Автономная некоммерческая образовательная организация высшего образования
«Европейский университет в Санкт-Петербурге»

Рабочая программа дисциплины
Безопасность искусственного интеллекта

образовательная программа
направление подготовки
09.04.03 Прикладная информатика

направленность (профиль)
«Прикладной анализ данных и искусственный интеллект»
программа подготовки – магистратура

язык обучения – русский
форма обучения - очная

квалификация (степень) выпускника
Магистр

Санкт-Петербург

Автор:
Левшун Д.С., к. тех. н., доцент, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Рецензент:
Котельников Евгений Вячеславович, д. тех. н., профессор, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Рабочая программа дисциплины **«Безопасность искусственного интеллекта»**, входящей в образовательную программу уровня магистратуры «Прикладной анализ данных и искусственный интеллект», утверждена на заседании Совета Школы вычислительных социальных наук.

Протокол заседания № 4 от 25.02.2025 года.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ
«Безопасность искусственного интеллекта»

Дисциплина **«Безопасность искусственного интеллекта»** является дисциплиной по выбору части, формируемой участниками образовательных отношений, основной профессиональной образовательной программы высшего образования «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика.

Дисциплина **«Безопасность искусственного интеллекта»** фокусируется на критически важных аспектах безопасности, возникающих при разработке и применении систем ИИ. Магистранты изучат вопросы защиты чувствительных данных, предотвращения утечек информации и обеспечения конфиденциальности. Будут рассмотрены методы анализа уязвимостей систем ИИ, а также этические и правовые аспекты, связанные с безопасностью. Особое внимание уделяется практическим кейсам и разработке стратегий минимизации рисков, связанных с ИИ. Цель курса — подготовить специалистов, способных создавать и внедрять надежные и безопасные системы искусственного интеллекта, учитывая как технические, так и социо-гуманитарные аспекты.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточный контроль в форме зачета.

Общая трудоемкость освоения дисциплины составляет 6 зачетных единиц, 216 часов.

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Безопасность искусственного интеллекта» является формирование у обучающихся системных знаний и компетенций в области обеспечения безопасности систем искусственного интеллекта, с учетом технических, этических и правовых аспектов.

Задачи освоения дисциплины**:**

1. Изучение принципов и методов защиты данных в системах искусственного интеллекта.
2. Освоение методик анализа уязвимостей систем ИИ и оценки рисков.
3. Формирование навыков разработки стратегий минимизации рисков безопасности ИИ.
4. Изучение нормативно-правовой базы и этических принципов при работе с ИИ-системами.
5. Развитие практических навыков обеспечения конфиденциальности и защиты информации в ИИ-приложениях.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: профессиональными (ПК). Планируемые результаты формирования компетенций и индикаторы их достижения в результате освоения дисциплины представлены в Таблице 1.

Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

Таблица 1

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ПК-3 Способен управлять бизнес-процессом по сбору данных в цифровой форме	ИД.ПК-3.1 Управляет деятельностью команды сборки, разметки и анализа цифрового следа ИД.ПК-3.2 Управляет действиями по отслеживанию процесса сбора данных ИД.ПК-3.3. Осуществляет проверку соответствия структуры и способов передачи данных цифрового следа для последующей обработки ИД.ПК-3.4 Осуществляет контроль потоковых данных цифрового следа ИД.ПК-3.5 Осуществляет контроль соответствия цифрового следа разметке согласно сформулированной модели ИД.ПК-3.6. Контролирует взаимодействия технических средств передачи, перекодирования, хранения и предобработки цифрового следа ИД.ПК-3.7. Контролирует выполнение функций участниками команды по сбору цифрового следа ИД.ПК-3.8. Контролирует соответствие процесса получения и обработки данных заданному алгоритму	Знать: основы управления проектами, основы контрольно-надзорной деятельности, виды контроля при управлении информационными проектами, методы контроля при управлении информационными проектами З (ПК-3)
		Уметь: определять валидность и достоверность цифрового следа, применять специализированные программы для контроля потоковых данных цифрового следа, применять методы верификации отчетной документации, определять соответствие представленных результатов техническому заданию У (ПК-3)
		Владеть: навыками управления ресурсами, процессами и работой команды для решения поставленной задачи В (ПК-3)

В результате освоения дисциплины магистрант должен:

Знать:

— Основные виды угроз безопасности систем искусственного интеллекта

— Методы защиты данных и обеспечения конфиденциальности в ИИ-системах
— Подходы к анализу уязвимостей и оценке рисков систем ИИ
— Нормативно-правовую базу в области безопасности искусственного интеллекта
— Этические принципы и стандарты при разработке и внедрении ИИ-систем
— Современные технологии и инструменты обеспечения безопасности ИИ

Уметь:

— Проводить анализ уязвимостей систем искусственного интеллекта
— Разрабатывать стратегии минимизации рисков безопасности ИИ
— Применять методы защиты чувствительных данных и предотвращения утечек информации
— Оценивать соответствие ИИ-систем нормативно-правовым требованиям
— Интегрировать механизмы безопасности в процессы разработки и внедрения ИИ
— Анализировать практические кейсы нарушения безопасности ИИ-систем

Владеть:

— Методиками анализа и оценки безопасности систем искусственного интеллекта
— Инструментами защиты данных и обеспечения конфиденциальности в ИИ-системах
— Навыками разработки политик и процедур обеспечения безопасности ИИ
— Техниками выявления и предотвращения уязвимостей в ИИ-приложениях
— Способами интеграции этических принципов в процесс разработки безопасных ИИ-систем
— Практическими подходами к управлению рисками при внедрении ИИ-технологий

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина **«Безопасность искусственного интеллекта»** является дисциплиной по выбору части, формируемой участниками образовательных отношений, Блока 1 «Дисциплины (модули)» образовательной программы «Прикладной анализ данных и искусственный интеллект». Курс читается в девятом модуле, форма промежуточной аттестации – зачет.

Для успешного освоения данной дисциплины требуются знания, полученные в рамках прохождения обучения на уровне бакалавриата/ специалитета.

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе прохождения Б2.О.01(У) Технологической (проектно-технологической) практики и выполнения выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 6 (шесть) зачетных единиц, 216 часов.

Таблица 2

Объем дисциплины

Типы учебных занятий и самостоятельная работа	**Объем дисциплины**										
	Всего	**Модуль**									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Контактная работа обучающихся с преподавателем в соответствии с УП:	**28**	-	-	-	-	-	-	-	-	28	-
Лекции (Л)	**14**	-	-	-	-	-	-	-	-	14	-
Лабораторные занятия (ЛЗ)	**14**	-	-	-	-	-	-	-	-	14	-

Типы учебных занятий и самостоятельная работа		Объем дисциплины										
		Всего	Модуль									
			1	2	3	4	5	6	7	8	9	10
Самостоятельная работа (СР)		**188**	-	-	-	-	-	-	-	-	188	-
Промежуточная аттестация	**форма**	**Зачет**	-	-	-	-	-	-	-	-	Зачет	-
	час.	**-**	-	-	-	-	-	-	-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)		**216/6**	-	-	-	-	-	-	-	-	216/6	-

5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине: через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) по средствам индикаторов достижения компетенций в соответствии с Таблицей 3.

5.1 Содержание дисциплины

Таблица 3

Содержание дисциплины

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
1	Технические аспекты безопасности искусственного интеллекта	Защита данных и конфиденциальность в системах ИИ, методы предотвращения утечек информации, технологии дифференциальной приватности, федеративное обучение, гомоморфное шифрование, уязвимости моделей машинного обучения, adversarial attacks и методы защиты, безопасность нейронных сетей, устойчивость к атакам на ИИ-системы, защита от data poisoning, безопасность инфраструктуры ИИ, мониторинг и аудит безопасности моделей ИИ, методы обнаружения аномалий в работе ИИ.	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)
2	Социо-гуманитарные аспекты безопасности искусственного интеллекта	Этические принципы разработки и применения систем ИИ, нормативно-правовое регулирование в области безопасности ИИ, предвзятость и дискриминация в ИИ-системах, прозрачность и объяснимость алгоритмов ИИ,	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
		управление рисками внедрения ИИ-технологий, социальные последствия уязвимостей ИИ, ответственность разработчиков ИИ, корпоративные политики безопасности ИИ, международные стандарты безопасности ИИ, баланс между инновациями и безопасностью, долгосрочные риски развития ИИ, принципы ответственной разработки ИИ.			

5.2 Структура дисциплины

Таблица 4

Структура дисциплины

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.				Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП		СР	
			Л	ЛЗ		
Очная форма обучения						
Тема 1	Технические аспекты безопасности искусственного интеллекта	108	7	7	94	КР
Тема 2	Социо-гуманитарные аспекты безопасности искусственного интеллекта	108	7	7	94	КР
Промежуточная аттестация		-	-	-	-	Зачет
Всего:		**216/6**	**14**	**14**	**152**	

**Примечание: формы текущего контроля успеваемости: контрольная работа (КР).*

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения текстов и исследовательской литературы (из списков основной и дополнительной литературы) и их анализа.

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответственным образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ

конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарским занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего семестра, концентрируясь на обусловленных программой курса темах.

Существенную часть самостоятельной работы магистранта представляет самостоятельное изучение вспомогательных учебно-методических изданий, лекционных конспектов, интернет-ресурсов и пр. Подготовка к семинарским занятиям, контрольному тесту также является важной формой работы магистранта. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя.

6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины

Тема 1. Технические аспекты безопасности искусственного интеллекта:

1.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 47 часов.

1.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 47 часов. Итого: 94 часов.

Тема 2. Социо-гуманитарные аспекты безопасности искусственного интеллекта:

2.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 47 часов.

2.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 47 часов. Итого: 94 часов.

6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине

Вопросы для самостоятельной подготовки по темам дисциплины:

1. Какие основные типы атак на системы машинного обучения существуют и как от них защищаться?
2. В чем заключаются преимущества и ограничения дифференциальной приватности при обработке данных в ИИ-системах?
3. Как реализовать федеративное обучение для защиты конфиденциальных данных?
4. Каковы основные этические принципы разработки безопасных систем ИИ?
5. Какие международные стандарты регулируют безопасность систем искусственного интеллекта?
6. Как выявить и минимизировать предвзятость в алгоритмах машинного обучения?
7. Каковы методики оценки рисков внедрения ИИ-технологий в критически важных отраслях?
8. Каким образом можно обеспечить прозрачность и объяснимость решений, принимаемых ИИ-системами?
9. Какие существуют подходы к защите от атак data poisoning при обучении моделей?
10. Как разработать корпоративную политику безопасности для проектов, связанных с ИИ?

11. Какие технические решения позволяют обеспечить конфиденциальность данных при обучении нейронных сетей?

12. В чем заключаются особенности правового регулирования безопасности ИИ в России и мире?

13. Какие методы аудита безопасности применяются для оценки надежности ИИ-систем?

14. Как обеспечить баланс между производительностью модели ИИ и уровнем ее защищенности?

15. Какие долгосрочные риски развития искусственного интеллекта требуют внимания с точки зрения безопасности?

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Баланов А. Н. Комплексная информационная безопасность: полный справочник специалиста: практическое пособие / А. Н. Баланов. Москва; Вологда: Инфра-Инженерия, 2024. 156 с. ISBN 978-5-9729-1771-6. Текст: электронный. URL: https://znanium.ru/catalog/product/2169705 . Режим доступа: по подписке.

2. Бессонов А. А. Изучение преступной деятельности с использованием искусственного интеллекта: монография / А.А. Бессонов. Москва: ИНФРА-М, 2025. 432 с.: ил. (Научная мысль). ISBN 978-5-16-020805-3. Текст: электронный. URL: https://znanium.ru/catalog/product/2195488 . Режим доступа: по подписке.

3. Веселов Г. Е. Менеджмент риска информационной безопасности: Учебное пособие / Веселов Г.Е., Абрамов Е.С., Шилов А.К. Таганрог: Южный федеральный университет, 2016. 107 с.: ISBN 978-5-9275-2327-5. Текст: электронный. URL: https://znanium.com/catalog/product/997108 . Режим доступа: по подписке.

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «**Безопасность искусственного интеллекта**» разработано учебно-методическое обеспечение в составе:

1. Контрольные задания для подготовки к процедурам текущего контроля (п. 7.2 Рабочей программы).

2. Типовые задания для подготовки к промежуточной аттестации (п. 7.4 Рабочей программы).

3. Рекомендуемые основная, дополнительная литература, Интернет-ресурсы и справочные системы (п. 8, 9 Рабочей программы).

4. Рабочая программа дисциплины размещена в электронной информационно-образовательной среде Университета на электронном учебно-методическом ресурсе АНООВО «ЕУСПб» — образовательном портале LMS Sakai — Sakai@EU.

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому лабораторному занятию, выполнение контрольных работ, активное слушание на лекциях. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные

вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных контрольных работ, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 5

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	**Коды компетенций**	**Индикаторы компетенций**	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Формы текущего контроля успеваемости**	**Результаты текущего контроля**
Технические аспекты безопасности искусственного интеллекта	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)	Контрольная работа 1	зачтено/ не зачтено
Социо-гуманитарные аспекты безопасности искусственного интеллекта	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)	Контрольная работа 2	зачтено/ не зачтено

Таблица 6

Критерии оценивания

Формы текущего контроля успеваемости	**Критерии оценивания**
Контрольная работа	магистрант выполняет задания контрольной работы частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, полное и правильное выполнение заданий контрольной работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

7.2 Контрольные задания для текущей аттестации

Примерные задания для контрольных работ

Тема 1: Технические аспекты безопасности искусственного интеллекта

1. Проведите анализ уязвимостей предложенной модели машинного обучения и предложите методы их устранения.
2. Разработайте схему применения дифференциальной приватности для защиты данных в системе рекомендаций.
3. Сравните эффективность различных методов защиты от adversarial attacks для классификационных моделей.
4. Спроектируйте архитектуру федеративного обучения для медицинского приложения, обеспечивающую конфиденциальность пациентских данных.
5. Разработайте методику тестирования нейронной сети на устойчивость к атакам методом обратного градиента.

6. Предложите алгоритм обнаружения попыток data poisoning при обучении модели на открытых данных.
7. Сравните методы гомоморфного шифрования и их применимость для защиты моделей машинного обучения.
8. Разработайте систему мониторинга аномального поведения ИИ-системы в режиме реального времени.
9. Проведите анализ и оценку рисков утечки конфиденциальных данных при использовании предобученных языковых моделей.
10. Предложите архитектуру безопасной инфраструктуры для развертывания критически важных ИИ-систем.
11. Разработайте протокол защиты модели машинного обучения от извлечения обучающих данных через API.
12. Проанализируйте уязвимости в системах компьютерного зрения и предложите методы их устранения.
13. Спроектируйте механизм защиты от membership inference attacks для модели машинного обучения.
14. Разработайте методику оценки конфиденциальности модели машинного обучения с точки зрения возможности восстановления обучающих данных.
15. Предложите технические решения для обеспечения безопасности при обмене моделями и данными между организациями.

Тема 2: Социо-гуманитарные аспекты безопасности искусственного интеллекта

1. Проанализируйте существующие этические кодексы разработки ИИ и предложите дополнения с точки зрения безопасности.
2. Разработайте методику оценки предвзятости алгоритма машинного обучения и план по ее минимизации.
3. Сравните нормативно-правовые подходы к регулированию безопасности ИИ в разных странах и предложите оптимальную модель.
4. Разработайте корпоративную политику ответственного использования ИИ-технологий с учетом аспектов безопасности.
5. Предложите методику оценки социальных рисков внедрения конкретной ИИ-системы.
6. Проанализируйте кейс нарушения безопасности ИИ-системы и разработайте план по предотвращению подобных инцидентов.
7. Разработайте стратегию обеспечения прозрачности и объяснимости решений, принимаемых ИИ-системой в критической области.
8. Предложите метрики для оценки уровня доверия пользователей к ИИ-системе и способы его повышения.
9. Спроектируйте процедуру этического аудита ИИ-системы с точки зрения безопасности.
10. Разработайте руководство по ответственному сбору и обработке данных для обучения моделей машинного обучения.
11. Проанализируйте долгосрочные риски развития ИИ и предложите превентивные меры безопасности.
12. Разработайте план внедрения принципов ответственного ИИ в существующие бизнес-процессы организации.
13. Предложите методику оценки соответствия ИИ-системы международным стандартам безопасности.
14. Проанализируйте баланс между инновациями и безопасностью на примере конкретной технологии ИИ.

15. Разработайте стратегию информирования и обучения пользователей по вопросам безопасности при взаимодействии с ИИ-системами.

7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет в форме тестирования.

Перед зачетом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Комбинированные задания

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутвует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

Задания открытого типа

Повышенный уровень сложности: ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F=\frac{100}{K}*\left(\frac{x_1}{k_1}+\frac{x_2}{k_2}+\ldots+\frac{x_n}{k_n}\right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

k_n – максимально возможное количество баллов за вопросы по компетенции,

x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Таблица 7

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	**Коды компетенций**	**Индикаторы компетенций** *(в соот. с Таблицей 1)*	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Критерии оценивания**	**Оценка**
Зачет / Тест	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)	41-100% правильных ответов	Зачтено
				0-40% правильных ответов	Не зачтено

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по стобалльной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 7а.

Таблица 7а

Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «зачтено», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «не зачтено», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

ПК-3 Способен управлять бизнес-процессом по сбору данных в цифровой форме

Комбинированные задания

Повышенный уровень сложности

Задание 1:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Согласно статье, традиционные методы защиты программного обеспечения не обеспечивают достаточную безопасность систем искусственного интеллекта. Какие новые аспекты необходимо внедрить в модели защищенного проектирования ИИ-систем?

Варианты ответа:

1) избыточность и прозрачность
2) устойчивость и дискреционность
3) масштабируемость и адаптивность
4) модульность и независимость
5) гибкость и отказоустойчивость

Поле для ответа:

Обоснование_____

Задание 2:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какую ключевую проблему современных ИИ-систем, затрудняющую доказательство правильности их результатов, выделяют авторы статьи?

Варианты ответа:

1) проблема "черного ящика" и непрозрачность процесса принятия решений
2) недостаточная вычислительная мощность для обработки больших объемов данных
3) отсутствие общепринятых стандартов в области ИИ
4) уязвимость к традиционным хакерским атакам
5) неспособность систем ИИ к самообучению

Поле для ответа:

Обоснование_____

Задание 3:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какая особенность должна быть внедрена в системы ИИ для предоставления клиентам прозрачности и подотчетности?

Варианты ответа:

1) встроенные функции аналитической экспертизы
2) усиленные механизмы шифрования
3) интегрированные системы резервного копирования
4) автоматическое обновление алгоритмов
5) многоуровневая система аутентификации

Поле для ответа:

Обоснование_____

Задание 4:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какую ключевую уязвимость обучающих данных для моделей машинного обучения отмечают авторы статьи?

Варианты ответа:

1) недостаточный объем данных для полноценного обучения

2) низкое качество данных из-за технических ограничений

3) неспособность отличать вредоносные входящие данные от безвредных нестандартных

4) слишком высокая стоимость сбора релевантных данных

5) отсутствие единого формата данных

Поле для ответа:

Обоснование_____

Задание 5:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какую способность должен иметь искусственный интеллект, чтобы защитить себя от социальных инженеров и атак на набор данных?

Варианты ответа:

1) способность блокировать все нестандартные запросы

2) умение различать намеренные отклонения в поведении, оставаясь непредвзятым

3) возможность работать только с проверенными источниками данных

4) способность ограничивать доступ к своим функциям

5) умение изолировать свою работу от внешних воздействий

Поле для ответа:

Обоснование_____

Задания открытого типа

Высокий уровень сложности:

Задание 1:

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Опишите основные проблемы безопасности, с которыми сталкиваются системы искусственного интеллекта и машинного обучения, и почему традиционные методы защиты программного обеспечения недостаточны для их решения.

Поле для ответа _____

Задание 2:

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Раскройте концепции устойчивости и дискреционности, которые должны быть внедрены в системы искусственного интеллекта для повышения их безопасности. Приведите конкретные примеры реализации этих концепций.
Поле для ответа _____

Задание 3:
Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ
Вопрос: Объясните, почему для систем искусственного интеллекта важно иметь встроенные функции аналитической экспертизы. Какие преимущества это дает и какие проблемы помогает решить?
Поле для ответа _____

Задание 4:
Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ
Вопрос: Опишите основные уязвимости в процессе обучения систем искусственного интеллекта и машинного обучения. Какие риски они создают и какие меры можно предпринять для их минимизации?
Поле для ответа _____

Задание 5:
Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ
Вопрос: Какие новые векторы атак возникают в связи с развитием систем искусственного интеллекта, и почему традиционные методы моделирования угроз недостаточны для их выявления? Предложите подходы к защите от таких атак.
Поле для ответа _____

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

Коды компетенций	**Индикаторы компетенций** *(в соот.с Таблицей 1)*	**Средства оценки** *(в соот. с Таблицами 5, 7)*
ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	Контрольная работа, тест

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки *(в соот. С Таблицами 5, 7)*	**Рекомендованный план выполнения работы**
Контрольная работа	Магистрант в ходе подготовки и выполнения контрольной работы показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Выполнять функции управляющего команды сборки, разметки и анализа цифрового следа, а также управляющего действиями по отслеживанию процесса сбора данных, выполнять проверку соответствия структуры и способов передачи данных

Средства оценки *(в соот. С Таблицами 5, 7)*	**Рекомендованный план выполнения работы**
	цифрового следа для последующей обработки, контролировать потоковые данные, соответствие цифрового следа разметке согласно сформулированной модели, взаимодействие технических средств передачи, перекодирования, хранения и предобработки цифрового следа, выполнение функций участниками команды по сбору цифрового следа, соответствие процесса получения и обработки данных заданному алгоритму
Тест	Магистрант в ходе подготовки и выполнения тестов показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Выполнять функции управляющего команды сборки, разметки и анализа цифрового следа, а также управляющего действиями по отслеживанию процесса сбора данных, выполнять проверку соответствия структуры и способов передачи данных цифрового следа для последующей обработки, контролировать потоковые данные, соответствие цифрового следа разметке согласно сформулированной модели, взаимодействие технических средств передачи, перекодирования, хранения и предобработки цифрового следа, выполнение функций участниками команды по сбору цифрового следа, соответствие процесса получения и обработки данных заданному алгоритму

8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1. Основная литература

1. Баланов А. Н. Комплексная информационная безопасность: полный справочник специалиста: практическое пособие / А. Н. Баланов. Москва; Вологда: Инфра-Инженерия, 2024. 156 с. ISBN 978-5-9729-1771-6. Текст: электронный. URL: https://znanium.ru/catalog/product/2169705 . Режим доступа: по подписке.

8.2 Дополнительная литература

1. Бессонов А. А. Изучение преступной деятельности с использованием искусственного интеллекта: монография / А.А. Бессонов. Москва: ИНФРА-М, 2025. 432 с.: ил. (Научная мысль). ISBN 978-5-16-020805-3. Текст: электронный. URL: https://znanium.ru/catalog/product/2195488 . Режим доступа: по подписке.

2. Веселов Г. Е. Менеджмент риска информационной безопасности: Учебное пособие / Веселов Г.Е., Абрамов Е.С., Шилов А.К. Таганрог: Южный федеральный университет, 2016. 107 с.: ISBN 978-5-9275-2327-5. Текст: электронный. URL: https://znanium.com/catalog/product/997108 . Режим доступа: по подписке.

9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

9.1 Программное обеспечение

При осуществлении образовательного процесса магистрантами и профессорско-преподавательским составом используется следующее лицензионное программное обеспечение:

1. OS Microsoft Windows (OVS OS Platform)
2. MS Office (OVS Office Platform)
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. ABBYY FineReader 11 Corporate Edition
6. ABBYY Lingvo x5
7. Adobe Photoshop Extended CS6 13.0 MLP AOO License RU
8. Adobe Acrobat Reader DC /Pro – бесплатно

9. Google Chrome – бесплатно
10. Opera – бесплатно
11. Mozilla – бесплатно
12. VLC – бесплатно
13. Яндекс Браузер – бесплатно
14. Anaconda - бесплатно

9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:

Информационно-справочные системы

1. Гарант.Ру. Информационно-правовой портал: http://www.garant.ru
2. Информационная система «Единое окно доступа к образовательным ресурсам»: http://window.edu.ru/
3. Открытое образование. Ассоциация «Национальная платформа открытого образования»: http://npoed.ru
4. Официальная Россия. Сервер органов государственной власти Российской Федерации: http://www.gov.ru
5. Официальный интернет-портал правовой информации. Государственная система правовой информации: http://pravo.gov.ru
6. Правовой сайт КонсультантПлюс: http://www.consultant.ru/sys
7. Российское образование. Федеральный портал: http://www.edu.ru

Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:

1. ЕНИП — Электронная библиотека «Научное наследие России»: http://e-heritage.ru/
2. Интелрос. Интеллектуальная Россия: http://www.intelros.ru/
3. Национальная электронная библиотека НЭБ: http://www.rusneb.ru
4. Президентская библиотека: http://www.prlib.ru
5. Российская государственная библиотека: http://www.rsl.ru/
6. Российская национальная библиотека: http://www.nlr.ru/poisk/

9.3 Лицензионные электронные ресурсы библиотеки Университета

Профессиональные базы данных:

Полный перечень доступных обучающимся профессиональных баз данных представлен на официальном сайте Университета https://eusp.org/library/electronic-resources, включая следующие базы данных:

1. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, наукометрическая база данных: http://elibrary.ru;
2. Электронные журналы по подписке (текущие номера научных зарубежных журналов)

Электронные библиотечные системы:

1. **Znanium.com** – Электронная библиотечная система **(ЭБС)** – http://znanium.com/;
2. Университетская библиотека онлайн – Электронная библиотечная система (ЭБС) – http://biblioclub.ru/

9.4 Электронная информационно-образовательная среда Университета

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая включает в себя электронный учебно-методический ресурс АНООВО «ЕУСПб» — образовательный портал LMS Sakai — Sakai@EU, лицензионные электронные ресурсы библиотеки Университета, официальный сайт Университета (Европейский университет в Санкт-Петербурге [https://eusp.org/]), локальную сеть и корпоративную электронную почту Университета, и обеспечивает:

— доступ к учебным планам, рабочим программам дисциплин (модулей), практик и к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах;

— фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;

— формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок за эти работы со стороны любых участников образовательного процесса;

— взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет» (электронной почты и т.д.).

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронным ресурсам библиотеки Университета, содержащей издания учебной, учебно-методической и иной литературы по изучаемой дисциплине.

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

В ходе реализации образовательного процесса используются специализированные многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа (практических занятий, лабораторных работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране ПК. Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для самостоятельной работы организовано одно место (ПК) с возможностями бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к электронным ресурсам библиотеки Университета с возможностями для слабовидящих увеличения текста на экране ПК. Лица с ограниченными возможностями здоровья могут при необходимости воспользоваться имеющимся в университете креслом-коляской. В учебном корпусе имеется адаптированный лифт. На первом этаже оборудован специализированный туалет. У входа в здание университета для инвалидов оборудована специальная кнопка, входная среда обеспечена информационной доской о режиме работы университета, выполненной рельефно-точечным тактильным шрифтом (азбука Брайля).

ПРИЛОЖЕНИЕ 1

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

«Безопасность искусственного интеллекта»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому лабораторному занятию, выполнение контрольных работ, активное слушание на лекциях. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных контрольных работ, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 1

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	**Коды компетенций**	**Индикаторы компетенций**	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Формы текущего контроля успеваемости**	**Результаты текущего контроля**
Технические аспекты безопасности искусственного интеллекта	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)	Контрольная работа 1	зачтено/ не зачтено
Социо-гуманитарные аспекты безопасности искусственного интеллекта	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)	Контрольная работа 2	зачтено/ не зачтено

Таблица 2

Критерии оценивания

Формы текущего контроля успеваемости	**Критерии оценивания**
Контрольная работа	магистрант выполняет задания контрольной работы частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, полное и правильное выполнение заданий контрольной работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

2 Контрольные задания для текущей аттестации

Задания для контрольных работ

Тема 1: Технические аспекты безопасности искусственного интеллекта

1. Проведите анализ уязвимостей предложенной модели машинного обучения и предложите методы их устранения.
2. Разработайте схему применения дифференциальной приватности для защиты данных в системе рекомендаций.
3. Сравните эффективность различных методов защиты от adversarial attacks для классификационных моделей.
4. Спроектируйте архитектуру федеративного обучения для медицинского приложения, обеспечивающую конфиденциальность пациентских данных.
5. Разработайте методику тестирования нейронной сети на устойчивость к атакам методом обратного градиента.
6. Предложите алгоритм обнаружения попыток data poisoning при обучении модели на открытых данных.
7. Сравните методы гомоморфного шифрования и их применимость для защиты моделей машинного обучения.
8. Разработайте систему мониторинга аномального поведения ИИ-системы в режиме реального времени.
9. Проведите анализ и оценку рисков утечки конфиденциальных данных при использовании предобученных языковых моделей.
10. Предложите архитектуру безопасной инфраструктуры для развертывания критически важных ИИ-систем.
11. Разработайте протокол защиты модели машинного обучения от извлечения обучающих данных через API.
12. Проанализируйте уязвимости в системах компьютерного зрения и предложите методы их устранения.
13. Спроектируйте механизм защиты от membership inference attacks для модели машинного обучения.
14. Разработайте методику оценки конфиденциальности модели машинного обучения с точки зрения возможности восстановления обучающих данных.
15. Предложите технические решения для обеспечения безопасности при обмене моделями и данными между организациями.

Тема 2: Социо-гуманитарные аспекты безопасности искусственного интеллекта

1. Проанализируйте существующие этические кодексы разработки ИИ и предложите дополнения с точки зрения безопасности.
2. Разработайте методику оценки предвзятости алгоритма машинного обучения и план по ее минимизации.
3. Сравните нормативно-правовые подходы к регулированию безопасности ИИ в разных странах и предложите оптимальную модель.
4. Разработайте корпоративную политику ответственного использования ИИ-технологий с учетом аспектов безопасности.
5. Предложите методику оценки социальных рисков внедрения конкретной ИИ-системы.
6. Проанализируйте кейс нарушения безопасности ИИ-системы и разработайте план по предотвращению подобных инцидентов.
7. Разработайте стратегию обеспечения прозрачности и объяснимости решений, принимаемых ИИ-системой в критической области.
8. Предложите метрики для оценки уровня доверия пользователей к ИИ-системе и способы его повышения.

9. Спроектируйте процедуру этического аудита ИИ-системы с точки зрения безопасности.

10. Разработайте руководство по ответственному сбору и обработке данных для обучения моделей машинного обучения.

11. Проанализируйте долгосрочные риски развития ИИ и предложите превентивные меры безопасности.

12. Разработайте план внедрения принципов ответственного ИИ в существующие бизнес-процессы организации.

13. Предложите методику оценки соответствия ИИ-системы международным стандартам безопасности.

14. Проанализируйте баланс между инновациями и безопасностью на примере конкретной технологии ИИ.

15. Разработайте стратегию информирования и обучения пользователей по вопросам безопасности при взаимодействии с ИИ-системами.

3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет в форме тестирования.

Перед зачетом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Комбинированные задания

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутствует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

Задания открытого типа

Повышенный уровень сложности: ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ

значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F=\frac{100}{K}*\left(\frac{x_1}{k_1}+\frac{x_2}{k_2}+\ldots+\frac{x_n}{k_n}\right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

k_n – максимально возможное количество баллов за вопросы по компетенции,

x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Таблица 3

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	**Коды компетенций**	**Индикаторы компетенций** *(в соот. с Таблицей 1)*	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Критерии оценивания**	**Оценка**
Зачет / Тест	ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5 ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	З (ПК-3) У (ПК-3) В (ПК-3)	41-100% правильных ответов	Зачтено
				0-40% правильных ответов	Не зачтено

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по стобалльной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 3а.

Таблица 3а

Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «зачтено», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «не зачтено», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

4 Задания к промежуточной аттестации

ПК-3 Способен управлять бизнес-процессом по сбору данных в цифровой форме

Комбинированные задания

Повышенный уровень сложности

Задание 1:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Согласно статье, традиционные методы защиты программного обеспечения не обеспечивают достаточную безопасность систем искусственного интеллекта. Какие новые аспекты необходимо внедрить в модели защищенного проектирования ИИ-систем?

Варианты ответа:

1) избыточность и прозрачность
2) устойчивость и дискреционность
3) масштабируемость и адаптивность
4) модульность и независимость
5) гибкость и отказоустойчивость

Поле для ответа:

Обоснование_____

Задание 2:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какую ключевую проблему современных ИИ-систем, затрудняющую доказательство правильности их результатов, выделяют авторы статьи?

Варианты ответа:

1) проблема "черного ящика" и непрозрачность процесса принятия решений
2) недостаточная вычислительная мощность для обработки больших объемов данных
3) отсутствие общепринятых стандартов в области ИИ
4) уязвимость к традиционным хакерским атакам
5) неспособность систем ИИ к самообучению

Поле для ответа:

Обоснование_____

Задание 3:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какая особенность должна быть внедрена в системы ИИ для предоставления клиентам прозрачности и подотчетности?

Варианты ответа:

1) встроенные функции аналитической экспертизы
2) усиленные механизмы шифрования
3) интегрированные системы резервного копирования
4) автоматическое обновление алгоритмов
5) многоуровневая система аутентификации

Поле для ответа:

Обоснование_____

Задание 4:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какую ключевую уязвимость обучающих данных для моделей машинного обучения отмечают авторы статьи?

Варианты ответа:

1) недостаточный объем данных для полноценного обучения
2) низкое качество данных из-за технических ограничений
3) неспособность отличать вредоносные входящие данные от безвредных нестандартных
4) слишком высокая стоимость сбора релевантных данных
5) отсутствие единого формата данных

Поле для ответа:

Обоснование_____

Задание 5:

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Какую способность должен иметь искусственный интеллект, чтобы защитить себя от социальных инженеров и атак на набор данных?

Варианты ответа:

1) способность блокировать все нестандартные запросы
2) умение различать намеренные отклонения в поведении, оставаясь непредвзятым
3) возможность работать только с проверенными источниками данных
4) способность ограничивать доступ к своим функциям
5) умение изолировать свою работу от внешних воздействий

Поле для ответа:

Обоснование_____

Задания открытого типа

Высокий уровень сложности:

Задание 1:

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Опишите основные проблемы безопасности, с которыми сталкиваются системы искусственного интеллекта и машинного обучения, и почему традиционные методы защиты программного обеспечения недостаточны для их решения.

Поле для ответа ___

Задание 2:

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Раскройте концепции устойчивости и дискреционности, которые должны быть внедрены в системы искусственного интеллекта для повышения их безопасности. Приведите конкретные примеры реализации этих концепций.

Поле для ответа _____

Задание 3:

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Объясните, почему для систем искусственного интеллекта важно иметь встроенные функции аналитической экспертизы. Какие преимущества это дает и какие проблемы помогает решить?

Поле для ответа _____

Задание 4:

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Опишите основные уязвимости в процессе обучения систем искусственного интеллекта и машинного обучения. Какие риски они создают и какие меры можно предпринять для их минимизации?

Поле для ответа _____

Задание 5:

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Какие новые векторы атак возникают в связи с развитием систем искусственного интеллекта, и почему традиционные методы моделирования угроз недостаточны для их выявления? Предложите подходы к защите от таких атак.

Поле для ответа _____

5 Средства оценки индикаторов достижения компетенций

Таблица 4

Средства оценки индикаторов достижения компетенций

Коды компетенций	**Индикаторы компетенций** *(в соот.с Таблицей 1)*	**Средства оценки** *(в соот. с Таблицами 5, 7)*
ПК-3	ИД.ПК-3.1 ИД.ПК-3.2 ИД.ПК-3.3. ИД.ПК-3.4 ИД.ПК-3.5	Контрольная работа, тест

Коды компетенций	**Индикаторы компетенций** *(в соот.с Таблицей 1)*	**Средства оценки** *(в соот. с Таблицами 5, 7)*
	ИД.ПК-3.6. ИД.ПК-3.7. ИД.ПК-3.8.	

Таблица 5

Описание средств оценки индикаторов достижения компетенций

Средства оценки *(в соот. С Таблицами 5, 7)*	**Рекомендованный план выполнения работы**
Контрольная работа	Магистрант в ходе подготовки и выполнения контрольной работы показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Выполнять функции управляющего команды сборки, разметки и анализа цифрового следа, а также управляющего действиями по отслеживанию процесса сбора данных, выполнять проверку соответствия структуры и способов передачи данных цифрового следа для последующей обработки, контролировать потоковые данные, соответствие цифрового следа разметке согласно сформулированной модели, взаимодействие технических средств передачи, перекодирования, хранения и предобработки цифрового следа, выполнение функций участниками команды по сбору цифрового следа, соответствие процесса получения и обработки данных заданному алгоритму
Тест	Магистрант в ходе подготовки и выполнения тестов показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Выполнять функции управляющего команды сборки, разметки и анализа цифрового следа, а также управляющего действиями по отслеживанию процесса сбора данных, выполнять проверку соответствия структуры и способов передачи данных цифрового следа для последующей обработки, контролировать потоковые данные, соответствие цифрового следа разметке согласно сформулированной модели, взаимодействие технических средств передачи, перекодирования, хранения и предобработки цифрового следа, выполнение функций участниками команды по сбору цифрового следа, соответствие процесса получения и обработки данных заданному алгоритму