МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ "РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ В.Ф. УТКИНА"

СОГЛАСОВАНО
Зав. выпускающей кафедры

УТВЕРЖДАЮ

Защита информации в системах ИИ

рабочая программа дисциплины (модуля)

Закреплена за кафедрой **Вычислительной и прикладной математики**

Учебный план 09.03.04_25_00_ ИИ_В2.plx
09.03.04 Программная инженерия

Квалификация **бакалавр**

Форма обучения **очная**

Общая трудоемкость **3 ЗЕТ**

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	**8 (4.2)**		Итого	
Недель	8			
Вид занятий	УП	РП	УП	РП
Лекции	24	24	24	24
Лабораторные	24	24	24	24
Иная контактная работа	0,25	0,25	0,25	0,25
Итого ауд.	48,25	48,25	48,25	48,25
Контактная работа	48,25	48,25	48,25	48,25
Сам. работа	51	51	51	51
Часы на контроль	8,75	8,75	8,75	8,75
Итого	108	108	108	108

г. Рязань

Программу составил(и):
к.ф-м.н., доц, Бубнов Сергей Алексеевич;к.т.н., доц., Тишкина Валерия Валентиновна

Рабочая программа дисциплины
Защита информации в системах ИИ

разработана в соответствии с ФГОС ВО:
ФГОС ВО - бакалавриат по направлению подготовки 09.03.04 Программная инженерия (приказ Минобрнауки России от 19.09.2017 г. № 920)

составлена на основании учебного плана:
09.03.04 Программная инженерия
утвержденного учёным советом вуза от 29.08.2025 протокол № 1.

Рабочая программа одобрена на заседании кафедры
Вычислительной и прикладной математики

Протокол от 01.09.2025 г. № 1
Срок действия программы: 20252029 уч.г.
Зав. кафедрой Овечкин Геннадий Владимирович

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2026-2027 учебном году на заседании кафедры
Вычислительной и прикладной математики

Протокол от __ _____ 2026 г. № __

Зав. кафедрой _____

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2027-2028 учебном году на заседании кафедры
Вычислительной и прикладной математики

Протокол от __ _____ 2027 г. № __

Зав. кафедрой _____

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2028-2029 учебном году на заседании кафедры
Вычислительной и прикладной математики

Протокол от __ _____ 2028 г. № __

Зав. кафедрой _____

Визирование РПД для исполнения в очередном учебном году

Рабочая программа пересмотрена, обсуждена и одобрена для исполнения в 2029-2030 учебном году на заседании кафедры

Вычислительной и прикладной математики

Протокол от __ _____ 2029 г. № __

Зав. кафедрой _____

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	
1.1	Цель: Сформировать у студентов системное представление о методологии, принципах и практиках обеспечения информационной безопасности в системах искусственного интеллекта, с углубленным изучением угроз и механизмов защиты, характерных для генеративных моделей и больших языковых моделей (LLM).
1.2	В результате изучения дисциплины студент должен быть способен решать следующие профессиональные задачи:
1.3	Аналитико-теоретические задачи:
1.4	Анализировать архитектуру систем искусственного интеллекта, в частности генеративных моделей, на предмет уязвимостей и потенциальных векторов атак.
1.5	Классифицировать угрозы безопасности для систем ИИ, включая атаки на обучающие данные (Data Poisoning), атаки на вывод модели (Adversarial Attacks), кражу конфиденциальной информации из модели (Model Inversion, Membership Inference), и манипуляцию поведением генеративных моделей (prompt injection, jailbreaking).
1.6	Формулировать требования к системе защиты информации для проекта, связанного с разработкой или развертыванием систем ИИ.
1.7	Технологические и практические задачи:
1.8	Применять методы и инструменты для оценки защищенности и проведения тестирования на проникновение (pen-testing) моделей машинного обучения, в том числе с использованием фреймворков для генерации состязательных примеров (adversarial examples).
1.9	Разрабатывать и реализовывать базовые механизмы защиты для систем ИИ, включая методы дифференциальной приватности (Differential Privacy) для защиты данных обучения, механизмы устойчивости к состязательным атакам (Adversarial Training), и контроль доступа к API и весам моделей.
1.10	Оценивать риски, связанные с развертыванием генеративных моделей (таких как LLM), и предлагать архитектурные и процедурные меры по их минимизации.
1.11	Интеграционно-системные задачи:
1.12	Интегрировать рассмотренные методы защиты в общий жизненный цикл разработки программного обеспечения (SDLC) и DevOps/MLOps-процессы.
1.13	Соотносить угрозы для систем ИИ с классическими угрозами информационной безопасности и выстраивать комплексную систему защиты.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	
Цикл (раздел) ОП:	Б1.О
2.1	**Требования к предварительной подготовке обучающегося:**
2.1.1	Учебная практика
2.1.2	Физика
2.1.3	Дискретная математика
2.1.4	Программирование
2.1.5	Технологическая практика
2.1.6	Ознакомительная практика
2.1.7	Алгоритмы и структуры данных
2.1.8	Большие языковые модели и генеративный ИИ
2.1.9	Цифровая обработка сигналов
2.1.10	Введение в профессиональную деятельность
2.1.11	Тестирование ПО
2.1.12	Вычислительные алгоритмы для систем ИИ
2.1.13	Производственная практика
2.2	**Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:**
2.2.1	Выполнение и защита выпускной квалификационной работы

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)
ОПК-3: Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности;
ОПК-3.1. Владеет информационной и библиографической культурой

Знать
Специализированные источники информации: ключевые международные и российские научные журналы, конференции и репозитории (такие как arXiv, IEEE Xplore, ACM Digital Library, Springer LNCS), публикующие передовые исследования в области безопасности искусственного интеллекта и машинного обучения.
Базы знаний и стандарты: основные профильные организации и базы знаний (например, OWASP AI Security & Privacy Guide, MITRE ATLAS — Adversarial Threat Landscape for Artificial-Intelligence Systems, NIST guidelines on AI risk management).
Библиографические требования: правила оформления научных ссылок, цитирования и составления библиографических списков по стандартам, принятым в вузе (например, ГОСТ), при работе с научно-технической литературой.
Критерии оценки информации: принципы верификации информации, критерии оценки достоверности и актуальности научных и технических источников в быстроразвивающейся области ИИ.
Уметь
Эффективно искать информацию: проводить целенаправленный поиск научных статей, технических отчетов, документации к фреймворкам и нормативных документов по темам дисциплины (защита генеративных моделей, атаки на ИИ, дифференциальная приватность и т.д.).
Анализировать и синтезировать информацию: критически оценивать найденные источники, выделять ключевые идеи, сравнивать различные подходы к решению задач защиты информации в системах ИИ и формулировать на их основе собственные выводы.
Систематизировать информацию: структурировать полученные знания для использования в учебной (подготовка рефератов, презентаций) и практической деятельности (выполнение курсовых проектов, выпускной квалификационной работы).
Работать с источниками на английском языке: читать и понимать основную смысловую нагрузку технических и научных текстов на английском языке, являющемся основным для данной предметной области.
Владеть
Навыками библиографирования: практическими навыками оформления цитат, ссылок и библиографических списков в соответствии с заданным стандартом.
Навыками информационного поиска: технологиями и методами эффективного поиска и отбора релевантной информации с использованием специализированных поисковых систем, баз данных и научных библиотек.
Навыками академического письма: навыками корректного использования информации из найденных источников в своих письменных работах (отчетах, рефератах) без нарушения академической добросовестности.
Информационной культурой: способностью самостоятельно поддерживать и обновлять свои знания в области защиты информации в системах ИИ, отслеживая новые публикации и тенденции.

ОПК-3.2. Решает задачи профессиональной деятельности с применением информационно-коммуникационных технологий с учетом основных требований информационной безопасности

Знать
Основы ИКТ: принципы работы и особенности современных информационно-коммуникационных технологий (ИКТ), используемых в жизненном цикле систем ИИ (облачные платформы, системы версионного контроля, контейнеризация, инструменты MLOps).
Угрозы информационной безопасности: ключевые угрозы ИБ, актуальные для каждого этапа работы с системами ИИ: от сбора данных и обучения до развертывания и эксплуатации модели.
Требования ИБ: основные требования, политики и лучшие практики обеспечения информационной безопасности при использовании ИКТ в профессиональной деятельности, в том числе: при работе с данными (конфиденциальность, целостность), при доступе к вычислительным ресурсам и моделям (аутентификация, авторизация), при сетевом взаимодействии компонентов системы.
Нормативную базу: базовые положения требований, стандартов и рекомендаций (как международных, так и российских), касающихся защиты информации и данных при разработке и использовании систем ИИ.
Уметь
Выбирать инструменты с учетом ИБ: аргументированно выбирать программные и аппаратные средства (ИКТ) для решения профессиональных задач (например, выбор фреймворка для машинного обучения, облачного провайдера, инструмента развертывания) с учетом их возможностей для обеспечения безопасности.
Анализировать риски: проводить первичный анализ рисков информационной безопасности, возникающих при применении конкретных ИКТ для построения систем ИИ (например, риски утечки данных из облачного хранилища, риски инъекции вредоносного кода).
Реализовывать базовые меры защиты: настраивать и применять базовые механизмы защиты при использовании ИКТ.
Организовывать безопасное хранение и передачу данных и моделей (шифрование, VPN).
Настраивать управление доступом к ресурсам и API моделей.
Использовать инструменты для безопасной разработки и развертывания (например, сканирование контейнеров на уязвимости).
Эксплуатировать системы ИИ с учетом ИБ: обеспечивать безопасную эксплуатацию развернутых моделей, включая мониторинг аномальной активности, подозрительных запросов и попыток атак.
Владеть
Навыками безопасной настройки ИКТ: практическими навыками безопасной конфигурации программных сред и инструментов, используемых для разработки и исследований в области ИИ (настройка прав доступа, использование токенов и ключей, работа в изолированных средах).
Навыками применения защитных механизмов: навыками использования встроенных и сторонних средств защиты (например, механизмы шифрования в СУБД, средства аутентификации в веб-фреймворках, инструменты для обеспечения дифференциальной приватности в библиотеках машинного обучения) при решении практических задач.
Культурой безопасного проектирования: способностью проектировать и реализовывать решения в области ИИ, изначально учитывающие требования информационной безопасности, а не добавляющие их постфактум.

ОПК-6: Способен разрабатывать алгоритмы и программы, пригодные для практического использования, применять основы информатики и программирования к проектированию, конструированию и тестированию программных продуктов;

ОПК-6.1. Выполняет разработку алгоритмического и программного обеспечения для решения прикладных задач

Знать
Архитектурные принципы безопасных систем: принципы проектирования алгоритмов и программного обеспечения для систем ИИ, устойчивых к атакам (например, принцип минимальных привилегий, безопасная обработка ввода, разделение ответственности между компонентами).
Алгоритмы и методы защиты: конкретные алгоритмы и методики, используемые для защиты систем ИИ:
Алгоритмы повышения устойчивости моделей (Adversarial Training, детектирование состязательных примеров).
Методы защиты конфиденциальности данных (Дифференциальная приватность, гомоморфное шифрование, федеративное обучение — на концептуальном уровне).
Алгоритмы мониторинга и анализа поведения модели на предмет аномалий.
Уязвимости на уровне кода: типовые уязвимости, возникающие при реализации компонентов системы ИИ (например, инъекция в промпты, небезопасная десериализация моделей, ошибки в функциях обработки пользовательского ввода).
Инструменты безопасной разработки: назначение и возможности использования инструментов статического и динамического анализа кода, сканеров уязвимостей в зависимостях (SAST, DAST, SCA) при разработке ПО для ИИ.
Уметь
Проектировать защищенные алгоритмы: разрабатывать и формализовывать алгоритмы ключевых компонентов системы ИИ, учитывающие требования безопасности (например, алгоритм предобработки входных данных для фильтрации вредоносных входов, алгоритм логирования и аудита запросов к модели).
Реализовывать механизмы защиты: переводить теоретические методы защиты в работающий программный код на языках программирования, таких как Python, используя соответствующие библиотеки и фреймворки (например, реализовать базовый механизм атаки или защиты для учебной модели).
Интегрировать средства безопасности: встраивать готовые библиотеки и инструменты обеспечения безопасности в общий конвейер разработки и эксплуатации (MLOps pipeline) системы ИИ.
Тестировать безопасность реализаций: проводить базовое тестирование разработанного программного обеспечения на устойчивость к известным типам атак (например, тестирование модели на состязательных примерах с использованием библиотек like Adversarial Robustness Toolbox).
Владеть
Навыками защитного программирования (Secure Coding): практическими навыками написания кода, устойчивого к эксплуатации, для компонентов, взаимодействующих с моделью ИИ (веб-интерфейсы, API, обработчики данных).
Навыками использования специализированных библиотек: технологиями работы с фреймворками и библиотеками для безопасности машинного обучения (например, IBM Adversarial Robustness Toolbox, Microsoft Counterfit, CleverHans) в процессе разработки.
Методами верификации и валидации защитных свойств: способностью применять методы тестирования и оценки для проверки того, что реализованные алгоритмы и программное обеспечение выполняют свои защитные функции.

В результате освоения дисциплины (модуля) обучающийся должен

3.1	**Знать:**
3.1.1	Ключевые угрозы безопасности систем ИИ: атаки на данные, модели (включая генеративные) и эксплуатационную инфраструктуру.
3.1.2	Принципы и методы защиты систем ИИ на всех этапах жизненного цикла (дифференциальная приватность, мониторинг).
3.1.3	Основы безопасной разработки и эксплуатации ПО для ИИ (архитектура, инструменты, лучшие практики).
3.2	**Уметь:**
3.2.1	Анализировать архитектуру систем ИИ на предмет уязвимостей и формулировать требования к защите.
3.2.2	Применять методы и инструменты для оценки защищенности и тестирования моделей машинного обучения.
3.2.3	Разрабатывать и реализовывать базовые механизмы защиты для компонентов систем ИИ.
3.3	**Владеть:**
3.3.1	Навыками защитного программирования (Secure Coding) для компонентов, взаимодействующих с моделями ИИ.
3.3.2	Технологиями работы с специализированными фреймворками и инструментами для безопасности ML/AI.
3.3.3	Навыками поиска и критического анализа актуальной информации в области безопасности систем ИИ.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетен-ции	Литература	Форма контроля
	Раздел 1. Защита информации					
1.1	Теоретические основы защиты информации. /Тема/	8	0			

1.2	Введение в предметную область, изложение основных понятий, определения, освещение исторического контекста защиты информации. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1	Зачёт
1.3	Защита программ от нелегального копирования. Проектирование информационных систем с точки зрения безопасности. Моделирование угроз и нарушителей. Риски.Требования к алгоритмам шифрования. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.4	Правило Огюста Керкхоффа. Симметричные алгоритмы. Поточное шифрование. Блочные алгоритмы. Шифры перестановки. Шифры замены. Моно- и полиалфавитные шифры. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.5	Индекс соответствия. Шифровальная машина Энигма. Шифрование с открытым ключом. Применение генераторов случайных чисел. Алгоритм Фон Неймана. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.6	Реализация алгоритма шифрования с открытым ключом (RSA) /Лаб/	8	3	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.7	Реализация алгоритма шифрования с открытым ключом (RSA) /Лаб/	8	3	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.8	Введение в предметную область, изложение основных понятий, определения, освещение исторического контекста защиты информации. /Ср/	8	3	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт

1.9	Защита программ от нелегального копирования. Проектирование информационных систем с точки зрения безопасности. Моделирование угроз и нарушителей. Риски.Требования к алгоритмам шифрования. /Ср/	8	3	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.10	Правило Огюста Керкхоффа. Симметричные алгоритмы. Поточное шифрование. Блочные алгоритмы. Шифры перестановки. Шифры замены. Моно- и полиалфавитные шифры. /Ср/	8	3	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.11	Индекс соответствия. Шифровальная машина Энигма. Шифрование с открытым ключом. Применение генераторов случайных чисел. Алгоритм Фон Неймана. /Ср/	8	4	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.12	Алгоритмы защиты информации и их практическая реализация. /Тема/	8	0			
1.13	Алгоритм на числах Фибоначи. Линейный конгруэнтный генератор. Квадратичный конгруэнтный генератор. Нормальное (Гауссово) распределение. Особенности практического применения симметричного шифрования. Примеры алгоритмов шифрования. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.14	Алгоритм DataEncryptionStandard (DES). Этапы шифрования DES. Функция Фейстеля. Взлом шифра DES. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.15	Открытый конкурс NationalInstituteofStandardsandTechnology (NIST). Алгоритм AdvancedEncryptionStandard. История и особенности асимметричного шифрования. Алгоритм RSA. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.16	Исследование устойчивости модели классификации изображений к состязательным атакам (Adversarial Attacks). Анализ рисков конфиденциальности в моделях машинного обучения с помощью Membership Inference Attack. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт

1.17	Реализация алгоритма шифрования с открытым ключом (RSA) /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.18	Реализация алгоритма симметричного шифрования (DES) /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.19	Реализация алгоритма симметричного шифрования (AES) /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.20	Исследование устойчивости модели классификации изображений к состязательным атакам (Adversarial Attacks) /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.21	Анализ рисков конфиденциальности в моделях машинного обучения с помощью Membership Inference Attack /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.22	Алгоритм на числах Фибоначи. Линейный конгруэнтный генератор. Квадратичный конгруэнтный генератор. Нормальное (Гауссово) распределение. /Ср/	8	3	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.23	Особенности практического применения симметричного шифрования. Примеры алгоритмов шифрования. Алгоритм DataEncryptionStandard (DES). Этапы шифрования DES. Функция Фейстеля. Взлом шифра DES. Усовершенствования DES. /Ср/	8	5	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.24	Защита информации в системах искусственного интеллекта. Вопросы, смежные с криптографией. /Тема/	8	0			

1.25	Методы и алгоритмы сжатия данных. Алгоритм сжатия Хафмана. Алгоритм сжатия LZW. Математическое сжатие. Криптоатаки и методы криптоанализа. Стеганография. Правовые основы защиты информации. Механизмы защиты для систем ИИ, включая методы дифференциальной приватности (Differential Privacy) для защиты данных обучения, механизмы устойчивости к состязательным атакам (Adversarial Training), и контроль доступа к API и весам моделей. /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.26	Классификация угроз безопасности для систем ИИ, включая атаки на обучающие данные (Data Poisoning), атаки на вывод модели (Adversarial Attacks), кражу конфиденциальной информации из модели (Model Inversion, Membership Inference), и манипуляцию поведением генеративных моделей (prompt injection, jailbreaking). /Лек/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.27	Применение дифференциальной приватности для защиты тренировочных данных. /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.28	Создание и проверка электронной подписи для документа /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.29	Анализ уязвимостей в MLOps-конвейере: атака через подмену модели (Model Replacement Attack) /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.30	Алгоритм сжатия Хафмана. Алгоритм сжатия LZW. Математическое сжатие. Классификация угроз безопасности для систем ИИ, включая атаки на обучающие данные (Data Poisoning), атаки на вывод модели (Adversarial Attacks), кражу конфиденциальной информации из модели (Model Inversion, Membership Inference), и манипуляцию поведением генеративных моделей (prompt injection, jailbreaking). Применение дифференциальной приватности для защиты тренировочных данных. /Ср/	8	10	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт

1.31	Безопасность систем искусственного интеллекта. Безопасность больших языковых моделей /Лек/	8	4	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.32	Криптоатаки и методы криптоанализа.Защита информации систем искусственного интеллекта и больших языковых моделей. Анализ уязвимостей в MLOps-конвейере: атака через подмену модели (Model Replacement Attack). /Ср/	8	10	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.33	Исследование уязвимости генеративной языковой модели к атакам типа "Инъекция промптов" (Prompt Injection) /Лаб/	8	2	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В		
1.34	Механизмы защиты для систем ИИ, включая методы дифференциальной приватности (Differential Privacy) для защиты данных обучения, механизмы устойчивости к состязательным атакам (Adversarial Training), и контроль доступа к API и весам моделей. /Ср/	8	10	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В		
1.35	Промежуточная аттестация /Тема/	8	0			
1.36	Подготовка к зачёту /Зачёт/	8	8,75	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1 Э1	Зачёт
1.37	Прием зачета /ИКР/	8	0,25	ОПК-3.1-З ОПК-3.1-У ОПК-3.1-В ОПК-3.2-З ОПК-3.2-У ОПК-3.2-В ОПК-6.1-З ОПК-6.1-У ОПК-6.1-В	Л1.1 Л1.2 Л1.3Л2.2 Л2.1Л3.2 Л3.1	

5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Основные материалы приведены в приложении к рабочей программе дисциплины (см. документ "Оценочные материалы по дисциплине "Защита информации в системах искусственного интеллекта")

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Рекомендуемая литература

6.1.1. Основная литература				
№	Авторы, составители	Заглавие	Издательство, год	Количество/ название ЭБС
Л1.1	Башлы П. Н., Бабаш А. В., Баранова Е. К.	Информационная безопасность и защита информации : учебное пособие	Москва: Евразийский открытый институт, 2012, 311 с.	978-5-374-00301-7, http://www.iprbookshop.ru/10677.html
Л1.2	Евдокимова Л.М., Корябкин В.В., Пылькин А.Н., Швечкова О.Г.	Электронный документооборот и обеспечение безопасности стандартными средствами WINDOWS : учеб. пособие	М.: КУРС, 2017, 294с.; прил.	978-5-906923-24-0,978-5-16-012741-5, 1
Л1.3	Швечкова О.Г., Пылькин А.Н., Марчев Д.В.	Базовые криптографические алгоритмы защиты информации: учеб. пособие : Учебное пособие	Рязань: КУРС, 2023,	, https://elib.rsreu.ru/ebs/download/3643
6.1.2. Дополнительная литература				
№	Авторы, составители	Заглавие	Издательство, год	Количество/ название ЭБС
Л2.1	Омарова С. А., Искакова К. А., Тойганбаева Н. А.	Информационная безопасность и защита информации : учебно-методический комплекс	Алматы: Нур-Принт, 2012, 98 с.	9965-756-05-8, http://www.iprbookshop.ru/67055.html
Л2.2	Нерсесянц А. А.	Защита информации : учебное пособие	Ростов-на-Дону: Северо-Кавказский филиал Московского технического университета связи и информатики, 2010, 61 с.	2227-8397, http://www.iprbookshop.ru/61295.html
6.1.3. Методические разработки				
№	Авторы, составители	Заглавие	Издательство, год	Количество/ название ЭБС
Л3.1	Малинин Ю.И.	Информационная безопасность и защита информации : Методические указания	Рязань: РИЦ РГРТУ, 2009,	, https://elib.rsreu.ru/ebs/download/851
Л3.2	Каторин Ю. Ф., Разумовский А. В., Спивак А. И., Каторин Ю. Ф.	Техническая защита информации : лабораторный практикум	Санкт-Петербург: Университет ИТМО, 2013, 113 с.	2227-8397, http://www.iprbookshop.ru/68715.html
6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"				
Э1	Электронная библиотека РГРТУ			

6.3 Перечень программного обеспечения и информационных справочных систем

6.3.1 Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

Наименование	Описание
PyCharm Community	Свободное ПО
PyCharm	Свободное ПО
Интерпретатор Python	Свободное ПО
Python	Свободно распространяемое программное обеспечение под лицензиями

Python	Свободное ПО
Среда разработки Qt Creator	Свободное ПО
Qt Creator Community	Свободное ПО
Qt	Лицензия Opensource
6.3.2 Перечень информационных справочных систем	
6.3.2.1	Система КонсультантПлюс http://www.consultant.ru

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)	
1	206-1 учебно-административный корпус. Учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации 42 мест, 1 ПК: ЦП: Intel i7 6700 ОЗУ: 8 Гб ПЗУ: 120Gb ssd + 500 Гб (1 шт. преподавательский) Телевизор: PHILIPS U7PEL4606H/60 документ-камера: AVER Media POB3 (AverVision 330) документ-камера: AverVisionF33 POE7D;
2	206-2 учебно-административный корпус. Аудитория для самостоятельной работы 18 мест, Телевизор PHILIPS 46PFL3208T/60; 18 ПК с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду: ЦП: AMD a10 7800 ОЗУ: 8 Gb ПЗУ: 500-1000 Gb (18 шт.)
3	206-3 учебно-административный корпус. Учебная аудитория для проведения практический занятий, лабораторных работ Проектор: InFocus LP640 31 ПК с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду: ЦП: Intel i7 6700 ОЗУ: 8 Гб ПЗУ: 120Gb ssd + 500 Гб (31 шт.))
4	206-4 учебно-административный корпус. Учебная аудитория для проведения практический занятий, лабораторных работ 30 ПК с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду: ЦП: Intel i5-12400 ОЗУ: 16 Гб ПЗУ: ssd 512 Гб (18 шт. моноблок) ЦП: Intel i7 6700 ОЗУ: 8 Гб ПЗУ: 120Gb ssd + 500 Гб (12 шт.)

8. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)
Методическое обеспечение дисциплины приведено в приложении к рабочей программе дисциплины (см. документ "Методические материалы по дисциплине "Защита информации в системах искусственного интеллекта")

Оператор ЭДО ООО "Компания "Тензор"

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

ПОДПИСАНО ЗАВЕДУЮЩИМ КАФЕДРЫ — **ФГБОУ ВО "РГРТУ", РГРТУ,** Овечкин Геннадий Владимирович, Заведующий кафедрой ВПМ — **03.11.26** 16:20 (MSK) — Простая подпись

ПОДПИСАНО ЗАВЕДУЮЩИМ ВЫПУСКАЮЩЕЙ КАФЕДРЫ — **ФГБОУ ВО "РГРТУ", РГРТУ,** Овечкин Геннадий Владимирович, Заведующий кафедрой ВПМ — **03.11.26** 16:20 (MSK) — Простая подпись