Лебедева
Екатерина
Александровна

Подписано цифровой подписью:
Лебедева Екатерина
Александровна
Дата: 2026.06.25 15:48:09 +03'00'

Приложение 4

к образовательной программе

# РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.ДЭ.03.01 Тестирование, валидация и интерпретируемость (Explainable AI – XAI) моделей
(индекс, наименование дисциплины в соответствии с учебным планом)

40.04.01 Юриспруденция
(код, наименование направления подготовки/специальности)

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:**

Чайка Юлия Валерьевна, старший преподаватель кафедры гражданского и международного права

**Заведующий кафедрой:**

Калинина Евгения Валерьевна, д-р. юрид. наук., доцент, заведующий кафедрой гражданского и международного права

Рабочая программа дисциплины Б1.В.ДЭ.03.01 Тестирование, валидация и интерпретируемость (Explainable AI – XAI) моделей одобрена на заседании кафедры гражданского и международного права

протокол № 6 от «4» марта 2026 г

## 3. СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Типы оценочных материалов, показатели и критерии их оценивания
5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам
6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине
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.ДЭ.03.01 Тестирование, валидация и интерпретируемость (Explainable AI – XAI) моделей обеспечивает формирование у обучающихся следующих универсальных, общепрофессиональных и профессиональных компетенций*:

| **ОТФ/ТФ и реквизиты ПС** *(при наличии)* ** | **Код компетенции** ** | **Наименование Компетенции** ** | **Код индикатора достижения компетенций** ** | **Наименование индикатора достижения компетенций** ** | **Образовательный результат** ** |
|---|---|---|---|---|---|
| | УК-1 | Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий | ИУК-1.1 | Определяет стратегию действий на основе критического анализа и системного подхода | **Знать:** методы тестирования, валидации и интерпретируемости моделей ИИ, их сильные и слабые стороны, ограничения; правовые требования к объяснимости (GDPR, 152-ФЗ). **Уметь:** критически оценивать качество и надёжность ИИ-моделей, выявлять риски непрозрачности и дискриминации, выбирать адекватные методы XAI для конкретной правовой ситуации. **Владеть:** навыкам и системного анализа правовых проблем, связанных с «чёрным ящиком» ИИ, и выработки стратегии |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | внедрения XAI для соблюдения законодательства и защиты прав субъектов. |
| | ПКс-ИИ-1 | Способен анализировать правовые риски и разрабатывать юридические модели внедрения и эксплуатации систем искусственного интеллекта (АИ) в бизнес-процессы, включая оценку соответствия законодательству о персональных данных, интеллектуальной собственности и конкуренции | ИПКс-ИИ-1.1 | Разрабатывает юридические модели и рекомендации по внедрению и эксплуатации систем АИ в бизнес-процессы, обеспечивающие минимизацию правовых рисков и соответствие требованиям законодательства о персональных данных, интеллектуальной собственности и конкуренции | **Знать:** требования к валидации и тестированию ИИ-моделей, предусмотренные стандартами качества (ISO, NIST) и отраслевыми регламентами; правовые последствия использования неверифицированных или ненадёжных моделей.<br>**Уметь:** проводить правовой аудит процессов тестирования и валидации ИИ-проекта, разрабатывать рекомендации по включению требований к интерпретируемости в политики и регламенты организации.<br>**Владеть:** методами юридической оценки достаточности тестирования и валидации для целей комплаенса, навыками интеграции XAI в политики обработки данных и управления |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | рисками. |
| | ПКс-ИИ-3 | | 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 АИ и нарушением интеллектуальных прав на AI-модели и данные | **Знать:** практику судов по делам о вреде, причинённом алгоритмическими решениями; роль экспертизы интерпретируемости (XAI) как доказательства; стандарты доказывания причинно-следственной связи при «чёрном ящике».<br>**Уметь:** формировать правовую позицию по спорам, используя результаты тестирования и XAI для обоснования надлежащего уровня прозрачности или, напротив, необъяснимости модели; готовить ходатайства о назначении компьютерной экспертизы с акцентом на интерпретируемость.<br>**Владеть:** навыкам и подготовки процессуальных документов, включающих ссылки на методы XAI и валидации, и взаимодействия с экспертами в |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | области ИИ для доказывания причинно-следственных связей. |

** Дисциплина может формировать компетенцию полностью или частично.*

***Должно соответствовать Приложению 1 к образовательной программе*

## 2. Объем и место дисциплины в структуре образовательной программы

Общая трудоемкость дисциплины составляет 108 а.ч. (3 з.е.)

Количество академических часов, выделенных на контактную работу с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся составляет 32 часов, в том числе:

- очная форма обучения:

3 сем. - лекции – 14 а.ч., лабораторные занятия – 2 а.ч., практические занятия – 12 а.ч., самостоятельная работа – 76 а.ч., контактная работа на аттестацию в период экзаменационных сессий – 4 часа.

Дисциплина «*Б1.В.ДЭ.03.01 Тестирование, валидация и интерпретируемость (Explainable AI – XAI) моделей»* относится к элективной части (дисциплины по выбору). Изучается в 3 семестре 2 курса магистерской программы.

Для освоения дисциплины «*Б1.В.ДЭ.03.01 Тестирование, валидация и интерпретируемость (Explainable AI – XAI) моделей»* обучающиеся используют знания, умения, навыки, способы деятельности и установки, сформированные в ходе изучения дисциплин: *«Информационные технологии в юридической деятельности», «Информационное право», «Правовые режимы данных «Data Law)»*, «Блокчейн, смарт-контракты и цифровые активы в контексте ИИ».

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 | ВСЕГО | Объем дисциплины, ак.час | | | | | | | | | | | | Форма текущего контроля успеваемости, промежуточной аттестации |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Контактная работа обучающихся с преподавателем по видам учебных занятий | | | | | | | | | Самостоятельная работа | | | |
| | | | Период теоретического обучения | | | | | | Период промежуточной аттестации (сессия) | | | | | | |
| | | | Занятия лекционного типа | | Занятия семинарского типа | | ИК | КСР | КЭ | Каттэк | Контроль | СРкр | СРэк | СР | |
| | | | Л | ВЛ | ЛР | ПЗ | | | | | | | | | |
| Тема 1 | Введение в тестирование и валидацию моделей ИИ: правовой контекст | 18 | 2 | | | 2 | | | | | | | | 14 | Устный опрос |
| Тема 2 | Методы тестирования моделей ИИ: метрики, робастность, безопасность | 18 | 2 | | | 2 | | | | | | | | 12 | Устный опрос |
| Тема 3 | Валидация моделей ИИ: процедуры, стандарты, правовое | 18 | 2 | | | 2 | | | | | | | | 12 | Устный опрос |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | значение | | | | | | | | | | | | | |
| Тема 4 | Интерпретируемость и объяснимость (XAI): понятие, методы | 18 | 2 | | | 2 | | | | | | | 12 | Устный опрос |
| Тема 5 | Правовое значение XAI: GDPR, 152-ФЗ, отраслевые требования | 18 | 4 | | | 2 | | | | | | | 14 | Устный опрос |
| Тема 6 | XAI в судебных спорах и экспертизе: доказывание, ответственность за вред от ИИ | 18 | 4 | | 2 | 2 | 2 | | | | | | 12 | Устный опрос |
| Промежуточная аттестация | | 4 | | | | | | | | 4 | | | | Зачет |
| **Итого** | | 108 | 14 | | 2 | 12 | | | | 4 | | | 76 | |

*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

Контроль - контактная работа на аттестацию в период экзаменационных сессий для заочной формы обучения

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. Введение в тестирование и валидацию моделей ИИ: правовой контекст.**

Понятие тестирования и валидации в жизненном цикле ИИ-моделей. Отличие от обучения (training) и оценки (evaluation). Правовые стимулы: требования надёжности и безопасности в GDPR (ст. 35, DPIA), в 152-ФЗ (ст. 19 – достаточные меры), в проектах регулирования ИИ (ЕС AI Act). Риски использования непроверенных моделей: дискриминация, ошибки, вред. Роль валидации в доказывании добросовестности оператора. (УК-1).

**Тема 2. Методы тестирования моделей ИИ: метрики, робастность, безопасность.**

Основные метрики качества: точность (accuracy), полнота (recall), F1-мера, AUC-ROC, среднеквадратичная ошибка (MSE). Тестирование на робастность (устойчивость к аномалиям, adversarial attacks). Безопасностное тестирование: граничные значения, выявление скрытых уязвимостей. Правовое значение выбора метрик: соответствие заявленным характеристикам, ответственность за вводящие в заблуждение метрики. (ПКс-АИ-1).

**Тема 3. Валидация моделей ИИ: процедуры, стандарты, правовое значение.**

Офлайн-валидация (кросс-валидация, hold-out) vs онлайн-валидация (A/B тестирование, канареечное развёртывание). Стандарты: ISO/IEC 25000 (SQuaRE), NIST AI Risk Management Framework, ГОСТ Р 59277-2020. Валидация как элемент due diligence при внедрении ИИ. Документирование валидации для регуляторов (Роскомнадзор, ФАС) и суда. (ПКс-АИ-1).

**Тема 4. Интерпретируемость и объяснимость (XAI): понятие, методы.**

Определение: интерпретируемость (interpretability) vs объяснимость (explainability). Проблема «чёрного ящика» в сложных моделях (нейросети, ансамбли). Методы XAI: глобальные и локальные. LIME (Local Interpretable Model-agnostic Explanations), SHAP (Shapley Additive Explanations), частичные зависимости (PDP), карты внимания (attention maps) для трансформеров. Выбор метода в зависимости от правовой задачи (опровержение дискриминации, обоснование отказа, выявление ошибки). (УК-1, ПКс-АИ-1).

**Тема 5. Правовое значение XAI: GDPR, 152-ФЗ, отраслевые требования.**

Право на объяснение (right to explanation) в GDPR (ст. 13–15, 22, Преамбула 71). Разъяснения EDPB (Guidelines on Automated individual decision-making). Соответствие XAI требованиям 152-ФЗ: право субъекта знать об обработке, недопустимость решений на основе исключительно автоматизированной обработки (ст. 16 152-ФЗ). Отраслевые требования: банковское регулирование (Положение Банка России № 719-П об ИИ в скоринге), медицинские стандарты (Clinical Decision Support Systems). (ПКс-АИ-1, ПКс-АИ-3).

**Тема 6. XAI в судебных спорах и экспертизе: доказывание, ответственность за вред от ИИ.**

Структура исков о вреде, причинённом решением ИИ: необходимость установить причинно-следственную связь между дефектом модели и вредом. XAI как источник доказательств: объяснения решений (SHAP, LIME) могут подтвердить дискриминацию

или ошибку. Назначение компьютерной экспертизы с вопросами об интерпретируемости. Правовые позиции: достаточность объяснений для освобождения от ответственности (due diligence) или, наоборот, их отсутствие как признак вины. Примеры судебных дел (США: State v. Loomis (COMPAS), ЕС: Schufa case (C-634/21)). Роль юриста в формулировании вопросов эксперту и оспаривании XAI-выводов. (ПКс-АИ-3).

## 4. Типы оценочных материалов, показатели и критерии оценивания

4.1. Оценочные материалы по дисциплине Б1.В.ДЭ.03.01 Тестирование, валидация и интерпретируемость (Explainable AI – XAI) моделей входят в состав оценочных материалов по образовательной программе. Совокупность оценочных материалов по всем дисциплинам (модулям) образовательной программы составляет 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.<br>2. Внимательно прочитать предложенные вариант-ты ответа.<br>3. Выбрать один верный ответ.<br>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.<br>2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.;<br>список 2 – утверждения, свойства объектов и т.д.<br>3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов.<br>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.<br>2. Внимательно прочитать предложенные вариант-ты ответа.<br>3. Выбрать несколько правильных ответов.<br>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.<br>2. Внимательно прочитать предложенные варианты ответа.<br>3. Построить верную последовательность из предложенных элементов.<br>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 выбором одного правильного ответа из предложенных и | Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор | 1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.<br>2. Внимательно прочитать предложенные варианты ответа.<br>3. Выбрать один верный ответ. | Ответ считается верным, если правильно указана цифра или буква и приведены корректные аргументы, используемые при выборе ответа |

| | | | |
|---|---|---|---|
| обоснованием выбора | ответа | 4. Записать только номер (или букву) выбранного варианта ответа.<br>5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования). | |
| Задание открытого типа с развернутым ответом | Прочитайте текст и запишите развернутый обоснованный ответ | 1. Внимательно прочитать текст задания и понять суть вопроса.<br>2.Продумать логику и полноту ответа.<br>3.Записать ответ, используя четкие компактные формулировки.<br>4.В случае расчетной задачи, записать решение и ответ | Ответ считается верным:<br>1. Отсутствие фактических ошибок.<br>2. Раскрытие объема используемых понятий (полнота ответа).<br>3.Обоснованность ответа (наличие аргументов).<br>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 | 100баллов |

**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.ДЭ.03.01 Тестирование, валидация и интерпретируемость (Explainable AI – XAI) моделей используются следующие формы текущего контроля успеваемости обучающихся (в том числе, задания к контрольным точкам):

Устный опрос, тестирование, реферат, эссе, практическое задание (кейс).

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

*Включаются типовые оценочные материалы по формам текущего контроля, перечисленным соответственно в п. 3.1 и 4.1, вне контрольных точек*

**Тема 1. Введение в тестирование и валидацию моделей ИИ: правовой контекст. УК-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 для опроса:*

1. Понятие тестирования и валидации в жизненном цикле ИИ-моделей. Отличие от обучения (training) и оценки (evaluation).
2. Правовые стимулы: требования надёжности и безопасности в GDPR (ст. 35, DPIA), в 152-ФЗ (ст. 19 – достаточные меры), в проектах регулирования ИИ (EC AI Act).
3. Риски использования непроверенных моделей: дискриминация, ошибки, вред.
4. Роль валидации в доказывании добросовестности оператора.

**Тема 2. Методы тестирования моделей ИИ: метрики, робастность, безопасность. ПКс-АИ-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 для опроса:*

1. Основные метрики качества: точность (accuracy), полнота (recall), F1-мера, AUC-ROC, среднеквадратичная ошибка (MSE).
2. Тестирование на робастность (устойчивость к аномалиям, adversarial attacks). Безопасностное тестирование: граничные значения, выявление скрытых уязвимостей.
3. Правовое значение выбора метрик: соответствие заявленным характеристикам, ответственность за вводящие в заблуждение метрики.

**Тема 3. Валидация моделей ИИ: процедуры, стандарты, правовое значение. ПКс-АИ-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 для опроса:*

1. Офлайн-валидация (кросс-валидация, hold-out) vs онлайн-валидация (A/B тестирование, канареечное развёртывание). Стандарты: ISO/IEC 25000 (SQuaRE), NIST AI Risk Management Framework, ГОСТ Р 59277-2020.
2. Валидация как элемент due diligence при внедрении ИИ.
3. Документирование валидации для регуляторов (Роскомнадзор, ФАС) и суда.

**Тема 4. Интерпретируемость и объяснимость (XAI): понятие, методы. УК-1, ПКс-АИ-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 для опроса:*

1. Определение: интерпретируемость (interpretability) vs объяснимость (explainability). Проблема «чёрного ящика» в сложных моделях (нейросети, ансамбли).
2. Методы XAI: глобальные и локальные. LIME (Local Interpretable Model-agnostic Explanations), SHAP (Shapley Additive Explanations), частичные зависимости (PDP), карты внимания (attention maps) для трансформеров.
3. Выбор метода в зависимости от правовой задачи (опровержение дискриминации, обоснование отказа, выявление ошибки)

**Тема 5. Правовое значение XAI: GDPR, 152-ФЗ, отраслевые требования. ПКс-АИ-1, ПКс-АИ-3.**

**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 для опроса:*

1. Право на объяснение (right to explanation) в GDPR (ст. 13–15, 22, Преамбула 71). Разъяснения EDPB (Guidelines on Automated individual decision-making).
2. Соответствие XAI требованиям 152-ФЗ: право субъекта знать об обработке, недопустимость решений на основе исключительно автоматизированной обработки (ст. 16 152-ФЗ).
3. Отраслевые требования: банковское регулирование (Положение Банка России № 719-П об ИИ в скоринге), медицинские стандарты (Clinical Decision Support Systems).

**Тема 6. XAI в судебных спорах и экспертизе: доказывание, ответственность за вред от ИИ. ПКи-АИ-3.**

**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 для опроса:*

1. Структура исков о вреде, причинённом решением ИИ: необходимость установить причинно-следственную связь между дефектом модели и вредом. XAI как источник доказательств: объяснения решений (SHAP, LIME) могут подтвердить дискриминацию или ошибку.
2. Назначение компьютерной экспертизы с вопросами об интерпретируемости.
3. Правовые позиции: достаточность объяснений для освобождения от ответственности (due diligence) или, наоборот, их отсутствие как признак вины.
4. Примеры судебных дел (США: State v. Loomis (COMPAS), ЕС: Schufa case (C-634/21)). Роль юриста в формулировании вопросов эксперту и оспаривании XAI-выводов.

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 студент | 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 1-3.**

Доклады.

Темы докладов:

1. Применение SHAP для объяснения решений кредитного скоринга: юридическая значимость
2. XAI как доказательство в деле о дискриминации (анализ зарубежного кейса)
3. Право на объяснение vs коммерческая тайна алгоритма: конфликт и пути решения

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.**

**Тема 4-6**

*Тестовые задания с инструкцией по выполнению:*

*В каждом вопросе один правильный ответ.*

*Пример тестового задания:*

**Вопрос:** Какая метрика наиболее чувствительна к дисбалансу классов (например, 95% одобрений и 5% отказов) и лучше отражает способность модели находить редкие положительные события?
а) Accuracy
б) Precision
в) Recall (чувствительность)
г) F1-мера

Критерии оценивания тестовых заданий:

| Диапазон баллов | Описание критерия | |
|---|---|---|
| 85-100 | Свыше 80% правильных ответов. | Обучающийся демонстрирует глубокое познание в освоенном материале. |
| 65-84 | Свыше 70% правильных ответов. | Обучающимся материал освоен полностью, без существенных ошибок. |
| 55-64 | Свыше 50% правильных ответов. | Обучающимся материал освоен не полностью, имеются значительные пробелы в знаниях. |
| 0-54 | Менее 50% правильных ответов. | Обучающимся материал не освоен, знания обучающегося ниже базового уровня. |

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

6.1. Промежуточная аттестация проводится в форме зачета с оценкой.
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. Введение в тестирование и валидацию моделей ИИ: правовой контекст. УК-1.**

Вопросы открытого типа:

| № п.п. | Вопрос |
|---|---|
| 1. | Понятие тестирования и валидации в жизненном цикле ИИ-моделей. Отличие от обучения (training) и оценки (evaluation). |
| 2. | Правовые стимулы: требования надёжности и безопасности в GDPR (ст. 35, DPIA), в 152-ФЗ (ст. 19 – достаточные меры), в проектах регулирования ИИ (EC AI Act). |
| 3. | Риски использования непроверенных моделей: дискриминация, ошибки, вред. |
| 4. | Роль валидации в доказывании добросовестности оператора. |

Задания комбинированного типа:

Тестовые задания с обоснованием выбора*:*

Задание 1. Что из перечисленного является отличием валидации (validation) ИИ-модели от её обучения (training)? Выберите один правильный ответ и кратко обоснуйте свой выбор.

1. Валидация всегда проводится на тех же данных, что и обучение.
2. Валидация использует метрики, оценивающие способность модели к обобщению на новых данных.
3. Валидация выполняется только после промышленного развёртывания модели.
4. Валидация направлена исключительно на проверку соответствия модели законодательству.

Обоснование: Обучение модели заключается в подборе параметров на тренировочных данных, тогда как валидация — это независимая оценка качества работы модели на ранее не использовавшихся данных с целью проверки её способности корректно работать в реальных условиях (обобщение). Варианты 1, 3 и 4 неверны: валидация не проводится на обучающих данных, выполняется до развёртывания, а проверка законодательства — лишь часть валидации, но не её единственная цель.

Задание 2. Какие из следующих нормативных актов прямо или косвенно вводят требования к тестированию и валидации ИИ-моделей как к мере обеспечения надёжности и безопасности? Выберите все правильные варианты и обоснуйте свой выбор.

1. GDPR (Общий регламент по защите данных), статья 35 (DPIA).
2. Федеральный закон № 152-ФЗ «О персональных данных», статья 19.
3. Проект Европейского акта об искусственном интеллекте (EU AI Act).
4. Федеральный закон «О рекламе» (№ 38-ФЗ).

Обоснование:

GDPR (ст. 35) обязывает проводить оценку воздействия на защиту данных (DPIA) для систем, использующих автоматизированную обработку, включая ИИ, что предполагает тестирование и валидацию для выявления рисков.
152-ФЗ (ст. 19) требует принятия «достаточных мер» для обеспечения безопасности персональных данных, что в контексте ИИ-моделей включает их валидацию.
EU AI Act (проект) прямо устанавливает требования к системам управления качеством данных, тестированию и валидации для высокорисковых ИИ-систем.
Закон о рекламе не содержит норм о тестировании или валидации ИИ-моделей.

Задания закрытого типа.

*Тестовые задания.*

Внимательно прочитайте вопрос и выберите один или несколько правильных, на ваш взгляд, вариантов ответа. Запишите номера выбранных ответов.

Вопрос 1. Что из перечисленного является риском использования непроверенной ИИ-модели? Выберите один правильный ответ.

1. Гарантированное увеличение точности прогнозов.
2. Полное отсутствие ошибок в работе.

3. Дискриминация отдельных групп пользователей из-за невыявленных смещений (bias).
4. Автоматическое соответствие модели всем правовым нормам.

Вопрос 2. Какие из перечисленных действий относятся непосредственно к этапу валидации ИИ-модели (а не к оценке или обучению)? Выберите все правильные варианты.

1. Проверка модели на независимом тестовом наборе данных.
2. Настройка гиперпараметров модели.
3. Оценка устойчивости модели к малым изменениям входных данных (робастность).
4. Расчет точности (accuracy) на данных, не участвовавших в обучении.

Вопрос 3. Согласно проекту EU AI Act, требования к тестированию и валидации в первую очередь предъявляются к:

1. ИИ-системам с открытым исходным кодом.
2. ИИ-системам с минимальным риском (например, спам-фильтрам).
3. Высокорисковым ИИ-системам (например, для трудоустройства, кредитования, управления критической инфраструктурой).
4. Любым чат-ботам независимо от сферы применения.

Вопрос 4. В чём заключается роль валидации ИИ-модели для доказывания добросовестности оператора при возможных юридических претензиях?

1. Валидация позволяет полностью исключить ответственность оператора за любые ошибки модели.
2. Наличие документированной процедуры валидации служит доказательством принятия «достаточных мер» для обеспечения безопасности и снижения рисков.
3. Валидация заменяет собой необходимость соблюдения требований о защите персональных данных.
4. Валидация требуется только для государственных органов, для частных компаний она не имеет правового значения.

**Тема 2. Методы тестирования моделей ИИ: метрики, робастность, безопасность. ПКс-АИ-1.**

*Вопросы открытого типа:*

| № п.п. | Вопрос |
|---|---|
| 1. | Основные метрики качества: точность (accuracy), полнота (recall), F1-мера, AUC-ROC, среднеквадратичная ошибка (MSE). |
| 2. | Тестирование на робастность (устойчивость к аномалиям, adversarial attacks). Безопасностное тестирование: граничные значения, выявление скрытых уязвимостей. |
| 3. | Правовое значение выбора метрик: соответствие заявленным характеристикам, ответственность за вводящие в заблуждение метрики. |

Задания комбинированного типа:

Тестовые задания с обоснованием выбора:

Задание 1. Вам необходимо оценить качество модели ИИ для задачи обнаружения редкого опасного заболевания (положительный класс – «заболевание есть», отрицательный – «здоров»). Частота заболевания в популяции – 0,1%. Модель показывает accuracy = 99,9%. Следует ли считать такую модель приемлемой с точки зрения задачи?
Выберите одинправильный ответ и кратко обоснуйте.

1. Да, accuracy – универсальная метрика, 99,9% – отличный результат.
2. Нет, при сильном дисбалансе классов accuracy может быть обманчива; нужны метрики полноты (recall) и F1-меры.
3. Нет, потому что accuracy не может быть выше 95% для медицинских систем по закону.
4. Да, но только при условии, что модель прошла тесты на робастность.

Задание 2. Какие из перечисленных методов тестирования направлены на проверку робастности модели ИИ? Выберите все правильные варианты и обоснуйте свой выбор.

1. Вычисление AUC-ROC на тестовой выборке.
2. Подача на вход модели примеров с малыми аддитивными шумами (adversarial perturbations).
3. Тестирование на граничных значениях входных параметров.
4. Измерение среднеквадратичной ошибки (MSE) на валидационной выборке.

*Задания закрытого типа.*

*Тестовые задания.*

Внимательно прочитайте вопрос и выберите один или несколько правильных, на ваш взгляд, вариантов ответа. Запишите номера выбранных ответов.

Вопрос 1. Какая метрика наиболее информативна для задачи бинарной классификации с сильно несбалансированными классами, если важнее минимизировать ложноположительные срабатывания (ошибочное обнаружение редкого события)?
Выберите один правильный ответ.

1. Accuracy
2. Precision (точность)
3. Recall (полнота)
4. F1-мера

Вопрос 2. Какие из перечисленных действий являются примерами безопасностного тестирования ИИ-модели? Выберите все правильныеварианты.

1. Проверка работы модели при подаче на вход значений, выходящих за пределы обучающей выборки (граничные значения).
2. Тестирование с состязательными примерами (adversarial examples).
3. Расчёт средней абсолютной ошибки (MAE) на отложенной выборке.
4. Анализ поведения модели при одновременном изменении нескольких параметров в экстремальные значения.

Вопрос 3. Согласно правовым требованиям (в том числе с учётом возможной ответственности за вводящие в заблуждение метрики), оператор ИИ-системы обязан:

1. Всегда использовать только метрику accuracy для отчётности.

2. Выбирать метрики, адекватно отражающие качество модели в условиях её реального применения, и документировать этот выбор.
3. Гарантировать значение F1-меры не менее 0,99 для любых систем.
4. Скрывать низкие значения метрик, чтобы не вводить пользователей в заблуждение.

Вопрос 4. Для задачи регрессии (прогнозирование цены акции) модель показала MSE = 0,01, но при этом систематически ошибается в 20% случаев более чем на 50% от реальной цены. Какое утверждение верно?

1. MSE полностью характеризует качество модели, дополнительные проверки не нужны.
2. MSE чувствителен к выбросам, но не показывает систематическое смещение (bias), поэтому необходимы дополнительные метрики (например, MAE, MAPE) и анализ остатков.
3. MSE не может быть равен 0,01 при таких ошибках – значит, расчёты неверны.
4. Для регрессионных задач не требуется тестирование на безопасность.

**Тема 3. Валидация моделей ИИ: процедуры, стандарты, правовое значение. ПКс-АИ-1.**

*Вопросы открытого типа:*

| № п.п. | Вопрос |
|---|---|
| 1. | Офлайн-валидация (кросс-валидация, hold-out) vs онлайн-валидация (A/B тестирование, канареечное развёртывание). Стандарты: ISO/IEC 25000 (SQuaRE), NIST AI Risk Management Framework, ГОСТ Р 59277-2020. |
| 2. | Валидация как элемент due diligence при внедрении ИИ. |
| 3. | Документирование валидации для регуляторов (Роскомнадзор, ФАС) и суда. |

Задания комбинированного типа:

Тестовые задания с обоснованием выбора:

Задание 1. В компании разработана новая модель ИИ для скоринга кредитных заявок. Модель успешно прошла офлайн-валидацию (hold-out, кросс-валидация) с высокими метриками. Однако при онлайн-валидации (A/B-тестирование) на реальных клиентах модель показала значительное снижение качества. Выберите один наиболее вероятный объясняющий фактор и кратко обоснуйте.

1. Офлайн-валидация всегда даёт завышенные оценки качества по сравнению с онлайн.
2. Распределение данных в реальном времени (data drift) отличается от исторических данных, использованных при офлайн-валидации.
3. Офлайн-валидация запрещена регуляторами для кредитных систем.
4. Валидация модели не требуется, достаточно обучения.

Задание 2. Какие из перечисленных стандартов и нормативных документовнепосредственно содержат требования к процессам валидации ИИ-моделей или могут служить основой для доказательства добросовестности оператора?
Выберите все правильные варианты и обоснуйте.

1. ISO/IEC 25000 (серия SQuaRE).
2. NIST AI Risk Management Framework.
3. ГОСТ Р 59277-2020 «Системы искусственного интеллекта. Классификация».
4. Методические рекомендации Банка России по оценке рисков при использовании ИИ (при наличии).

*Задания закрытого типа.*

*Тестовые задания.*

Внимательно прочитайте вопрос и выберите один или несколько правильных, на ваш взгляд, вариантов ответа. Запишите номера выбранных ответов.

Вопрос 1. Что из перечисленного является примером онлайн-валидации ИИ-модели? Выберите один правильный ответ.

1. Кросс-валидация по K-блокам на исторических данных.
2. Hold-out с разделением выборки 80/20.
3. Канареечное развёртывание (canary deployment) с мониторингом метрик на 1% трафика.
4. Расчёт F1-меры на тестовой выборке.

Вопрос 2. Какие из перечисленных действий являются обязательными элементами документирования валидации для предоставления регулятору (например, Роскомнадзор или ФАС) при внедрении ИИ-системы? Выберите все правильные варианты.

1. Описание процедуры валидации (метод, метрики, критерии успеха).
2. Протоколы тестирования с датами, результатами и подписями ответственных лиц.
3. Коммерческая тайна на алгоритмы – документирование не требуется.
4. Анализ расхождений между ожидаемой и фактической производительностью модели.

Вопрос 3. Валидация ИИ-модели в рамках due diligence при внедрении (например, при покупке готового ИИ-решения) выполняет следующую функцию:

1. Заменяет собой юридическую экспертизу договоров.
2. Позволяет выявить риски, связанные с недостаточным качеством, робастностью и безопасностью модели, и учесть их в условиях сделки или в процессе эксплуатации.
3. Является добровольной и не имеет правовых последствий.
4. Проводится только после того, как модель уже внедрена в продуктивную среду.

Вопрос 4. Какое утверждение о правовом значении валидации ИИ-модели является верным согласно российской и международной практике?

1. Наличие валидации полностью освобождает оператора от ответственности за вред, причинённый моделью.
2. Отсутствие валидации не может быть использовано против оператора в суде.
3. Корректно проведённая и задокументированная валидация служит доказательством принятия «достаточных мер» для обеспечения надёжности и может снизить риск привлечения к ответственности.
4. Валидация требуется только для моделей, работающих в режиме реального времени.

**Тема 4. Интерпретируемость и объяснимость (XAI): понятие, методы. УК-1, ПКс-АИ-1.**

1.Задания открытого типа.

| № п.п. | Вопрос |
|---|---|
| 1. | Определение: интерпретируемость (interpretability) vs объяснимость (explainability). Проблема «чёрного ящика» в сложных моделях (нейросети, ансамбли). |
| 2. | Методы XAI: глобальные и локальные. LIME (Local Interpretable Model-agnostic Explanations), SHAP (Shapley Additive Explanations), частичные зависимости (PDP), карты внимания (attention maps) для трансформеров. |
| 3. | Выбор метода в зависимости от правовой задачи (опровержение дискриминации, обоснование отказа, выявление ошибки) |

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1. В чём заключается ключевое различие между интерпретируемостью (interpretability) и объяснимостью (explainability)модели ИИ? Выберите один правильный ответ и кратко обоснуйте.

1. Интерпретируемость – это способность человека понять внутреннее устройство модели целиком, а объяснимость – способность модели предоставить понятные пояснения к отдельным решениям.
2. Интерпретируемость и объяснимость – это полные синонимы.
3. Интерпретируемость относится только к линейным моделям, а объяснимость – к нейросетям.
4. Интерпретируемость – это точность модели, а объяснимость – её робастность.

Задание 2. Для рассмотрения жалобы гражданина, которому банк отказал в кредите на основании решения нейросетевой модели, суд потребовал объяснить, какие признаки больше всего повлияли на отказ. Выберите все подходящие методы XAI из списка и обоснуйте их применимость в данной правовой задаче.

1. LIME (Local Interpretable Model-agnostic Explanations).
2. SHAP (Shapley Additive Explanations).
3. Глобальные частичные зависимости (PDP) без локальной детализации.
4. Карты внимания (attention maps) для трансформеров.

3. Задания закрытого типа.

Тестовые задания.

Вопрос 1. Какая из перечисленных проблем характерна для сложных моделей «чёрного ящика» (глубокие нейросети, градиентный бустинг) с точки зрения правового регулирования?

1. Их невозможно обучить на больших данных.
2. Они всегда менее точны, чем линейные модели.

3. Отсутствие прозрачности затрудняет доказывание отсутствия дискриминации и обоснование решений, влияющих на права граждан.
4. Они не требуют валидации.

Вопрос 2. Какие из перечисленных утверждений о методах XAI являются верными? Выберите все правильные варианты.

1. SHAP основан на теории кооперативных игр (значения Шепли).
2. LIME работает только с линейными моделями.
3. LIME является model-agnostic (не зависит от типа модели).
4. Карты внимания (attention maps) можно применять к любым нейронным сетям, включая свёрточные.

Вопрос 3. В задаче выявления скрытой дискриминации при найме персонала (решения принимает ансамбль деревьев решений) наиболее целесообразно использовать:

1. Глобальную интерпретацию через веса признаков (если бы модель была линейной).
2. Локальный метод SHAP для анализа отдельных отказов кандидатам.
3. Только метрику accuracy.
4. Отказ от объяснения, ссылаясь на коммерческую тайну.

Вопрос 4. При выборе метода XAI для обоснования врачебного решения, принятого нейросетью при диагностике заболевания, главным требованием должно быть:

1. Максимальная скорость работы метода.
2. Способность предоставить локальное, понятное врачу и пациенту объяснение (например, какие области на снимке повлияли на диагноз).
3. Использование только глобальных методов, чтобы оценить модель в целом.
4. Отсутствие необходимости в интерпретации, так как врач доверяет ИИ.

## Тема 5. Правовое значение XAI: GDPR, 152-ФЗ, отраслевые требования. ПКс-АИ-1, ПКс-АИ-3.

1. Задания открытого типа.

| № п.п. | Вопрос |
|---|---|
| 1. | Право на объяснение (right to explanation) в GDPR (ст. 13–15, 22, Преамбула 71). Разъяснения EDPB (Guidelines on Automated individual decision-making). |
| 2. | Соответствие XAI требованиям 152-ФЗ: право субъекта знать об обработке, недопустимость решений на основе исключительно автоматизированной обработки (ст. 16 152-ФЗ). |
| 3. | Отраслевые требования: банковское регулирование (Положение Банка России № 719-П об ИИ в скоринге), медицинские стандарты (Clinical Decision Support Systems). |

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1. В соответствии с GDPR (ст. 22, Преамбула 71) и разъяснениями EDPB, субъект данных имеет право не подвергаться решению, основанному исключительно на автоматизированной обработке, включая профилирование, если это решение порождает юридические последствия или аналогичным образом существенно затрагивает его права. Какое из следующих действий оператора ИИ-системы наиболее полно соответствует требованию предоставить «право на объяснение» (right to explanation)? Выберите один правильный ответ и кратко обоснуйте.

1. Предоставить субъекту общее описание архитектуры нейросети (количество слоёв, функция активации).
2. По запросу субъекта сформировать локальное объяснение конкретного решения с помощью метода SHAP или LIME, понятное человеку без специальных знаний.
3. Сообщить субъекту, что решение принято алгоритмом, и отказаться от дальнейших разъяснений ввиду коммерческой тайны.
4. Предоставить субъекту весь исходный код модели для самостоятельного анализа.

Задание 2. Какие из перечисленных нормативных актов и разъяснений прямо или косвенно вводят требования к объяснимости (XAI) решений, принимаемых ИИ-системами? Выберите все правильные варианты и обоснуйте.

1. GDPR, статьи 13–15, 22, Преамбула 71.
2. Федеральный закон № 152-ФЗ «О персональных данных», статья 16.
3. Положение Банка России № 719-П о требованиях к системам кредитного скоринга с использованием ИИ.
4. Клинические рекомендации (стандарты) по применению систем поддержки принятия врачебных решений (CDSS).

3. Задания закрытого типа.

Тестовые задания.

Вопрос 1. Согласно ст. 22 GDPR, решение, основанное исключительно на автоматизированной обработке (включая профилирование), имеющее юридические последствия или аналогичным образом существенно затрагивающее субъекта:

1. Всегда запрещено в любых случаях.
2. Допустимо, если необходимо для заключения или исполнения договора, или разрешено законодательством, и при этом оператор предоставляет объяснение и возможность оспорить решение.
3. Допустимо без каких-либо ограничений.
4. Разрешено только государственным органам.

Вопрос 2. Какие из перечисленных требований к объяснимости ИИ-решений прямо вытекают из статьи 16 Федерального закона № 152-ФЗ? Выберите все правильные варианты.

1. Решение, порождающее юридические последствия для субъекта, не может быть принято исключительно на основе автоматизированной обработки его персональных данных без согласия или без участия человека.
2. Оператор обязан разъяснить субъекту порядок принятия решения на доступном языке.
3. Оператор обязан предоставить субъекту возможность заявить возражение против такого решения.

4. Оператор обязан опубликовать полный исходный код модели в открытом доступе.

Вопрос 3. Положение Банка России № 719-П (о системах кредитного скоринга с использованием ИИ) требует, чтобы модель ИИ обеспечивала:

1. Максимальную сложность архитектуры (не менее 100 слоёв).
2. Интерпретируемость факторов, влияющих на отказ в кредите, и возможность предоставления заёмщику понятного объяснения.
3. Полное отсутствие ошибок в предсказаниях.
4. Использование только методов глубокого обучения.

Вопрос 4. В медицинской системе поддержки принятия врачебных решений (CDSS) на основе ИИ правовое значение объяснимости (XAI) заключается прежде всего в том, чтобы:

1. Заменить врача полностью автоматической диагностикой.
2. Обеспечить врачу и пациенту понимание логики рекомендаций для обоснования клинического решения и распределения ответственности.
3. Скрыть возможные ошибки модели.
4. Ускорить обработку медицинских данных.

**Тема 6. XAI в судебных спорах и экспертизе: доказывание, ответственность за вред от ИИ. ПКи-АИ-3.**

1. Задания открытого типа.

| № п.п. | Вопрос |
|---|---|
| 1. | Структура исков о вреде, причинённом решением ИИ: необходимость установить причинно-следственную связь между дефектом модели и вредом. XAI как источник доказательств: объяснения решений (SHAP, LIME) могут подтвердить дискриминацию или ошибку. |
| 2. | Назначение компьютерной экспертизы с вопросами об интерпретируемости. |
| 3. | Правовые позиции: достаточность объяснений для освобождения от ответственности (due diligence) или, наоборот, их отсутствие как признак вины. |
| 4. | Примеры судебных дел (США: State v. Loomis (COMPAS), ЕС: Schufa case (C-634/21)). Роль юриста в формулировании вопросов эксперту и оспаривании XAI-выводов. |

2. Задания комбинированного типа:

Тестовые задания с обоснованием выбора.

Задание 1. В суде рассматривается иск гражданина к банку, использующему ИИ-модель для скоринга. Гражданин утверждает, что получил отказ в кредите по дискриминационным основаниям, но банк отказывается раскрывать логику принятия решения, ссылаясь на коммерческую тайну. Какую роль в этом споре могут сыграть методы XAI (например, SHAP или LIME) с точки зрения доказывания?
Выберите один правильный ответ и кратко обоснуйте.

1. XAI не имеет юридической силы, поскольку его методы не признаны доказательствами в суде.
2. Банк может ссылаться на XAI, чтобы доказать, что его модель является «чёрным ящиком», и тем самым освободить себя от ответственности.

3. Локальные объяснения XAI (например, SHAP) могут быть использованы истцом или назначенным судом экспертом, чтобы установить, какие именно признаки модели повлияли на отказ, и выявить возможную дискриминацию.
4. XAI полностью заменяет собой необходимость в компьютерной экспертизе при рассмотрении таких споров.

Задание 2 Какие из перечисленных позиций относительно роли XAI в доказывании добросовестности (due diligence) оператора ИИ-системы и распределения ответственности за причинённый вред являются верными? Выберите все правильные варианты и обоснуйте.

1. Наличие документально подтверждённого применения методов XAI для аудита модели до её внедрения является сильным, но не абсолютным аргументом в пользу того, что оператор принял «достаточные меры» для обеспечения надёжности и снижения рисков.
2. Использование XAI полностью освобождает оператора от гражданско-правовой ответственности за вред, причинённый ИИ-системой.
3. При судебном рассмотрении спора о вреде от ИИ отсутствие в модели механизмов XAI может быть истолковано как признак вины оператора (непринятие разумных мер по обеспечению прозрачности и контролю за решениями алгоритма).
4. Валидация ИИ-модели и её документирование не имеют значения для суда, если модель не оснащена XAI.

3. Задания закрытого типа.

Тестовые задания.

Вопрос 1. В каком из перечисленных дел суд рассмотрел вопрос о допустимости использования закрытого алгоритма COMPAS (риск рецидива) в уголовном судопроизводстве и пришёл к выводу, что его использование не нарушает принцип надлежащей правовой процедуры (due process), однако признал необходимость определённых ограничений? Выберите один правильный ответ.

1. Дело Schufa (C-634/21).
2. Дело State v. Loomis (Wisconsin Supreme Court, 2016).
3. Дело ProPublica против COMPAS.
4. Дело «Google против CNIL».

Вопрос 2. Какие из перечисленных процессуальных действий и задач относятся к роли юриста при работе с XAI-выводами в судебном споре о вреде, причинённом ИИ-системой? Выберите все правильные варианты.

1. Понимание базовых принципов работы методов XAI (например, SHAP, LIME) для корректной интерпретации предоставленных объяснений.
2. Формулирование вопросов судебному эксперту, которые позволят получить объяснения, юридически значимые для установления причинно-следственной связи.
3. Самостоятельное проведение технического аудита модели и генерация XAI-объяснений без привлечения специалистов.
4. Критическая оценка XAI-выводов, представленных противоположной стороной, на предмет их полноты, достоверности и релевантности для дела.

Вопрос 3. В решении Суда Европейского союза по делу Schufa (C-634/21) установлено, что:

1. Передача кредитного скоринга третьим лицам никогда не может считаться автоматизированным решением по смыслу статьи 22 GDPR.
2. Формирование и передача кредитного рейтинга (скоринга) могут сами по себе считаться автоматизированным решением, если этот рейтинг является определяющим для заключения договора третьей стороной.
3. Кредитные организации не обязаны информировать граждан о логике принятия решений при использовании ИИ-скоринга.
4. Дело не имеет отношения к GDPR, так как рассматривало исключительно вопросы национального немецкого права.

Вопрос 4. Какие элементы должны быть установлены истцом при предъявлении иска о вреде, причинённом решением ИИ-системы (структура иска)? Выберите один правильный ответ.

1. Наличие самого факта причинения вреда; этого достаточно, поскольку ответственность за действия ИИ всегда безвиновная.
2. Причинно-следственная связь между конкретным дефектом модели (ошибка, дискриминационный паттерн) и наступившим вредом, а также доказательства того, что ответчик (разработчик/оператор) не принял разумных мер для предотвращения вреда.
3. Только факт использования ИИ-системы, без необходимости доказывать какие-либо дополнительные обстоятельства.
4. Наличие письменного согласия истца на обработку его данных ИИ-системой.

6.3. Критерии и шкала оценивания на основе БРС.

*Критерии и балльная шкала определяются преподавателем*

| 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* | 30-39 |

| | |
|---|---|
| *студент 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.* | |
| *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. студент не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.* | 0-19 |

## 7. Методические материалы по освоению дисциплины (модуля)

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. Понятие тестирования и валидации ИИ-моделей. Правовые основания для обязательного тестирования (GDPR, 152-ФЗ, проекты регулирования).
2. Основные метрики качества моделей классификации и регрессии. Правовое значение выбора метрики для доказывания добросовестности оператора.
3. Робастность и безопасностное тестирование: методы и правовые последствия недостаточной робастности.
4. Офлайн- и онлайн-валидация (A/B тестирование). Примеры правовых споров, вызванных ошибками валидации.
5. Стандарты валидации ИИ: ISO/IEC 25000, NIST AI RMF, ГОСТ Р. Применение в юридической практике.
6. Интерпретируемость vs объяснимость. Проблема «чёрного ящика» в глубоком обучении.
7. Методы XAI: LIME (принцип, преимущества, ограничения для судебного использования).
8. Методы XAI: SHAP (математическая основа, интерпретация значений, примеры).
9. Право на объяснение (right to explanation) в GDPR: содержание, пределы, связь с XAI.
10. Требования 152-ФЗ к автоматизированной обработке персональных данных и роль XAI.
11. Отраслевые требования к объяснимости ИИ (банковский скоринг, медицинские системы).
12. XAI как доказательство в судебных спорах о дискриминации или ошибочных решениях ИИ.

13. Назначение компьютерной экспертизы интерпретируемости: формулировка вопросов, выбор методов.
14. Ответственность разработчика и оператора ИИ за непредоставление объяснений решения (ст. 14.8 КоАП РФ? – обсуждение).
15. Конфликт между объяснимостью и коммерческой тайной алгоритма: возможные компромиссы (независимый аудит, предоставление объяснений регулятору).

## 8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет

### 8.1. Основная литература

1. 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.

2. Ищейнов, В. Я. Основные положения информационной безопасности: учебное пособие / В.Я. Ищейнов, М.В. Мецатунян. — Москва : ФОРУМ : ИНФРА-М, 2027. — 208 с. — (Среднее профессиональное образование). - ISBN 978-5-16-022356-8. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2262621 (дата обращения: 10.06.2026). – Режим доступа: по подписке.

3. Клименко, И. С. Информационная безопасность и защита информации: модели и методы управления : монография / И.С. Клименко. — Москва : ИНФРА-М, 2026. — 180 с. — (Научная мысль). — DOI 10.12737/monography_5d412ff13c0b88.75804464. - ISBN 978-5-16-015149-6. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2238628 (дата обращения: 10.06.2026). – Режим доступа: по подписке.

### 8.2 Дополнительная литература

1. OECD (2020). "Privacy-Enhancing Technologies: A Policy Primer". OECD Digital Economy Papers. (Режим доступа: www.oecd.org).
2. EDPB (2022). "Guidelines 05/2022 on the use of synthetic data". (Режим доступа: edpb.europa.eu).
3. Документы Роскомнадзора по разъяснению анонимизации и псевдонимизации (раздел «Разъяснения» на портале pd.rkn.gov.ru).

### *8.3. Нормативные правовые документы и иная правовая информация*

В рамках данной дисциплины не предусмотрено изучение нормативных правовых актов в качестве обязательной литературы. При изучении модуля 5 (регулирование и этика) рекомендуется обращаться к актуальным документам:

- Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных»
- Федеральный закон от 28.06.2021 № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве»
- Этические принципы и рекомендации ЮНЕСКО по искусственному интеллекту (2021)

Указанные документы могут быть найдены в справочно-правовых системах «КонсультантПлюс» или «Гарант» (доступ через университетскую сеть).

### *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 Роскомнадзора (rkn.gov.ru).
5. Официальный сайт Министерства цифрового развития, связи и массовых коммуникаций РФ (digital.gov.ru).
6. ГАС «Правосудие» (sudrf.ru) и Картотека арбитражных дел (kad.arbitr.ru). (Ресурсы для мониторинга судебной практики по спорам, связанным с обработкой данных и защитой информации, включая новые дела по ч. 10 ст. 13.11 КоАП РФ о неуведомлении Роскомнадзора).
7. Портал Европейского Союза EUR-Lex (eur-lex.europa.eu). (Официальный источник для изучения GDPR и иных актов европейского законодательства в сфере данных).
8. Научная электронная библиотека eLIBRARY.RU (elibrary.ru). (Доступ к полным текстам статей из российских научных журналов по информационному праву и правовым режимам данных).
9. Профессиональные ресурсы по комплаенсу в сфере данных (comply.ru и аналогичные). (Площадки для отслеживания актуальных позиций регуляторов, разъяснений и практики применения законодательства, включая обзоры мероприятий Роскомнадзора).

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

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 используется 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, средстваработы 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. https://znanium.ru - 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 (EastView). 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-правоваясистема «Консультант».
9. http://www.garant.ru/Справочно-правоваясистема «Гарант».