводящих к кардинально разным предсказаниям.
4. Поиск границ принятия решений	D) Обнаружение неожиданных сбоев путем подачи на вход случайных, невалидных или искаженных данных.

Задание № 8. Установите соответствие между ключевым документом, формируемым в рамках проверки, и фазой жизненного цикла тестирования

Документ / Артефакт	Фаза жизненного цикла тестирования
1. План тестирования (Test Plan)	A) Фаза подготовки и планирования. Содержит цели, объем, критерии входа/выхода, подход и график.
2. Сценарии тестирования (Test Cases)	B) Фаза исполнения. Описывает последовательность шагов, входные данные и ожидаемые результаты для конкретной проверки.
3. Отчет об инциденте (Bug Report)	C) Фаза анализа результатов и отчетности. Фиксирует обнаруженные дефекты, их критичность и шаги для воспроизведения.
Документ / Артефакт	Фаза жизненного цикла тестирования

Задание № 9. Определите последовательность этапов эксперимента по проверке модели на устойчивость к атаке отравления данных Data Poisoning.

1. Подготовить контрольную "чистую" выборку данных для оценки эталонного качества модели.
2. Обучить новую версию модели на отравленном наборе данных.

3. Зафиксировать разницу в метриках качества и эффективности на целевых классах между двумя версиями модели.
4. Внедрить в часть обучающего набора скомпрометированные данные по заданному сценарию атаки.
5. Протестировать обе версии модели на контрольной и целевых выборках.

Задание № 10. Определите порядок действий при проведении практического теста на извлечение модели Model Extraction.

1. Проанализировать полученную копию, сравнив её архитектуру и точность с оригиналом.
2. На основе собранных пар "вход-выход" обучить локальную модель-копию.
3. Задокументировать факт извлечения и оценить потенциальный ущерб интеллектуальной собственности.
4. Отправить к целевой модели через её публичный API серию запросов для сбора предсказаний.
5. Создать или выбрать подходящий синтетический набор входных данных.

Перечень заданий закрытого типа

Задание № 1. Как называется основной количественный показатель успешности состязательной атаки, рассчитываемый как доля входных данных, на которых атака привела к целевому ошибочному предсказанию?

Задание № 2. Какой метод экспериментальной проверки предполагает подачу на вход модели заведомо некорректных, искаженных или случайных данных с целью вызова её сбоя или получения неожиданного поведения?

Задание № 3. Как называется практика проведения экспериментов по проверке безопасности на изолированной, идентичной production, копии системы?

Задание № 4. При проверке приватности модели с использованием дифференциальной приватности (DP), какой ключевой параметр (ε, "эпсилон") количественно измеряет?

Задание № 5. Дополните определение, вставляя пропущенное слово:
Процесс подтверждения путем объективного исследования, что программная система соответствует заданным спецификациям и требованиям, называется _____.

Задание № 6. Дополните определение, вставляя пропущенное слово:
Экспериментальная проверка, имитирующая поведение реального злоумышленника с целью найти и эксплуатации уязвимостей в работающей системе, называется тестированием на _____.

5. КРИТЕРИИ ОЦЕНКИ

5.1. Критерии оценки текущего контроля и промежуточной аттестации

В ФГБОУ ВО «ДГТУ» внедрена модульно-рейтинговая система оценки учебной деятельности обучающихся. В соответствии с этой системой применяются пятибальная, двадцатибальная и стобальная шкалы знаний, умений, навыков.

Таблица 3.

Шкалы оценивания			**Критерии оценивания**
пятибальная	двадцатибальная	стобальная	
«Отлично» - 5 баллов	«Отлично» - 18-20 баллов	«Отлично» - 85 – 100 баллов	Показывает высокий уровень сформированности компетенций, т.е.: – продемонстрирует глубокое и прочное усвоение материала; – исчерпывающе, четко, последовательно, грамотно и логически стройно излагает теоретический материал; – правильно формирует определения; – демонстрирует умения самостоятельной работы с нормативно-правовой литературой; – умеет делать выводы по излагаемому материалу.
«Хорошо» - 4 баллов	«Хорошо» - 15 - 17 баллов	«Хорошо» - 70 - 84 баллов	Показывает достаточный уровень сформированности компетенций, т.е.: – демонстрирует достаточно полное знание материала, основных теоретических положений; – достаточно последовательно, грамотно логически стройно излагает материал; – демонстрирует умения ориентироваться в нормальной литературе; – умеет делать достаточно обоснованные выводы по излагаемому материалу.
«Удовлетворительно» - 3 баллов	«Удовлетворительно» - 12 - 14 баллов	«Удовлетворительно» - 56 – 69 баллов	Показывает пороговый уровень сформированности компетенций, т.е.: – демонстрирует общее знание изучаемого материала; – испытывает серьезные затруднения при ответах на дополнительные вопросы; – знает основную рекомендуемую литературу; – умеет строить ответ в соответствии со структурой излагаемого материала.
«Неудовлетворительно» - 2 баллов	«Неудовлетворительно» - 1-11 баллов	«Неудовлетворительно» - 1-55 баллов	Ставится в случае: – незнания значительной части программного материала; – не владения понятийным аппаратом дисциплины; – допущения существенных ошибок при изложении учебного материала; – неумение строить ответ в соответствии со структурой излагаемого вопроса; – неумение делать выводы по излагаемому материалу.

Критерии оценки тестовых заданий

Таблица 4.

Процент выполненных тестовых заданий	**Оценка**
до 50%	неудовлетворительно
50-69%	удовлетворительно
70-84%	хорошо
85-100%	отлично

Критерии оценки тестовых заданий, заданий на дополнение, с развернутым ответом и на установление правильной последовательности

Верный ответ - 2 балла.
Неверный ответ или его отсутствие - 0 баллов.

Критерии оценки заданий на сопоставление

Верный ответ - 2 балла
1 ошибка - 1 балл
более 1-й ошибки или ответ отсутствует - 0 баллов.

КЛЮЧИ К ЗАДАНИЯМ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ

Таблица 5.

Формируемые компетенции	№ задания	Ответ
ПК-2	**Задания закрытого типа**	
	№ 1	C
	№ 2	B
	№ 3	1-C, 2-A, 3-B, 4-D
	№ 4	1-B, 2-A, 3-C, 4-D
	№ 5	5 1 4 2 3
	Задания открытого типа	
	№ 1	Состязательная атака
	№ 2	Adversarial Training
	№ 3	Дифференциальная приватность
	№ 4	CVE
	№ 5	MITRE ATLAS
ПК-6	**Задания закрытого типа**	
	№ 1	A
	№ 2	C
	№ 3	1 — B, 2 — D, 3 — A, 4 — C
	№ 4	1 — A, 2 — B, 3 — C, 4 — D
	№ 5	1 4 2 5 3
	Задания открытого типа	
	№ 1	ASR
	№ 2	Фаззинг-тестирование
	№ 3	Sandbox Testing
	№ 4	Валидация
	№ 5	Проникновение

КЛЮЧИ К ЗАДАНИЯМ ДЛЯ ОЦЕНКИ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

Таблица 6.

Формируемые компетенции	№ задания	Ответ
ПК-2	**Задания закрытого типа**	
	№ 1	C
	№ 2	B
	№ 3	C
	№ 4	C
	№ 5	B
	№ 6	C
	№ 7	1-C, 2-A, 3-B, 4-D
	№ 8	1-B, 2-A, 3-C, 4-D
	№ 9	5 1 4 2 3
	№ 10	5 4 3 2 1
	Задания открытого типа	
	№ 1	Состязательная атака
	№ 2	Adversarial Training
	№ 3	Дифференциальная приватность
	№ 4	Model Stealing
	№ 5	CVE
	№ 6	MITRE ATLAS
ПК-6	**Задания закрытого типа**	
	№ 1	A
	№ 2	C
	№ 3	C
	№ 4	C
	№ 5	C
	№ 6	B
	№ 7	1 — B, 2 — D, 3 — A, 4 — C
	№ 8	1 — A, 2 — B, 3 — C, 4 — D
	№ 9	1 4 2 5 3
	№ 10	5 4 2 1 3
	Задания открытого типа	
	№ 1	ASR
	№ 2	Фаззинг-тестирование
	№ 3	Sandbox Testing
	№ 4	Гарантию приватности
	№ 5	Валидация
	№ 6	Проникновение