МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(МОСКОВСКИЙ ПОЛИТЕХ)

Факультет информационных технологий

УТВЕРЖДАЮ
Декан факультета
«Информационных технологий»
_____ / Д.Г. Демидов /
«26» февраля 2026 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Технологии искусственного интеллекта в информационной безопасности»

Направление подготовки
10.04.01 «Информационная безопасность»

Профиль/специализация
«Системы управления информационной безопасностью»

Квалификация
Магистр

Формы обучения
Очная

Москва, 2026 г.

Разработчик(и):

Доцент, к.т.н., доцент  /И.В. Калуцкий/

Согласовано:
Заведующий кафедрой «Информационная безопасность»  И.В.Калуцкий

Содержание

1. Цели, задачи и планируемые результаты обучения по дисциплине

К основным **целям** освоения дисциплины «Технологии искусственного интеллекта в информационной безопасности» следует отнести:

- изучение современных методов и технологий искусственного интеллекта и машинного обучения для решения ключевых задач обеспечения информационной безопасности, включая обнаружение, анализ, прогнозирование и противодействие киберугрозам;
- формирование навыков проектирования, разработки и внедрения интеллектуальных систем и подсистем защиты информации с использованием ИИ;
- развитие компетенций по обеспечению безопасности самих систем искусственного интеллекта;
- подготовка специалистов, способных интегрировать технологии ИИ в системы информационной безопасности критически важных объектов и инфраструктур.

К основным **задачам** освоения дисциплины «Технологии искусственного интеллекта в информационной безопасности» следует отнести:

- овладение теоретическими основами и практическими инструментами машинного обучения, глубокого обучения и других методов ИИ, применимых в задачах кибербезопасности;
- освоение методов анализа больших объёмов данных для выявления аномалий, инцидентов и признаков кибератак;
- изучение подходов к построению и оценке эффективности интеллектуальных систем обнаружения вторжений (IDS/IPS), анализа вредоносного ПО, поведенческого анализа и SIEM-систем нового поколения;
- приобретение навыков защиты моделей ИИ от специализированных атак и обеспечения их устойчивости, интерпретируемости и соответствия требованиям регуляторов;
- формирование умений разрабатывать технические проекты интеллектуальных компонентов систем обеспечения информационной безопасности.

Обучение по дисциплине «Технологии искусственного интеллекта в информационной безопасности» направлено на формирование у обучающихся следующих компетенций:

Код и наименование компетенций	Индикаторы достижения компетенции
ОПК-2. Способен разрабатывать технический проект системы (подсистемы либо компонента системы) обеспечения информационной безопасности	уметь: • разрабатывать технический проект системы (подсистемы либо компонента системы) обеспечения информационной безопасности.
ОПК-5. Способен проводить научные исследования, включая экспериментальные, обрабатывать результаты исследований, оформлять научно-технические отчеты, обзоры, готовить по результатам выполненных исследований научные доклады и статьи	уметь: • проводить научные исследования, включая экспериментальные, обрабатывать результаты исследований, оформлять научно—технические отчеты, обзоры,

	готовить по результатам выполненных исследований научные доклады и статьи.
ПК-7. Способен разрабатывать программы и методики испытаний средств и систем обеспечения информационной безопасности	знать: • методы экспериментальных исследований защищенности объектов с применением соответствующих физических и математических методов, технических и программных средств обработки результатов эксперимента; уметь: • проводить экспериментальные исследования защищенности объектов с применением соответствующих физических и математических методов, технических и программных средств обработки результатов эксперимента; владеть: • навыками проведения экспериментальных исследований защищенности объектов с применением соответствующих физических и математических методов, технических и программных средств обработки результатов эксперимента.

2. Цели, задачи и планируемые результаты обучения по дисциплине

Дисциплина «Технологии искусственного интеллекта в информационной безопасности» относится к числу учебных дисциплин блока 1 обязательной части (Б1.1) основной образовательной программы (Б1.1.7).

Дисциплина логически и содержательно-методически взаимосвязана со следующими дисциплинами и практиками ООП: «Математические методы информационной безопасности», «Управление информационной безопасностью», «Построение и совершенствование систем управления информационной безопасностью».

3. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы (144 часа).

3.1. Виды учебной работы и трудоемкость

(по формам обучения)

3.1.1 Очная форма обучения

№ п/п	Вид учебной работы	Количество часов	Семестры	
			Семестр	Неделя семестра
1	**Аудиторные занятия**	**72**	3	
	В том числе:			
1.1	Лекции	18		
1.2	Семинарские/практические занятия	-		
1.3	Лабораторные занятия	**54**	3	
2	**Самостоятельная работа**	**72**	3	
3	**Промежуточная аттестация**			
	Экзамен		3	По расписанию
	Итого	**144**		

3.2. Тематический план изучения дисциплины

(по формам обучения)

3.2.1 Очная форма обучения

№ п/п	Разделы/темы дисциплины	Трудоемкость, час					
		Всего	Аудиторная работа				Самостоятельная работа
			Лекции	Семинарские/ практические занятия	Лабораторные занятия	Практическая подготовка	
1	Раздел 1. Основы искусственного интеллекта и машинного обучения в задачах ИБ						
1.1	Тема 1. Введение в ИИ и ML: история, типы задач (supervised, unsupervised, reinforcement) в кибербезопасности	10	1		4		5
1.2	Тема 2. Основные математические основы: векторные пространства, метрики расстояний, вероятностные модели	12	2		5		5
1.3	Тема 3. Классификация и регрессия в задачах ИБ (логистическая регрессия, SVM, деревья решений, случайный лес)	10	1		4		5
1.4	Тема 4. Нейронные сети и глубокое обучение: основы, архитектуры (MLP, CNN, RNN/LSTM) для анализа угроз	13	2		5		6
1.5	Тема 5. Обучение без учителя: кластеризация, обнаружение аномалий (k-means, DBSCAN, автоэнкодеры, Isolation Forest)	9	1		3		5
2	Раздел 2. Применение технологий ИИ для обнаружения и противодействия угрозам						
2.1	Тема 1. Системы обнаружения вторжений (IDS/IPS) на основе ML: анализ трафика, поведенческий анализ	9	2		2		5
2.2	Тема 2. Обнаружение и классификация вредоносного ПО с использованием ИИ (статический/динамический анализ, ВПО family classification)	9	1		2		6
2.3	Тема 3. Анализ больших данных и SIEM нового поколения: обработка логов, threat hunting с ML	8	1		2		5

2.4	Тема 4. Обнаружение фишинга, спама и социальной инженерии с помощью NLP и ИИ	9	1		3		5
2.5	Тема 5. Прогнозирование и предиктивная аналитика киберугроз (time-series forecasting, графовые нейросети)	11	1		5		5
3	Раздел 3. Безопасность самих систем искусственного интеллекта						
3.1	Тема 1. Угрозы ИИ-систем: adversarial attacks, data poisoning, model inversion, extraction	12	2		5		5
3.2	Тема 2. Методы защиты моделей ML: adversarial training, defensive distillation, сертифицированная robustness	11	1		5		5
3.3	Тема 3. Обеспечение интерпретируемости в задачах ИБ (SHAP, LIME, counterfactuals)	11	1		5		5
3.4	Тема 4. Стандарты и регуляции безопасности ИИ (NIST AI RMF, ISO/IEC, российские требования к ИИ в КИИ)	10	1		4		5
Итого		**144**	**18**		**54**		**72**

3.3. Содержание дисциплины

Раздел 1. Основы искусственного интеллекта и машинного обучения в задачах информационной безопасности

Тема 1.1. Введение в ИИ и машинное обучение: история, типы задач в кибербезопасности

История развития искусственного интеллекта и машинного обучения. Основные этапы: от экспертных систем до глубокого обучения. Современные тенденции применения ИИ в информационной безопасности. Типы задач машинного обучения: обучение с учителем, без учителя, с подкреплением. Примеры применения в ИБ: классификация угроз, обнаружение аномалий, поведенческий анализ.

Тема 1.2. Основные математические основы

Векторные пространства и линейная алгебра в задачах ML. Метрики расстояний: евклидова, манхэттенская, косинусное сходство. Вероятностные модели и байесовский подход. Основы оптимизации: градиентный спуск, стохастический градиентный спуск. Функции потерь и метрики качества моделей в задачах информационной безопасности.

Тема 1.3. Классификация и регрессия в задачах информационной безопасности

Методы классификации: логистическая регрессия, метод опорных векторов, деревья решений, ансамбли: Random Forest, Gradient Boosting. Применение для классификации вредоносного ПО(ВПО), обнаружения фишинга, категоризации сетевых атак. Методы регрессии для оценки рисков и предсказания времени жизни уязвимостей.

Тема 1.4. Нейронные сети и глубокое обучение: основы, архитектуры для анализа угроз

Принципы работы искусственных нейронных сетей. Основные архитектуры: многослойный перцептрон, свёрточные нейронные сети, рекуррентные сети, трансформеры.

Применение в задачах ИБ: анализ сетевого трафика, последовательный анализ логов и сессий, обработка естественного языка в фишинговых письмах и вредоносных скриптах.

Тема 1.5. Обучение без учителя: кластеризация, обнаружение аномалий

Методы кластеризации: k-means, иерархическая кластеризация, DBSCAN. Алгоритмы обнаружения аномалий: Isolation Forest, One-Class SVM, автоэнкодеры, GAN-based anomaly detection. Применение для выявления неизвестных угроз, обнаружения вторжений нулевого дня, поиска аномалий в поведении пользователей и систем.

Раздел 2. Применение технологий ИИ для обнаружения и противодействия угрозам

Тема 2.1. Системы обнаружения вторжений на основе машинного обучения

Гибридные системы IDS/IPS: сигнатурные и ML-компоненты. Анализ сетевого трафика. Поведенческий анализ пользователей и сущностей. Примеры современных решений.

Тема 2.2. Обнаружение и классификация вредоносного ПО с использованием ИИ

Статический анализ и динамический анализ. Семейственная классификация ВПО. Глубокое обучение на сырых байтах. Проблема концептуального дрейфа и методы адаптивного переобучения моделей.

Тема 2.3. Анализ больших данных и SIEM нового поколения

Обработка и обогащение больших объёмов логов и телеметрии. Threat hunting с применением ML. Автоматическая корреляция событий и построение цепочек атак. Интеграция ИИ в SIEM-системы.

Тема 2.4. Обнаружение фишинга, спама и социальной инженерии с помощью NLP и ИИ

Обработка естественного языка: токенизация, эмбеддинги. Классификация фишинговых сообщений и URL. Детекция дипфейков в текстах и мультимодальных атаках. Анализ социальной инженерии в мессенджерах и соцсетях.

Тема 2.5. Прогнозирование и предиктивная аналитика киберугроз

Анализ временных рядов. Графовые нейронные сети для моделирования распространения угроз и инфраструктуры злоумышленников. Предсказание целей атак, вероятности эксплуатации уязвимостей, динамики киберкампаний.

Раздел 3. Безопасность самих систем искусственного интеллекта

Тема 3.1. Угрозы системам искусственного интеллекта и машинного обучения

Adversarial attacks. Отравление данных. Инверсия модели, извлечение модели, членство в обучающей выборке. Атаки на федеративное обучение и другие распределённые сценарии.

Тема 3.2. Методы защиты моделей машинного обучения

Adversarial training. Defensive distillation. Сжатие градиентов и добавление шума. Сертифицированная устойчивость. Техники против отравления данных. Watermarking моделей.

Тема 3.3. Обеспечение интерпретируемости AI в задачах ИБ

Методы интерпретируемости: LIME, SHAP, Integrated Gradients, counterfactual explanations. Построение правил и деревьев решений из чёрных ящиков.

Тема 3.4. Стандарты и регуляции безопасности ИИ

NIST AI Risk Management Framework. ISO/IEC 42001. Европейский AI Act. Российские требования к ИИ в объектах КИИ. Рекомендации по аудиту и сертификации ИИ-систем в сфере информационной безопасности.

3.4. Тематика семинарских/практических и лабораторных занятий

3.4.1 Семинарские/практические занятия

Не предусмотрены программой.

3.4.2 Лабораторные занятия

Лабораторная работа 1.1: реализовать и сравнить логистическую регрессию, SVM и деревья решений на датасетах фишинга, вредоносного ПО или сетевых аномалий. Оценить accuracy, precision, recall, F1, ROC-AUC

Лабораторная работа 1.2: реализовать Random Forest, XGBoost, LightGBM, CatBoost для классификации вредоносного ПО и обнаружения вторжений. Обработать дисбаланс классов. Проанализировать важность признаков

Лабораторная работа 1.3: создать и обучить MLP и CNN в PyTorch, TensorFlow или Keras. Применить CNN к сырым байтам PE-файлов или изображениям вредоносного ПО

Лабораторная работа 1.4: реализовать RNN, LSTM, GRU для анализа последовательностей сетевых сессий, логов или цепочек поведения. Обнаружить аномалии во временных рядах

Лабораторная работа 1.5: реализовать Isolation Forest, One-Class SVM и автоэнкодеры для обнаружения аномалий в сетевом трафике, UEBA или SIEM-логах. Оценить по ROC-AUC и Precision@K

Лабораторная работа 2.1: построить ML-систему обнаружения вторжений с извлечением признаков из сетевого трафика и обучением на датасетах CIC-IDS2017 или CIC-IDS2018

Лабораторная работа 2.2: выполнить статический анализ PE-файлов и классификацию с помощью ML или глубоких моделей на сырых байтах на датасетах EMBER или Microsoft Malware

Лабораторная работа 2.3: извлечь признаки из URL, писем или HTML. Дообучить BERT, DistilBERT или RuBERT для классификации фишинга на открытых датасетах

Лабораторная работа 3.1: реализовать adversarial-атаки FGSM, PGD, CW на модели обнаружения вредоносного ПО или IDS. Оценить уязвимость моделей

Лабораторная работа 3.2: применить adversarial training, добавление шума, defensive distillation или randomized smoothing. Сравнить устойчивость защищённой и базовой модели

Лабораторная работа 3.3: использовать SHAP, LIME или Integrated Gradients для объяснения предсказаний моделей в задачах информационной безопасности. Визуализировать важные признаки

Лабораторная работа 4.1: интегрировать ML-модель в пайплайн обработки логов SIEM. Настроить автоматическое обнаружение аномалий и генерацию алертов

Лабораторная работа 5.1: протестировать все разработанные модели, сравнить метрики качества, время обучения и инференса, устойчивость к атакам. Провести кросс-валидацию

Лабораторная работа 5.2: нарисовать блок-схемы пайплайнов предобработки, обучения и инференса, а также архитектуры нейронных сетей

Лабораторная работа 5.3: создать простой интерфейс на Streamlit, Flask или в Jupyter для демонстрации работы моделей с визуализацией предсказаний и объяснений

3.5. Тематика курсовых проектов (курсовых работ)

Не предусмотрены программой.

4. Учебно-методическое и информационное обеспечение

4.1. Нормативные документы и ГОСТы

1. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных»: Текст: электронный // КонсультантПлюс. – URL: http://www.consultant.ru/document/cons_doc_LAW_61801/ (дата обращения: 23.02.2026).

2. Приказ ФСТЭК России от 11.04.2025 № 117 «Об утверждении Требований к защите информации, содержащейся в государственных информационных системах, информационных системах государственных органов, органов местного самоуправления, государственных и муниципальных учреждений, иных организаций, создаваемых для выполнения государственных и муниципальных функций» (вступ. в силу 01.03.2026): Текст: электронный // Официальный сайт ФСТЭК России. – URL: https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/trebovaniya-utverzhdeny-prikazom-fstek-rossii-ot-11-aprelya-2025-g-n-117 (дата обращения: 23.02.2026).

4.2. Основная литература

1. Рассел С., Норвиг П. Искусственный интеллект: современный подход = Artificial Intelligence: A Modern Approach. — 4-е изд. — М.: Вильямс, 2021 (переизд. 2024–2025). — 1408 с. — ISBN 978-5-907394-XX-X (или оригинал: 4th ed., Pearson, 2020).

2. Гудфеллоу Я., Бенджио И., Курвиль А. Глубокое обучение = Deep Learning. — СПб.: ДМК Пресс, 2018 (переизд. с обновлениями 2024). — 652 с. — ISBN 978-5-97060-618-6.

3. Чесалин А. Н. Основы искусственного интеллекта с приложениями в информационной безопасности. Практикум: учебное пособие. — М.: РТУ МИРЭА, 2023–2024. — ЭБС Лань. — URL: https://e.lanbook.com/book/163838 (дата обращения: 23.02.2026).

4. Калинин М. О., Ложников П. С. и др. Основы искусственного интеллекта. Безопасность искусственного интеллекта: учебное пособие. — СПб.: СПбПУ, 2024. — 150 с. — ЭБ СПбПУ. — URL: https://elib.spbstu.ru/dl/2/i24-263.pdf (дата обращения: 23.02.2026).

4.3. Дополнительная литература

1. Банк данных угроз безопасности информации ФСТЭК России: Текст: электронный // Официальный сайт ФСТЭК. — URL: https://bdu.fstec.ru/ (дата обращения: 23.02.2026).

4.4. Электронные образовательные ресурсы

Электронный образовательный ресурс в разработке

4.5 Лицензионное и свободно распространяемое программное обеспечение

1. Операционная система Microsoft Windows от 7 версии и выше; Microsoft Office или Open Office, браузер (Firefox/Google Chrome /Explorer).
2. Python 3.10–3.12 (устанавливается через официальный сайт python.org или через Anaconda/Miniconda)
3. Anaconda / Miniconda
4. Git (для скачивания датасетов и репозиториев с GitHub)
5. Visual Studio Code (бесплатный редактор кода с расширениями Python, Jupyter, Git) или PyCharm Community Edition (бесплатная версия)

4.6 Современные профессиональные базы данных и информационные справочные системы

1. eLIBRARY.RU: Научная электронная библиотека: сайт. - Москва, 2000 -. - URL: https://www.elibrary.ru (дата обращения: 23.02.2026). – Текст: электронный.

2. ЛАНЬ: электронно-библиотечная система: сайт. – Санкт-Петербург,2010 -. - URL: https://e.lanbook.com (дата обращения: 23.02.2026). - Текст: электронный.

3. ФСТЭК России: Государственный реестр сертифицированных средств защиты информации. – Москва, 2014. – URL: https://fstec.ru/tekhnicheskaya-zashchita-informatsii/dokumenty-po-sertifikatsii/153-sistema-sertifikatsii (дата обращения: 23.02.2026). - Текст: электронный.

5. Материально-техническое обеспечение

Для проведения всех видов занятий необходимо презентационное оборудование (мультимедийный проектор, ноутбук, экран или интерактивная доска) – 1 комплект.

Для проведения практических занятий необходимо наличие компьютерных классов, оборудованных современной вычислительной техникой из расчета одно рабочее место на одного обучаемого.

6. Методические рекомендации

6.1 Методические рекомендации для преподавателя по организации обучения

При подготовке к лабораторным работам следует предварительно проработать теоретический материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия.

При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.

При проверке работ и отчетов следует учитывать правильность выполнения лабораторных работ на всех этапах.

6.2 Методические указания для обучающихся по освоению дисциплины

Изучение дисциплины осуществляется в строгом соответствии с целевой установкой в тесной взаимосвязи учебным планом. Основой теоретической подготовки студентов являются *лекции.*

В процессе самостоятельной работы студенты закрепляют и углубляют знания, полученные во время аудиторных занятий, дорабатывают конспекты лекций, готовятся к экзамену, а также самостоятельно изучают отдельные темы учебной программы.

Лабораторные работы проводятся по наиболее важным темам дисциплины. Осуществляется закрепление знаний, полученных студентами на лекциях и в процессе самостоятельной работы. Особое внимание обращается на развитие умений и навыков установления связи положений теории с профессиональной деятельностью будущего специалиста по ИБ. Лабораторные работы проводятся по теоретическим и проблемным вопросам ИБ. Лабораторные работы предполагает творческие дискуссии, активный обмен мнениями по поставленным вопросам, заслушивание и обсуждение докладов по предложенным преподавателем темам.

Важным обстоятельством является привлечение внимания студентов к обсуждаемой проблеме, стимулирование интереса к ней и организация активного обсуждения, как структуры проблемы, так и составляющих ее наиболее актуальных тем. Для повышения эффективности проведения занятия требуется предварительная подготовка всех его участников. В этой связи рекомендуется заблаговременно (не менее, чем за неделю) оповестить студентов о теме занятия, дать перечень литературы по теме, назначить из числа студентов докладчиков и содокладчиков.

При проведении лабораторной работы преподаватель выполняет, в основном, функции ведущего - следит за регламентом времени, помогает уточнить формулировки, обобщает результаты дискуссии, подводит итог занятию в целом. При высоком уровне подготовки студенческой группы отдельные функции ведущего можно поручить одному из студентов. В случае необходимости, преподаватель оказывает ему поддержку, а при подведении итогов - дает оценку работе ведущего.

Активная работа студента на лабораторной работе учитывается при определении итоговой оценки его знаний по дисциплине на экзамене.

Самостоятельная работа студентов предполагает изучение теоретического и практического материала по актуальным вопросам дисциплины. Рекомендуется самостоятельное изучение учебной и научной литературы, использование справочной литературы и др..

При выдаче заданий на самостоятельную работу используется дифференцированный подход к студентам. Перед выполнением студентами самостоятельной внеаудиторной работы преподаватель проводит инструктаж по выполнению задания, который включает: цель здания, его содержание, сроки выполнения, ориентировочный объем работы, основные требования к результатам работы, критерии оценки. В процессе инструктажа преподаватель предупреждает студентов о возможных типичных ошибках, встречающихся при выполнении задания. Инструктаж проводится преподавателем за счет объема времени, отведенного на изучение дисциплины.

Текущий контроль осуществляется на практических занятиях, промежуточный контроль осуществляется на экзамене в письменной (устной) форме.

Самостоятельная работа осуществляется индивидуально.

Контроль самостоятельной работы организуется в двух формах:

- самоконтроль и самооценка студента;
- контроль со стороны преподавателей (текущий и промежуточный).

Текущий контроль осуществляется на практических занятиях, промежуточный контроль осуществляется на экзамене в письменной (устной) форме.

Критериями оценки результатов самостоятельной работы студента являются:

- уровень освоения студентом учебного материала;
- умения студента использовать теоретические знания при выполнении практических задач;
- сформированность компетенций;
- оформление материала в соответствии с требованиями.

7 Фонд оценочных средств

7.1 Методы контроля и оценивания результатов обучения

В процессе обучения используются следующие оценочные формы самостоятельной работы студентов, оценочные средства текущего контроля успеваемости и промежуточных аттестаций:

- проведение лабораторных работ (практических занятий с использованием спецтехники) и их защита;
- самостоятельная подготовка и проведение презентаций по темам дисциплины;
- экзамен.

7.2 Шкала и критерии оценивания результатов обучения

Форма промежуточной аттестации: экзамен.

По итогам промежуточной аттестации по дисциплине (модулю) выставляется оценка «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно».

Шкала оценивания	Описание
Отлично	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателей, оперирует приобретенными знаниями, умениями, навыками, применяет их в ситуациях повышенной сложности. При этом могут быть допущены незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
Хорошо	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует неполное, правильное соответствие знаний, умений, навыков приведенным в таблицах показателей, либо если при этом были допущены 2-3 несущественные ошибки.
Удовлетворительно	Выполнены все виды учебной работы, предусмотренные учебным планом. Студент демонстрирует соответствие знаний, в котором освещена основная, наиболее важная часть материала,

	но при этом допущена одна значительная ошибка или неточность.
Неудовлетворительно	Не выполнен один или более видов учебной работы, предусмотренных учебным планом. Студент демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателей, допускаются значительные ошибки, проявляется отсутствие знаний, умений, навыков по ряду показателей, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

7.3 Оценочные средства

7.3.1 Текущий контроль
Оценочные средства для текущей аттестации
- Защита отчетов о выполнении лабораторных работ

7.3.2 Промежуточная аттестация

Оценочные средства для промежуточной аттестации
- Экзамен

Список вопросов для экзамена по дисциплине:

1. Какие основные типы задач машинного обучения чаще всего применяются в задачах информационной безопасности?
2. В чём разница между обучением с учителем, обучением без учителя и обучением с подкреплением в контексте обнаружения киберугроз?
3. Перечислите и кратко опишите ключевые метрики качества моделей классификации, используемые при сильном дисбалансе классов (например, в задачах обнаружения атак).
4. Почему метрика accuracy часто непригодна для оценки моделей обнаружения вторжений и вредоносного ПО?
5. Что такое аномалии в системах информационной безопасности? Назовите 4 популярных метода их обнаружения без учителя.
6. Опишите основные архитектуры нейронных сетей, применяемые в задачах ИБ (MLP, CNN, LSTM/GRU, трансформеры), и укажите по одной типичной задаче для каждой.
7. Приведите примеры атак белого и чёрного ящика.
8. Опишите общий пайплайн построения системы обнаружения вторжений на основе машинного обучения.
9. Какие основные источники данных используются для обучения моделей сетевого обнаружения вторжений? Какие проблемы возникают при работе с реальными данными?
10. В чём заключаются различия между статическим и динамическим анализом вредоносного ПО с применением методов ИИ?

11. Какие преимущества и недостатки у глубоких моделей (типа MalConv) по сравнению с классическими алгоритмами (Random Forest, XGBoost) при классификации вредоносного ПО?
12. Как работает алгоритм Isolation Forest? Почему он популярен для задач обнаружения аномалий в больших объёмах логов?
13. Что такое UEBA (User and Entity Behavior Analytics)? Как в нём используется машинное обучение?
14. Перечислите основные категории угроз самим системам искусственного интеллекта.
15. Что такое отравление данных? Как эта атака может быть реализована на этапе обучения модели?
16. Опишите атаку model extraction. Каковы её цели и возможные последствия?
17. В чём состоит принцип adversarial training? Какие у него основные ограничения?
18. Назовите и кратко опишите три популярных метода интерпретируемости моделей машинного обучения (XAI), используемых в задачах ИБ.
19. Почему обеспечение интерпретируемости особенно важно для моделей ИИ, применяемых в критически важных системах безопасности?
20. Каковы основные требования Приказа ФСТЭК России № 117 (2025) к защите информации при использовании технологий искусственного интеллекта?
21. Перечислите меры защиты моделей машинного обучения от отравления данных.
22. Что такое certified adversarial robustness? Приведите пример метода достижения сертифицированной устойчивости.
23. Как генеративные модели (GAN, диффузионные модели, большие языковые модели) могут использоваться для повышения уровня кибербезопасности?
24. Приведите примеры злоумышленного применения генеративного ИИ в кибератаках (deepfakes, prompt injection и др.).
25. Опишите возможную архитектуру современной SIEM-системы с интеграцией технологий искусственного интеллекта.
26. Как можно использовать графовые нейронные сети в задачах анализа и прогнозирования киберугроз?
27. Что такое концептуальный дрейф в контексте моделей обнаружения угроз? Как с ним бороться?
28. Каковы основные требования ГОСТ Р ИСО/МЭК 42001 к системам менеджмента искусственного интеллекта?
29. Опишите возможную схему эксперимента по оценке устойчивости модели классификации фишинговых сообщений к текстовым adversarial атакам.
30. Какие направления развития технологий ИИ в области информационной безопасности наиболее перспективны на период 2026–2030 годов? Обоснуйте.