Приложение № 4
к образовательной программе высшего образования – программе магистратуры
по направлению подготовки 09.04.01 Информатика и вычислительная техника
(направленность (профиль) «Разработка и управление
критической информационной инфраструктурой»)

**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ «НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»
(АНОО ВО «УНИВЕРСИТЕТ «СИРИУС»)**

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат: 7835bcdac1c89a3da835a18f6045494f3f404337
Владелец: Некрасова Людмила Андреевна
Действителен: с 08.07.2026 до 08.07.2027

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Безопасность технологий искусственного интеллекта»

1. Трудоемкость дисциплины (модуля): 4 з.е.

2. Место дисциплины (модуля) в учебном плане:

Часть, формируемая участниками образовательных отношений, «Профессиональной подготовки» - «Дисциплины (модули) специализации» (Блок дисциплин (модулей) по выбору «Математические модели для разработки и обеспечения безопасности объектов критической информационной инфраструктуры») в рамках Блока 1. Дисциплины (модули), реализуется в 3 семестре.

3. Цель дисциплины (модуля): формирование у обучающихся представления об угрозах безопасности систем искусственного интеллекта и машинного обучения, методах их защиты от состязательных атак, отравления данных и компрометации моделей, а также навыков построения доверенных и устойчивых ИИ-систем.

4. Задачи дисциплины (модуля):

– сформировать представление о специфических угрозах и моделях нарушителя для систем искусственного интеллекта и машинного обучения;

– обеспечить владение методами защиты данных, моделей и конвейеров машинного обучения от компрометации;

– научить оценивать защищенность ИИ-систем и применять методы построения доверенного и устойчивого искусственного интеллекта.

5. Перечень разделов (тем) дисциплины (модуля) и их краткое содержание:

Наименование раздела (темы) дисциплины (модуля)	**Краткое содержание**
Раздел 1. Угрозы безопасности систем искусственного интеллекта	Жизненный цикл ИИ-системы как объект защиты. Классификация угроз. Модель нарушителя. Поверхность атаки систем машинного обучения.
Раздел 2. Состязательные атаки на модели машинного обучения	Состязательные примеры (adversarial examples). Атаки уклонения и инверсии модели. Извлечение модели и данных. Методы генерации состязательных воздействий.
Раздел 3. Атаки на данные и конвейеры обучения	Отравление обучающих данных (data poisoning). Закладки и троянские модели (backdoor). Атаки на цепочку поставок ML. Контроль целостности данных и моделей.
Раздел 4. Методы защиты моделей и данных	Состязательное обучение и регуляризация. Обнаружение аномальных входов. Дифференциальная приватность. Федеративное обучение и защита конфиденциальности.
Раздел 5. Доверенный и устойчивый искусственный интеллект	Принципы доверенного ИИ. Объяснимость и интерпретируемость моделей. Робастность и

Наименование раздела (темы) дисциплины (модуля)	Краткое содержание
	верификация. Нормативные и этические требования к ИИ.
Раздел 6. Оценка защищенности ИИ-систем	Методики тестирования устойчивости моделей. Метрики робастности. Инструменты анализа безопасности ML. Подготовка рекомендаций по защите ИИ-системы.

6. Образовательные результаты освоения дисциплины (модуля):

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
ПК-5. Способен анализировать и реагировать на угрозы информационной безопасности	ИПК-5.1. Знает основы анализа и классификации угроз информационной безопасности, виды вредоносного программного обеспечения
	ИПК-5.2. Умеет выявлять и устранять угрозы информационной безопасности, а также обнаруживать и реагировать на вредоносное программное обеспечение
	ИПК-5.3. Владеет методами оперативного реагирования на угрозы, включая обнаружение и устранение вредоносного программного обеспечения, а также способен создавать собственные методологии для обеспечения защиты информационной безопасности
ПК-7. Способен разрабатывать и применять решения на основе математических моделей для обеспечения информационной безопасности	ИПК-7.1. Знает принципы построения математических моделей для обеспечения информационной безопасности
	ИПК-7.2. Умеет разрабатывать и внедрять математические модели для обеспечения информационной безопасности
	ИПК-7.3. Владеет методами сбора, анализа и использования данных для математического моделирования в составе решений, обеспечивающих информационную безопасность в информационных системах и объектах критической информационной инфраструктуры