Лебедева Екатерина Александровна
Подписано цифровой подписью: Лебедева Екатерина Александровна
Дата: 2026.06.25 14:45:32 +03'00'

Приложение 4

к образовательной программе

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.09 Интерпретируемость, объяснимость и доверие к ИИ (XAI)

(индекс, наименование дисциплины в соответствии с учебным планом)

40.04.01 "Юриспруденция"
(код, наименование направления подготовки)

Правовое обеспечение использования искусственного интеллекта
(наименование образовательной программы)

Очная форма обучения
(форма обучения)

Год набора – 2026
Нижний Новгород, 2026 г.

Автор(ы)-составитель(и) РПД:

Калинина Евгения Валерьевна, доктор юридических наук, доцент, заведующая кафедрой гражданского и международного права

Заведующий кафедрой:

Калинина Евгения Валерьевна, доктор юридических наук, доцент, заведующая кафедрой гражданского и международного права

Рабочая программа дисциплины Б1.В.09 «Интерпретируемость, объяснимость и доверие к ИИ (XAI)» одобрена на заседании кафедры гражданского и международного права

протокол № 6 от «4» марта 2026 г.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Типы оценочных материалов, показатели, критерии, шкалы оценивания
5. Формы аттестации и типовые оценочные материалы для текущего контроля успеваемости обучающихся
6. Формы промежуточной аттестации по дисциплине, типы оценочных материалов, показатели, критерии, шкалы оценивания
7. Методические материалы по освоению дисциплины
8. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет»
9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Дисциплина Б1.В.09 «Интерпретируемость, объяснимость и доверие к ИИ (XAI)» обеспечивает формирование у обучающихся следующих профессиональных компетенций:

ОТФ/ТФ и реквизиты ПС (при наличии)	**Код компетенции**	**Наименование компетенции**	**Код индикатора достижения компетенций**	**Наименование индикатора достижения компетенций**	**Образовательный результат**
(не предусмотрено для УК)	**УК-1**	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	**ИУК-1.2**	Выявляет и оценивает комплексные технологические и правовые риски в проектах с использованием искусственного интеллекта.	**ИУК-1.2 З-1.** Знает: методологии идентификации и оценки рисков (risk assessment) в сфере высоких технологий, включая риски кибербезопасности, нарушения защиты данных и несоответствия регуляторным требованиям. **ИУК-1.2 У-1.** Умеет: проводить декомпозицию ИИ-проекта на компоненты (данные, модель, интерфейс, инфраструктура) для системного выявления уязвимостей и

ОТФ/ТФ и реквизиты ПС (при наличии)	**Код компетенции**	**Наименование компетенции**	**Код индикатора достижения компетенций**	**Наименование индикатора достижения компетенций**	**Образовательный результат**
					правовых пробелов; ранжировать риски по вероятности и потенциальному ущербу в сфере кибербезопасности и защиты информации.
40.001 Специалист по патентованию (ОТФ С/7, С/01.7, С/02.7)	**ПКс-ИИ-3**	Способен представлять интересы клиентов в спорах, связанных с причинением вреда решениями ИИ, защитой интеллектуальных прав на AI-модели и данные, а также в административных процедурах, связанных с регулированием инновационных технологий	**ИПКс-ИИ-3.1**	Разрабатывает правовые позиции и процессуальные документы (исковые заявления, отзывы, ходатайства, жалобы) для защиты интересов клиентов в спорах, связанных с причинением вреда решениями ИИ и нарушением интеллектуальных прав на AI-модели и данные. Представляет интересы клиентов во взаимодейст	**ИПКс-ИИ-3.1. З-1.** Знает: основания и особенности гражданско-правовой, административной и уголовной ответственности за нарушения в сфере кибербезопасности (утечки данных, неправомерный доступ, нарушение правил обработки ПДн), включая практику привлечения к ответственности операторов ИИ-систем. **ИПКс-ИИ-3.1. У-1.** Умеет:

ОТФ/ТФ и реквизиты ПС (при наличии)	Код компетенции	Наименование компетенции	Код индикатора достижения компетенций	Наименование индикатора достижения компетенций	Образовательный результат
				вии с регуляторами (Роскомнадзор, ФАС, ФСТЭК и др.) в рамках административных процедур, связанных с расследованием инцидентов и регулированием инновационных технологий.	анализировать инциденты кибербезопасности в ИИ-проектах (утечка данных, атаки на модель), определять возможные правовые последствия для клиента, выбирать оптимальную стратегию защиты (досудебное урегулирование с регулятором, судебная защита).

2. Объем и место дисциплины (модуля) в структуре образовательной программы

Общая трудоемкость дисциплины составляет 72 а.ч. (2 з.е.)

Количество академических часов, выделенных на контактную работу с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся: 4 сем. - лекции – 16 а.ч., практические занятия – 16 а.ч., самостоятельная работа – 72 а.ч., контактная работа на аттестацию в период экзаменационных сессий – 4 часа.

Дисциплина *Б1.В.09 «Интерпретируемость, объяснимость и доверие к ИИ (XAI)»* относится к дисциплинам части, формируемой участниками образовательных отношений. Изучается на очной форме обучения - на 2 курсе.

Для освоения дисциплины Б1.В.09 «Интерпретируемость, объяснимость и доверие к ИИ (XAI)» обучающиеся используют знания, умения, навыки, способы деятельности и

установки, сформированные в ходе изучения дисциплины *«Информационное право»*, *«Информационные технологии в юридической деятельности»*.

Форма промежуточной аттестации в соответствии с учебным планом:

Очная форма обучения – зачет.

3. Содержание и структура дисциплины

3.1. Структура дисциплины

Очная форма обучения

№ п/п	Наименование тем и (или) разделов		Объем дисциплины, ак.час											Форма текущего контроля успеваемости, промежуточной аттестации
		ВСЕГО	Контактная работа обучающихся с преподавателем по видам учебных занятий						Период промежуточной аттестации (сессия)			Самостоятельная работа		
			Период теоретического обучения											
			Занятия лекционного типа		Занятия семинарского типа		ИК	КСР	КЭ	Кат тэк	Контроль	СРкр	СРэк	СР
			Л	ВЛ	ЛР	ПЗ								
Тема 1.	Понятие, предпосылки и границы объяснимого ИИ (XAI)	19	2	0	0	2	0	0	0	0		0	3	12	Устный опрос, Ситуационная задача
Тема 2.	Классификация методов интерпретируемости и объяснения моделей	19	2	0	0	2	0	0	0	0		0	3	12	Устный опрос, Ситуационная задача
Тема 3.	Построение	19	3	0	0	3	0	0	0	0		0	3	12	Устный опрос,

	объяснений для правовых задач: LIME, SHAP, деревья решений													Ситуационная задача
Тема 4.	Объяснимость генеративного ИИ и больших языковых моделей (LLM)	19	3	0	0	3	0	0	0	0	0	3	12	Устный опрос, Ситуационная задача
Тема 5.	Доверие к ИИ, алгоритмическая справедливость и борьба с предвзятостью	19	3	0	2	3	0	0	0	0	0	3	12	Устный опрос, Ситуационная задача
Тема 6.	Использование XAI в судебных и регуляторных процедурах	19	3	0	0	3	0	0	0	0	0	3	12	Устный опрос, Ситуационная задача
Промежуточная аттестация		4								4				Зачет
Итого		108	16	0	0	16	0	0	0	4	0	0	72	

Используемые сокращения:

Л – лекции - занятия, предусматривающие преимущественную передачу учебной информации обучающимся педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях,).

ВЛ – видео лекции.

ЛР – лабораторные работы.

ПЗ – практические занятия (за исключением лабораторных работ).

ИК – индивидуальные консультации.

КСР – контроль самостоятельной работы

КЭ – консультации перед экзаменом

Каттэк – контактная работа на аттестацию в период экзаменационных сессий

СРкр – самостоятельная работа на подготовку курсовой работы/ курсового проекта.

СРэк – самостоятельная работа на подготовку к экзамену.СР – самостоятельная работа в семестре на подготовку к учебным занятиям.

3.2 Содержание дисциплины

Тема 1. Понятие, предпосылки и границы объяснимого ИИ (XAI). ИУК-1.1

Причины «чёрного ящика» в ML и необходимость объяснимости для права.
Определение XAI, интерпретируемость vs объяснимость.
Правовые и этические драйверы доверия к ИИ.
Ограничения объяснимости: точность vs прозрачность.

Тема 2. Классификация методов интерпретируемости и объяснения моделей. ИПКс-ИИ-3.1

Intrinsic vs post-hoc объяснения.
Model-specific vs model-agnostic методы.
Глобальная и локальная интерпретируемость.
Выбор метода в зависимости от юридической задачи (доказывание, аудит, спор).

Тема 3. Построение объяснений для правовых задач: LIME, SHAP, деревья решений. ИУК-1.1

Принципы LIME (локальные аппроксимации).
SHAP-значения и интерпретация важности признаков.
Деревья решений и rule-based модели как понятные для суда.
Практический разбор: объяснение кредитного или кадрового AI-решения.

Тема 4. Объяснимость генеративного ИИ и больших языковых моделей (LLM). ИПКс-ИИ-3.1

Проблема объяснения ответов LLM (галлюцинации, источники).
Механизмы внимания (attention) и их доказательственная ценность.
Правовые риски непроверяемого вывода GenAI.
Требования к explainability в AI-актах (ЕС, draft РФ).

Тема 5. Доверие к ИИ, алгоритмическая справедливость и борьба с предвзятостью. ИУК-1.1

Связь объяснимости и доверия (user trust vs legal reliance).
Выявление дискриминации через XAI-инструменты.
Fairness metrics и их интерпретация для юриста.
Правовые последствия необъяснимой предвзятости.

Тема 6. Использование XAI в судебных и регуляторных процедурах. ИПКс-ИИ-3.1

Стандарты доказывания и бремя объяснения (right to explanation).
Назначение экспертизы объяснимости ИИ.
Практика регуляторов: требования к отчётам об интерпретируемости.
Построение правовой позиции на основе XAI-отчёта.

4. Типы оценочных материалов, показатели и критерии оценивания

4.1. Оценочные материалы по дисциплине Б1.В.09 «Интерпретируемость, объяснимость и доверие к ИИ (XAI)» входят в состав оценочных материалов по образовательной программе. Совокупность оценочных материалов по всем дисциплинам (модулям) образовательной программы составляют фонд оценочных средств (далее – ФОС). ФОС используется при проведении текущего контроля успеваемости и промежуточной аттестации обучающихся с целью оценивания достижения обучающимися планируемых результатов обучения.

4.2. ФОС разработан как комплекс проверочных заданий различного типа и уровня сложности, включает критерии и шкалы оценивания, а также «ключи» правильных ответов. ФОС формируется как отдельный документ и хранится в электронном виде, доступ к ФОС предоставлен ограниченному кругу лиц.

4.3. Для самостоятельной работы обучающихся при подготовке к текущему контролю успеваемости и промежуточной аттестации в рабочих программах дисциплин размещены типовые проверочные задания, которые можно условно разделить на задания закрытого, комбинированного и открытого типов.

Задания закрытого типа — это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных.

Задания комбинированного типа – это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных и обосновать свой выбор.

Задания открытого типа — это задания, в которых на каждый вопрос должен быть предложен развернутый обоснованный ответ.

В зависимости от типа задания рекомендованы определенная последовательность выполнения и система оценивания выполнения заданий.

4.4. Типы заданий, сценарии выполнения, критерии оценивания

ТИП ЗАДАНИЯ	ИНСТРУКЦИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	КРИТЕРИИ ОЦЕНИВАНИЯ
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	Прочитайте текст, выберите правильный ответ	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).	Ответ считается верным, если правильно указана цифра или буква
Задание закрытого типа на установление соответствия	Прочитайте текст и установите соответствие	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	Ответ считается верным, если правильно указаны цифры или буквы
Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	Прочитайте текст, выберите правильные ответы	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или А Г).	Ответ считается верным, если правильно установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого)
Задание закрытого типа на установление последовательности	Прочитайте текст и установите последовательность	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).	Ответ считается верным, если правильно указана вся последовательность цифр
Задание комбинированного типа с	Прочитайте текст, выберите правильный	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных	Ответ считается верным, если правильно указана цифра или буква

выбором одного правильного ответа из предложенных и обоснованием выбора	ответ и запишите аргументы, обосновывающие выбор ответа	вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).	и приведены корректные аргументы, используемые при выборе ответа
Задание открытого типа с развернутым ответом	Прочитайте текст и запишите развернутый обоснованный ответ	1. Внимательно прочитать текст задания и понять суть вопроса. 2.Продумать логику и полноту ответа. 3.Записать ответ, используя четкие компактные формулировки. 4.В случае расчетной задачи, записать решение и ответ	Ответ считается верным: 1. Отсутствие фактических ошибок. 2. Раскрытие объема используемых понятий (полнота ответа). 3.Обоснованность ответа (наличие аргументов). 4. Логическая последовательность излагаемого материала.

4.5. Общая шкала оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся с применением БРС

Итоговая балльная оценка	**Традиционная система**	**Бинарная система**	**ECTS**	
			Для традиционной системы	**Для бинарной системы**
95-100	Отлично	Зачтено	A	P/ Passed
85-94			B	P/ Passed
75-84	Хорошо		C	P/ Passed
65-74			D	P/ Passed
55-64	Удовлетворительно		E	P/ Passed
0-54	Неудовлетворительно	Не зачтено	F	F/Failed

Соотношение баллов за текущий контроль успеваемости и промежуточную аттестацию, а также повторную промежуточную аттестацию:

Максимальная сумма баллов за текущий контроль успеваемости	Максимальная сумма баллов за промежуточную аттестацию	Максимальная итоговая балльная оценка	Максимальная сумма баллов за повторную промежуточную аттестацию
60 баллов	40 баллов	100 баллов	100 баллов

5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам

5.1. В ходе реализации дисциплины Б1.В.09 «Интерпретируемость, объяснимость и доверие к ИИ (XAI)» используются следующие формы текущего контроля успеваемости обучающихся (в том числе, задания к контрольным точкам):

Устный опрос, ситуационная задача, тестирование.

5.2. Типовые оценочные материалы для текущего контроля успеваемости обучающихся (вне контрольных точек):

Тема 1. Понятие, предпосылки и границы объяснимого ИИ (XAI). ИУК-1.1

Подготовка к практическому занятию:

– внимательно прочитать материал лекций, относящихся к данному занятию, ознакомиться с учебными материалами, включая электронные в соответствии с предложенным списком литературы в рабочей программе учебной дисциплины;

– подготовить развернутые ответы на вопросы для проведения опроса по теме практического занятия;

– понять, что осталось неясными и постараться получить на них ответ заранее;

– готовиться к практическому занятию можно как индивидуально, так и в составе малой группы.

Вопросы к практическому занятию для проведения опроса:

Вопросы для устного опроса:

1. Причины «чёрного ящика» в ML и необходимость объяснимости для права.
2. Определение XAI, интерпретируемость vs объяснимость.
3. Правовые и этические драйверы доверия к ИИ.
4. Ограничения объяснимости: точность vs прозрачность.

Ситуационные задачи

Задание 1

Банк внедрил нейросеть для одобрения кредитов. Модель работает как «черный ящик» (1000 параметров). Клиенту отказали в кредите, он требует объяснения. Банк отвечает: «модель точная, объяснить решение не можем, это коммерческая тайна». *Вопрос: какую стратегию действий вы как юрист предложите банку? Какие правовые основания для требования объяснения существуют в РФ?*

Задание 2

Вы участвуете в закупке ИИ-системы для государственного органа. Два поставщика:

- Поставщик А: точность 97%, полный «черный ящик».
- Поставщик Б: точность 91%, полностью интерпретируемая модель (дерево решений).

Вопрос: какую стратегию вы выберете и почему? Какие риски несет выбор более точной, но непрозрачной модели?

Тема 2. Классификация методов интерпретируемости и объяснения моделей. ИПКс-ИИ-3.1

Подготовка к практическому занятию:

– внимательно прочитать материал лекций, относящихся к данному занятию, ознакомиться с учебными материалами, включая электронные в соответствии с предложенным списком литературы в рабочей программе учебной дисциплины;

– подготовить развернутые ответы на вопросы для проведения опроса по теме практического занятия;

– понять, что осталось неясными и постараться получить на них ответ заранее;

– готовиться к практическому занятию можно как индивидуально, так и в составе малой группы.

Вопросы к практическому занятию для проведения опроса:

Вопросы для устного опроса:

1. Intrinsic vs post-hoc объяснения.
2. Model-specific vs model-agnostic методы.
3. Глобальная и локальная интерпретируемость.
4. Выбор метода в зависимости от юридической задачи (доказывание, аудит, спор).

Ситуационные задачи.

Задание 1
Истец утверждает, что ИИ-система отбора резюме дискриминирует женщин (отказов на 30% больше). У ответчика — модель-«черный ящик» (глубокая нейросеть). *Вопрос: какой метод объяснения (глобальный или локальный, post-hoc или intrinsic) вы как юрист истца потребуете предоставить и почему?*

Задание 2
Вы представляете интересы страховой компании, которой предъявлен иск за отказ в выплате. Решение принято XGBoost-моделью (сложный ансамбль).
Вопрос: какой model-agnostic метод вы выберете для защиты в суде, чтобы показать, что решение было обоснованным? Какие ограничения этого метода нужно учитывать?

Тема 3. Построение объяснений для правовых задач: LIME, SHAP, деревья решений. ИУК-1.1

Подготовка к практическому занятию:

– внимательно прочитать материал лекций, относящихся к данному занятию, ознакомиться с учебными материалами, включая электронные в соответствии с предложенным списком литературы в рабочей программе учебной дисциплины;

– подготовить развернутые ответы на вопросы для проведения опроса по теме практического занятия;

– понять, что осталось неясными и постараться получить на них ответ заранее;

– готовиться к практическому занятию можно как индивидуально, так и в составе малой группы.

Вопросы к практическому занятию для проведения опроса:

Вопросы для устного опроса:

1. Принципы LIME (локальные аппроксимации).
2. SHAP-значения и интерпретация важности признаков.
3. Деревья решений и rule-based модели как понятные для суда.
4. Практический разбор: объяснение кредитного или кадрового AI-решения.

Ситуационные задачи.

Задание 1
Судья в арбитражном процессе спрашивает: «Почему ваша модель классифицировала этот договор как кабальный? Я не понимаю эти SHAP-графики». *Вопрос: какую стратегию вы предложите для представления объяснений в суде, чтобы они были понятны и убедительны?*

Задание 2
Вы проводите внутренний аудит HR-модели найма. Модель — случайный лес. Анализ SHAP показал, что признак «возраст» имеет высокую важность.

Вопрос: что вы будете делать? Какие шаги предпримете как юрист?

Тема 4. Объяснимость генеративного ИИ и больших языковых моделей (LLM). ИПКс-ИИ-3.1

Подготовка к практическому занятию:

– внимательно прочитать материал лекций, относящихся к данному занятию, ознакомиться с учебными материалами, включая электронные в соответствии с предложенным списком литературы в рабочей программе учебной дисциплины;

– подготовить развернутые ответы на вопросы для проведения опроса по теме практического занятия;

– понять, что осталось неясными и постараться получить на них ответ заранее;

– готовиться к практическому занятию можно как индивидуально, так и в составе малой группы.

Вопросы к практическому занятию для проведения опроса:

Вопросы для устного опроса:

1. Проблема объяснения ответов LLM (галлюцинации, источники).
2. Механизмы внимания (attention) и их доказательственная ценность.
3. Правовые риски непроверяемого вывода GenAI.
4. Требования к explainability в AI-актах (ЕС, draft РФ).

Ситуационные задачи.

Задание 1

Чат-бот на базе LLM дал пользователю неверный медицинский совет, что привело к вреду здоровью. В суде вы представляете интересы потерпевшего. Разработчик утверждает: «модель — черный ящик, мы не знаем, почему она так ответила».

Вопрос: какие аргументы вы приведете в суде о необходимости объяснения? Какие методы объяснения LLM существуют?

Задание 2

Компания разрабатывает LLM для юридических консультаций. Вы как юрист-консультант готовите требования к системе.

Вопрос: какие требования к объяснимости вы включите в ТЗ, чтобы минимизировать риски ответственности за галлюцинации?

Тема 5. Доверие к ИИ, алгоритмическая справедливость и борьба с предвзятостью. ИУК-1.1

Подготовка к практическому занятию:

– внимательно прочитать материал лекций, относящихся к данному занятию, ознакомиться с учебными материалами, включая электронные в соответствии с предложенным списком литературы в рабочей программе учебной дисциплины;

– подготовить развернутые ответы на вопросы для проведения опроса по теме практического занятия;

– понять, что осталось неясными и постараться получить на них ответ заранее;

– готовиться к практическому занятию можно как индивидуально, так и в составе малой группы.

Вопросы к практическому занятию для проведения опроса:

Вопросы для устного опроса:

1. Связь объяснимости и доверия (user trust vs legal reliance).
2. Выявление дискриминации через XAI-инструменты.
3. Fairness metrics и их интерпретация для юриста.
4. Правовые последствия необъяснимой предвзятости.

Ситуационные задачи:

Задание 1

Вы аудируете ИИ-систему для оценки риска рецидива в УИС. Модель показала высокую точность, но анализ fairness выявил, что ложноотрицательные ошибки (недооценка риска) в 3 раза чаще происходят для определенной этнической группы.

Вопрос: какую стратегию вы предложите руководству? Должны ли fairness metrics иметь приоритет над точностью?

Задание 2

К вам обратилась IT-компания: их ИИ-система подбора персонала обвинена в дискриминации по признаку пола. Внутренний аудит показал, что bias возник из-за исторических данных (в компании раньше работало 90% мужчин).

Вопрос: какую стратегию защиты вы предложите? Является ли исторический bias оправданием в суде?

Тема 6. Использование XAI в судебных и регуляторных процедурах. ИПКс-ИИ-3.1

Подготовка к практическому занятию:

– внимательно прочитать материал лекций, относящихся к данному занятию, ознакомиться с учебными материалами, включая электронные в соответствии с предложенным списком литературы в рабочей программе учебной дисциплины;

– подготовить развернутые ответы на вопросы для проведения опроса по теме практического занятия;

– понять, что осталось неясными и постараться получить на них ответ заранее;

– готовиться к практическому занятию можно как индивидуально, так и в составе малой группы.

Вопросы к практическому занятию для проведения опроса:

Вопросы для устного опроса:

1. Стандарты доказывания и бремя объяснения (right to explanation).
2. Назначение экспертизы объяснимости ИИ.
3. Практика регуляторов: требования к отчётам об интерпретируемости.
4. Построение правовой позиции на основе XAI-отчёта.

Ситуационные задачи.

Задание 1

Вы представляете истца в деле о дискриминации ИИ-системой кредитного скоринга. У вас есть XAI-отчет (SHAP-значения), показывающий, что признак «пол» имел высокую

важность. Ответчик утверждает, что XAI-отчет не является допустимым доказательством. *Вопрос: как обосновать допустимость XAI-отчета? Какие требования к его оформлению?*

Задание 2

Регулятор (Роскомнадзор) проводит проверку ИИ-системы, обрабатывающей ПД. Требует предоставить отчет об интерпретируемости.

Вопрос: какие разделы должен содержать отчет об интерпретируемости по требованиям регулятора? На какой нормативный акт можно ссылаться (РФ, ЕС)?

5.3. Один или несколько тематических блоков дисциплины завершаются контрольной точкой (далее – КТ). Текущий контроль успеваемости по дисциплине предусматривает не менее 2 (двух) и не более 10 (десяти) КТ в течение периода освоения дисциплины.

Максимальное количество баллов за любой тип работ в рамках КТ составляет 100 (сто) баллов.

Распределение весовых коэффициентов по КТ в рамках текущего контроля успеваемости по дисциплине и формулы расчета:

Наименование контрольной точки	Максимальное количество баллов за работу в рамках КТ, которое может набрать обучающийся	Коэффициент веса контрольной точки	Результат контрольной точки, участвующий в формировании итоговой балльной оценки по дисциплине (отражается в журнале БРС в СДО)
КТ 1	100	0,3	30
КТ 2	100	0,3	30
Итого:	x	0,6	60

Формула расчета результата контрольной точки:

Результат контрольной точки = Количество баллов за работу в рамках КТ X Коэффициент веса контрольной точки.

5.4. Формы текущего контроля успеваемости обучающихся в рамках КТ и типовые оценочные материалы:

КТ – 1.

Тема 1-6.

Доклад.

Тематика докладов:

1. Причины «чёрного ящика» в ML и необходимость объяснимости для права.
2. Определение XAI, интерпретируемость vs объяснимость.
3. Правовые и этические драйверы доверия к ИИ.
4. Ограничения объяснимости: точность vs прозрачность.
5. Intrinsic vs post-hoc объяснения.
6. Model-specific vs model-agnostic методы.
7. Глобальная и локальная интерпретируемость.
8. Выбор метода в зависимости от юридической задачи (доказывание, аудит, спор).

9. Принципы LIME (локальные аппроксимации).
10. SHAP-значения и интерпретация важности признаков.
11. Деревья решений и rule-based модели как понятные для суда.
12. Практический разбор: объяснение кредитного или кадрового AI-решения.
13. Проблема объяснения ответов LLM (галлюцинации, источники).
14. Механизмы внимания (attention) и их доказательственная ценность.
15. Правовые риски непроверяемого вывода GenAI.
16. Требования к explainability в AI-актах (EC, draft РФ).
17. Связь объяснимости и доверия (user trust vs legal reliance).
18. Выявление дискриминации через XAI-инструменты.
19. Fairness metrics и их интерпретация для юриста.
20. Правовые последствия необъяснимой предвзятости.
21. Стандарты доказывания и бремя объяснения (right to explanation).
22. Назначение экспертизы объяснимости ИИ.
23. Практика регуляторов: требования к отчётам об интерпретируемости.
24. Построение правовой позиции на основе XAI-отчёта.

Методические рекомендации по подготовке доклада.

Подготовка доклада способствует формированию навыков исследовательской работы, расширяет познавательные интересы, приучает критически мыслить. При написании доклада по заданной теме составляется план, подбираются основные источники. В процессе работы с источниками, систематизируют полученные сведения, делают выводы и обобщения.

Подготовка доклада требует от обучающегося большой самостоятельности и серьезной интеллектуальной работы, которая принесет наибольшую пользу, если будет включать с себя следующие этапы: изучение наиболее важных научных работ по данной теме, перечень которых дает сам преподаватель; анализ изученного материала, выделение наиболее значимых для раскрытия темы фактов, мнений разных ученых и научных положений; обобщение и логическое построение материала доклада, например, в форме развернутого плана; написание текста доклада с соблюдением требований научного стиля.

Построение доклада включает три части: вступление, основную часть и заключение. Во вступлении указывается тема доклада, устанавливается логическая связь ее с другими темами или место рассматриваемой проблемы среди других проблем, дается краткий обзор источников, на материале которых раскрывается тема и т. п. Основная часть должна иметь четкое логическое построение, в ней должна быть раскрыта тема доклада. В заключении обычно подводятся итоги, формулируются выводы, подчеркивается значение рассмотренной проблемы и т. п.

Критерии оценивания доклада:

Критерии оценки	Диапазон баллов	Описание критерия
Содержание и раскрытие темы	0-20	Детальное, последовательное описание всех этапов с конкретными примерами
Грамотность изложения	0-20	Соблюдены все правила грамматики, орфографии и пунктуации

Стилистика	0-20	Единый стиль изложения, точные формулировки, уместное использование терминов, лаконичность
Логика изложения	0-20	Чёткая последовательность изложения, логические связи между частями текста, аргументы подтверждают выводы
Оригинальность	0-20	Уникальный подход к теме, нестандартные решения, инновационные идеи, собственная позиция автора
Итого максимально:	100	

КТ – 2.

Темы 1-6
Устный опрос:

Вопросы для опроса:

№ п.п.	Содержание вопроса
1.	Причины «чёрного ящика» в ML и необходимость объяснимости для права.
2.	Определение XAI, интерпретируемость vs объяснимость.
3.	Правовые и этические драйверы доверия к ИИ.
4.	Ограничения объяснимости: точность vs прозрачность.
5.	Intrinsic vs post-hoc объяснения.
6.	Model-specific vs model-agnostic методы.
7.	Глобальная и локальная интерпретируемость.
8.	Выбор метода в зависимости от юридической задачи (доказывание, аудит, спор).
9.	Принципы LIME (локальные аппроксимации).
10.	SHAP-значения и интерпретация важности признаков.
11.	Деревья решений и rule-based модели как понятные для суда.
12.	Практический разбор: объяснение кредитного или кадрового AI-решения.
13.	Проблема объяснения ответов LLM (галлюцинации, источники).
14.	Механизмы внимания (attention) и их доказательственная ценность.

15.	Правовые риски непроверяемого вывода GenAI.
16.	Требования к explainability в AI-актах (ЕС, draft РФ).
17.	Связь объяснимости и доверия (user trust vs legal reliance).
18.	Выявление дискриминации через XAI-инструменты.
19.	Fairness metrics и их интерпретация для юриста.
20.	Правовые последствия необъяснимой предвзятости.
21.	Стандарты доказывания и бремя объяснения (right to explanation).
22.	Назначение экспертизы объяснимости ИИ.
23.	Практика регуляторов: требования к отчётам об интерпретируемости.
24.	Построение правовой позиции на основе XAI-отчёта.

Критерии оценивания опроса:

Диапазон баллов	Описание критерия
85-100	Обучающийся полно излагает материал (отвечает на вопрос), дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из учебника, но и самостоятельно составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.
65-84	Обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1–2 ошибки, которые сам же исправляет, и 1–2 недочета в последовательности и языковом оформлении излагаемого.
55-64	Обучающийся обнаруживает знание и понимание основных положений данной темы, но излагает материал неполно и допускает неточности в определении понятий или формулировке правил; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.
0-54	Обучающийся обнаруживает незнание вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал.

5.5. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (*при необходимости*).

Для ответа на вопросы студенту разрешается использование нормативно-правовых актов.

6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

6.1. Промежуточная аттестация (зачет, экзамен) проводится в устной форме. Обучающийся получает экзаменационный билет и готовится к устному ответу.

6.2. Типовые оценочные материалы промежуточной аттестации

Типовые проверочные задания для самоподготовки обучающегося к промежуточной аттестации:

Тема 1. Понятие, предпосылки и границы объяснимого ИИ (XAI). ИУК-1.1

1. Задания открытого типа.

№ п.п.	Вопрос
1.	Почему сложные ML-модели называют «чёрным ящиком»?
2.	Дайте определение XAI. Чем объяснимость отличается от интерпретируемости?
3.	Назовите основные правовые драйверы внедрения XAI.
4.	В чём заключается компромисс между точностью и объяснимостью?

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1	Что понимается под «чёрным ящиком» (black box) в контексте машинного обучения? а) Модель, которая физически защищена от несанкционированного доступа; б) Модель, внутренние механизмы принятия решений которой непрозрачны и труднообъяснимы для человека; в) Модель, использующая только анонимные данные; г) Модель, работающая только на квантовых вычислениях.		

Задание 2

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
2	В чём заключается компромисс между точностью (accuracy) и объяснимостью (explainability) модели? а) Точные модели всегда полностью объяснимы; б) Объяснимые модели всегда точнее сложных; в) Высокоточные модели часто менее объяснимы, а простые объяснимые модели могут уступать в точности; г) Компромисс отсутствует — можно одновременно достичь идеальной точности и полной объяснимости.		

3. Задания закрытого типа.

Тестовые задания.

Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

Внимательно прочитать предложенные варианты ответа.

Выбрать один верный ответ.

Записать только букву выбранного варианта ответа.

1. Почему сложные ML-модели (например, глубокие нейронные сети) называют «чёрным ящиком»?

а) Потому что они имеют чёрный корпус

б) Потому что их внутренние механизмы принятия решений непрозрачны и труднообъяснимы для человека

в) Потому что они работают только в тёмное время суток

г) Потому что они не используют математические вычисления

2. Что такое XAI (Explainable Artificial Intelligence)?

а) Это раздел права, регулирующий ответственность за ошибки ИИ

б) Это набор методов, делающих решения ИИ понятными для человека

в) Это язык программирования для создания нейросетей

г) Это международный стандарт безопасности ИИ

3. Какой из перечисленных актов закрепляет «право на объяснение» (right to explanation) в ЕС?

а) Бернская конвенция
б) GDPR (Общий регламент о защите данных)
в) Стокгольмская декларация
г) Парижское соглашение

4. В чём заключается компромисс между точностью и объяснимостью модели?
а) Точные модели всегда полностью объяснимы
б) Объяснимые модели всегда точнее сложных
в) Высокоточные модели часто менее объяснимы, а простые объяснимые модели могут уступать в точности
г) Компромисс отсутствует, можно достичь и высокой точности, и полной объяснимости

Тема 2. Классификация методов интерпретируемости и объяснения моделей. ИПКс-ИИ-3.1

1. Задания открытого типа.

№ п.п.	Вопрос
1.	В чём разница между intrinsic и post-hoc методами объяснения?
2.	Что означает model-specific и model-agnostic подходы?
3.	Чем глобальная интерпретируемость отличается от локальной?
4.	Как выбрать метод объяснения в зависимости от юридической задачи?

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1	В чём состоит различие между intrinsic и post-hoc методами объяснения? а) Intrinsic методы применяются только для текстовых данных, post-hoc — для изображений; б) Intrinsic — модель прозрачна «изнутри» (линейная регрессия), post-hoc — объяснения создаются после обучения сложной модели; в) Intrinsic методы всегда точнее post-hoc; г) Intrinsic и post-hoc — это		

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
	синонимы, обозначающие одно и то же.		

Задание 2

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
2	Какой метод объяснения предпочтительнее для судебного спора по конкретному случаю (например, обжалование отказа в кредите)? а) Глобальная интерпретируемость; б) Локальные методы (например, LIME, SHAP); в) Только intrinsic модели; г) Никакие методы не нужны, суд сам разберётся.		

3. Задания закрытого типа.

Тестовые задания.

Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

Внимательно прочитать предложенные варианты ответа.

Выбрать один верный ответ.

Записать только букву выбранного варианта ответа.

1. В чём разница между intrinsic и post-hoc методами объяснения?

а) Intrinsic применяются только в России, post-hoc — за рубежом

б) Intrinsic — модель прозрачна «изнутри» (линейная регрессия), post-hoc — объяснения создаются после обучения сложной модели

в) Intrinsic объясняют только изображения, post-hoc — только текст

г) Разницы нет, это синонимы

2. Что означает model-agnostic подход к объяснению?

а) Метод, работающий только с нейронными сетями

б) Метод, работающий только с деревьями решений

в) Универсальный метод, применимый к любой модели («чёрному ящику»)

г) Метод, который не требует данных

3. Чем глобальная интерпретируемость отличается от локальной?

а) Глобальная — объяснение одного решения, локальная — всей модели в целом
б) Глобальная — для изображений, локальная — для текста
в) Глобальная — понимание работы модели в целом, локальная — объяснение конкретного единичного решения
г) Глобальная применяется в суде, локальная — в досудебном порядке

4. Какой метод объяснения предпочтительнее для судебного спора по конкретному случаю?
а) Глобальная интерпретируемость
б) Локальные методы (например, LIME, SHAP)
в) Только intrinsic модели
г) Никакие методы не нужны

Тема 3. Построение объяснений для правовых задач: LIME, SHAP, деревья решений. ИУК-1.1

1. Задания открытого типа.

№ п.п.	Вопрос
1.	В чём заключается принцип работы LIME?
2.	Что такое SHAP-значения и как их интерпретировать?
3.	Почему деревья решений и rule-based модели считаются понятными для суда?
4.	Приведите пример правового разбора с использованием LIME или SHAP.

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1	В чём заключается принцип работы метода LIME? а) LIME создаёт точную копию исходной модели с меньшим числом параметров; б) LIME аппроксимирует сложную модель локально простой интерпретируемой моделью на основе возмущений данных; в) LIME удаляет все признаки, снижающие точность; г) LIME заменяет любую модель на линейную регрессию глобально.		

Задание 2

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
2	Что показывают SHAP-значения применительно к предсказанию модели? а) Только общую точность модели на тестовой выборке; б) Вклад каждого признака в отклонение предсказания от среднего значения; в) Вероятность того, что модель ошиблась; г) Время, затраченное моделью на обработку одного объекта.		

3. Задания закрытого типа.

Тестовые задания.

Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

Внимательно прочитать предложенные варианты ответа.

Выбрать один верный ответ.

Записать только букву выбранного варианта ответа.

1. В чём заключается принцип работы LIME?

а) LIME создаёт копию модели для каждого нового объекта

б) LIME аппроксимирует сложную модель локально простой интерпретируемой моделью на основе возмущений данных

в) LIME удаляет все сложные признаки из модели

г) LIME заменяет модель на случайный лес

2. Что показывают SHAP-значения?

а) Только точность модели

б) Вклад каждого признака в отклонение предсказания от среднего значения

в) Скорость работы модели

г) Объём обучающей выборки

3. Почему деревья решений считаются понятными для суда?

а) Потому что они всегда точнее нейросетей

б) Потому что логика принятия решений представлена в виде наглядных правил «если — то»

в) Потому что они не используют математику

г) Потому что они запрещены законом

4. Какое SHAP-значение указывает на то, что признак увеличил предсказание?

а) Отрицательное
б) Нулевое
в) Положительное
г) Любое, кроме нуля

Тема 4. Объяснимость генеративного ИИ и больших языковых моделей (LLM). ИПКс-ИИ-3.1

1. Задания открытого типа.

№ п.п.	Вопрос
1.	В чём заключается проблема объяснения ответов LLM?
2.	Что такое механизмы внимания (attention) и какова их доказательственная ценность?
3.	Назовите правовые риски использования необъяснимого GenAI.
4.	Какие требования к объяснимости GenAI содержатся в зарубежных AI-актах?

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1	Что такое «галлюцинации» (hallucinations) больших языковых моделей (LLM)? а) Специальный режим энергосбережения модели; б) Генерация моделью ложных, но правдоподобных фактов, не соответствующих действительности; в) Способ шифрования выходных данных для защиты конфиденциальности; г) Функция визуализации внимания модели.		

Задание 2

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
2	Какова доказательственная ценность механизмов внимания		

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
	(attention) в судебном процессе? а) Attention является безусловным доказательством причинно-следственной связи; б) Attention — вспомогательное доказательство, которое не всегда коррелирует с истинной причинностью; в) Attention не может использоваться в суде ни при каких условиях; г) Attention полностью заменяет экспертизу объяснимости ИИ.		

3. Задания закрытого типа.

Тестовые задания.

Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

Внимательно прочитать предложенные варианты ответа.

Выбрать один верный ответ.

Записать только букву выбранного варианта ответа.

1. Что такое «галлюцинации» (hallucinations) LLM?

а) Специальный режим энергосбережения модели

б) Генерация моделью ложных, но правдоподобных фактов, не соответствующих действительности

в) Способ шифрования выходных данных

г) Функция визуализации внимания модели

2. Что такое механизмы внимания (attention) в LLM?

а) Компонент, определяющий, на какие части входного текста модель «обращает внимание» при генерации

б) Функция, отправляющая уведомления разработчику

в) Модуль для сохранения конфиденциальности данных

г) Алгоритм сжатия модели

3. Какова доказательственная ценность attention в суде?

а) Attention является безусловным доказательством причинно-следственной связи

б) Attention — вспомогательное доказательство, которое не всегда коррелирует с истинной причинностью

в) Attention не может использоваться в суде ни при каких условиях

г) Attention заменяет экспертизу

4. Какой из перечисленных актов ЕС предъявляет требования к прозрачности систем генеративного ИИ общего назначения (GPAI)?

а) Маастрихтский договор
б) Лиссабонский договор
в) EU AI Act (Закон об искусственном интеллекте ЕС)
г) Шенгенское соглашение

Тема 5. Доверие к ИИ, алгоритмическая справедливость и борьба с предвзятостью. ИУК-1.1

1. Задания открытого типа.

№ п.п.	Вопрос
1.	Как связаны объяснимость и доверие к ИИ? Разграничьте user trust и legal reliance.
2.	Как с помощью XAI-инструментов выявить дискриминацию в алгоритмах?
3.	Что такое fairness metrics и как юристу их интерпретировать?
4.	Каковы правовые последствия использования необъяснимой предвзятой модели?

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1	В чём разница между user trust и legal reliance применительно к ИИ? а) User trust — доверие к одному пользователю, legal reliance — ко всем; б) User trust — психологическое доверие пользователя, legal reliance — юридическая возможность ссылаться на результат ИИ в суде; в) User trust применяется в США, legal reliance — в Европе; г) Разницы нет, это синонимы.		

Задание 2

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
2	Какая fairness-метрика измеряет равную долю положительных решений (например, одобрение		

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
	кредита) в разных группах (например, среди мужчин и женщин)? а) Равенство возможностей (Equal Opportunity); б) Индивидуальная справедливость (Individual Fairness); в) Демографический паритет (Demographic Parity); г) Точность модели (Accuracy).		

3. Задания закрытого типа.

Тестовые задания.

Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

Внимательно прочитать предложенные варианты ответа.

Выбрать один верный ответ.

Записать только букву выбранного варианта ответа.

1. В чём разница между user trust и legal reliance?

а) User trust — психологическое доверие пользователя, legal reliance — юридическая возможность ссылаться на результат ИИ в суде

б) User trust — доверие к одному пользователю, legal reliance — ко всем

в) User trust применяется в США, legal reliance — в Европе

г) Разницы нет, это синонимы

2. Какая fairness-метрика измеряет равную долю положительных решений в разных группах?

а) Равенство возможностей

б) Индивидуальная справедливость

в) Демографический паритет

г) Точность модели

3. Как с помощью SHAP можно выявить дискриминацию в алгоритме?

а) Сравнить SHAP-значения защищённого признака для разных групп

б) Измерить скорость работы модели

в) Посчитать количество параметров модели

г) Увеличить обучающую выборку

4. Какое правовое последствие НЕ наступает при использовании необъяснимой предвзятой модели?

а) Административная ответственность по 152-ФЗ

б) Гражданско-правовая ответственность за причинение вреда

в) Обязательное лицензирование модели

г) Судебные иски о дискриминации

Тема 6. Использование XAI в судебных и регуляторных процедурах. ИПКс-ИИ-3.1

1. Задания открытого типа.

№ п.п.	Вопрос
1.	Что означает «право на объяснение» (right to explanation) в GDPR и иных актах?
2.	Как назначается и проводится экспертиза объяснимости ИИ в судебном процессе?
3.	Какие требования к отчётам об интерпретируемости предъявляют регуляторы?
4.	Как на основе XAI-отчёта построить правовую позицию для суда или регулятора?

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
1	Что означает «право на объяснение» (right to explanation) согласно GDPR? а) Право требовать удаления персональных данных; б) Право требовать содержательного объяснения логики работы автоматизированного решения, принятого в отношении субъекта; в) Право на бесплатный доступ к исходному коду ИИ-системы; г) Право на автоматическое обжалование любого решения ИИ.		

Задание 2

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
2	Какой документ является основным для демонстрации соответствия ИИ-системы требованиям прозрачности и объяснимости при проверке регулятором (например, Роскомнадзором)? а) Договор с разработчиком; б) Отчёт об интерпретируемости (XAI-		

№ п.п.	Содержание задания	Правильный ответ	Аргументы, обосновывающие выбор ответа
	отчёт); в) Бухгалтерский баланс организации; г) Рекламные материалы системы.		

3. Задания закрытого типа.

Тестовые задания.

Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

Внимательно прочитать предложенные варианты ответа.

Выбрать один верный ответ.

Записать только букву выбранного варианта ответа.

1. Что означает «право на объяснение» (right to explanation) по GDPR?

а) Право требовать удаления персональных данных

б) Право требовать содержательного объяснения логики работы автоматизированного решения

в) Право на бесплатный доступ к интернету

г) Право на смену оператора данных

2. Кто назначает экспертизу объяснимости ИИ в судебном процессе?

а) Эксперт самостоятельно

б) Истец

в) Суд

г) Регулятор по своему усмотрению

3. Какой документ является основным для демонстрации соответствия ИИ-системы требованиям прозрачности регулятору?

а) Договор с разработчиком

б) Отчёт об интерпретируемости (XAI-отчёт)

в) Бухгалтерский баланс

г) Штатное расписание

4. На основе XAI-отчёта можно построить правовую позицию для:

а) Обоснования причинно-следственной связи в деле о причинении вреда

б) Доказательства отсутствия дискриминации

в) Оспаривания необоснованного решения регулятора

г) Все перечисленные варианты

6.3. Критерии и шкала оценивания на основе БРС.

КРИТЕРИИ ОЦЕНИВАНИЯ	РЕЗУЛЬТАТ В БАЛЛАХ

Дан полный, в логической последовательности развернутый ответ на поставленный вопрос, где он продемонстрировал знания предмета в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину, самостоятельно, и исчерпывающе отвечает на дополнительные вопросы, приводит собственные примеры по проблематике поставленного вопроса, решил предложенные практические задания без ошибок	40
Дан развернутый ответ на поставленный вопрос, где обучающийся демонстрирует знания, приобретенные на лекционных и семинарских занятиях, а также полученные посредством изучения обязательных учебных материалов по курсу, дает аргументированные ответы, приводит примеры, в ответе присутствует свободное владение монологической речью, логичность и последовательность ответа. Однако допускается неточность в ответе. Решил предложенные практические задания с небольшими неточностями.	30-39
Дан ответ, свидетельствующий в основном о знании процессов изучаемой дисциплины, отличающийся недостаточной глубиной и полнотой раскрытия темы, знанием основных вопросов теории, слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры, недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа и решении практических заданий.	20-29
Дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Выводы поверхностны. Решение практических заданий не выполнено, т.е. обучающийся не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.	0-19

6.4. Для решения практических заданий обучающемуся разрешается использование нормативно-правовых актов.

7. Методические материалы по освоению дисциплины

Подготовка к лекциям.

Главное в период подготовки к лекционным занятиям – научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин. Каждому обучающемуся следует составлять еженедельный и семестровый планы работы, а также план

на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

Самостоятельная работа на лекции.

Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность обучающегося. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим обучающимся. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось просить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Подготовка к практическим занятиям.

Подготовку к каждому практическому занятию каждый обучающийся должен начать с ознакомления с планом практического занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. На основе индивидуальных предпочтений обучающемуся необходимо самостоятельно выбрать тему доклада по проблеме практического занятия и по возможности подготовить по нему презентацию. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции (устно или 10 письменно). Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности обучающегося свободно ответить на теоретические вопросы практического занятия, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Структура практического занятия:

В зависимости от содержания и количества отведенного времени на изучение каждой темы может практическое занятие состоять из четырех-пяти частей:

1. Обсуждение теоретических вопросов, определенных программой дисциплины.
2. Доклад и/ или выступление с презентациями по проблеме практического занятия.
3. Обсуждение выступлений по теме – дискуссия.

4. Выполнение практического задания с последующим разбором полученных результатов или обсуждение практического задания, выполненного дома, если это предусмотрено программой.

5. Подведение итогов занятия.

Первая часть – обсуждение теоретических вопросов - проводится в виде фронтальной беседы со всей группой и включает выборочную проверку преподавателем теоретических знаний обучающихся. Примерная продолжительность — до 15 минут. Вторая часть — выступление обучающихся с докладами, которые должны сопровождаться презентациями с целью усиления наглядности восприятия, по одному из вопросов практического занятия. Обязательный элемент доклада – представление и анализ статистических данных, обоснование социальных последствий любого экономического факта, явления или процесса. Примерная продолжительность — 20-25 минут. После докладов следует их обсуждение – дискуссия. В ходе этого этапа практического занятия могут быть заданы уточняющие вопросы к докладчикам. Примерная продолжительность – до 15-20 минут. Если программой предусмотрено выполнение практического задания в рамках конкретной темы, то преподавателями определяется его содержание и дается время на его выполнение, а затем идет обсуждение результатов. Если практическое задание должно было быть выполнено дома, то на практическом занятии преподаватель проверяет его выполнение (устно или письменно). Примерная продолжительность – 15-20 минут. Подведением итогов заканчивается практическое занятие. Обучающимся должны быть объявлены оценки за работу и даны их четкие обоснования. Примерная продолжительность — 5 минут.

Работа с литературными источниками.

В процессе подготовки к практическим занятиям, обучающимся необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем, что позволяет обучающимся проявить свою индивидуальность в рамках выступления на занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Вопросы для самостоятельной подготовки к лекционным и практическим занятиям:

1. Причины «чёрного ящика» в ML и необходимость объяснимости для права.
2. Определение XAI, интерпретируемость vs объяснимость.
3. Правовые и этические драйверы доверия к ИИ.
4. Ограничения объяснимости: точность vs прозрачность.
5. Intrinsic vs post-hoc объяснения.
6. Model-specific vs model-agnostic методы.
7. Глобальная и локальная интерпретируемость.
8. Выбор метода в зависимости от юридической задачи (доказывание, аудит, спор).
9. Принципы LIME (локальные аппроксимации).
10. SHAP-значения и интерпретация важности признаков.
11. Деревья решений и rule-based модели как понятные для суда.
12. Практический разбор: объяснение кредитного или кадрового AI-решения.
13. Проблема объяснения ответов LLM (галлюцинации, источники).
14. Механизмы внимания (attention) и их доказательственная ценность.

15. Правовые риски непроверяемого вывода GenAI.
16. Требования к explainability в AI-актах (ЕС, draft РФ).
17. Связь объяснимости и доверия (user trust vs legal reliance).
18. Выявление дискриминации через XAI-инструменты.
19. Fairness metrics и их интерпретация для юриста.
20. Правовые последствия необъяснимой предвзятости.
21. Стандарты доказывания и бремя объяснения (right to explanation).
22. Назначение экспертизы объяснимости ИИ.
23. Практика регуляторов: требования к отчётам об интерпретируемости.
24. Построение правовой позиции на основе XAI-отчёта.

8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет

8.1. Основная литература

1. Баланов А.Н. Искусственный интеллект. Понимание, применение и перспективы: учебник для вузов. — Санкт-Петербург: Лань, 2024. — 312 с. — ISBN 978-5-507-49392-0

2. Мишра, П. Объяснимые модели искусственного интеллекта на Python. Модель искусственного интеллекта. Объяснения с использованием библиотек, расширений и фреймворков на основе языка Python : практическое руководство / П. Мишра ; пер. с англ. С. В. Минца. - Москва : ДМК Пресс, 2022. - 298 с. - ISBN 978-5-93700-124-5. - Текст : электронный. - URL: https://znanium.com/catalog/product/2109490 (дата обращения: 07.06.2026). – Режим доступа: по подписке.

3. Di Cecco A., Gianfagna L. Explainable AI with Python. — 2nd ed. — Cham: Springer Nature Switzerland, 2025. — XX, 324 p. — ISBN 978-3-031-92229-9. — DOI: 10.1007/978-3-031-92229-9

Explainable Artificial Intelligence for Trustworthy Decisions in Smart Applications. — Cham: Springer Nature Switzerland, 2026. — ISBN 978-3-031-97007-8

4. Mahalle P.N., Ingle Y.S. Explainable Artificial Intelligence: A Practical Guide. — Gistrup: River Publishers, 2024. — 94 p. — ISBN 978-87-7004-713-5

5. Guidotti R., Schmid U., Longo L. (eds.). Explainable Artificial Intelligence: Third World Conference, xAI 2025, Istanbul, Turkey, July 9–11, 2025, Proceedings. — 5 volumes. — Cham: Springer Nature Switzerland, 2025. — (Communications in Computer and Information Science). — ISBN 978-3-031-08332-6.

6. Ziemba E.W., Grzenda W., Ramsza M. (eds.). Explainable Artificial Intelligence for Sustainable Development. — Abingdon: Taylor & Francis (Routledge), 2025. — ISBN 978-1-003-59922-7.

8.2. Дополнительная литература

1. Долинер Л.И. Искусственный интеллект: объяснимый ИИ (XAI) : учебное пособие. – Екатеринбург: Изд-во УрФУ, 2023. – 124 с.

2. Рассолов И.М., Чубукова С.Г. Право и искусственный интеллект: теория и практика : монография. – М.: Проспект, 2022. – 320 с.

3. Регулирование (ЕС) 2024/1689 Европейского парламента и Совета от 13 июня 2024 г. «Об установлении гармонизированных правил в области искусственного интеллекта (Закон об ИИ)» // Official Journal of the European Union. – 2024. – L series.

4. Sokol K., Flach P. Interpretable representations in explainable AI: from theory to practice // Data Mining and Knowledge Discovery. – 2024. – Vol. 38. – P. 3102–3140.

5. Minh D., Wang H.X., Li Y.F., Nguyen T.N. Explainable artificial intelligence: a comprehensive review // Artificial Intelligence Review. – 2022. – Vol. 55. – P. 3503–3568.

8.3. Нормативные правовые документы и иная правовая информация

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)
2. General Data Protection Regulation (GDPR) – Regulation (EU) 2016/679 (статья 22 – автоматизированное индивидуальное принятие решений)
3. Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»)
4. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных» (статья 16 – решения, принимаемые на основании исключительно автоматизированной обработки персональных данных)
5. Федеральный закон от 26.03.2025 № 57-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации в части регулирования отношений, связанных с использованием систем искусственного интеллекта, отнесенных к категории высокого риска». (Вступил в силу с 01.01.2026).
6. Постановление Правительства РФ от 15.10.2025 № 1789 «Об утверждении Правил классификации систем искусственного интеллекта по уровням риска и перечня высокорисковых систем искусственного интеллекта».
7. Постановление Правительства РФ от 28.11.2025 № 2156 «Об утверждении требований к системам управления рисками при разработке и эксплуатации высокорисковых систем искусственного интеллекта и порядка проведения оценки соответствия таких систем».
8. Приказ Минцифры России от 20.12.2025 № 987 «Об утверждении форм документов, необходимых для регистрации высокорисковых систем искусственного интеллекта, и порядка ведения реестра высокорисковых систем искусственного интеллекта».
9. Приказ Росстандарта от 25.12.2025 № 1542-ст «Об утверждении национального стандарта Российской Федерации ГОСТ Р 71026-2025 «Системы искусственного интеллекта. Высокорисковые системы. Требования к управлению рисками». (Национальный стандарт, детализирующий технические и организационные требования к управлению рисками высокорисковых ИИ-систем, гармонизированный с международными подходами).
10. Кодекс этики в сфере искусственного интеллекта (в ред. 2025 г.). (Обновленная версия документа, содержащая специальные положения о повышенных этических требованиях к разработке и применению высокорисковых систем ИИ, включая прозрачность, объяснимость и справедливость принимаемых решений) .
11. Проект федерального закона «О внесении изменений в Кодекс Российской Федерации об административных правонарушениях» (в части установления ответственности за нарушение требований к разработке и эксплуатации высокорисковых систем искусственного интеллекта). (Законопроект, предусматривающий введение административной ответственности для должностных и юридических лиц за нарушение требований, установленных Федеральным законом № 57-ФЗ и подзаконными актами).

8.4. Интернет-ресурсы

1. Официальный интернет-портал правовой информации (www.pravo.gov.ru).
2. СПС «КонсультантПлюс» (www.consultant.ru).
3. СПС «Гарант» (www.garant.ru).
4. Официальный сайт Министерства экономического развития РФ (www.economy.gov.ru). (Информация о реализации федерального проекта «Нормативное регулирование цифровой среды», включая разработку и внедрение механизмов регулирования высокорисковых систем ИИ).
5. Официальный сайт Министерства цифрового развития, связи и массовых коммуникаций РФ (digital.gov.ru). (Информация о государственной политике в сфере искусственного интеллекта, реестре высокорисковых систем ИИ, разъяснения по применению законодательства, проекты нормативных актов).

6. Сайт Альянса в сфере искусственного интеллекта (a-ai.ru). (Публикация актуальной редакции Кодекса этики в сфере ИИ, методические материалы по комплаенсу, аналитические доклады о практике применения законодательства, информация о саморегулировании отрасли).
7. Официальный портал Европейского Союза EUR-Lex (eur-lex.europa.eu). (Официальный источник для изучения полного текста EU AI Act и имплементационных регламентов на всех официальных языках ЕС).
8. Сайт Европейского совета по искусственному интеллекту (European Artificial Intelligence Board) (ai-board.europa.eu). (Официальный сайт наднационального органа, координирующего применение EU AI Act, публикующего руководства, разъяснения и рекомендации по применению законодательства, включая вопросы классификации высокорисковых систем).
9. Научная электронная библиотека eLIBRARY.RU (elibrary.ru). (Доступ к полным текстам статей из российских научных журналов по правовому регулированию искусственного интеллекта, комплаенсу и управлению рисками).
10. Профессиональные ресурсы по комплаенсу и регуляторике (comply.ru, regulation.gov.ru и аналогичные). (Площадки для отслеживания актуальных позиций регуляторов, разъяснений и практики применения законодательства о высокорисковых системах ИИ).

9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

9.1. Материально-техническая база

Перечень материально-технического обеспечения:

1. Учебные аудитории, оборудованные для проведения учебных занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, коллоквиумов, мероприятий текущего контроля и промежуточной аттестации, в том числе мультимедийным оборудованием для демонстрации электронных презентаций и аудио- и видеоматериалов;
2. Компьютерные классы для выполнения групповых тестовых и иных заданий, а также для самостоятельной работы обучающихся оснащенные компьютерной техникой и обеспечением доступа к сети «Интернет» и доступа в электронную информационно- образовательную среду организации;
3. Специализированные аудитории и лаборатории;
4. Библиотека с обеспечением печатными изданиями или электронно-библиотечная система обеспечивающая доступ к электронным изданиям (электронная библиотека);
5. Читальный зал;
6. Технические средства обучения: персональные компьютеры; компьютерные проекторы; звуковые динамики; программные средства, обеспечивающие просмотр видеофайлов в форматах AVI, MPEG-4, DivX, RMVB, WMV и др.

9.2. Информационные технологии, программное обеспечение:

При осуществлении образовательного процесса по дисциплине используются информационные технологии и программное обеспечение:

1. Современная операционная система.
2. Kaspersky Endpoint Security (или аналог).
3. Средство просмотра файлов формата pdf.
4. Современные офисные средства (текстовые и табличные редакторы,

средствработы с презентационными материалами и т.д.).

5. Архиватор 7-Zip.
6. Система дистанционного обучения.
7. Автоматизированная библиотечная система.

9.3.Информационные справочные системы:

1. https://www.iprbookshop.ru – Электронно-библиотечная система [ЭБС] «IPRSMART» (ранее – IPRBooks)
2. https://e.lanbook.com - Электронно-библиотечная система [ЭБС] «Лань».
3. Электронно-библиотечная система [ЭБС] «Znanium.com».
4. https://www.book.ru - Электронно-библиотечная система [ЭБС] «Book.ru».
5. https://ibooks.ru - Электронно-библиотечная система [ЭБС] «ibooks.ru».
6. https://ranepalib.miflib.ru - Электронная библиотека издательства «МИФ».
7. https://eivis.ru – Ивис. Полные тексты российских научных и практических журналов, а также газет центральной прессы России. Доступ с ip-адресов локальной сети Института.
8. http://www.consultant.ru/ - Справочно-правовая система «Консультант».
9. http://www.garant.ru/ Справочно-правовая система «Гарант».