МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ
И.о. декана факультета математики и компьютерных наук имени профессора Н. И. Червякова
Гробова Т. А.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ
«Атаки на системы искусственного интеллекта»

Направление подготовки	02.04.01 Математика и компьютерные науки
Профиль	Искусственный интеллект в кибербезопасности
Год начала обучения	2026
Форма обучения	очная
Реализуется в 3 семестре	

Введение

1. Назначение: Фонд оценочных средств предназначен для обеспечения методической основы для организации и проведения текущего контроля по дисциплине «Атаки на системы искусственного интеллекта». Текущий контроль по данной дисциплине – вид систематической проверки знаний, умений, навыков студентов. Задачами текущего контроля являются получение первичной информации о ходе и качестве освоения компетенций, а также стимулирование регулярной целенаправленной работы студентов. Для формирования определенного уровня компетенций.

2. ФОС является приложением к программе дисциплины «Атаки на системы искусственного интеллекта»

3. Разработчик: кафедры вычислительной математики и кибернетики.

4. Проведена экспертиза ФОС.

Члены экспертной группы:

Председатель

Бабенко М. Г. заведующий кафедрой вычислительной математики и кибернетики

Члены комиссии:

Пелешенко В.С. доцент кафедры вычислительной математики и кибернетики

Пелешенко Т. А. доцент кафедры вычислительной математики и кибернетики

Представитель организации-работодателя Корниенко С. А. начальник управления филиала ФГУП «Главный радиочастотный центр» в Южном и Северо-Кавказском федеральных округах

Экспертное заключение: фонд оценочных средств соответствует образовательной программе по направлению подготовки 02.04.01 Математика и компьютерные науки профиль «Искусственный интеллект в кибербезопасности» и рекомендуется для проведения текущего контроля успеваемости и промежуточной аттестации студентов.

5. Срок действия ФОС определяется сроком реализации образовательной программы.

1. Описание критериев оценивания компетенции на различных этапах их формирования, описание шкал оценивания

Компетенция (ии), индикатор (ы)	Уровни сформированности компетенци(ий),			
	Минимальный уровень не достигнут (Неудовлетворительно) 2 балла	Минимальный уровень (удовлетворительно) 3 балла	Средний уровень (хорошо) 4 балла	Высокий уровень (отлично) 5 баллов
Компетенция: ПК-9 С помощью современных пакетов прикладных программ моделирования, исследует вновь создаваемые математические модели				
Результаты обучения по дисциплине (модулю): *Индикатор: ИД-1. ПК-9.* С помощью современных пакетов прикладных программ моделирования, исследует вновь создаваемые математические модели	не знает как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	плохо знает как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	знает как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	отлично знает как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования
ИД-2. ПК -9 Использует современные методы разработки алгоритмов математических моделей	не умеет использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	плохо умеет использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	умеет использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	отлично умеет использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования
ИД-3. ПК -9 Использует современные языки программирования для реализации конкретных алгоритмов и	не владеет навыками как использовать современные методы разработки и реализации конкретных	плохо владеет навыками как использовать современные методы разработки и реализации конкретных	владеет навыками как использовать современные методы разработки и реализации	отлично владеет навыками как использовать современные методы

математических моделей	алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования	разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования

Оценивание уровня сформированности компетенции по дисциплине осуществляется на основе «Положения о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры - в федеральном государственном автономном образовательном учреждении высшего образования «Северо-Кавказский федеральный университет» в актуальной редакции.

ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОВЕРКИ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

Номер задания	Правильный ответ	Содержание вопроса	Компетенция
		Форма обучения очная/заочная/очно-заочная	
1.	B	Какая из следующих проблем является ключевой при обсуждении доверия к системам искусственного интеллекта? A) Высокая скорость обработки данных B) Прозрачность принятия решений C) Сложность алгоритмов D) Высокая точность прогнозов	ПК-9
2.	D	Какой аспект систем ИИ наиболее важен для обеспечения доверия? A) Эффективность B) Прозрачность C) Безопасность D) Все вышеперечисленное	ПК-9
3.	D	Какие методы используются для повышения прозрачности моделей ИИ? A) Explainable AI (XAI) B) Transfer Learning C) Reinforcement Learning D) Только A	ПК-9
4.	D	Какая из следующих атак на системы ИИ может повлиять на доверие к ним? A) Враждебные атаки B) Атаки на данные C) Атаки на модели D) Все вышеперечисленное	ПК-9
5.	C	Какой инструмент используется для анализа и объяснения решений моделей ИИ? A) SHAP (SHapley Additive exPlanations) B) LIME (Local Interpretable Model-agnostic Explanations) C) Both A and B D) Только A	ПК-9
6.	Основные проблемы включают	Опишите основные проблемы доверия к системам искусственного интеллекта.	ПК-9

	прозрачность, объяснимость и безопасность моделей ИИ.		
7.	Explainable AI (XAI), аудит моделей и тестирование на устойчивость.	Какие методы можно использовать для повышения доверия к моделям ИИ?	ПК-9
8.	Прозрачность позволяет понять, как модель принимает решения, что повышает доверие к ее выводам.	Как прозрачность моделей ИИ влияет на доверие к ним?	ПК-9
9.	XAI помогает объяснить решения моделей ИИ, что повышает доверие. Примеры включают использование SHAP и LIME для анализа решений в медицинских и финансовых приложениях.	Как индивидуальные задания могут помочь в обсуждении и решении проблемы доверия к системам ИИ? Приведите примеры таких заданий	ПК-9
10.	Индивидуальные задания могут включать разработку прозрачных моделей или анализ уязвимостей конкретных систем ИИ. Например, задание по реализации SHAP для объяснения решений нейронной сети.	Опишите роль Explainable AI (XAI) в повышении доверия к системам ИИ и приведите примеры его применения.	ПК-9

11.		Какие технологии используются для выявления потенциальных рисков в управлении рисками ИИ? A) Машинное обучение B) Анализ больших данных C) Обработка естественного языка D) Все вышеперечисленное	ПК-9
12.		Какой стандарт обеспечивает целостный подход к управлению рисками, связанными с ИИ? A) ISO 31000 B) ISO/IEC 23894:2023 C) ISO/IEC JTC 1/SC 42 D) Только B	ПК-9
13.		Какой метод используется для непрерывного мониторинга сетевого трафика и идентификации угроз? A) Обработка естественного языка B) Машинное обучение C) Анализ больших данных D) Только B	ПК-9
14.		Какие области управления рисками ИИ включают применение ИИ? A) Выявление потенциальных рисков B) Анализ сетевого трафика C) Обработка данных D) Все вышеперечисленное	ПК-9
15.		Какой подход к управлению рисками ИИ включает установление контекста определения рисков и категоризацию систем ИИ? A) Стандарт ISO 31000 B) Стандарт ISO/IEC 23894:2023 C) Система AI RMF D) Все вышеперечисленное	ПК-9

16.	Основные риски включают потенциальные ошибки в моделях ИИ, уязвимости в данных и непредвиденные последствия принятия решений.	Опишите основные риски, связанные с внедрением систем ИИ в управление рисками.	ПК-9
17.	Преимущества включают повышение точности и скорости выявления рисков, оптимизацию ресурсов и улучшение принятия решений.	Какие преимущества дает использование ИИ в управлении рисками?	ПК-9
18.	Используются инструменты машинного обучения и анализа больших данных для мониторинга сетевого трафика и выявления аномалий.	Какие инструменты и технологии используются для непрерывного мониторинга и идентификации угроз в системах ИИ?	ПК-9
19.	Машинное обучение играет ключевую роль в выявлении рисков, анализируя большие объемы данных для прогнозирования потенциальных угроз. Примеры включают обнаружение	Опишите основные риски, связанные с внедрением систем ИИ в управление рисками.	ПК-9

	мошенничества в финансовых транзакциях и прогнозирование сбоев в производстве.		
20.	Стандарт ISO/IEC 23894:2023 обеспечивает целостный подход к управлению рисками ИИ, включая установление контекста, выявление и оценку рисков. Примеры его использования включают разработку стратегий смягчения рисков в организациях, использующих ИИ.	Какие преимущества дает использование ИИ в управлении рисками?	ПК-9
21.	A	Какой тип атаки на системы ИИ направлен на остановку или подавление их работы? A) Атака с искажением B) Атака с подавлением C) Атака с переполнением D) Атака с отказом	ПК-9
22.	B	Какая атака включает преднамеренное искажение данных для обмана ИИ? A) Атака с подавлением B) Атака с искажением C) Атака с переполнением D) Атака с отказом	ПК-9

23.	C	Какой метод атаки включает переполнение входных данных для выявления слабостей? A) Атака с подавлением B) Атака с искажением C) Атака с переполнением D) Атака с отказом	ПК-9
24.	D	Какая атака направлена на организацию отказов и блокировку работы ИИ? A) Атака с подавлением B) Атака с искажением C) Атака с переполнением D) Атака с отказом	ПК-9
25.	D	Какой тип атаки включает манипуляцию для перенаправления работы ИИ в некорректное направление? A) Атака с подавлением B) Атака с искажением C) Атака с переполнением D) Атака с перенаправлением	ПК-9
26.	D	Какая атака включает проникновение в систему и введение нежелательного кода или команд? A) Атака с подавлением B) Атака с искажением C) Атака с переполнением D) Атака с введением нежелательных данных	ПК-9
27.	A	Какой тип атаки на ИИ включает внедрение вредоносных данных в систему? A) Отравление данных B) Атака с подавлением C) Атака с искажением D) Атака с отказом	ПК-9
28.	A	Какая атака на ИИ направлена на извлечение информации из моделей? A) Инверсионная атака B) Атака с подавлением C) Атака с искажением D) Атака с отказом	ПК-9

29.	D	Какой тип атаки включает внедрение команд в промпт для нарушения работы ИИ? A) Prompt Injection B) Token Truncation Attack C) Unauthorized Function Call Attack D) Все вышеперечисленное	ПК-9
30.	A	Какая атака на ИИ использует бинарный поиск для нахождения точки выборки вблизи границы классификации? A) BA (граничная атака) B) UP-атака C) Атака с подавлением D) Атака с отказом	ПК-9
31.	D	Какой тип атаки на ИИ включает использование уязвимостей «нулевого дня»? A) Атака с подавлением B) Атака с искажением C) Атака с отказом D) Атака с использованием уязвимостей «нулевого дня»	ПК-9
32.	A	Какая атака на ИИ направлена на перехват и изменение электронных сообщений? A) Атака «человек посередине» B) Атака с подавление C) Атака с искажением D) Атака с отказом	ПК-9
33.	D	Какой метод атаки включает внедрение вредоносных данных в промпт для манипулирования ответом ИИ? A) Prompt Injection B) Token Truncation Attack C) Unauthorized Function Call Attack D) Все вышеперечисленное	ПК-9
34.	A	Какая атака на ИИ использует постепенные изменения контекста для выполнения вредоносных действий? A) Многоступенчатая контекстная инъекция B) Атака с подавлением C) Атака с искажением D) Атака с отказом	ПК-9

35.	A	Какой тип атаки на ИИ включает скрытые команды внутри текста или документа? A) Скрытые команды (Embedded Prompts) B) Prompt Injection C) Token Truncation Attack D) Unauthorized Function Call Attack	ПК-9
36.	A	Какой тип атаки на ИИ включает внедрение вредоносных данных в систему для изменения ее поведения? A) Отравление данных B) Атака с подавлением C) Атака с искажением D) Атака с отказом	ПК-9
37.	A	Какая атака на ИИ направлена на создание фальшивых изображений и видео? A) ZQA-атака B) Атака с подавлением C) Атака с искажением D) Атака с отказом	ПК-9
38.	A	Какой метод атаки включает использование JSON HiJacking для перехвата данных? A) JSON HiJacking B) Атака «человек посередине» C) Атака с подавлением D) Атака с отказом	ПК-9
39.	A	Какой тип атаки на ИИ включает использование бинарного поиска для нахождения точки выборки вблизи границы классификации? A) BA (граничная атака) B) UP-атака C) Атака с подавлением D) Атака с отказом	ПК-9
40.	A	Какая атака на ИИ использует постепенные изменения контекста для выполнения вредоносных действий? A) Многоступенчатая контекстная инъекция B) Prompt Injection C) Token Truncation Attack	ПК-9

		D) Unauthorized Function Call Attack	
41.	Основные типы атак включают whitebox, blackbox и greybox-атаки, которые различаются по уровню доступа к модели ИИ.	Какие основные типы атак существуют на системы искусственного интеллекта?	ПК-9
42.	Whitebox-атаки предполагают полный доступ к структуре и параметрам модели ИИ, что позволяет точно манипулировать входными данными.	В чем заключается суть whitebox-атак на модели ИИ?	ПК-9
43.	Blackbox-атаки используют уязвимости, не требуя доступа к внутренней структуре модели, а только к входным и выходным данным.	Какие уязвимости используются при blackbox-атаках на ИИ?	ПК-9
44.	Используются методы, такие как adversarial attacks, которые вводят небольшие изменения в данные для обмана модели ИИ.	Какие методы используются для эксплуатации небольших возмущений во входных данных?	ПК-9
45.	Атаки могут привести к нарушению работы критической инфраструктуры, что может иметь серьезные	Какие последствия могут иметь атаки на системы ИИ в критической инфраструктуре?	ПК-9

	последствия для общества.		
46.	Меры включают регулярные обновления моделей, тестирование на устойчивость и использование методов защиты от adversarial attacks.	Какие меры безопасности можно принять для защиты от атак на ИИ?	ПК-9
47.	Использование ИИ в атаках позволяет автоматизировать и персонализировать вредоносные действия, что повышает их эффективность.	Какие новые киберугрозы несет в себе использование ИИ в атаках?	ПК-9
48.	Наиболее распространены adversarial attacks и атаки на данные, которые влияют на точность моделей ИИ.	Какие типы атак на ИИ наиболее распространены в современных условиях?	ПК-9
49.	Используются методы, такие как мониторинг входных данных, анализ поведения и обучение на данных в реальном времени.	Какие методы используются для обнаружения и предотвращения атак на ИИ?	ПК-9
50.	Атаки могут привести к распространению фейковых новостей и манипуляциям	Какие последствия могут иметь атаки на системы ИИ в социальных сетях?	ПК-9

	общественным мнением.		
51.	Уязвимости могут быть связаны с неадекватной защитой данных и отсутствием регулярных обновлений систем.	Какие уязвимости существуют в интерфейсах транспортной инфраструктуры умного города?	ПК-9
52.	Используются методы генеративных моделей ИИ для создания реалистичных видео и аудиозаписей.	Какие атаки на ИИ могут быть использованы для создания дипфейков?	ПК-9
53.	Ботнеты, управляемые ИИ, могут автоматизировать фишинговые атаки и распространение вредоносного ПО.	Какие ботнеты управляются с помощью ИИ?	ПК-9
54.	Атаки на IoT могут привести к утечке данных и нарушению работы устройств.	Какие угрозы представляют собой атаки на интернет вещей (IoT)?	ПК-9
55.	Используются алгоритмы машинного обучения для автоматизации поиска уязвимостей и эксплуатации.	Какие методы используются для автоматизации атак с помощью ИИ?	ПК-9
56.	Атаки могут поставить под угрозу жизнь пациентов, нарушая	Какие последствия могут иметь атаки на системы ИИ в медицинских устройствах?	ПК-9

	работу медицинского оборудования.		
57.	Меры включают регулярные обновления прошивки и использование безопасных протоколов связи.	Какие меры безопасности можно принять для защиты устройств IoT от атак?	ПК-9
58.	Используются adversarial attacks для обмана моделей и получения доступа к конфиденциальным данным.	Какие атаки на ИИ могут быть использованы для взлома систем машинного обучения?	ПК-9
59.	Уязвимости могут быть связаны с неадекватной защитой данных и отсутствием регулярных обновлений.	Какие уязвимости существуют в алгоритмах ИИ, используемых в умных домах?	ПК-9
60.	Используются алгоритмы для анализа поведения пользователей и создания персонализированных фишинговых писем.	Какие методы используются для создания персонализированных фишинговых атак с помощью ИИ?	ПК-9
61.	Атаки могут привести к финансовым потерям и нарушению доверия к финансовым учреждениям.	Какие последствия могут иметь атаки на системы ИИ в финансовой сфере?	ПК-9

62.	Меры включают регулярные аудиты безопасности и использование методов защиты от adversarial attacks.	Какие меры безопасности можно принять для защиты от атак на ИИ в финансовых учреждениях?	ПК-9
63.	Используются методы ИИ для создания более сложных и скрытных вредоносных программ.	Какие атаки на ИИ могут быть использованы для создания вредоносных программ?	ПК-9
64.	Уязвимости могут быть связаны с неадекватной защитой данных и отсутствием регулярных обновлений.	Какие уязвимости существуют в системах ИИ, используемых в автомобильной промышленности?	ПК-9
65.	Используются методы мониторинга и анализа поведения систем.	Какие методы используются для обнаружения и предотвращения атак на ИИ в автомобилях?	ПК-9
66.	Атаки могут поставить под угрозу безопасность полетов и жизни пассажиров.	Какие последствия могут иметь атаки на системы ИИ в авиационной промышленности?	ПК-9
67.	Меры включают регулярные обновления систем и использование методов защиты от adversarial attacks.	Какие меры безопасности можно принять для защиты от атак на ИИ в авиации?	ПК-9
68.	Используются методы ИИ для автоматизации создания и управления ботнетами.	Какие атаки на ИИ могут быть использованы для создания ботнет-сетей?	ПК-9

69.	Уязвимости могут быть связаны с неадекватной защитой данных и отсутствием регулярных обновлений.	Какие уязвимости существуют в системах ИИ, используемых в энергетической сфере?	ПК-9
70.	Используются методы мониторинга и анализа поведения систем.	Какие методы используются для обнаружения и предотвращения атак на ИИ в энергетике?	ПК-9
71.	Атаки могут привести к утечке данных и нарушению процесса обучения.	Какие последствия могут иметь атаки на системы ИИ в сфере образования?	ПК-9
72.	Меры включают регулярные обновления систем и использование методов защиты от adversarial attacks.	Какие меры безопасности можно принять для защиты от атак на ИИ в образовательных учреждениях?	ПК-9
73.	Используются методы генеративных моделей ИИ для создания реалистичных текстов и видео.	Какие атаки на ИИ могут быть использованы для создания фальшивых новостей?	ПК-9
74.	Уязвимости могут быть связаны с неадекватной защитой данных и отсутствием регулярных обновлений.	Какие уязвимости существуют в системах ИИ, используемых в правительственных учреждениях?	ПК-9
75.	Используются методы мониторинга и анализа поведения систем.	Какие методы используются для обнаружения и предотвращения атак на ИИ в правительстве?	ПК-9

76.	Атаки могут поставить под угрозу жизнь пациентов и нарушить работу медицинских учреждений.	Какие последствия могут иметь атаки на системы ИИ в сфере здравоохранения?	ПК-9
77.	Меры включают регулярные обновления систем и использование методов защиты от adversarial attacks.	Какие меры безопасности можно принять для защиты от атак на ИИ в медицинских учреждениях?	ПК-9
78.	Используются методы ИИ для автоматизации создания и управления вредоносными ботами.	Какие атаки на ИИ могут быть использованы для создания вредоносных ботов в социальных сетях?	ПК-9
79.	Уязвимости могут быть связаны с неадекватной защитой данных и отсутствием регулярных обновлений.	Какие уязвимости существуют в системах ИИ, используемых в логистике?	ПК-9
80.	Используются методы мониторинга и анализа поведения систем.	Какие методы используются для обнаружения и предотвращения атак на ИИ в логистике?	ПК-9
81.	Атаки могут привести к утечке данных и нарушению работы туристических сервисов.	Какие последствия могут иметь атаки на системы ИИ в сфере туризма?	ПК-9
82.	Меры включают регулярные обновления систем и использование методов защиты от adversarial attacks.	Какие меры безопасности можно принять для защиты от атак на ИИ в туристической индустрии?	ПК-9

83.	Перспективы включают использование ИИ для автоматизации защиты, развитие доверенного ИИ и мультимодальных моделей для обнаружения угроз.	Какие перспективы развития защиты от атак на системы ИИ существуют в ближайшем будущем?	ПК-9

2. Описание шкалы оценивания

В рамках рейтинговой системы успеваемость студентов по каждой дисциплине оценивается в ходе текущего контроля и промежуточной аттестации. Система оценки знаний студентов основана на использовании совокупности контрольных мероприятий по проверке пройденного материала, оптимально расположенных на всем временном интервале изучения дисциплины.

3. Критерии оценивания компетенций*

Оценка «отлично» выставляется студенту, если он, отлично знает как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования, отлично умеет использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования, отлично владеет навыками как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.

Оценка «хорошо» выставляется студенту, если он, знает как использовать современные методы разработки и реализации конкретных алгоритмов математически х моделей на базе языков программирования и пакетов прикладных программ моделирования, умеет использовать современные методы разработки и реализации конкретных алгоритмов математически х моделей на базе языков программирования и пакетов прикладных программ моделирования, владеет навыками как использовать современные методы разработки и реализации конкретных алгоритмов математически х моделей на базе языков программирования и пакетов прикладных программ моделирования.

Оценка «удовлетворительно» выставляется студенту, если он, плохо знает как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования, плохо умеет использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования, плохо владеет навыками как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.

Оценка «неудовлетворительно» выставляется студенту, если он, не знает как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования, не умеет использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования, не владеет навыками как использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования.