

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Б1. О.37 Этика и социальная ответственность в ИИ

Цели и задачи изучения дисциплины (модуля)

Цель освоения дисциплины Формирование у будущих специалистов в области искусственного интеллекта и аналитики данных системного понимания этических принципов, правовых норм и социальной ответственности, связанных с разработкой и применением ИИ, а также выработка практических навыков для выявления, анализа и минимизации этических рисков в профессиональной деятельности.

Задачи дисциплины

- Сформировать понимание ключевых этических проблем и дилемм, порождаемых современными ИИ-технологиями (предвзятость, приватность, безопасность, подотчетность).
- Изучить основные принципы, frameworks и регуляторные инициативы в области ответственного ИИ (на международном и национальном уровнях).
- Освоить методики выявления ценностных предпосылок, когнитивных искажений и культурно-обусловленных предвзятостей в данных и алгоритмах.
- Развить навыки применения инструментов и методов профессиональной коммуникации для обсуждения, презентации и аргументации этических аспектов ИИ-проектов.
- Научить интегрировать этические соображения и методики управления рисками на различных стадиях жизненного цикла ИИ-систем.

Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Этика и социальная ответственность в ИИ» относится к Обязательной части Блока 1 «Дисциплины (модули учебного плана).

Материал курса тесно связан с дисциплинами (пререквизиты) Машинное обучение, Глубокое обучение, Нейросетевые технологии, Обработка данных на Python, Компьютерная графика, Технологии обработки больших данных, Технологии компьютерного зрения, Технологии обработки языка, Генеративный искусственный интеллект.

Пререквизиты – Философия (SS-1.1 (Б), SS-1.2 (Б)), Введение в специальность (LC-8.1 (Б))

Постреквизиты – Математические модели защиты информации SS-1.1 (П), Правовые основы оценки проектных решений, SS-1.1 (П), SS-1.2 (П)), Правовые основы рынка программного обеспечения SS-1.1 (П), SS-1.2 (П).

Дисциплина напрямую определяет **качество и глубину** выпускной работы.

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Роль 1: Data Analyst (Аналитик данных)

Задачи:

1. Статистический анализ, визуализация данных, предварительная обработка.
2. Создание прогнозных моделей
3. Построение аналитических моделей для поддержки бизнес-решений.

Роль 2: MLOps (Специалист по эксплуатации ИИ)

Задачи:

1. DevOps для ML.
2. Автоматизация, мониторинг ML-систем.
3. Операционное управление жизненным циклом ML-моделей.

Роль 3: AI PM (Менеджер проектов ИИ)

Задачи:

1. Управление ИИ-проектами от идеи до внедрения
2. Анализ бизнес-требований и постановка задач
3. Оценка эффективности и ROI ИИ-решений

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора	Результаты обучения по дисциплине
ОПК-2 Способен применять современный математический аппарат, связанный с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности	
ОПК-2.1 Способен применять системный подход к анализу предметной (проблемной) области, выявлению требований к ИС	Знать основные понятия системного подхода, методы выявления требований к ИС (интервью, сценарии, анализ стейкхолдеров, risk-воркшопы)
	Уметь Проводить анализ проблемной области, выделять стейкхолдеров, формулировать требования к ИИ-системе с учетом этических аспектов
	Владеть методами системного анализа применительно к этическому аудиту ИИ-проектов
ЛС-8 Способен применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ	
ЛС-8.1 Позиционирует собственную задачу в заданной области знания с точки зрения трендов современного искусственного интеллекта	Средний уровень Знать: расширенные тренды Уметь: анализировать и сопоставлять задачу с современными трендами, выделять специфику задачи в контексте последних достижений ИИ Владеть: навыком классификации задач по отношению к трендам Продвинутый уровень Знать Тренды и историю ИИ Уметь Выявлять потенциал новых подходов, формулировать задачи на стыке дисциплин Владеть Методами позиционирования задач в контексте трендов
СС-1 Способен осуществлять свою трудовую деятельность с учетом определения корректной роли ИИ в различных процессах, критического анализа последствий применения ИИ-технологий, этических принципов	
СС-1.1 Учитывает в разработке и эксплуатации систем ИИ философские основания концепций интеллекта, языка, знания, агентности	Средний уровень Знать: основы эпистемологии машинного обучения (как «обучаются» алгоритмы и что они «знают») Уметь: обосновывать выбор терминологии и интерпретации при описании поведения ИИ, применять философские подходы для анализа конкретных моделей ИИ (сравнение символьного ИИ и нейросетей) Владеть: философским анализом архитектур ИИ

	Продвинутый уровень Знать Философские основы ИИ (интеллект, знание, агентность) Уметь Применять философские модели для анализа ИИ, определять границы машинного мышления Владеть Навыками философской рефлексии в инженерных решениях
SS-1.2 Применяет методики работы с этическими и социальными рисками, возникающими на разных стадиях жизненного цикла ИИ.	Средний уровень Знать: подходы к этической оценке ИИ и их ограничения (риск-ориентированный подход, бенчмаркинг, анкетирование) Уметь: <i>применять эти подходы к типовым системам</i> Владеть: навыком выбора методики для конкретного ЖЦ Продвинутый уровень Знать Методики работы с этическими рисками на ЖЦ ИИ Уметь Проводить этический аудит, участвовать в экспертизе и регуляторных инициативах Владеть Встраиванием этических принципов в проектирование и тестирование
LLM-1 Способен применять и (или) разрабатывать генеративные модели и БЯМ	
LLM-1.6 Учитывает этические аспекты генерации	Базовый уровень Знать: что такое галлюцинации, фильтрация, небезопасная генерация Уметь: распознавать этически проблемные ответы БЯМ на уровне пользователя Владеть: использованием готовых фильтров (например, moderation API) Средний уровень Знать: основные способы постобработки генераций и виды фильтров Уметь: настраивать фильтры и правила постобработки под задачу Владеть: применением фильтров в коде или конфигурации Продвинутый уровень Знать Этические аспекты генерации (галлюцинации, фильтрация) Уметь Разрабатывать и внедрять механизмы контроля этичности генеративных моделей Владеть Методами оценки этических рисков БЯМ

Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 2 зач. ед. (72 час.), их распределение по видам работ представлено в таблице

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 7 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Введение в этику ИИ. Основные концепции и вызовы	10	2	2		6
2.	Философские основания ИИ: интеллект, знание, агентность	12	2	2		6
3.	История ИИ и современные тренды. Позиционирование задач	8	2	2		4
4.	Предвзятость, справедливость, прозрачность и конфиденциальность	8	2	2		4
5.	Регулирование ИИ и управление социальными рисками	8	2	2		4
6.	Этические аспекты генеративных моделей и БЯМ	8	2	2		4
7.	Инструменты и практики внедрения ответственного ИИ. Этический аудит	7,8	2	2		3,8
8.	Профессиональная коммуникация в проектах ИИ: отчеты, презентации, дебаты	8	2	2		4
ИТОГО по разделам дисциплины		69,8	16	18		35,8
Контроль самостоятельной работы (КСР)		2				
Промежуточная аттестация (ИКР)		0,2				
Подготовка к текущему контролю						
Общая трудоемкость по дисциплине		72				

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Форма отчетности: зачет

Программу составила Г.В. Калайдина к.ф.-м.н., , доцент КАДИИ