Observatoire ARGA

МЕЖДУНАРОДНЫЙ АНАЛИТИЧЕСКИЙ ДОКЛАД

Исследование №16

Независимый аудит искусственного интеллекта

Стандарты и доказательственная методология
Гарантии независимости и качество проверок

Организация: Observatoire ARGA

Русская версия

21 rue de l’Aviation, 64600 Anglet, France

info@argaobservatory.org | www.argaobservatory.org

20 октября 2026 года

Содержание

1 Предмет исследования и пределы аудиторского вывода — 3

Научные основания и исследовательский дизайн — 5

2 Нормативная и институциональная карта — 7

3 Европейские процедуры и требования к независимости — 10

4 Сопоставление подходов США и Великобритании — 12

5 Стандарты и выбор критериев — 15

6 Независимость и конфликт интересов — 17

7 Компетентность команды и контроль качества — 20

8 Задание на аудит и условия доступа — 23

9 Объект аудита и цепочка поставщиков — 26

10 Происхождение и сохранность доказательств — 28

11 Качество и происхождение данных — 31

12 Выборка и проектирование испытаний — 33

13 Показатели качества и статистическая неопределённость — 36

Парное сравнение конфигураций на условном наборе досье — 38

14 Предвзятость и последствия для прав человека — 40

15 Объяснимость и человеческий контроль — 42

16 Аудит генеративного юридического помощника — 45

17 Устойчивость и информационная безопасность — 48

18 Персональные данные и анонимизация — 50

19 Воспроизводимость результатов — 53

20 Находки и итоговое заключение — 56

21 Мониторинг и пересмотр результата — 59

22 Применение к профильным задачам ARGA — 62

23 Пробелы регулирования и международные консультации — 64

24 Практическая программа ARGA и итоговые выводы — 67

Приложение 1 Задание на ограниченный аудит — 70

Приложение 2 Реестр доказательств — 71

Приложение 3 Оценка конфликта интересов — 73

Приложение 4 План выборки и редкие ошибки — 75

Приложение 5 Проверка показателей классификатора — 77

Приложение 6 Различия между группами — 78

Приложение 7 Протокол воспроизводимого испытания — 80

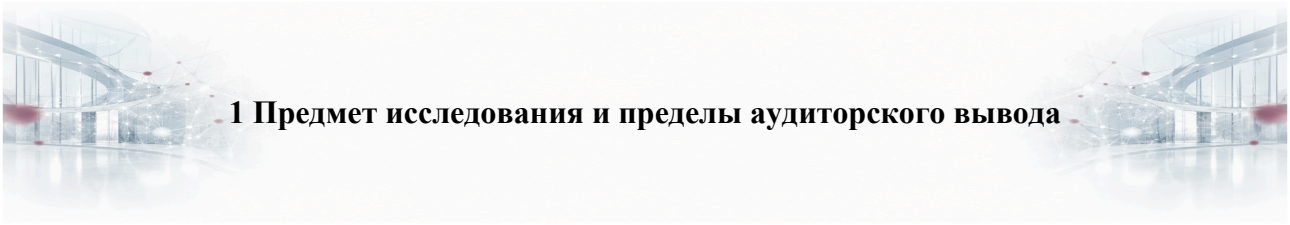
Приложение 8 Разбор ответов юридического помощника — 81

Приложение 9 Образец ограниченного заключения — 83

Приложение 10 Пакет предложений для ARGA — 85

Практический указатель рекомендаций — 86

Источники и указатель оснований — 88



1 Предмет исследования и пределы аудиторского вывода

Какой вопрос должен разрешить аудит

Независимый аудит искусственного интеллекта имеет практический смысл, когда позволяет проверить определённое утверждение о системе и назвать основания вывода. Предметом проверки может быть соблюдение требования, качество процесса, результат испытания или эффективность защитной меры. Формула «искусственный интеллект прошёл аудит» без указания объекта, версии, критериев и ограничений не сообщает, что именно установлено.

Исследовательский вопрос формулируется операционально: при каких условиях независимый исследователь может воспроизвести вывод о конкретном юридически значимом свойстве ИИ и определить допустимую область его использования? Проверяются четыре связанные предпосылки: свойство корректно определено; наблюдения позволяют его измерить; вывод устойчив к разумным изменениям оценки; процедура решения защищена от конфликта интересов. Сопоставление источников обосновывает предлагаемую методику; её эмпирическая результативность подлежит отдельной проверке по приведённому протоколу.

Для ARGA предмет особенно существенен в работе с юридическими досье, совпадениями по именам, финансовой аналитикой и подготовкой текстов. Ошибка может изменить оценку человека или обстоятельств дела. Поэтому предлагаемый подход связывает техническую проверку с последствиями для затронутого лица: доступностью исправления, понятностью причин и возможностью восстановить ход принятия решения.

В докладе слово «аудит» используется как общее обозначение систематической проверки по установленным критериям. Это рабочее определение не присваивает исполнителю статус нотифицированного органа, аккредитованной лаборатории или органа сертификации. Юридическое значение конкретного заключения определяется применимой процедурой. Разграничение таких процедур составляет исходное условие всей предлагаемой методологии.

Элемент задания	Что должно быть определено
Объект	Система, версия и конкретное применение
Утверждение	Проверяемое свойство или исполнение обязанности
Критерий	Норма, договорное условие либо утверждённая методика
Пользователь результата	Кто и для какого решения получает заключение

Что считается доказательством

Доказательство оценивают относительно утверждения и возможного альтернативного объяснения. Если ответ содержит верную цитату, это подтверждает существование

соответствующего фрагмента, но ещё не подтверждает полноту ответа или применимость нормы. Если журнал показывает обращение к документу, остаётся проверить, повлиял ли он на результат. Для каждого существенного утверждения поэтому фиксируются первичный объект, выполненная операция, полученное наблюдение и предел интерпретации. Эта запись позволяет отличить прямое подтверждение от косвенного признака.

Предлагается различать доказательства о составе системы, её поведении и последствиях применения. Паспорт версии отвечает на вопрос об объекте; воспроизводимый тест — о наблюдаемом поведении; материал реального процесса — о последствиях. Между уровнями требуется отдельное обоснование. Успешное испытание искусственного досье не доказывает уменьшения числа процессуальных ошибок в практике, а установленный вред не позволяет без исследования приписать причину конкретному компоненту.

Проверка строится от утверждения к процедуре. Если заявлено отсутствие передачи документов внешним сервисам, недостаточно прочитать договор. Понадобится исследовать маршруты передачи, настройки, разрешения и наблюдаемое поведение в разрешённой среде. Если заявлена точность распознавания лиц, правовая экспертиза условий применения не заменит испытание на подходящих данных.

Отдельное значение имеет отрицательный результат. Отсутствие найденного инцидента, отсутствие доступа к сведениям и подтверждённое отсутствие определённого действия — разные состояния. В заключении они отражаются раздельно. Иначе недостаток доказательств превращается в необоснованное подтверждение благополучия, а заказчик получает вывод, который проведённая работа фактически не поддерживает.

Материал	Поддерживаемый вывод	Существенное ограничение
Договор поставки	Согласованное обязательство	Не доказывает исполнение
Журнал событий	Зафиксированные операции	Нужна проверка полноты
Контрольный запуск	Наблюдаемое поведение	Ограничен условиями запуска
Интервью сотрудника	Понимание и описание процесса	Требует сопоставления с практикой

Метод исследования и статус предложений

Правовая часть опирается на официальные публикации ЕС, органов США и Великобритании. Стандарты ISO рассматриваются по открытым описаниям предмета и редакции. Полные платные тексты не используются для реконструкции обязательных пунктов. Поэтому предложенные формы проверки нельзя выдавать за полный перечень требований какого-либо стандарта или готовую программу сертификационного аудита.

Нормативная карта отражает состояние источников на 20 октября 2026 года. Применимость нормы устанавливается отдельно от её опубликования: необходимы территория, роль участника, категория системы, дата и предусмотренные исключения. Консультационные проекты выделяются как проекты. В частности, предлагаемые руководства EDPB по

анонимизации и сбору данных для генеративного ИИ не представлены как окончательно принятые правила. [27], [28], [30], [31]

Методические решения доклада являются предложениями для организации проверок. Они включают реестр доказательств, протокол испытаний, оценку независимости и форму заключения. В каждом практическом примере используются условные обстоятельства и учебные числа. Доклад не содержит результатов испытаний реального поставщика и не подтверждает наличие у ARGA статуса официального оценщика.

При применении методологии сначала выбирается допустимый объём работы. Для документальной проверки достаточно одного состава компетенций; техническая оценка модели требует другого. Если существенный вопрос выходит за возможности команды, его передают профильному специалисту либо прямо ограничивают вывод. Такое ограничение защищает качество исследования и позволяет заказчику понять, какую дополнительную работу ещё необходимо заказать.

Практический результат должен отвечать на три самостоятельных вопроса: что установлено, насколько надёжно это установлено и какое действие следует из результата. Ни положительная оценка организационного процесса, ни успешное испытание отдельной функции не освобождают пользователя от определения законности конкретного применения системы.

Научные основания и исследовательский дизайн

Научные направления используются как основания отдельных исследовательских операций. Теория измерений связывает понятие юридической ошибки с признаками ответа; статистика определяет единицу сравнения и неопределённость; программная инженерия обеспечивает идентификацию конфигурации и повторный анализ; исследования подотчётности связывают находку с ответственным действием. Предлагаемая ниже процедура адаптирует эти основания к юридическому помощнику и подлежит пилотной проверке. [M1–M3]

Научное направление	Операция в исследовании	Проверяемый результат
Теория измерений	Определение существенной ошибки и правил разметки	Классификатор с положительными и пограничными примерами
Прикладная статистика	Парный дизайн на уровне досье и анализ неопределённости	Исходные количества и воспроизводимый расчёт различия
Программная инженерия	Фиксация модели, поиска, инструкции и прав инструментов	Паспорт конфигурации и журнал воспроизведения
Алгоритмическая подотчётность	Связь находки с исправлением и ответственным решением	Мотивированное решение по каждому существенному дефекту
Правовая методология	Проверка применимости нормы и последствий ошибки	Раздельные фактический, технический и правовой выводы

Исследовательский вопрос и единица анализа

Основной вопрос пилота: меняет ли исследуемая конфигурация долю досье с хотя бы одной существенной ошибкой при сохранении полноты обязательных сведений? Объектом

является вся конфигурация помощника, включая извлечение документов и ограничения действий. Единицей первичного анализа служит досье. Детальные метки отдельных утверждений объясняют механизм дефекта, но не увеличивают автоматически число независимых наблюдений.

Определение ошибки должно быть связано с предполагаемым применением. Для внутренней хронологии существенна подмена даты или источника; для подготовки правовой позиции — утрата исключения, неверная юрисдикция или изменение статуса факта. Предлагается заранее составить карту «вид ошибки — возможное последствие — способ выявления». Научная валидность здесь означает обоснованность измерения конкретного свойства, а не присвоение ответу общего балла качества. [M1]

Гипотеза и план сравнительной проверки

Для основного сравнения задаются исходная конфигурация A, изменённая B и одинаковые досье. Нулевая гипотеза относится к отсутствию направленного различия парных исходов; исследовательская гипотеза — к снижению доли существенных ошибок. Практически значимое снижение и недопустимые новые дефекты определяются до запуска. Даже статистически различимые результаты могут не оправдывать внедрение, если изменение слишком мало или ухудшает критическую категорию.

Порядок предъявления задач и ответов перемешивается; экспертам по возможности скрывается принадлежность ответа конфигурации. Сохраняются все предусмотренные прогоны, отказы и технические сбои. Для нескольких запусков на одно досье заранее выбирается правило агрегирования либо модель вложенных наблюдений. Отдельно сообщаются качество на фиксированном наборе и основания для переноса на потенциальные новые задачи. [M3]

Проверка измерительного инструмента

До основной оценки два эксперта независимо размечают тренировочный материал. Обсуждение используется для уточнения инструкции, после чего её пригодность проверяется на новом поднаборе. Пригодность разметки оценивают по конкретным категориям ошибок и причинам расхождений. Достижение согласия после обсуждения не заменяет измерения исходной воспроизводимости; спорный случай сохраняется как наблюдаемое ограничение эталона.

Проверяются конкурирующие объяснения улучшения: более короткие ответы, увеличение числа отказов, исключение сложных досье или изменение поискового корпуса. Поэтому рядом с основным исходом приводятся полнота обязательных фактов, доля корректных отказов и распределение типов задач. Если одновременно изменены несколько компонентов, результат относится к их совокупности. Для вывода о вкладе одного компонента требуется отдельное сравнение.

Условия принятия и пересмотра вывода

Ограниченный допуск возможен после проверки данных, расчёта и существенных новых дефектов. В заключении указывают исследованную область, оставшиеся неопределённости и изменения, требующие новой оценки. К ним относятся смена модели, поискового корпуса, языка, полномочий инструмента и назначения ответа. Протокол допускает решение «данных

недостаточно»: при этом перечисляются конкретные сведения или испытания, необходимые для следующего решения.

Результат данного доклада — проект методики и проверяемого дизайна. Условные расчёты ниже демонстрируют применение правил, но не заменяют эмпирического исследования. Научная проверка самой методики должна оценить воспроизводимость экспертных выводов и её способность обнаруживать юридически значимые дефекты при ограниченных ресурсах. Такое разделение позволяет сформулировать проверяемое исследовательское задание для последующего пилота.

Практический разбор 16.01

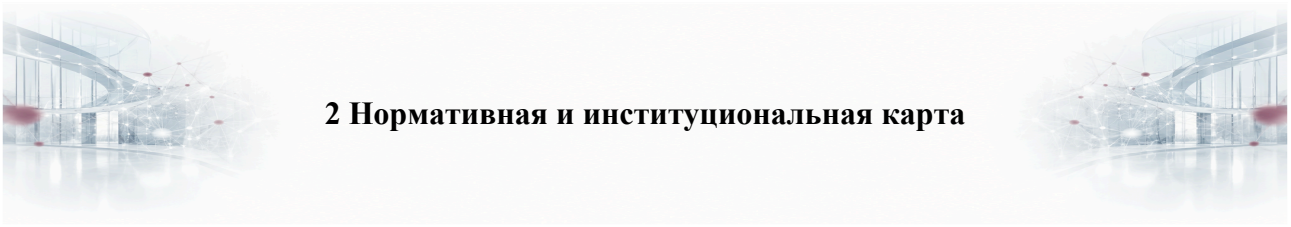
Рекомендация: Фиксировать узкое проверяемое утверждение

Условный пример: Заказчик просит подтвердить надёжность помощника, который составляет хронологию дела.

Разбор: Правильная дата в одном ответе не подтверждает полноту хронологии. Проверяются отдельно дата, источник и пропущенные события.

Практические действия: Руководитель аудита утверждает перечень обязательных событий и сравнивает ответы с исходными документами.

Критерий результата: Заключение относится к сохранению событий и источников в исследованном наборе.



2 Нормативная и институциональная карта

Разные основания для разных проверок

Обязанность провести проверку может возникать из закона, надзорного требования в установленной сфере, договора или добровольно принятой программы. Эти основания различаются по адресатам, содержанию и последствиям. Для начала работы требуется карточка применимости: кто обязан действовать, какой объект охватывается, когда наступает обязанность и какой документ должен быть получен.

В ЕС Регламент об искусственном интеллекте предусматривает несколько процедур оценки соответствия. Участие внешнего нотифицированного органа зависит от категории системы и предусмотренного маршрута. Само отнесение к высокому риску не позволяет утверждать, что любой поставщик обязан заказать универсальный независимый аудит у произвольной коммерческой организации. [3]

Отдельные режимы действуют за пределами этого регламента. Европейский режим цифровых услуг предусматривает независимые ежегодные аудиты особо крупных платформ и поисковых систем. В Нью-Йорке установлен специальный аудит определённых инструментов автоматизированного принятия кадровых решений. Ни одна из этих конструкций не распространяется на любой ИИ только потому, что в его работе применяется алгоритм. [6–9]

Основание	Что устанавливается перед проверкой
Обязательная процедура	Адресат, компетентный орган и допустимый исполнитель
Договорная проверка	Согласованные критерии, доступ и последствия результата
Добровольная оценка	Цель, метод и пределы заявляемого подтверждения
Внутренняя проверка	Подотчётность исполнителя и порядок устранения недостатков

Для ARGA эта карта позволяет предложить полезную услугу в пределах имеющейся компетенции: исследование документов, юридическую квалификацию и проверку согласованности утверждений. Если заказчику нужен документ установленной законом формы, сначала выясняется, кто вправе его выдавать. Подмена такого документа аналитическим заключением создаёт риск для обеих сторон.

Дата нормы и дата обязанности

Календарь проверки должен учитывать актуальные переходные положения. Статья 113 Регламента ЕС об ИИ сохраняет общую дату применения 2 августа 2026 года и специальные даты. Для разделов 1–3 главы III, кроме статьи 6(5), указаны 2 декабря 2027 года для систем по статье 6(2) и приложению III и 2 августа 2028 года для систем по статье 6(1) и приложению I. [1]

Из этого нельзя выводить, что до названных дат любая обработка данных или любое решение с ИИ допустимы. Параллельно могут действовать правила защиты данных, трудовое, потребительское и отраслевое законодательство. Проверка готовности к будущему требованию и оценка соблюдения уже применимой нормы имеют разный предмет и должны получать разные формулировки результата.

Полезно фиксировать дату не только у правового акта, но и у технического стандарта. ISO 19011 опубликован в новой редакции 2026 года. Если договор продолжает ссылаться на редакцию 2018 года, аудитор должен выяснить, намеренно ли стороны выбрали её и требуется ли изменение задания. Замена версии стандарта не должна происходить незаметно для адресата заключения. [16]

Запись в календаре	Значение для аудита
Дата действия нормы	Определяет юридическую оценку поведения
Дата начала обязательной процедуры	Определяет требуемый маршрут подтверждения
Редакция критерия	Позволяет восстановить содержание проверки

Запись в календаре	Значение для аудита
Дата испытания	Ограничивает техническую актуальность результата
Изменение конфигурации	Может потребовать пересмотра ранее сделанного вывода

Рабочий реестр должен сохранять прошлые записи. Иначе поздняя корректировка срока уничтожит объяснение того, почему прежнее решение считалось обоснованным. Для каждого изменения достаточно указать официальный источник, дату пересмотра и затронутые пункты задания.

Кто принимает решения после аудита

Аудитор устанавливает факты и оценивает их по выбранным критериям. Руководитель проверяемой организации определяет меры в пределах своих полномочий. Регулятор, суд или иной уполномоченный орган сохраняет собственную компетенцию. Передача отчёта внешнему адресату не превращает мнение аудитора в обязательное для этого адресата решение.

В предлагаемом порядке распределяются четыре роли: владелец процесса, исполнитель проверки, независимый рецензент и лицо, принимающее решение. Один сотрудник может совмещать некоторые функции в небольшой организации, но не должен подтверждать независимость оценки собственного спорного решения. Если разделение невозможно, ограничение раскрывается, а существенные выводы передаются внешнему рецензенту.

Особенно опасно размывание ответственности после привлечения поставщика. Заказчик может считать, что испытание полностью выполнила техническая команда, а техническая команда — что правовую допустимость подтвердил юрист. Поэтому каждое существенное утверждение связывается с конкретным исполнителем и конкретной процедурой. Общая подпись руководителя проекта не заменяет такую связь.

Для общественно значимых систем следует предусмотреть канал получения возражений от затронутых лиц. Обращение пользователя не всегда доказывает системный дефект, но может показать ситуацию, не попавшую в испытательную выборку. Аудиторская программа должна позволять исследовать такое свидетельство без автоматического раскрытия персонального досье всему составу команды.

Итоговый документ указывает, кому адресован вывод и какие решения он поддерживает. Например, подтверждение полноты журналирования может быть достаточным для следующего этапа испытаний, но недостаточным для запуска системы, принимающей значимые решения о человеке. Последовательность решений должна быть ясна до начала проверки, а не появляться после выдачи отчёта.

Практический разбор 16.02

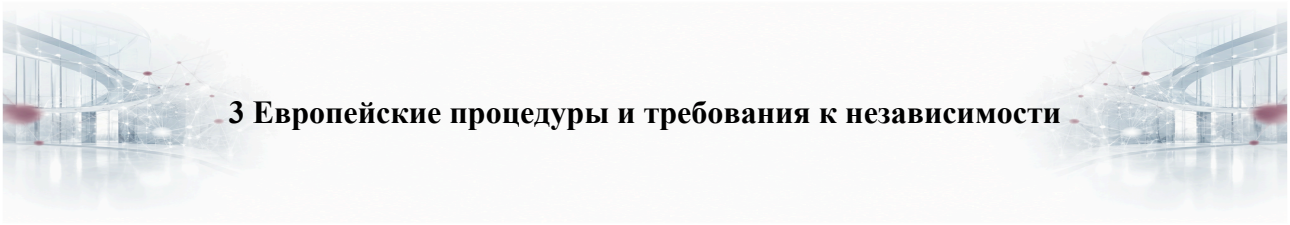
Рекомендация: Составлять карту применимых требований

Условный пример: Поставщик предлагает один сертификат для банковского скрининга и кадрового отбора.

Разбор: Одинаковая модель не означает одинаковые обязанности. Назначение, субъект и территория определяют отдельные вопросы проверки.

Практические действия: Юрист описывает оба применения и для каждого указывает основание, адресата обязанности и требуемое свидетельство.

Критерий результата: Каждое требование связано с конкретным применением и источником.



3 Европейские процедуры и требования к независимости

Нотифицированный орган и обычный внешний аудитор

Статья 31 Регламента ЕС об ИИ устанавливает специальные требования к нотифицированным органам. Они должны быть независимы от проверяемого поставщика и других экономически заинтересованных операторов. Предусмотрены организационные меры беспристрастности; участие в разработке и определённые консультационные услуги могут конфликтовать с оценочной функцией. Это требования к конкретному институциональному статусу. [2]

Внешний консультант, выполняющий договорную проверку, не становится нотифицированным органом вследствие регистрации юридического лица, наличия опытных сотрудников или использования международного стандарта. Для соответствующего статуса необходима предусмотренная законом процедура. В договоре и публичном описании услуги должны употребляться названия, которые точно отражают полномочия исполнителя.

Из специальной нормы можно извлечь методический ориентир для добровольной проверки, но нельзя переносить её юридическое действие на всех участников рынка. Предлагаемые в докладе ограничения совмещения консультирования и аудита применяются как внутренние гарантии качества, если иной режим не установил обязательные требования. Такое различение позволяет усиливать независимость без выдумывания всеобщей законодательной обязанности.

Проверяемый вопрос	Достаточное подтверждение
Заявлен официальный статус	Запись и область полномочий в применимой системе
Заявлена техническая компетентность	Состав специалистов и подтверждённый опыт
Заявлена независимость	Проверка отношений и действующие меры защиты
Заявлена сертификация	Схема, объект, орган и область сертификата

Для ARGA правильная исходная позиция — описывать фактически выполняемую исследовательскую или экспертную работу. Возможное развитие более формальной функции

требует отдельного анализа компетенции, ресурсов, ответственности и допустимого статуса. Настоящий доклад не объявляет такую функцию уже созданной.

Оценка соответствия и техническое досье

Статья 43 различает процедуры внутреннего контроля и процедуры с участием нотифицированного органа. Для предусмотренных ею групп систем значение имеют секторное законодательство, применение гармонизированных стандартов и общих спецификаций. Приложение VII соединяет оценку системы управления качеством и технической документации с установленным наблюдением. [3–4]

Для добровольного аудита полезен сам принцип разделения объекта на процесс и конкретную техническую реализацию. У организации может быть подробно описан порядок выпуска, но отсутствовать доказательство его выполнения для проверяемой версии. Обратная ситуация также возможна: единичный продукт хорошо испытан, а управление изменениями остаётся неустойчивым. Эти результаты не следует сводить в один недифференцированный балл.

Предлагается запросить два связанных набора материалов. Первый показывает правила управления: распределение ответственности, утверждение изменений и обработку несоответствий. Второй показывает реализацию: версии, журналы испытаний, состав данных и подтверждение устранения конкретных дефектов. Между наборами должны существовать ссылки, позволяющие восстановить путь от правила к выполненному действию.

Организационный документ	Проверяемая реализация
Порядок выпуска	Решение о выпуске определённой версии
Требование повторного испытания	Протокол после существенного изменения
Порядок исправления ошибок	Запись дефекта и подтверждение его устранения
Правило допуска сотрудников	Реальные права доступа и история их изменения

Если часть материалов недоступна по объективной причине, аудитор рассматривает альтернативные процедуры. Наблюдение в защищённой среде иногда даёт больше информации, чем обезличенная выдержка. Однако альтернативная процедура должна действительно отвечать на вопрос; она не используется как формальный способ обойти нехватку доказательств.

Что показывает опыт цифровых платформ

Европейская комиссия разъясняет, что особо крупные онлайн-платформы и поисковые системы в рамках режима цифровых услуг проходят независимый аудит не реже одного раза в год. Делегированный регламент 2024/436 уточняет порядок таких проверок и отчётности. Это отдельная конструкция контроля конкретных услуг, которую нельзя представлять как общий закон об аудите всех моделей ИИ. [6–7]

Её значение для настоящего исследования состоит в связи между проверкой, документированным заключением и последующими действиями проверяемой стороны. Для добровольной программы ARGA предлагается аналогичная прослеживаемость: существенный недостаток получает владельца, обоснованное исправление и отдельную проверку результата. Само обещание устранить проблему не закрывает находку.

Публичность требует самостоятельной работы. Отчёт для руководства может содержать технические сведения, персональные данные и описание уязвимостей, которые нельзя без оценки риска переносить в открытый доступ. Вместе с тем публичное резюме не должно исключать существенные ограничения так, чтобы читатель получал противоположное впечатление о результатах проверки.

В предлагаемом порядке готовятся полный отчёт, защищённые приложения и публичное резюме. Для каждой изъятой части фиксируется основание ограничения доступа. Читателю резюме сообщают предмет, период, виды процедур, существенные результаты и пределы вывода. Раскрытие деталей атаки можно ограничить, сохранив сообщение о наличии неустранённого критического недостатка.

Такое устройство не создаёт права публиковать чужую конфиденциальную информацию. Условия использования отчёта согласуются заранее и соотносятся с законом. Если заказчик требует рекламного заявления, несовместимого с установленными результатами, аудитор должен иметь возможность отказаться от соответствующего использования своего имени и заключения.

Практический разбор 16.03

Рекомендация: Проверять полномочия внешнего оценщика

Условный пример: Исполнитель показывает сертификат своей системы качества вместо подтверждения полномочий на требуемую процедуру.

Разбор: Свидетельство об организации работы не доказывает компетенцию по конкретному предмету и маршруту оценки.

Практические действия: Ответственный за закупку запрашивает область полномочий и сопоставляет её с заданием до заключения договора.

Критерий результата: В договоре указан допустимый вид заключения без расширения статуса исполнителя.

4 Сопоставление подходов США и Великобритании

Нью-Йорк как пример узкой обязательной проверки

Правила Нью-Йорка об отдельных инструментах автоматизированных кадровых решений связывают их использование с аудитом на предвзятость, публикацией сведений и уведомлениями. Официальная страница указывает на аудит в пределах года до использования. Сфера определяется установленным понятием инструмента и условиями применения, а не любой автоматизацией кадровой функции. [8]

Окончательные правила раскрывают независимость аудитора: участие в использовании, разработке или распространении инструмента, трудовая связь и определённые финансовые интересы препятствуют признанию независимости. Они также предусматривают расчёты по установленным категориям и требования к исходным данным. Эти положения нельзя заменять общей декларацией исполнителя о профессиональной объективности. [9]

Методический вывод состоит в необходимости проверить связи команды с самим инструментом. Организационная дистанция от работодателя ещё не исключает участия аудитора в разработке продукта поставщика. Для проверки требуется история отношений, а не только текущая должность. Предложенная в приложении форма охватывает прямые и косвенные связи, субподряд и зависимость оплаты от результата.

Узкий результат	**Почему он не даёт общего разрешения**
Расчёт различий между группами	Не устанавливает всю законность найма
Отсутствие выявленной разницы	Не исключает ошибочность критериев отбора
Публикация кратких результатов	Не подтверждает полноту всех проверок
Выполнение специального аудита	Не отменяет иные применимые обязанности

За пределами соответствующей юрисдикции этот опыт используется для сравнения. Если международная организация добровольно применяет похожие показатели, в задании нужно объяснить основания выбора категорий и законность обработки необходимых данных, не утверждая, что нью-йоркские правила стали для неё обязательными.

Американские методические и надзорные инструменты

Рамка подотчётности Счётной палаты США GAO 2021 года организована вокруг управления, данных, результативности и мониторинга. Она содержит вопросы, полезные для проверки использования ИИ государственными организациями и другими субъектами. Это методический источник для построения программы, а не универсальная схема допуска частных аудиторов на рынок. [12]

Пересмотренное американское руководство по модельному риску 2026 года имеет самостоятельный банковский контекст. Его полезность для сравнительного анализа состоит в внимании к назначению модели, проверке и управлению ограничениями. Генеративные и агентные модели прямо исключены из области этого руководства. Его нельзя представлять как специальный обязательный режим их валидации; использование отдельных подходов за пределами установленной сферы является самостоятельным методическим решением. [13]

Для ARGA предлагается использовать эти материалы как проверку полноты вопросов: кто владеет процессом, насколько пригодны данные, что измеряет испытание и как

обнаруживается ухудшение. Каждый вопрос преобразуется в конкретную процедуру и ожидаемое доказательство. Нельзя заявлять соответствие рамке лишь потому, что четыре её темы перечислены в отчёте.

В международном проекте особенно важны разные значения одинаковых слов. Валидация модели, аудит организации и юридическая оценка имеют различный предмет. Если поставщик присылает отчёт о валидации, аудитор должен прочитать фактический объём выполненных действий: были ли независимые данные, проверялись ли ограничения применения и рассматривались ли реальные последствия ошибок.

Сравнение показывает риск ложной взаимозаменяемости: подробная проверка по одной рамке может оставить без ответа существенный вопрос другой юрисдикции. Вместо общего заявления о международном соответствии полезнее подготовить карту переносимых доказательств и отдельно обозначить дополнительные процедуры для конкретного адресата.

Британский рынок подтверждения надёжности

Британские материалы по AI assurance рассматривают разные способы получения обоснованной уверенности в свойствах систем. Дорожная карта независимой внешней оценки, опубликованная 3 сентября 2025 года, посвящена развитию качества услуг, компетенций и рынка. Это программный документ; его публикация не означает введения единой обязательной сертификации любого ИИ. [10–11]

Для заказчика существенен выбор услуги. Проверка документации, оценка воздействия, техническое испытание и сертификация отвечают на разные вопросы. Если задача состоит в проверке корректности обработки юридических текстов, одна только оценка зрелости управления не даст достаточного ответа. Напротив, большой набор тестовых запросов не покажет, кто отвечает за исправление ошибок после внедрения.

Подход	Полезный результат	Предел переноса
ЕС и специальные процедуры	Установленная форма оценки определённого объекта	Требуются применимость и надлежащий исполнитель
Нью-Йорк и кадровые инструменты	Узкая проверка предвзятости по местным правилам	Не охватывает всю правомерность решения
Методические рамки США	Организация вопросов управления и испытаний	Не являются общим разрешением на эксплуатацию
Британская политика assurance	Развитие качественных внешних услуг	Не заменяет отраслевые обязанности

Для ARGA предлагается выбирать модель работы от потребности клиента, а не от наиболее убедительного названия. Предмет, квалификация исполнителей, доступ и вид итогового документа определяются совместно. Когда требуются несколько видов проверки, их результаты можно объединить, сохраняя отдельные основания и ограничения каждого вывода.

Практический разбор 16.04

Рекомендация: Разделять юрисдикции и назначения проверки

Условный пример: Отчёт о кадровом инструменте из одной юрисдикции предлагается банку в другой стране.

Разбор: Объект, критерии и последствия решений различаются. Перенос результата требует самостоятельного обоснования.

Практические действия: Юрист составляет таблицу совпадающих и отсутствующих критериев; технический эксперт назначает дополнительные тесты.

Критерий результата: Для каждого перенесённого вывода указано основание либо ограничение.



5 Стандарты и выбор критериев

Актуальная редакция и проверяемая область

Официальная карточка ISO 19011 указывает редакцию 2026 года и предмет руководства по аудиту систем менеджмента, включая управление программой и компетентность аудиторов. Сам стандарт не является самостоятельной схемой сертификации. Для настоящего доклада он служит ориентиром постановки аудиторской программы; конкретные процедуры далее предлагаются как авторская методология. [16]

ISO/IEC 42001:2023 относится к системе управления ИИ в организации. ISO/IEC 42006:2025 описывает дополнительные требования к органам, проводящим аудит и сертификацию таких систем, и дополняет ISO/IEC 17021-1. По открытым описаниям эти документы имеют разные адресаты. Сертификат организационной системы нельзя автоматически представлять доказательством безошибочности каждого продукта. [17–18]

При заключении договора фиксируются обозначение стандарта, редакция и источник его обязательности для задания. Если стороны согласуют соответствие конкретным пунктам, команде необходим законный доступ к полному тексту. Открытая карточка позволяет проверить название и предмет, но не заменяет профессиональное применение всех требований документа.

Документ	Объект открытого описания
ISO 19011:2026	Руководство по аудиту систем менеджмента
ISO/IEC 42001:2023	Система управления ИИ организации
ISO/IEC 42006:2025	Органы аудита и сертификации таких систем
ISO/IEC 23894:2023	Руководство по управлению рисками ИИ

Задание не должно содержать обещание соответствия «всем международным стандартам». Проверяемый набор всегда ограничен. Если критерий изменяется во время работы, стороны документируют последствия: какие процедуры нужно повторить, какие данные дополнить и какие выводы больше нельзя выдавать в прежней формулировке.

Испытание и подтверждение утверждений

ISO/IEC 17025:2017 касается компетентности испытательных и калибровочных лабораторий, а ISO/IEC 17029:2019 — органов валидации и верификации. Их открытые описания показывают, что разные виды подтверждения требуют разных организационных и методических оснований. Наличие одного статуса не доказывает способность выполнять любую проверку ИИ. [19–20]

Для выбора исполнителя полезно запросить область подтверждённой компетентности и сопоставить её с реальной задачей. Лаборатория может хорошо измерять определённый технический показатель, но не проводить анализ правомерности обработки персональных данных. Юридическая команда может установить обязанности, но не быть подготовленной к статистическому проектированию испытания. Письменное распределение этих задач предотвращает скрытые пробелы.

В добровольной программе необходимо отдельно определить, подтверждается ли утверждение о уже выполненном действии или оценивается обоснованность предполагаемого применения. Прошлая успешная работа в одной среде не даёт уверенности в будущем использовании в другой. Аудитор указывает, какие данные относятся к наблюдаемому опыту, а какие — к прогнозу или условному предположению.

Предлагается оформлять каждый критерий как короткую запись: проверяемое утверждение, основание, процедура, ожидаемый материал и допустимый предел вывода. Это делает видимой разницу между фактом и оценкой. Например, «существует журнал» проверяется иначе, чем «журнал позволяет расследовать каждое существенное действие». Второй вопрос требует исследования содержания, полноты и доступности записи.

Польза стандарта возникает при его точном включении в такую программу. Перечень номеров на титульном листе не заменяет выбора методов и не подтверждает, что конкретный тест был выполнен компетентной и независимой стороной.

Управление риском и оценка воздействия

ISO/IEC 23894:2023 посвящён управлению рисками ИИ, а ISO/IEC 42005:2025 — оценке воздействия систем ИИ. По открытым описаниям это взаимодополняющие направления: организация исследует неопределённость и возможные последствия для людей и общества. В докладе не воспроизводятся недоступные полные требования этих стандартов. [21–22]

Рамка NIST AI RMF 1.0 является добровольной и организует работу через функции управления, определения контекста, измерения и обработки риска. Она не устанавливает универсальный допустимый уровень риска. Для аудита это означает необходимость выяснить, кто выбрал критерий приемлемости и чем он обоснован в конкретной сфере. [14]

Предлагается отделять измеримость от значимости. Не каждый существенный вред удобно выразить одним числом. Разглашение конфиденциального досье, невозможность оспорить

ошибочный вывод или использование недопустимого основания могут требовать качественной правовой оценки. Низкая частота таких событий не делает их автоматически приемлемыми.

Вопрос оценки	Требуемый результат
Кто может пострадать	Перечень затронутых групп и ситуаций
Каким путём возникает вред	Описание причинной последовательности
Что измеряется	Показатель с областью применимости
Кто принимает остаточный риск	Решение уполномоченного лица с основанием
Когда вывод пересматривается	Событие или изменение, запускающее проверку

Аудитор проверяет качество этого рассуждения и доказательств. Он не должен незаметно подменять решение организации собственным предпочтением или принимать любое её решение как достаточное. Если критерий допуска противоречит применимой норме, согласие руководства не исправляет правовой дефект.

Практический разбор 16.05

Рекомендация: Выбирать критерий до испытания

Условный пример: Поставщик после теста выбирает показатель, по которому его продукт выглядит лучше.

Разбор: Такое решение скрывает неудачные исходы и делает оценку зависимой от результата.

Практические действия: Аудитор заранее фиксирует существенную ошибку, полноту ответа, отказы и правило допуска; изменения протоколирует.

Критерий результата: Итог содержит все заранее выбранные показатели, включая неблагоприятные.



6 Независимость и конфликт интересов

От заявления к проверке отношений

Независимость включает способность выбирать необходимые процедуры, оценивать материалы и сообщать неудобные результаты без недопустимого влияния. Внешний юридический адрес сам по себе этого не обеспечивает. Проверку могут ограничивать финансовая зависимость, предшествующее участие в проекте, личные связи или контроль заказчика над публикацией заключения.

В предлагаемой программе перед принятием задания исследуются отношения организации, руководителя проекта, технических исполнителей и рецензента с проверяемой системой. Особое внимание уделяется разработке, настройке, продаже и сопровождению продукта. Если сотрудник ранее принимал спорное решение, последующая оценка этого решения создаёт риск проверки собственной работы.

Угроза	Проверяемое обстоятельство	Возможное действие
Собственный интерес	Доля в бизнесе или зависимый платёж	Отказ либо устранение интереса
Проверка собственной работы	Участие в разработке проверяемой функции	Другой исполнитель и отдельная рецензия
Защита позиции клиента	Публичное продвижение спорного продукта	Разделение ролей и ограничение задания
Давление	Угроза оплаты или доступа	Эскалация и право прекращения работы

Эти категории используются как методическая карта, а не как утверждение о едином международном перечне запретов. При наличии обязательной процедуры сначала применяются её специальные правила. В добровольной проверке решение должно учитывать значимость угрозы и реальную эффективность защитных мер.

Одной декларации сотрудника недостаточно для сложного проекта. Её сопоставляют с договорами, составом команды и историей работ в доступном объёме. Скрытый конфликт, обнаруженный позднее, требует пересмотра затронутых процедур, даже если первоначальное заключение уже выдано.

Оплата и контроль над содержанием

Оплата заказчиком является обычной экономической моделью внешней проверки, но договор не должен превращать положительный вывод в условие вознаграждения. Предлагается связывать оплату с выполненными этапами и согласованным объёмом процедур. Отрицательная находка не считается недостатком услуги, если она обоснована и получена в пределах задания.

Право заказчика исправить фактическую неточность необходимо отличать от права удалить неблагоприятный вывод. Перед завершением отчёта проверяемая сторона получает возможность представить документы и возражения. Аудитор фиксирует рассмотрение каждого существенного возражения. Принятое замечание меняет текст по доказательственной причине, а не вследствие переговоров о репутационном удобстве.

Если одна группа одновременно продаёт внедрение и проверку внедрённой системы, внутреннее разделение команд не всегда устраняет угрозу. Нужно оценить общую финансовую заинтересованность, руководство, премирование и обмен материалами. При

существенном риске наиболее убедительным решением может быть привлечение другой организации к проверке определённых утверждений.

Условие договора	Предлагаемая защита
Порядок оплаты	Вознаграждение не зависит от положительного вывода
Доступ к материалам	Право запросить необходимые сведения и фиксировать отказ
Согласование проекта	Исправление фактов без права скрывать обоснованные находки
Использование имени	Запрет вводящего в заблуждение рекламного пересказа
Прекращение задания	Документирование причин и уже обнаруженных существенных рисков

Объём раскрытия условий оплаты определяется законом и договором. Даже если суммы конфиденциальны, итоговый документ может сообщать, кто заказал проверку и какие меры защищали независимость, не создавая у читателя ложного впечатления о происхождении финансирования.

Проверка независимости в ходе проекта

Независимость оценивается не только при подписании договора. В процессе могут измениться состав команды, условия оплаты, собственник поставщика или характер консультационной помощи. Поэтому предлагается повторять проверку при существенном изменении и перед подписанием заключения. Новый конфликт рассматривается до того, как его влияние будет встроено в окончательный результат.

Решение по конфликту должно содержать обстоятельства, затронутые задачи, меру защиты и лицо, которое оценило её достаточность. Формула «конфликт урегулирован» без объяснения не позволяет понять, что произошло. Например, замена рецензента устраняет его личную связь, но не обязательно исправляет уже выполненное им тестирование.

Для небольшого проекта возможна пропорциональная процедура: короткая декларация, проверка очевидных отношений и независимое рассмотрение существенного вопроса. Сложный проект с чувствительными последствиями потребует более подробной оценки. Масштаб документа не должен подменять результат: важно установить, способен ли исполнитель вынести неблагоприятное заключение и сохранить его обоснование.

При невозможности обеспечить независимость допустим другой вид работы, например консультационный анализ с открытым описанием роли. Однако переименование не должно скрывать ту же проблему от пользователя результата. Нельзя выдать консультационный отчёт за независимое подтверждение, а затем сослаться на мелкую оговорку в приложении.

Практический критерий прост: внешний читатель должен понимать, кто проверял систему, какие связи могли повлиять на оценку и почему принятые меры считаются достаточными. Если такое объяснение невозможно без умолчаний, независимый статус заключения не обоснован.

Практический разбор 16.06

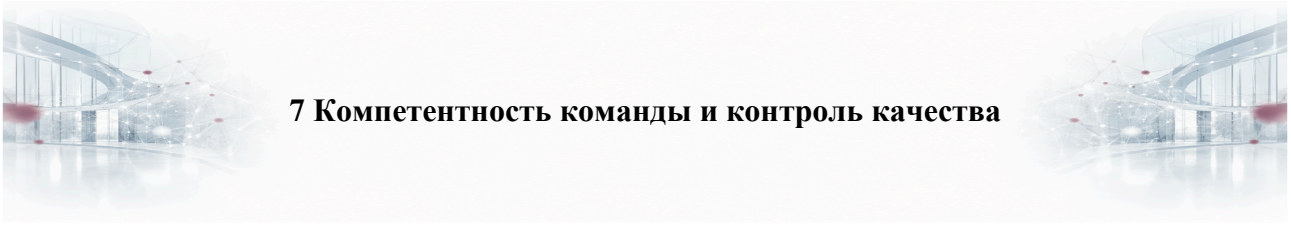
Рекомендация: Устранять конфликт интересов до назначения

Условный пример: Ведущий аудитор ранее проектировал проверяемый механизм фильтрации.

Разбор: Он оценивает собственное решение; декларация сама по себе не устраняет зависимость.

Практические действия: Координатор передаёт проверку механизма независимому специалисту и сохраняет решение об отводе.

Критерий результата: Автор механизма не утверждает вывод о его эффективности.



7 Компетентность команды и контроль качества

Компетенции под предмет проверки

Команда формируется после определения объекта, а не наоборот. Для проверки юридического помощника могут понадобиться правовая экспертиза, знание обработки документов, статистика и информационная безопасность. Для системы распознавания изображения состав будет иным. Общий опыт работы с ИИ не подтверждает компетентность во всех этих областях.

Предлагается разделить обязательные для проекта умения и вспомогательные. Руководитель отвечает за согласованность программы, но не обязан самостоятельно выполнять каждую специализированную процедуру. При этом он должен понимать границы получаемых заключений и уметь заметить, когда технический ответ не разрешает правовой вопрос. Подрядчик со специальными знаниями получает чётко сформулированное задание.

Роль	Проверяемое умение	Подтверждение в проекте
Руководитель	Связать цель, методы и вывод	Утверждённая программа и решения по ограничениям
Юрист	Установить применимые требования	Мотивированная карта обязанностей
Технический специалист	Воспроизвести наблюдаемое поведение	Протокол и сохранённые результаты
Специалист по данным	Обосновать выборку и показатели	Расчёт и анализ неопределённости

Роль	Проверяемое умение	Подтверждение в проекте
Рецензент	Проверить логику заключения	Замечания и след их разрешения

Заявление о квалификации оценивается по отношению к конкретной задаче. Сертификат обучения может показать прохождение курса, но не выполнение сопоставимых проверок. Опыт желательно подтверждать допустимыми примерами работ, описанием роли и способностью объяснить применённый метод. Конфиденциальность прежних проектов не оправдывает необоснованное обещание компетентности.

Если нужного специалиста нет, программа изменяется до начала испытаний. Существенный пробел нельзя скрывать общей формулой о междисциплинарном подходе. Ограниченная, но корректно выполненная проверка полезнее широкого заключения, для которого команда не располагала необходимыми знаниями.

Независимое рецензирование

Содержательная рецензия проверяет цепочку от вывода к данным. Рецензент получает протокол, исходные количества, классификатор ошибок и примеры спорной разметки; затем независимо восстанавливает основной расчёт и выбирает хотя бы одну правдоподобную альтернативную интерпретацию. Результат рецензии — перечень проверенных утверждений, существенных возражений и условий, при которых вывод остаётся обоснованным. Исправление стиля само по себе не выполняет эту функцию.

Предлагается выделить обязательные предметы рецензии: применимость критериев, полноту значимых процедур, расчёты, противоречащие сведения и формулировку итогового мнения. Для технических результатов полезна независимая перепроверка части расчётов. Для правовых — сопоставление спорного вывода с текстом нормы и фактическими условиями её применения.

Замечания рецензента сохраняются вместе с ответами исполнителя. Возможны исправление, дополнительная процедура или мотивированное несогласие. Сам факт обсуждения не закрывает проблему. Если спор относится к существенному выводу и не разрешён, его передают компетентному руководителю либо отражают как ограничение. Нельзя завершать рецензию простым удалением неудобного замечания.

Вопрос рецензента	Пример выявляемого дефекта
Совпадает ли объект с заданием	Испытана другая версия модели
Верен ли знаменатель	Ошибки поделены только на успешные ответы
Учтены ли исключения	Сложные документы удалены без объяснения
Есть ли противоречащий материал	Жалоба пользователя не включена в анализ

Вопрос рецензента	Пример выявляемого дефекта
Соразмерен ли вывод	Тест функции превращён в оценку всей системы

Рецензия не устраняет исходный конфликт интересов автоматически. Если рецензент получает только отобранные положительные материалы или зависит от того же коммерческого решения, организационное название проверки не обеспечивает её качества.

Работа с внешними специалистами

Передача части задач подрядчику требует определения предмета, доступа и формы результата. Внешний специалист должен знать, на какой вопрос его работа отвечает и какие утверждения она будет поддерживать. Иначе отчёт о техническом тесте может впоследствии использоваться для правового вывода, которого специалист не делал и не мог сделать.

Предлагается включать в задание описание системы, критерии, разрешённую среду, правила сохранения материалов и условия сообщения о критической находке. Для доступа к чувствительным данным отдельно определяются правовые основания и меры защиты. Подрядчик не должен самостоятельно передавать такие данные ещё одному сервису ради удобства обработки без предусмотренного разрешения и основания.

Руководитель проекта сохраняет ответственность за согласованность итогового документа. Он проверяет, что результаты подрядчика относятся к нужной версии, получены допустимым способом и сопровождаются необходимыми ограничениями. При отсутствии компетенции оценить метод следует привлечь дополнительного рецензента, а не принимать техническую сложность текста за доказательство качества.

Рабочие материалы подрядчика должны оставаться доступными в согласованном объёме после завершения проекта. Передача одной презентации без протоколов может сделать результат непроверяемым. Если полное раскрытие невозможно, необходимо заранее согласовать защищённый просмотр, независимое подтверждение или иной достаточный механизм доступа.

Для ARGA такая модель позволяет развивать экспертную работу без неподтверждённого заявления о собственной универсальной лаборатории. Организация может координировать юридическую постановку вопроса и проверку доказательств, привлекая специализированные ресурсы в пределах конкретного задания и описывая их фактическую роль.

Практический разбор 16.07

Рекомендация: Назначать компетенции по рискам задания

Условный пример: Юрист оценивает корректность ссылок, но никто не проверяет статистический расчёт.

Разбор: Юридическая квалификация не компенсирует неверный знаменатель или зависимость наблюдений.

Практические действия: Руководитель привлекает статистика для плана и расчёта, юриста для существенности ошибок; назначает рецензента.

Критерий результата: У каждой существенной части заключения есть компетентный проверяющий.



8 Задание на аудит и условия доступа

Проверяемые утверждения вместо общей цели

Задание «проверить безопасность ИИ» слишком неопределённо. Его следует преобразовать в перечень утверждений: какие сведения система получает, какие действия выполняет, какие ошибки имеют значение и какие меры должны их обнаруживать. Для каждого утверждения указываются объект, критерий и решение, которое заказчик намерен принять по результатам проверки.

Например, в системе обработки юридических материалов можно отдельно проверить корректность ссылок на документы, сохранение отрицаний, разделение пользовательских досье и обязательность человеческого подтверждения перед внешней отправкой. Эти свойства требуют разных процедур. Успех в одном тесте не должен компенсировать провал другого, если последний затрагивает существенный правовой интерес.

Утверждение	Процедура	Предел результата
Ответ связан с источником	Проверка текста и указанного фрагмента	Не подтверждает полноту всех источников
Досье разделены	Разрешённый тест доступа и настройки	Относится к испытанным ролям
Отправка требует согласования	Проверка маршрута и попытки обхода	Не охватывает неизвестные интеграции
Ошибку можно исправить	Прослеживание обращения до результата	Не доказывает одинаковое качество всех обращений

Критерии приемки определяются до просмотра итоговых результатов. Иначе заказчик или исполнитель может подобрать удобный показатель после обнаружения проблем. Допустимые изменения программы документируются с объяснением; первоначальный план и ранние результаты сохраняются, чтобы было видно, как менялась логика проверки.

Существенный вопрос не должен исчезать из задания только потому, что его трудно исследовать. Если достаточная процедура недоступна, это основание ограничить вывод или продолжить работу другим способом, а не заменить исходный вопрос более простым.

Границы разрешённого доступа

Доступ должен быть достаточным для выбранных утверждений и одновременно ограниченным необходимостью. Аудитор не получает общего права извлекать любые сведения из системы. В задании определяются среда, учётные записи, разрешённые операции, данные, время работы и порядок действий при неожиданном обнаружении чувствительной информации.

Различаются документальный доступ, наблюдение за исполнением, доступ к интерфейсу и доступ к внутренним компонентам. Ни один уровень не гарантирует полноты автоматически. При доступе только через интерфейс можно исследовать наблюдаемое поведение, но трудно установить все причины ошибки. Доступ к коду помогает понять реализацию, однако сам по себе не показывает производственные настройки и реальные данные.

Уровень доступа	Возможная польза	Требуемая оговорка
Документы	Правила и заявленная архитектура	Исполнение ещё не подтверждено
Наблюдение	Реальный процесс под контролем владельца	Возможен эффект подготовленного показа
Интерфейс	Поведение доступных функций	Внутренние причины могут быть неизвестны
Компоненты и журналы	Связь поведения с реализацией	Нужна проверка полноты и версии

Если заказчик ограничивает доступ, фиксируется конкретный недоступный материал и затронутый вывод. Общая фраза «проверка проведена на основании представленных документов» недостаточна, когда скрывает отсутствие ключевой информации. Читателю нужно понимать практическое последствие ограничения.

Для чувствительных досье предпочтительны искусственные примеры и контролируемые копии, когда они действительно позволяют проверить нужное свойство. Реальные материалы используются только при обоснованной необходимости и надлежащих условиях. Замена данных не должна менять предмет теста незаметно для аудитора.

Прекращение работы и существенная находка

Договор должен предусматривать действия при обнаружении ситуации, требующей немедленного внимания. Это может быть доступ к чужим документам, выполнение неразрешённой операции или сохранение чувствительных данных во внешней среде. Продолжение теста ради увеличения числа подтверждений способно усилить вред и выйти за пределы разрешения.

Предлагается заранее определить канал уведомления, ответственное лицо и безопасный порядок остановки. Аудитор сохраняет достаточные сведения о находке, избегая ненужного копирования чувствительных материалов. Затем согласуется ограниченная проверка

причины и исправления. Подробные действия зависят от правового режима и характера системы, поэтому универсальный сценарий отключения не устанавливается.

Прекращение задания не должно уничтожать след уже выполненной работы. В рабочем деле сохраняются причина, момент остановки, проведённые процедуры и статус существенных результатов. Итоговый документ может быть ограниченным отчётом о выполненном объёме. Его нельзя выдавать за положительное заключение только потому, что весь план не был завершён.

Если заказчик отказывается предоставить существенные материалы, аудитор оценивает возможность альтернативной процедуры. При её отсутствии соответствующее утверждение остаётся неподтверждённым. Коммерческая заинтересованность в завершении проекта не создаёт доказательств, которых нет.

Для ARGA предлагается включать эти условия в проект задания до начала любой технической проверки. Это уменьшает риск спора о том, вправе ли исполнитель сообщить о критическом недостатке, потребовать дополнительную информацию или отказаться от чрезмерно широкого использования своего заключения.

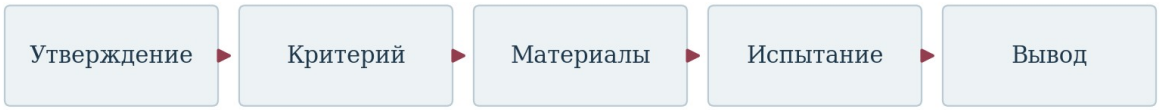


Рисунок 16.1. От проверяемого утверждения к решению. Схема методики раздела 8. Свидетельства относятся к определённой версии и назначению системы.

Практический разбор 16.08

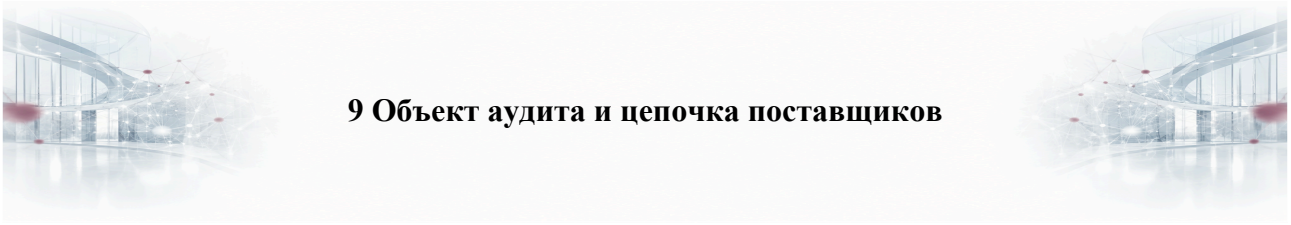
Рекомендация: Закреплять разрешённый доступ и остановку

Условный пример: Для аудита хронологии исполнитель получает доступ к отправке клиентских писем.

Разбор: Право изменять внешнее состояние не требуется для чтения досье и увеличивает последствия ошибки.

Практические действия: Заказчик ограничивает учётную запись чтением и задаёт остановку при обнаружении чувствительных данных вне задания.

Критерий результата: Журнал подтверждает отсутствие неразрешённых действий и соблюдение границ доступа.



9 Объект аудита и цепочка поставщиков

Система шире отдельной модели

В реальном применении результат создаётся сочетанием модели, входных данных, инструкций, поисковых компонентов, прав доступа и действий человека. Проверка только базовой модели может не охватить ошибку, возникающую при извлечении документов или последующей передаче ответа. Поэтому объект описывается через полный путь от получения запроса до значимого результата.

Для юридического помощника такая карта включает загрузку файла, распознавание текста, разбиение на фрагменты, поиск, генерацию и проверку пользователем. На каждом переходе возможна потеря смысла. Например, неверное распознавание даты может привести к корректно сформулированному, но фактически ошибочному выводу. Оценка одного последнего ответа не всегда позволяет локализовать причину.

Компонент	Вопрос аудитора
Приём документа	Сохранены ли содержание и принадлежность досье
Предварительная обработка	Не потеряны ли таблицы, отрицания и примечания
Поиск материалов	Какие источники доступны и что исключено
Генерация ответа	На каких данных и инструкциях построен результат
Последующее действие	Кто подтверждает решение и куда оно передаётся

В предлагаемом паспорте объекта фиксируются владельцы компонентов и точки, в которых система обращается к внешнему поставщику. Это позволяет связать найденный дефект с ответственным участником. Внешний компонент не освобождает интегратора от понимания его роли в конечном процессе.

Граница проверки должна быть видна в отчёте. Если аудит завершался на выдаче черновика, нельзя утверждать, что исследован весь процесс принятия решения. Если человек редактировал результат, отдельно рассматривается его вклад и возможное влияние интерфейса на выбор.

Идентификация версии и конфигурации

Коммерческое название системы недостаточно для повторной проверки. Требуются доступные идентификаторы версии, дата использования, параметры, системные инструкции и состояние внешних источников. Для облачного сервиса часть сведений может быть скрыта.

Тогда аудитор сохраняет доступные идентификаторы и прямо описывает невозможность установить некоторые внутренние изменения.

Изменение поведения возможно без замены базовой модели. Достаточно изменить шаблон запроса, права инструмента, список документов или правила фильтрации. Поэтому паспорт должен охватывать не только веса или название модели, но и конфигурацию её использования. Для важных компонентов полезны контрольные суммы файлов, если их законно можно сохранить.

Элемент паспорта	Зачем он нужен
Версия приложения	Связать интерфейс и логику интеграции
Идентификатор модели	Уточнить доступный объект тестирования
Настройки и инструкции	Восстановить условия получения ответа
Состав источников	Объяснить информационные границы системы
Права инструментов	Определить возможные действия
Дата и среда	Учесть внешние изменения и ограничения

Контрольная сумма подтверждает совпадение байтов, но не истинность содержания и не законность происхождения файла. Её значение состоит в обнаружении изменения уже зафиксированного материала. Аудитор должен сохранить связь между суммой, самим объектом и способом его получения; один список технических значений без этой связи бесполезен.

Если версия недоступна для повторного запуска, проверка может оставаться документально воспроизводимой в ограниченном смысле: другой специалист способен проследить расчёт и вывод по сохранённым материалам. Однако это не равнозначно независимому повторению поведения исходной системы.

Зависимость от поставщика

Поставщик может предоставить отчёт, описание модели, журнал изменений или результат собственной оценки. Такие материалы полезны, но их предмет необходимо сопоставить с проверяемой интеграцией. Успешная оценка общего назначения не обязательно включает язык, тип документов и характер ошибок, значимые для конкретного пользователя.

Предлагается составить карту зависимости: какой вывод основан на собственном наблюдении, какой — на сведениях поставщика, а какой — на предположении. Для каждого существенного предположения указывается действие при его нарушении. Например, обещание неизменности модели должно сопровождаться механизмом получения уведомлений либо самостоятельного обнаружения значимого изменения.

Наличие нескольких субподрядчиков усложняет доступ к доказательствам. Контракт основного поставщика может разрешать техническое обслуживание, но не предоставлять

заказчику достаточные сведения для аудита. Поэтому права на получение материалов и наблюдение за проверкой следует согласовать до внедрения, когда ещё возможно выбрать другой вариант интеграции.

Проверка должна учитывать прекращение услуги. Если договор заканчивается, организация должна понимать, какие журналы, настройки и подтверждения можно выгрузить и сохранить. Иначе позднее расследование ошибки окажется зависимым от добровольного содействия бывшего поставщика. Срок хранения определяется применимыми обязанностями и необходимостью, а не желанием сохранять всё бессрочно.

Для ARGA полезен принцип ограниченного доверия к документам поставщика: принимать их как один из источников, проверять внутреннюю согласованность и добиваться прямого подтверждения наиболее значимых утверждений. Это не означает заранее считать поставщика недобросовестным; это способ обосновать вывод независимо от его маркетингового описания.

Практический разбор 16.09

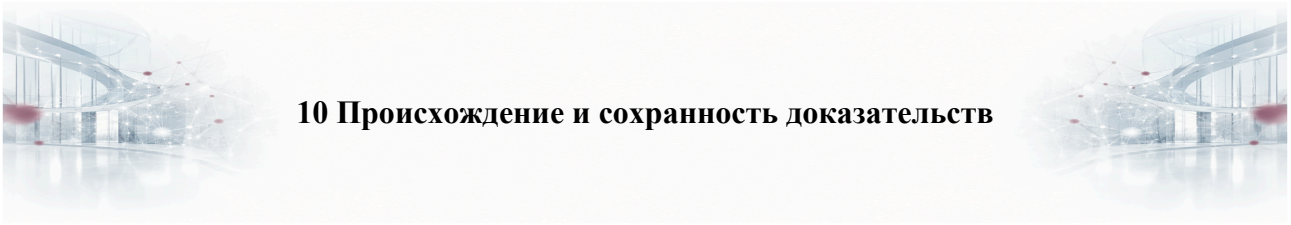
Рекомендация: Описывать всю проверяемую конфигурацию

Условный пример: Модель осталась прежней, но поисковый индекс обновили перед повторным тестом.

Разбор: Изменившийся результат нельзя приписать только модели: входной контекст тоже изменился.

Практические действия: Инженер фиксирует версии модели, индекса, инструкций и инструментов; сравнение повторяет на сохранённых сборках.

Критерий результата: Паспорт запуска позволяет отличить изменение модели от изменения поиска.



10 Происхождение и сохранность доказательств

Реестр материалов

Рабочий реестр связывает каждый материал с источником, датой получения, объектом и проверяемым утверждением. В него включаются не только файлы, но и наблюдения, интервью, результаты запросов и решения об исключении данных. Читатель должен понимать, какой материал реально использовался и где находится его сохранённая версия.

Предлагаемый минимальный состав записи включает идентификатор, название, происхождение, ответственного за получение, условия доступа и краткое описание доказательственного значения. При изменении файла создаётся новая запись или версия с прослеживаемой связью. Нельзя заменять исходный документ отредактированной копией без сохранения различия между ними.

Поле	Пример условной записи
Идентификатор	E-014
Материал	Журнал запросов тестовой среды
Источник	Выгрузка администратором в присутствии аудитора
Период	Установленный тестовый интервал
Проверяемый вопрос	Сохранение вызовов внешних инструментов
Ограничение	Полнота до начала интервала не исследована

Реестр не обязан раскрывать всем участникам полный путь к чувствительному файлу. Можно разделить описание материала и защищённое местонахождение. При этом уполномоченный рецензент должен иметь возможность получить доступ к доказательству, когда это необходимо для проверки вывода.

Неполный реестр создаёт риск незаметной подмены: новая версия отчёта поставщика может отличаться от той, на которой основана оценка. Поэтому при завершении задания фиксируется состав рабочего дела и связь между находками и конкретными версиями материалов.

Цепочка обработки материала

Материал нередко проходит преобразования: распознавание, перевод, очистку, обезличивание и расчёт показателей. Каждое преобразование может изменить доказательственное значение. Поэтому требуется связь между исходным объектом и производными результатами, включая используемый способ обработки и существенные решения исполнителя.

Например, при переводе юридического документа проверяется сохранение модальности: обязанность, возможность и рекомендация не взаимозаменяемы. При распознавании таблицы важно сохранить соответствие строк и столбцов. При обезличивании нельзя удалить признаки, от которых зависит исследуемая ошибка, не объяснив влияние такого удаления на тест.

Преобразование	Риск	Контроль
Распознавание	Потеря отрицания или номера	Сопоставление значимых фрагментов с изображением
Перевод	Усиление или ослабление смысла	Проверка специалистом по исходному языку
Очистка	Исключение неудобных наблюдений	Реестр правил и удалённых записей

Преобразование	Риск	Контроль
Обезличивание	Изменение проверяемого сценария	Оценка сохранности существенных свойств
Агрегирование	Скрытие различий между группами	Сохранение допустимой детализации

Контрольная сумма помогает обнаружить изменение сохранённого объекта после фиксации, но не показывает, был ли он полно представлен первоначально. Для выгрузки журнала дополнительно проверяются условия формирования, фильтры и временные границы. Несоответствие числа событий ожидаемому объёму требует объяснения до расчёта показателей.

Описанная цепочка является рабочей методикой сохранения прослеживаемости. Она не гарантирует допустимость материала в любом судебном процессе: процессуальные требования, полномочия получения и форма представления устанавливаются отдельно по применимому праву.

Противоречащие сведения и недоступные материалы

Аудитор обязан по смыслу выбранной методологии искать не только подтверждение исходного предположения, но и сведения, способные его опровергнуть. Для каждого существенного утверждения полезно задать вопрос: какое наблюдение заставило бы изменить вывод? Если ответ отсутствует, проверка рискует превратиться в оформление заранее принятой позиции.

Противоречие между документами не всегда означает нарушение. Политика могла быть обновлена, отчёт мог относиться к другой версии, а журнал — к иной среде. Сначала уточняются даты и предмет. Если различие остаётся необъяснённым, его сохраняют как находку, а не выбирают более удобный документ без мотивировки.

Недоступность материала фиксируется отдельно от его отсутствия. Заказчик может не иметь права передать персональные сведения либо не контролировать внутренний компонент поставщика. Это обстоятельство влияет на выбор процедуры и предел вывода, но само по себе не доказывает ни правомерность, ни неправомерность работы системы.

Предлагается вести журнал запросов: что запрошено, зачем, кому, какой ответ получен и как он повлиял на программу. Для существенного отказа указывают возможную альтернативу и причину её достаточности или недостаточности. Так рецензент видит, что ограничение действительно исследовалось.

В финальном отчёте существенные противоречия описываются рядом с соответствующим выводом. Перенос всех ограничений в длинный общий список в конце документа ухудшает понимание: пользователь может прочитать положительное утверждение, не заметив, что ключевой материал для него оказался недоступен.

Практический разбор 16.10

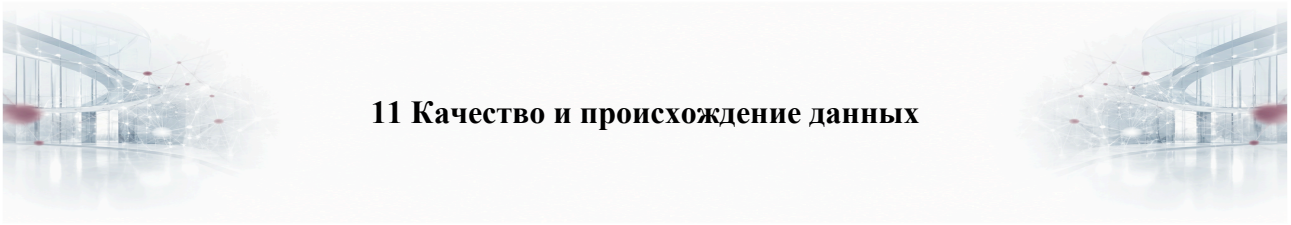
Рекомендация: Сохранять оригинал и цепочку преобразований

Условный пример: После распознавания скана в таблице поменялись местами даты двух событий.

Разбор: Ошибка может быть внесена OCR, а не помощником. Один итоговый текст не позволяет различить источники дефекта.

Практические действия: Специалист хранит скан, результат OCR и исправленную копию отдельно; указывает преобразования и проверяющего.

Критерий результата: Любую спорную дату можно проследить до исходной страницы.



11 Качество и происхождение данных

Пригодность данных для вопроса

Данные оцениваются относительно назначения проверки. Набор, подходящий для демонстрации интерфейса, может быть непригоден для измерения ошибок. Архив успешно завершённых дел не представляет все входящие обращения. Для аудита необходимо определить, какую совокупность предполагается описать и почему доступные наблюдения относятся именно к ней.

Предлагается проверять происхождение, полноту, актуальность, структуру пропусков и связь с реальной средой. Пропуск сам по себе является информацией: отсутствие даты рождения может встречаться чаще в определённой категории документов и влиять на ложные совпадения. Автоматическое удаление неполных записей способно сделать испытание искусственно благополучным.

Свойство данных	Проверочный вопрос
Происхождение	Из какого процесса получены записи
Полнота	Какие случаи отсутствуют и почему
Актуальность	Соответствуют ли данные периоду применения
Структура	Сохранены ли связи между полями и документами
Представительность	Какие группы и сценарии недопредставлены

Если реальных данных мало, можно использовать специально подготовленные сценарии. Их назначение должно быть ясным: они помогают обнаруживать определённый дефект, но не обязательно оценивают частоту этого дефекта в эксплуатации. Нельзя превращать долю ошибок на намеренно сложных примерах в оценку общего риска без дополнительного основания.

Для ARGA приоритетны свойства, влияющие на смысл: транслитерация имени, несколько алфавитов, качество скана, неполные даты, противоречащие документы и различие между утверждением стороны и установленным фактом. Эти признаки включаются в план данных до проведения испытаний.

Эталонная разметка

Сравнение ответа с «правильным» результатом требует проверки самого эталона. В юридических и социальных задачах ответ иногда зависит от контекста или допускает несколько обоснованных вариантов. Если разметчик не видел всех материалов либо использовал устаревшую норму, измеренная ошибка модели может фактически быть ошибкой эталона.

Предлагается разработать инструкцию разметки с определениями категорий, примерами и порядком разрешения спорных случаев. Часть материала независимо оценивают два специалиста. Их расхождения анализируются до окончательного расчёта, поскольку они могут указывать на неясность критерия. Простое голосование не всегда устраняет правовую неоднозначность.

Ситуация	**Корректная обработка**
Однозначный факт	Сопоставление с первичным материалом
Несколько допустимых формулировок	Оценка сохранения существенного смысла
Правовой спор	Отдельная категория и мотивировка
Недостаток сведений	Отметка неопределённости без выдуманного ответа
Ошибка разметчика	Исправление эталона с сохранением истории

Разметчики по возможности не должны знать, какой системе принадлежит ответ, если это может повлиять на оценку. При этом им необходимо предоставить контекст, без которого нельзя проверить содержание. Слепая оценка полезна только тогда, когда не лишает эксперта существенных данных.

После исправления эталона результаты пересчитываются по одинаковым правилам для всех сравниваемых вариантов. Нельзя менять критерий только для ответов предпочтительного поставщика. В рабочем деле сохраняются исходная разметка, причина изменения и влияние на итоговый показатель.

Утечка тестовых примеров и выбор удобных данных

Высокий результат может отражать знакомство модели или разработчика с тестовыми примерами. Если набор использовался для настройки, он перестаёт быть независимым основанием окончательной оценки в прежнем смысле. Необходимо установить, кто имел

доступ к данным и какие изменения выполнялись после просмотра промежуточных результатов.

Предлагается разделять данные для разработки, уточнения методики и окончательной проверки. Граница зависит от проекта, но её нужно документировать. После доработки системы на выявленных ошибках полезно проверить исправление на известных примерах и отдельно оценить перенос результата на новые случаи. Эти две процедуры отвечают на разные вопросы.

Другой риск — выбор только удобных документов. Например, исключение сканов низкого качества может быть оправдано, если система официально не предназначена для них и интерфейс надёжно отклоняет такие файлы. Если пользователи регулярно загружают их без предупреждения, исключение скрывает реальную часть процесса.

В отчёте приводится схема движения наблюдений: сколько записей было доступно, сколько исключено по каждому основанию и сколько вошло в расчёт. Для конфиденциальных данных достаточно агрегированных чисел и защищённого реестра исключений. Важно, чтобы рецензент мог проверить отсутствие произвольного удаления неблагоприятных случаев.

Ни большой объём, ни известное происхождение набора не гарантируют его пригодности. Основной вопрос остаётся предметным: какие выводы позволяют сделать именно эти данные о именно этой системе и условиях её использования.

Практический разбор 16.11

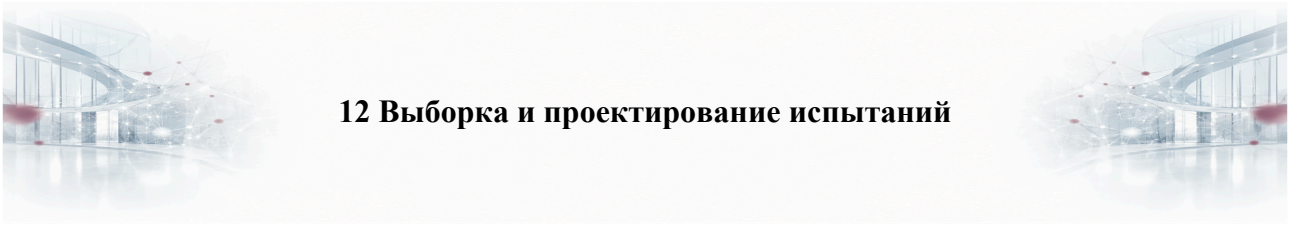
Рекомендация: Проверять эталон на неоднозначных фактах

Условный пример: Два эксперта по-разному определили статус заявления клиента: факт или неподтверждённое сообщение.

Разбор: Расхождение эталона может выглядеть как ошибка модели. Сначала нужно согласовать правило квалификации.

Практические действия: Эксперты размечают материал независимо, разбирают расхождение и проверяют уточнённую инструкцию на новом досье.

Критерий результата: Сохранены исходные метки, причина расхождения и версия инструкции.



12 Выборка и проектирование испытаний

Единица наблюдения и генеральная совокупность

Единицей отбора в предлагаемом юридическом пилоте является досье с одним заранее определённым заданием. Несколько документов внутри досье, несколько утверждений в ответе и повторные генерации относятся к вложенным наблюдениям. Основной показатель рассчитывается по досье, а детализация по утверждениям используется для анализа механизма ошибки. Если досье принадлежат одному делу или построены по одному

шаблону, эта зависимость дополнительно учитывается при отборе и оценке неопределённости.

Рамку отбора составляют допустимые типы задач, языки, объём документов и характер юридического действия. До испытания фиксируются включение, исключение и причина каждого исключения. Для эксплуатационной оценки нужен отбор, соответствующий заявленной совокупности; для поиска уязвимостей используется отдельный набор трудных сценариев. Смешивание этих наборов допускается для описания проведённой работы, но не для единой оценки частоты ошибки в практике.

Решение до отбора	**Пример для юридического помощника**
Единица	Одно досье с одним заранее заданным вопросом
Период	Согласованный интервал формирования документов
Страты	Язык, тип документа и качество распознавания
Исключения	Только заранее обоснованные ограничения назначения
Основной результат	Доля досье с существенной смысловой ошибкой

Если нужен вывод о средней частоте ошибки, случайный или обоснованный вероятностный отбор имеет особое значение. Если цель — найти уязвимость, целевой отбор сложных случаев может быть эффективнее. Эти дизайны допустимы для разных задач; их результаты нельзя смешивать в один показатель без объяснения.

Предлагается сохранять правило отбора и исходный список доступных объектов. Для компьютерной выборки фиксируются используемые параметры и реализация. Это позволяет проверить, не изменялась ли выборка после того, как стали известны ответы системы.

Редкие ошибки и размер испытания

Отсутствие ошибки в небольшой выборке не доказывает её невозможность. При независимых одинаково распределённых испытаниях с постоянной вероятностью ошибки p вероятность не увидеть ни одной ошибки за n попыток равна $(1 - p)^n$. Эта простая модель помогает объяснить, почему редкие события требуют большого числа подходящих наблюдений.

Если ошибок не найдено, односторонняя верхняя 95-процентная граница для p в этой биномиальной модели равна 1 − 0,05^(1/n). Для 100 испытаний она составляет около 2,95%, для 300 — около 0,99%, для 1000 — около 0,30%. Это расчёт неопределённости при указанных предпосылках, а не гарантия безопасности системы. [32]

Число независимых испытаний	**Найдено ошибок**	**Верхняя граница около**
100	0	2,95%

Число независимых испытаний	Найдено ошибок	Верхняя граница около
300	0	0,99%
1000	0	0,30%

В юридическом применении даже малая вероятность может быть существенной из-за тяжести последствий. Поэтому решение не должно основываться только на частоте: дополнительно оцениваются тип ошибки, возможность её обнаружить и механизм предотвращения вредного действия. Отсутствие утечки на тесте не заменяет ограничение прав доступа.

Если наблюдения зависимы, одинаковая арифметика может создавать ложную уверенность. Повторение одного запроса тысячу раз исследует вариативность ответа на этот запрос, но не охватывает тысячу разных ситуаций. Для такой задачи нужен иной план и отдельное объяснение того, что именно измерено.

Сочетание выборки и целевых сценариев

Рабочая программа разделяет эксплуатационный и диагностический наборы. В первом сохраняется обоснованный состав будущих задач; во втором намеренно увеличивается доля противоречий, отсутствующих документов и неоднозначных идентификаторов. По первому набору оценивают качество в заявленной области при выполнении предпосылок отбора. По второму выявляют механизм дефекта и условия его возникновения. Для обоих сохраняют исходные количества, версии и долю неопределённой экспертной разметки.

Целевые сценарии выбираются по предполагаемому пути вреда. Для поиска совпадений это могут быть однофамильцы и различия транслитерации; для обобщения — противоречащие документы и отрицания; для инструмента с действиями — попытка выполнить операцию без предусмотренного подтверждения. Сценарий должен проверять конкретное утверждение, а не просто демонстрировать необычный ответ.

Набор	Основной вопрос	Как представляется результат
Представительный	Как часто возникает заданная ошибка	Оценка с неопределённостью и областью переноса
Целевой	Срабатывает ли конкретная защита	Сценарий, результат и условия воспроизведения
Регрессионный	Сохранилось ли исправление	Сопоставление до и после изменения
Новый контрольный	Переносится ли улучшение на новые случаи	Независимая проверка после настройки

План должен учитывать ресурсное ограничение. Если невозможно проверить все языки и типы документов, выбирают приоритетные группы по риску и явно ограничивают итог. Честное ограничение позволяет принять следующий шаг: расширить испытание, сузить использование или усилить человеческий контроль.

Окончательный объём определяется требуемым выводом и характером риска. Универсальное число тестов для любого ИИ методологически необоснованно; оно должно быть выведено из вопроса, структуры данных и допустимой неопределённости.

Практический разбор 16.12

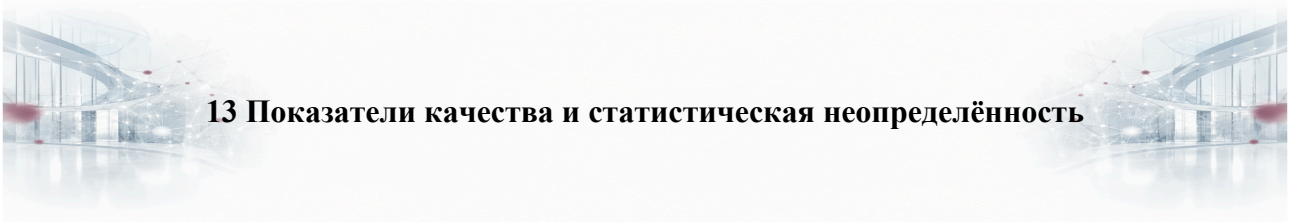
Рекомендация: Разделять выборочную и целевую проверку

Условный пример: В случайной выборке нет ни одного документа с нечитаемой печатью, хотя такие дела встречаются.

Разбор: Средний результат не показывает устойчивость на редком трудном сценарии. Добавление целевых случаев меняет состав набора.

Практические действия: Аудитор проводит отдельный стресс-набор с плохими сканами и сообщает оба результата раздельно.

Критерий результата: Доля ошибок случайной выборки не смешана с долей в стресс-наборе.



13 Показатели качества и статистическая неопределённость

Матрица ошибок

Показатель качества должен соответствовать решению, которое система помогает принять. Общая доля правильных ответов может скрывать неприемлемое число ложных обвинительных совпадений или пропусков значимых событий. Поэтому для бинарного результата сначала составляется матрица ошибок, а затем выбираются показатели, отражающие последствия каждого вида ошибки.

В условном примере проверены 10 000 случаев, из которых 100 действительно относятся к искомой категории. Система верно обнаружила 90, пропустила 10 и ошибочно пометила 99 остальных случаев. Ещё 9801 случай правильно оставлен без отметки. Из 189 сигналов верными являются 90: доля подтверждённых сигналов составляет около 47,62%, хотя общая правильность достигает 98,91%.

Результат системы	Категория есть	Категории нет
Положительный сигнал	90 верных обнаружений	99 ложных сигналов
Отрицательный результат	10 пропусков	9801 верное отклонение

Чувствительность здесь равна 90/100 = 90%. Частота ложноположительных результатов среди отрицательных случаев равна 99/9900 = 1%. Эти показатели имеют разные знаменатели и отвечают на разные вопросы. Называть любой из них просто «точностью» без определения опасно: пользователь может понять число как вероятность истинности конкретного сигнала.

Для ARGA практическое следствие состоит в проверке каждого юридически значимого совпадения по независимым признакам. Высокий общий процент не превращает сигнал в установленный факт. В отчёте приводятся исходные количества, формулы и округление, чтобы расчёт можно было проверить без доступа к программному инструменту автора.

Порог решения и цена ошибки

Изменение порога может уменьшить ложные сигналы и одновременно увеличить число пропусков. В том же условном наборе более строгий порог даёт 80 верных обнаружений, 20 пропусков и 20 ложных сигналов. Доля подтверждённых сигналов возрастает до 80%, но чувствительность снижается до 80%. Выбор лучшего варианта зависит от назначения и допустимых последствий.

Аудитор не должен выбирать порог исключительно по удобству итогового показателя. Нужно установить, кто уполномочен принять компромисс, на каких данных он исследован и какие защитные меры компенсируют остаточный риск. Для предварительного поиска документов и для блокирования операции одинаковый показатель может иметь совершенно разное значение.

Показатель	Исходный порог	Более строгий порог
Верные обнаружения	90	80
Пропуски	10	20
Ложные сигналы	99	20
Доля подтверждённых сигналов	47,62%	80%
Чувствительность	90%	80%

Порог желательно выбирать на данных настройки, а окончательно оценивать на отдельном наборе. Если его оптимизировали непосредственно по финальному тесту, результат может оказаться чрезмерно благоприятным. История выбора порога сохраняется вместе с отвергнутыми вариантами и мотивировкой решения.

В заключении следует избегать безусловного слова «лучше». Более точная формулировка указывает, какой вид ошибки уменьшился, какой вырос и для какого применения такое изменение признано допустимым. Решение о значимом правовом последствии дополнительно проверяется по применимым правилам.

Интервалы и честное представление результата

Оценка по выборке отличается от неизвестного показателя всей совокупности. Доверительный интервал показывает статистическую неопределённость при определённых предпосылках, но не учитывает автоматически ошибку разметки, несоответствие среды и скрытую зависимость наблюдений. Поэтому численный интервал необходимо сопровождать описанием дизайна и существенных ограничений. [32]

До доступа к итоговому набору регистрируются основной исход, способ оценки различия, правила учёта отказов и пропусков, а также число основных сравнений. Существенная ошибка и корректный отказ от ответа кодируются отдельно: исключение отказов из знаменателя может искусственно улучшить оценку. Для серии вариантов модели заранее задаётся порядок сравнения либо корректировка множественных проверок. Исследовательский анализ дополнительных срезов допускается, но обозначается отдельно от основной проверки.

При сравнении конфигураций A и B на одних досье анализ должен сохранять пары. Помимо общих долей ошибок показывают четыре количества: ошибка обеих конфигураций, ошибка только A, ошибка только B, отсутствие ошибки у обеих. Именно несовпадающие пары позволяют оценить направление изменения; их статистическое сравнение связано с подходом Мак-Немара [M4]. Если новые ошибки тяжелее устранённых, благоприятная средняя разность не служит достаточным основанием для допуска.

Элемент публикации	Зачем он нужен
Число наблюдений	Позволяет оценить информационную основу
Определение ошибки	Делает показатель содержательным
Знаменатель	Предотвращает смешение разных долей
Метод неопределённости	Уточняет условия статистического вывода
Исключения и пропуски	Показывает границы фактической проверки

Числа в докладе служат учебными примерами и рассчитываются из явно приведённых исходных данных. Для реального проекта расчёт должен быть независимо перепроверен, а используемый набор и код сохранены в допустимом объёме. Согласованность арифметики необходима, но сама по себе ещё не подтверждает пригодность всего исследования.

Парное сравнение конфигураций на условном наборе досье

Учебный пример включает 200 досье, каждое обработано конфигурациями A и B. Существенная ошибка обнаружена у A в 16 досье, у B — в 6. Этих итогов недостаточно для парного анализа: необходимо знать, на каких именно досье изменился результат. Предположим следующую совместную таблицу.

Результат конфигурации A	У B есть ошибка	У B нет ошибки
Есть ошибка	4	12
Нет ошибки	2	182

На данном наборе доля ошибочных досье снизилась с 8% до 3%, то есть на 5 процентных пунктов. Исправлены 12 исходно ошибочных случаев, но в 2 случаях возникли новые ошибки. Для условного точного парного теста при независимости досье анализируются 14 несовпадающих пар: двустороннее значение p равно $2 \times (1 + 14 + 91) / 2^{14} \approx 0{,}0129$. Это расчёт для заданных учебных предпосылок; он не является вероятностью истинности гипотезы и не доказывает юридическую пригодность системы. [M4]

До решения о внедрении разбирают два ухудшившихся случая. Если в них возникла ранее отсутствовавшая утечка или вымышленное существенное основание, среднее улучшение не закрывает проблему. Если досье подобраны вручную как трудные примеры, 3% характеризуют этот набор, а не реальную частоту ошибок у пользователей. Числа и вывод должны сохранять эту область действия.

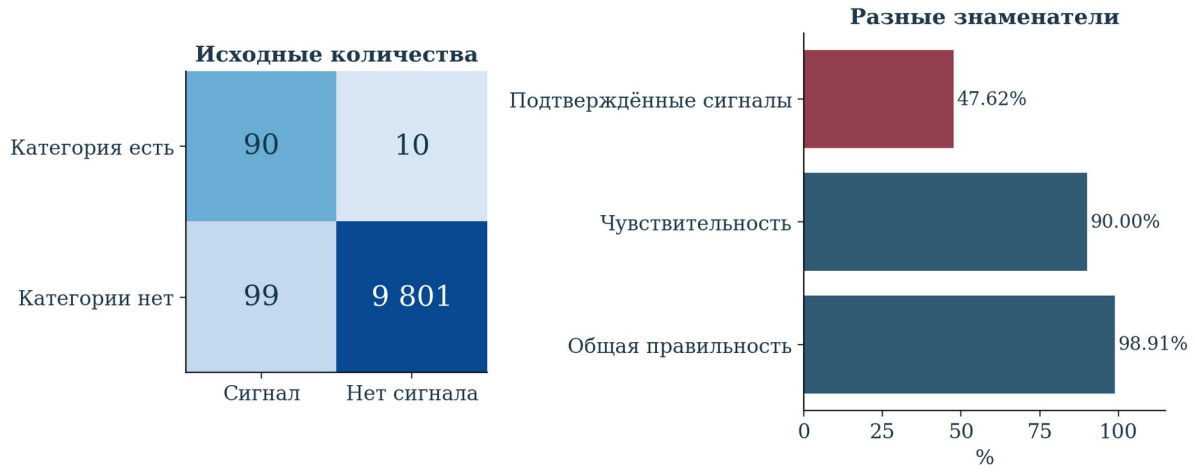


Рисунок 16.2. Почему общая правильность скрывает ложные сигналы. Условные данные раздела 13: 10 000 случаев; 189 сигналов, из них 90 подтверждены. Это учебный расчёт, не результаты ARGA.



Рисунок 16.3. Улучшение результата и новые дефекты. Условный парный пример раздела 13: доля ошибочных досье снизилась с 8% до 3%. Два новых дефекта анализируются отдельно.

Практический разбор 16.13

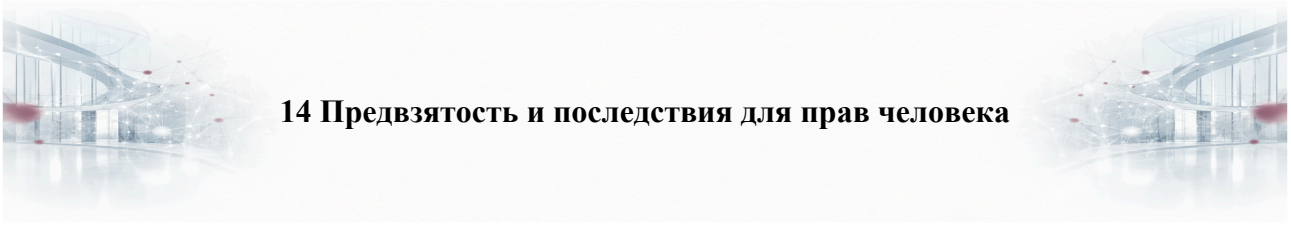
Рекомендация: Показывать исходные количества и новые дефекты

Условный пример: На 100 одинаковых досье версия A ошиблась в 10, версия B в 8; у B появились 3 новых ошибочных досье.

Разбор: Итоговое улучшение на два досье скрывает новые дефекты: пять ошибок устранены, три внесены.

Практические действия: Аналитик составляет парную таблицу и отдельно квалифицирует три новых дефекта по последствиям.

Критерий результата: Решение учитывает новые существенные ошибки, а не только среднюю разницу.



14 Предвзятость и последствия для прав человека

Различие показателей и правового вывода

Разница результатов между группами может указывать на проблему, но не устанавливает автоматически её причину и юридическую квалификацию. Одновременно отсутствие разницы по одному показателю не доказывает отсутствие дискриминации. Необходимо определить затронутое право, сравниваемые ситуации, релевантные признаки и то, как система влияет на фактическое решение.

NIST SP 1270 рассматривает предвзятость шире одного технического дефекта, включая человеческие и организационные источники. Для аудита это полезное напоминание: проблема может возникать при постановке задачи, выборе данных, интерпретации результата или устройстве процедуры. Измерение модели не охватывает весь социальный процесс. [23]

В предлагаемой программе сначала исследуется механизм возможного вреда. Например, система может чаще ошибаться на документах определённого языка, а сотрудники — реже проверять её отрицательные ответы. Итоговое неравенство возникает из сочетания технических и организационных условий. Для исправления важно выявить это сочетание, а не только посчитать разницу.

Уровень анализа	Пример вопроса
Постановка задачи	Не используется ли неподходящий признак успеха
Данные	Представлены ли затронутые группы и ситуации
Модель	Различаются ли ошибки при сопоставимых условиях
Интерфейс	Не направляет ли он пользователя к поспешному выводу
Процедура	Есть ли доступный способ исправить результат

Сбор чувствительных признаков для такого исследования требует отдельной правовой оценки. Благой исследовательский интерес не создаёт автоматического разрешения на любую обработку. Если необходимые данные недоступны, ограничения анализа раскрываются, а допустимые альтернативы рассматриваются с профильным специалистом.

Сопоставление групп

В условном примере из 200 заявлений группы А положительно отобраны 100, а из 100 заявлений группы Б — 30. Доли составляют 50% и 30%, отношение второй к первой равно 0,60. Это описание наблюдаемого различия. Для вывода о причине нужны дополнительные сведения о критериях, сопоставимости заявлений и действиях участников процесса.

Проверка должна учитывать пересечение признаков. Среднее по большой группе может скрыть проблему у небольшой подгруппы. Вместе с тем слишком мелкое деление уменьшает надёжность оценок и повышает риск идентификации людей. Поэтому детализация определяется предметом и правовыми условиями, а не стремлением получить максимальное число таблиц.

Группа в условном примере	Всего заявлений	Положительный отбор	Доля
А	200	100	50%
Б	100	30	30%

Предлагается сохранять данные о неизвестной принадлежности к группе и причинах отсутствия признака. Удаление таких наблюдений может изменить картину, если пропуски распределены неслучайно. В отчёте необходимо указать, какие записи вошли в сравнение и насколько оно охватывает реальную совокупность.

Исправление не должно сводиться к механическому выравниванию одного числа. Оно может потребовать изменения данных, критериев, интерфейса или процедуры обжалования. Аудитор проверяет, действительно ли мера воздействует на выявленный механизм вреда и не создаёт ли новую проблему для другой группы.

Участие затронутых лиц

Люди, на которых влияет система, могут сообщить о последствиях, не видимых из технических метрик. Например, формально доступное обжалование может требовать непонятных документов, а исправление записи — занимать столько времени, что первоначальная ошибка продолжает причинять вред. Поэтому исследование процедуры должно включать реальную доступность защиты.

В предлагаемой методологии используются интервью, анализ обращений и прохождение условного пользовательского маршрута. Эти материалы дополняют испытания и не заменяют статистическую выборку. Порядок участия должен исключать давление на

заявителей и ненужное раскрытие обстоятельств их дел. Информация собирается в объёме, необходимом для проверяемого вопроса.

Аудитор отдельно оценивает последствия для людей, чьи обращения не дошли до официальной жалобы. Отсутствие жалоб может быть связано с незнанием использования ИИ, языковым барьером или недоверием к процедуре. Поэтому один показатель количества жалоб не должен служить подтверждением отсутствия вреда.

Проверяемое свойство защиты	**Возможное свидетельство**
Понятность уведомления	Проверка текста с предполагаемыми пользователями
Доступность обращения	Прохождение маршрута без специальных знаний
Содержательность пересмотра	Запись проверки исходных обстоятельств
Исправление последствий	Изменение результата и связанных записей

Для ARGA связь с защитой прав составляет самостоятельную ценность исследования. Практическая рекомендация должна показывать, как человек сможет добиться проверки конкретной ошибки, а не ограничиваться общим обещанием ответственного применения технологий.

Практический разбор 16.14

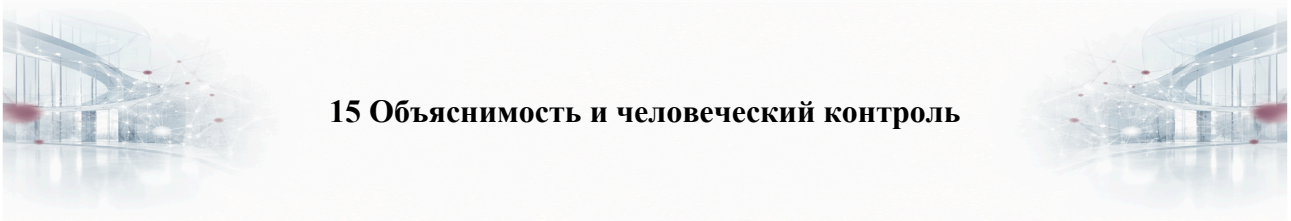
Рекомендация: Проверять различия групп на сопоставимых задачах

Условный пример: Для одного языка используются короткие тексты, для другого длинные сканы; ошибки различаются.

Разбор: Наблюдаемое различие может объясняться качеством входа. Оно не устанавливает причину или правовую дискриминацию.

Практические действия: Исследователь сопоставляет сложность, качество сканов и типы задач; повторяет анализ на сопоставимом поднаборе.

Критерий результата: В отчёте показаны состав групп и устойчивость различия после сопоставления.



15 Объяснимость и человеческий контроль

Какое объяснение нужно адресату

Разработчику, оператору, регулятору и затронутому человеку нужны разные сведения. Техническое описание архитектуры не обязательно помогает понять, почему конкретное решение оказалось неблагоприятным. А краткая пользовательская формулировка может

быть недостаточной для расследования системной ошибки. Поэтому аудит определяет адресата объяснения и проверяет его пригодность для конкретного действия.

В решении по делу Dun & Bradstreet Austria C-203/22 Суд ЕС рассматривал содержательную информацию о логике автоматизированной обработки в контексте GDPR. Существенна возможность понять использованные процедуру и принципы, а не просто получить сложную математическую формулу. Применение этого вывода требует учитывать факты дела и соответствующее право на доступ. [35]

Для добровольной проверки предлагается исследовать объяснение как отдельный результат. Оно должно быть связано с фактически использованными данными и процедурой. Правдоподобный текст, сгенерированный после решения без доступа к его реальным основаниям, может описывать вымышленную причинную историю. Красивое объяснение тогда повышает уверенность пользователя, не повышая проверяемость.

Адресат	**Основная потребность**
Пользователь системы	Понять ограничения и выбрать действие
Затронутое лицо	Установить основания и возможность возражения
Техническая команда	Найти причину и воспроизвести дефект
Уполномоченный орган	Проверить соблюдение требований

Аудитор сопоставляет объяснение с журналами, входными материалами и фактической последовательностью. Если прямая причинная связь не установлена, в заключении нельзя утверждать, что текст достоверно объясняет внутреннее решение модели.

Содержательная проверка человеком

Присутствие кнопки подтверждения не доказывает человеческий контроль. Сотрудник должен иметь время, знания, доступ к исходным материалам и полномочие изменить результат. Если интерфейс показывает только уверенную рекомендацию без основания, формальное согласование может стать механическим этапом.

Предлагается исследовать несколько условных ситуаций: правильный ответ, убедительно сформулированную ошибку, противоречие между источниками и отсутствие достаточных данных. Проверяется, замечает ли сотрудник проблему и что происходит после отказа согласовать ответ. Цель — оценить реальную способность процедуры предотвращать вред, а не склонность сотрудника нажимать нужную кнопку на обучающем примере.

Условие контроля	**Проверочная процедура**
Доступ к источнику	Наблюдение за проверкой значимого утверждения
Полномочие изменить результат	Проверка маршрута отказа и исправления

Условие контроля	Проверочная процедура
Достаточное время	Сопоставление нагрузки и сложности задания
Компетентность	Разбор ошибок без подсказки поставщика
Прослеживаемость	Запись основания окончательного решения

Оценка человека должна учитывать организационные стимулы. Если сотрудников премируют только за скорость, требование глубокой проверки может не выполняться. Аудитор рассматривает инструкции вместе с реальной нагрузкой и системой контроля качества. Нельзя возложить ответственность на пользователя, которому организация не дала условий для содержательной проверки.

Для ARGA полезен ограниченный первоначальный режим: система готовит вспомогательный материал, а значимые фактические и правовые утверждения проверяются по первоисточникам до внешнего использования. Расширение полномочий системы требует отдельного обоснования.

Исправление и обжалование

Процедура исправления должна позволять установить не только факт ошибки, но и её распространение. Неверная запись может попасть в несколько документов, поисковый индекс или последующее решение. Исправление одного видимого ответа тогда не устраняет последствия. Аудитор проверяет, как организация находит связанные материалы и доводит корректировку до необходимых участников.

В предлагаемой программе исследуется полный маршрут: получение обращения, фиксация предмета, проверка исходных данных, принятие решения и подтверждение исполнения. Для каждого этапа должен быть понятен ответственный. Сроки оцениваются по применимому праву и условиям конкретной процедуры; единый произвольный срок для всех случаев не устанавливается.

Если решение подтверждено после пересмотра, пользователь должен получить содержательное объяснение в допустимом объёме. Повторение первоначального вывода без проверки доводов не является убедительным свидетельством эффективного контроля. Для рецензии полезно сохранить, какие новые материалы были рассмотрены и почему они изменили или не изменили результат.

Состояние обращения	Что должно быть видно в записи
Получено	Предмет и способ связи с заявителем
Проверяется	Материалы и поставленные вопросы
Решено	Основание подтверждения либо исправления
Исполнено	Какие записи и действия изменены

Состояние обращения	Что должно быть видно в записи
Закрыто	Сообщение результата и дальнейшие возможности

Отчёт об аудите должен различать наличие процедуры и её эффективность. Одного утверждённого положения недостаточно, если рассмотренные обращения показывают невозможность добиться исправления очевидной ошибки.

Практический разбор 16.15

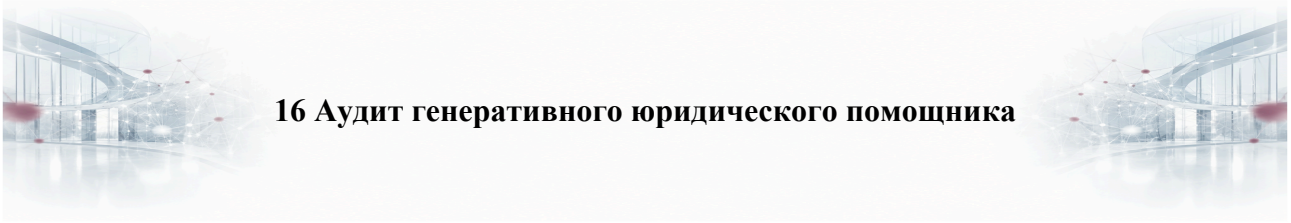
Рекомендация: Проверять возможность изменить результат

Условный пример: Сотрудник подтверждает санкционный флаг, видя только красный индикатор без даты рождения.

Разбор: Формальное нажатие кнопки не обеспечивает содержательную проверку идентичности.

Практические действия: Владелец процесса открывает сотруднику первичные идентификаторы, разрешает отмену флага и проверяет учебный спорный случай.

Критерий результата: Сотрудник мотивированно отменяет ошибочный флаг; решение отражено в журнале.



16 Аудит генеративного юридического помощника

Фактическая и юридическая достоверность

Юридическая достоверность в пилоте определяется по проверяемым компонентам ответа: фактическое утверждение, его документальное основание, статус достоверности, применимая юрисдикция и дата. Существенной признаётся ошибка, способная изменить юридическую оценку или рекомендуемое действие: например, подмена заявления стороны установленным фактом, утрата отрицания, использование неприменимой нормы либо вымышленная ссылка. Критерий тяжести закрепляют до разметки; стилистическая убедительность в него не входит.

Эталон формируют по исходному досье и первичным правовым материалам. Два специалиста независимо отмечают обязательные положения и допустимые варианты вывода; разногласия сохраняют до согласования. Если эксперты не могут обосновать единственную квалификацию, задача получает статус неоднозначной с описанием допустимого диапазона ответов. Модель не штрафуют за отличие от одной формулировки при сохранении фактического и правового смысла.

Вид дефекта	Что проверяет аудитор
Несуществующий источник	Наличие официальной публикации
Неверная ссылка	Соответствие документа названному вопросу

Вид дефекта	Что проверяет аудитор
Потерянное условие	Полный контекст правила и исключений
Устаревшая редакция	Дата и применимая версия
Ложный факт	Происхождение и подтверждение утверждения

Для испытания используются вопросы с проверяемыми ответами и заранее установленными критериями существенности. Оценка не должна поощрять уверенную формулировку при недостатке данных. Иногда правильным поведением является указание неопределённости или запрос необходимого документа.

Итоговый показатель лучше сопровождать разбором критических ошибок. Десять стилистических неточностей и одна ложная ссылка, от которой зависит правовая позиция, имеют разное значение. Простое усреднение способно скрыть главный риск.

Полнота ответа и отказ от ответа

Система может повысить долю правильных ответов, отвечая только на лёгкие вопросы. Поэтому необходимо одновременно измерять правильность среди выданных ответов и охват заданий. В условном наборе из 100 вопросов система ответила на 60, из них 54 верно. Правильность среди ответов составляет 90%, охват — 60%, а доля всех заданий с верным ответом — 54%.

Эти числа не противоречат друг другу. Они описывают разные свойства. Для вспомогательного инструмента осторожный отказ может быть полезен, если пользователь понимает причину и имеет рабочий способ продолжить задачу. Для автоматического процесса отказ без маршрута обработки может создавать задержку или скрытый неблагоприятный результат.

Состояние в условном наборе	Количество	Значение
Верный ответ	54	Задание выполнено в установленных пределах
Ошибочный ответ	6	Требуется анализ вида и последствий ошибки
Отказ или неопределённость	40	Нужен отдельный маршрут обработки

Предлагается фиксировать, какие темы и группы документов чаще приводят к отказу. Общий охват может скрывать систематическую непригодность для определённого языка или

типа досье. Если эти случаи входят в заявленное назначение продукта, ограничение должно быть отражено в решении о внедрении.

Для ARGA полезно требование к интерфейсу: неопределённость должна быть заметна и не оформляться как окончательный правовой вывод. Сотруднику необходимо видеть, что именно не установлено и какой дополнительный источник позволит продолжить работу.

Источник внутри документа и полномочия инструмента

Материал, загруженный для анализа, может содержать текст, пытающийся изменить поведение помощника: например, требование раскрыть другое досье или выполнить внешнее действие. Для аудита важно проверить, отделяет ли система содержание документа от доверенных инструкций. Испытания проводятся в разрешённой изолированной среде и не должны затрагивать реальные чужие материалы.

Программа включает безопасные искусственные сценарии: противоречивое указание в приложении, ложную ссылку, текст от имени администратора и попытку вызвать неразрешённый инструмент. Фиксируется не только ответ, но и совершённые действия. Отказ в финальном тексте не доказывает отсутствие промежуточной передачи данных.

Контроль	Проверяемый результат
Разделение инструкций и данных	Документ не расширяет полномочия системы
Ограничение доступа	Запрос не получает чужое досье
Подтверждение внешнего действия	Отправка не выполняется без предусмотренного решения
Журналирование	Видны существенные вызовы и их результат

Если помощник способен отправлять сообщения или изменять записи, аудит должен охватывать цепочку действия, а не только генерацию текста. Особое значение имеют подтверждение адресата, показ пользователю конкретного результата перед отправкой и возможность остановки. Эти меры являются предложениями для выбранного сценария и настраиваются под реальный риск.

Техническая защита не освобождает от проверки правового основания обработки документов. Допустимость доступа к досье, передачи поставщику и хранения журналов устанавливается отдельно от успешного прохождения тестов.

Практический разбор 16.16

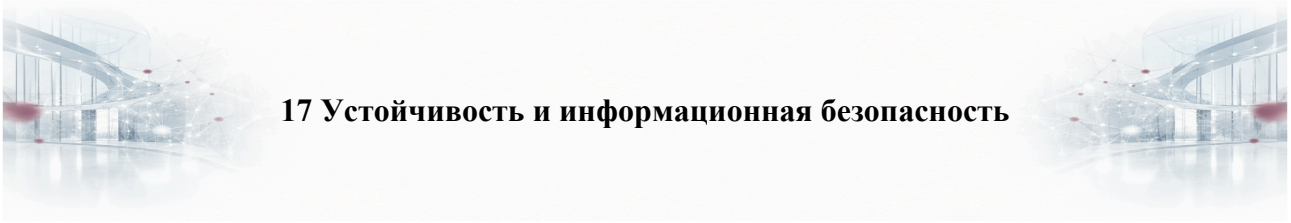
Рекомендация: Проверять факты и полноту независимо

Условный пример: Ответ правильно цитирует решение, но пропускает абзац об исключении из общего правила.

Разбор: Точность цитаты не означает применимость вывода. Пропуск исключения меняет правовую позицию.

Практические действия: Юрист сравнивает ответ с перечнем обязательных условий и исключений; систему повторно испытывают после исправления.

Критерий результата: Ответ сохраняет исключение и указывает соответствующий источник.



17 Устойчивость и информационная безопасность

Модель угроз для ограниченного задания

Аудит безопасности начинается с определения того, какие активы защищаются и кто способен воздействовать на систему. Для юридического процесса это содержание досье, полномочия пользователей, достоверность результата и возможность восстановить действия. Проверка без модели угроз может тратить ресурсы на эффектные, но малозначимые сценарии и пропускать простой путь вреда.

Публикация NIST AI 100-2e2025 систематизирует атаки на машинное обучение и меры противодействия, включая современные риски генеративных систем. Она используется как источник терминологии и ориентир полноты вопросов. Конкретная программа испытаний должна учитывать доступ нарушителя и фактическую архитектуру, а не механически воспроизводить весь каталог. [24]

Актив	Возможное воздействие	Предмет проверки
Конфиденциальные данные	Получение через неверные права	Изоляция и разрешения
Входные материалы	Подмена или изменение	Происхождение и целостность
Решение системы	Манипулирование контекстом	Устойчивость защитной процедуры
Внешнее действие	Использование лишних полномочий	Подтверждение и ограничение функций
Журналы	Удаление следа	Доступ, сохранность и полнота

Предлагается описывать сценарий через условия, действие и ожидаемый защитный результат. Это позволяет повторить проверку после исправления. Общая запись «проведено тестирование безопасности» не показывает, какие пути были рассмотрены и какие остались вне работы.

Детали потенциально опасных испытаний согласуются с владельцем среды и профильными специалистами. Настоящая методология направлена на разрешённую проверку защиты и не создаёт полномочий исследовать чужую инфраструктуру без основания.

Жизненный цикл и поставка изменений

Риск зависит не только от поведения модели, но и от способа разработки, выпуска и обновления. NIST SP 800-218A рассматривает практики безопасной разработки применительно к генеративному ИИ и моделям широкого назначения. Для аудита полезно исследовать, как организация связывает требования безопасности с фактическими изменениями продукта. [25]

Предлагается выбрать несколько существенных изменений и проследить их путь: постановка задачи, проверка кода и конфигурации, испытание, разрешение выпуска и наблюдение после внедрения. Так можно установить, работает ли процесс на практике. Наличие общего регламента не подтверждает, что срочный выпуск проходил те же проверки.

Этап изменения	Необходимое свидетельство
Запрос	Причина и затронутые риски
Подготовка	Версия и ответственный исполнитель
Проверка	Протокол по значимым свойствам
Выпуск	Решение уполномоченного лица
Наблюдение	Данные о поведении после внедрения

Особое внимание уделяется внешним библиотекам, настройкам облачного сервиса и поисковым данным. Изменение такого компонента может повлиять на результат без обычного выпуска приложения. Поэтому карта изменений должна отражать фактическую цепочку зависимости, а не только внутренний репозиторий разработчика.

Для критических функций заранее определяется безопасный вариант восстановления или ограничения работы. Аудитор проверяет его пригодность на допустимом сценарии. Наличие резервной копии не обязательно означает способность быстро вернуть работоспособный и правомерный процесс.

От находки к подтверждённому исправлению

После обнаружения дефекта необходимо отделить устранение конкретного симптома от исправления причины. Изменение ответа на один известный запрос может не решить проблему разграничения доступа. Поэтому повторная проверка включает исходный пример и новые варианты, проверяющие тот же механизм.

Предлагается сохранять идентификатор находки, затронутую версию, возможные последствия, согласованное исправление и результат проверки после изменения. Если исправление меняет другую функцию, её также исследуют в необходимом объёме. Иначе закрытие одной проблемы способно создать новую, не отражённую в отчёте.

Состояние	Что оно означает
Обнаружено	Есть воспроизводимое либо достаточное документальное основание
Анализируется	Причина и область воздействия уточняются
Исправлено поставщиком	Заявлено выполненное изменение
Перепроверено	Аудитор исследовал результат изменения
Закрыто	Выполнен согласованный критерий с указанными ограничениями

Раскрытие информации о дефекте должно учитывать риск для пользователей и действующие обязанности. Публичное резюме может описывать значимость и статус исправления без технических деталей, позволяющих причинить вред. При этом нельзя скрывать неустранённую проблему за общей фразой о непрерывном улучшении.

Вывод аудитора относится к исследованному исправлению и версии. Если после закрытия находки меняются полномочия, модель или среда, прежняя проверка может потерять актуальность. Условия такого пересмотра фиксируются вместе с первоначальным результатом.

Практический разбор 16.17

Рекомендация: Подтверждать исправление повторным испытанием

Условный пример: Поставщик исправил инструкцию после обнаружения утечки тестового идентификатора.

Разбор: Изменение текста настройки не доказывает устранение сбоя и может ухудшить полезную функцию.

Практические действия: Инженер повторяет исходный разрешённый тест и независимые близкие сценарии; проверяет штатную обработку документов.

Критерий результата: Дефект не воспроизводится в проверенном наборе, штатная задача выполняется.



18 Персональные данные и анонимизация

Законность обработки в ходе аудита

Аудит сам является процессом обработки информации. Необходимо определить роли участников, цели, правовое основание, получателей и сроки хранения. Нельзя считать, что передача данных аудитору автоматически разрешена потому, что проверка направлена на

соблюдение закона. Содержание договора и фактическое распределение решений должны быть сопоставлены. [5]

В зависимости от обстоятельств исполнитель может действовать как обработчик, самостоятельный контролёр или в иной предусмотренной законом конструкции. Название стороны в договоре не заменяет анализа фактической роли. Для конкретного проекта правовая квалификация проводится до предоставления доступа, особенно если материалы включают чувствительные сведения и документы доверителей.

Предлагается принцип минимально достаточного доступа: использовать искусственные данные, ограниченные копии или защищённый просмотр, когда они позволяют ответить на вопрос. Если без реальных данных нельзя проверить распределение ошибок, необходимость документируется и сопровождается соответствующими мерами. Экономия времени не является самостоятельным обоснованием передачи полного архива.

Решение до доступа	**Что фиксируется**
Цель	Какой вопрос нельзя разрешить без данных
Роль сторон	Кто определяет цели и средства обработки
Основание	Применимое правовое и договорное условие
Объём	Поля, документы и круг получателей
Завершение	Возврат, удаление или обоснованное хранение

Оценка воздействия на защиту данных и независимый аудит могут использовать общие материалы, но не являются взаимозаменяемыми процедурами. Выполненный аудит не отменяет обязанности провести необходимую оценку воздействия, если её условия установлены применимым правом.

Проверка заявления об анонимности

Удаление имени не означает автоматической анонимности. Идентификация может стать возможной через сочетание дат, места, профессии и внешних сведений. Проверка должна рассматривать реальный контекст доступа и вероятные способы связывания информации. Псевдонимизация и анонимизация имеют разные последствия и не должны смешиваться в документах проекта.

Проект руководства EDPB 02/2026 предлагает рассматривать анонимность с учётом перспективы соответствующего участника и средств, которые с разумной вероятностью могут использоваться для идентификации. На дату доклада это версия для публичного обсуждения, а не окончательное руководство. Она используется для постановки исследовательских вопросов и подготовки предложений. [28; 31]

В предлагаемой проверке фиксируются предполагаемые получатели, доступные им дополнительные сведения, технические возможности и срок актуальности оценки.

Испытание проводится только законными и согласованными способами. Отсутствие успешного связывания в одном ограниченном тесте не доказывает невозможность идентификации всеми потенциальными участниками.

Вопрос	Свидетельство для оценки
Какие признаки остались	Описание полей и редких сочетаний
Кто получает данные	Модель доступа и ограничения использования
Какие сведения можно сопоставить	Обоснованный перечень дополнительных источников
Что проверялось	Разрешённые сценарии и результаты
Когда требуется пересмотр	Новые данные, технологии или круг получателей

Договорный запрет идентификации может быть частью защиты, но не заменяет технический и правовой анализ сам по себе. В заключении необходимо объяснить, какие свойства данных установлены и какие предположения о контексте остаются существенными.

Журналы и трансграничный доступ

Для воспроизводимости часто нужны журналы запросов и ответов, однако они могут содержать наиболее чувствительную часть работы. Поэтому требуется определить, какие сведения действительно необходимы, кому доступны и как долго сохраняются. Полное бессрочное журналирование всех досье не является нейтральным техническим решением.

Предлагается разделить сведения о выполнении процедуры и содержание документа. Иногда достаточно сохранить идентификаторы, параметры, результат проверки и защищённую ссылку на исходный материал. В других случаях для разбора ошибки необходим точный текст. Выбор обосновывается доказательственной потребностью и правовыми ограничениями, а не единым правилом «сохранять всё».

Если к журналам обращается внешний поставщик или специалист из другой страны, отдельно исследуются правила международной передачи и фактические получатели. Шифрование снижает определённые технические риски, но само по себе не устанавливает законность передачи. Обещание хранить данные в одном регионе также требует проверки доступа и субподрядчиков. [5]

Материал	Возможная организация доступа
Общий протокол	Доступ команды проекта
Чувствительное приложение	Ограниченный круг уполномоченных лиц
Технический журнал	Сокращённые поля и контроль выгрузки

Материал	Возможная организация доступа
Публичное резюме	Исключение персональных и опасных технических деталей

При завершении проекта проверяется исполнение согласованного порядка хранения и удаления. Удаление копий не должно уничтожать обязательные доказательства, а сохранение рабочего дела не должно превращаться в неограниченный архив чужих данных. Решение принимается по конкретным основаниям и отражается в документе закрытия задания.

Практический разбор 16.18

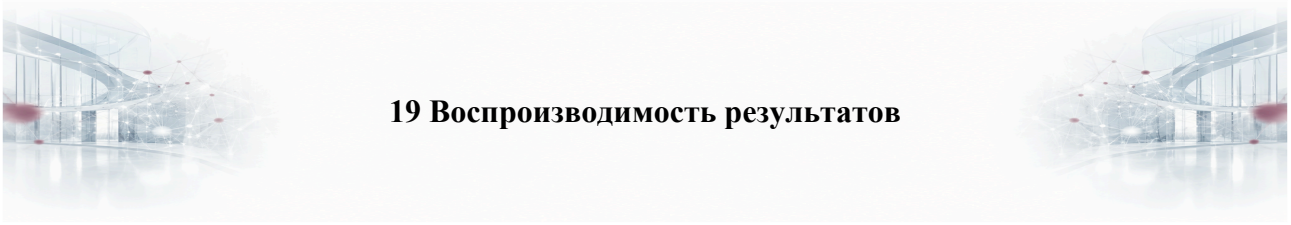
Рекомендация: Минимизировать данные без разрушения теста

Условный пример: Из копии для аудитора удалили все имена, хотя проверяется различение однофамильцев.

Разбор: Обезличивание уничтожило признак исследуемой ошибки. Нужен безопасный эквивалент с сохранением структуры совпадений.

Практические действия: Ответственный готовит условные согласованные идентификаторы и отдельно ограничивает доступ к исходнику.

Критерий результата: Тест сохраняет однофамильцев и различающиеся даты без раскрытия реальных лиц.



19 Воспроизводимость результатов

Что другой специалист должен суметь повторить

Воспроизводимость проверяют на трёх уровнях. Другой исследователь должен восстановить состав входных данных и конфигурацию, повторить обработку сохранённых ответов и получить те же основные количества. Повторный запрос к облачной модели дополнительно проверяет устойчивость поведения, но не заменяет воспроизводимость самого расчёта. Различие между этими процедурами обязательно отражается в отчёте: сохранённый ответ можно перепроверить даже тогда, когда прежняя модель уже недоступна.

Предлагается сохранять входные материалы в допустимом объёме, версии используемых компонентов, параметры, инструкции, последовательность действий и необработанные результаты. Итоговая таблица без промежуточных данных не позволяет проверить, как обрабатывались отказы и ошибки. Скриншот успешного ответа не заменяет полный протокол испытания.

Уровень повторения	Что проверяется	Ограничение
Повторный расчёт	Формулы и итоговые показатели	Не подтверждает получение исходных данных
Повторная обработка	Правила подготовки и оценки	Зависит от сохранности исходного набора
Повторный запуск	Поведение системы в заданной среде	Возможна недоступность прежней версии
Независимое исследование	Устойчивость вывода на новых данных	Требует сопоставимости назначения

Официальная документация инструмента Inspect показывает пример структурированного сохранения журналов оценивания. Наличие такого инструмента помогает организовать записи, но не гарантирует корректность самих заданий и критериев. Выбор программного средства не должен подменять методологическую проверку. [33]

Для ARGA полезен минимальный воспроизводимый пакет: описание вопроса, паспорт объекта, реестр данных, инструкция испытания, результаты и расчёт. Если техническое повторение невозможно, другой специалист всё равно должен видеть, на чём основано заключение и где заканчивается проверенное знание.

Вариативность генерации

Генеративная система может выдавать разные ответы при одинаковом видимом запросе. На результат влияют параметры генерации, скрытые обновления, внешние источники и последовательность взаимодействия. Поэтому один удачный запуск не подтверждает устойчивость поведения, а один неудачный запуск не всегда позволяет оценить частоту ошибки.

Предлагается заранее определить число повторов для значимых сценариев и сохранять все ответы. Нельзя выбирать наиболее убедительный вариант после испытания и представлять его как единственный результат. Если целью является демонстрация конкретной возможности, это нужно назвать; если оценка устойчивости — необходим другой дизайн.

Важен учёт общей истории диалога. Ответ после нескольких уточнений не сопоставим с первым ответом другой системы, если задание предполагает одинаковые условия. Для каждого запуска фиксируются доступный контекст, используемые инструменты и вмешательство пользователя. Удаление неудачной попытки с последующим перезапуском должно оставлять запись.

Фактор вариативности	Предлагаемый контроль
Параметры генерации	Зафиксировать доступные настройки

Фактор вариативности	Предлагаемый контроль
История взаимодействия	Сохранить существенный контекст
Внешний поиск	Указать найденные материалы и время
Обновление сервиса	Сохранить доступный идентификатор версии
Повторные попытки	Учесть все запуски по принятому правилу

Даже при фиксированных настройках нельзя обещать полное совпадение, если поставщик не гарантирует соответствующее поведение инфраструктуры. Правильный вывод описывает наблюдаемую устойчивость и условия её оценки. При существенной неопределённости решение о применении должно учитывать возможность другого ответа в будущем.

Автоматическая оценка ответов

Автоматический оценщик рассматривается как измерительный инструмент с собственной областью пригодности. Его проверяют на экспертно размеченных ответах, включающих убедительно сформулированные ошибки, корректные отказы и неоднозначные случаи. Приоритет имеет частота пропуска существенного дефекта, а не среднее совпадение общего балла. Порядок ответов, длина текста и упоминание названия модели не должны становиться скрытыми основаниями оценки; соответствующие эффекты проверяются на контрольных парах.

Для подтверждающей оценки заранее фиксируются версия оценщика, инструкция, шкала и правила передачи спорного случая человеку. Настройку проводят на отдельной части материала. Итоговый набор не используется для подбора удачного промпта оценщика. При систематическом пропуске критического вида ошибки автоматическая оценка сохраняет вспомогательную роль, а окончательная классификация этого вида передаётся профильному специалисту.

Часть результатов должна проходить человеческую перепроверку по заранее определённому правилу. Все критические находки и значимые расхождения требуют отдельного разбора. Если модель-оценщик участвует в подготовке окончательного мнения, в рабочем деле сохраняются её версия, инструкция и способ проверки её пригодности.

Проверка оценщика	Почему она нужна
Сопоставление с экспертной разметкой	Оценить содержательную пригодность
Проверка одинаковых ответов в разном порядке	Обнаружить влияние порядка
Безопасные вводящие в заблуждение инструкции	Проверить устойчивость процедуры оценки

Проверка оценщика	Почему она нужна
Анализ расхождений	Найти категории систематической ошибки

Экономия времени оправдана, когда качество оценки достаточно для выбранной задачи и ограничения известны. Если автоматический балл невозможно объяснить или проверить, он не должен становиться единственным основанием юридически значимого заключения.

Практический разбор 16.19

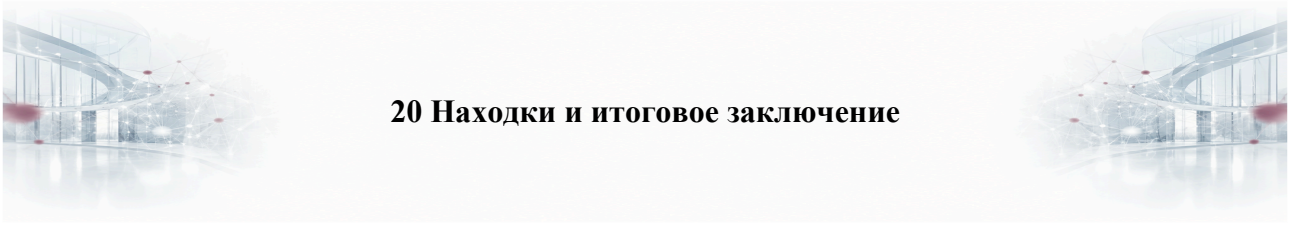
Рекомендация: Передавать воспроизводимый пакет запуска

Условный пример: Рецензент получает только итоговый процент качества без ответов модели и правил разметки.

Разбор: Он не может проверить расчёт, критерий или полноту учтённых запусков.

Практические действия: Исполнитель передаёт версии, задания, ответы, метки и расчёт; второй аналитик воспроизводит показатель.

Критерий результата: Независимый пересчёт совпадает, расхождения объяснены в журнале.



20 Находки и итоговое заключение

Структура отдельной находки

Находка должна связывать критерий, наблюдение и последствие. Фраза «система недостаточно прозрачна» не позволяет понять, что исправлять. Полезнее указать: для определённого класса ответов невозможно установить использованный документ, вследствие чего проверяющий сотрудник не может подтвердить значимое утверждение до внешнего использования.

Предлагается фиксировать идентификатор, объект, критерий, материал, наблюдаемое расхождение, последствия и необходимые действия. Рекомендация должна быть отделена от установленного факта. Если причина ещё не известна, её нельзя представлять как доказанную только потому, что она кажется наиболее вероятной.

Элемент находки	Условный пример
Критерий	Существенные утверждения связаны с исходным документом
Наблюдение	В 6 из 60 ответов указанный фрагмент не поддерживает вывод
Доказательство	Протокол, ответы и сопоставленные документы
Значение	Риск переноса неподтверждённого утверждения в досье

Элемент находки	Условный пример
Предлагаемое действие	Изменить проверку ссылок и повторить испытание

В этом примере число характеризует только исследованный набор. Без подходящего отбора его нельзя объявлять частотой ошибки всей эксплуатации. Если часть ответов содержит разные виды дефекта, они описываются отдельно, поскольку способы исправления могут различаться.

Находка проходит обсуждение с проверяемой стороной и рецензию. Исправление формулировки допустимо, если оно делает вывод точнее. Смягчение значимости ради удобства заказчика без новых доказательств несовместимо с заявленной независимостью.

Значимость и приоритет устранения

Значимость определяется последствиями, вероятностью, масштабом воздействия и возможностью обнаружения до вредного действия. Предлагается использовать описательные категории с определёнными критериями, а не произвольный общий балл. Для правового запрета или утечки чувствительных данных средняя хорошая результативность других функций не является компенсацией.

Возможна четырёхуровневая рабочая классификация: критическая проблема препятствует выбранному применению; существенная требует исправления или доказанных компенсирующих мер; ограниченная влияет на отдельный процесс; наблюдение указывает на улучшение без установленного значимого дефекта. Это проектная методика, а не универсальная нормативная шкала.

Категория	Предлагаемое управленческое действие
Критическая	Ограничить соответствующее применение до разрешения вопроса
Существенная	Установить исправление и проверку его эффективности
Ограниченная	Включить в контролируемый план улучшения
Наблюдение	Рассмотреть целесообразность изменения

Категория должна сопровождаться объяснением. Одинаковый технический дефект может иметь разную значимость в учебной среде и при подготовке документа, от которого зависит судьба человека. Аудитор учитывает контекст, но не должен снижать категорию только потому, что до проверки ещё не поступила жалоба.

Если факт установлен, а масштаб неизвестен, эти состояния отражаются раздельно. Дополнительная проверка масштаба может быть срочным действием сама по себе. Неопределённость не является основанием автоматически считать проблему незначительной или, наоборот, утверждать максимальный возможный вред как уже произошедший.

Формулировка мнения и ограничения

Итоговое мнение должно содержать исследованный объект, проверяемое утверждение, процедуру, результат и границу переноса. Для каждой существенной характеристики указывается один из обоснованных исходов: критерий выполнен; критерий не выполнен; сведений недостаточно; вопрос не входил в задание. Недостаточность данных не подменяет установленный дефект и одновременно не допускает положительного заверения. К заключению привязывается действие: ограничить применение, исправить конкретный механизм, получить недостающие данные либо продолжить работу в указанной области.

Ограничение должно быть содержательным. Вместо общего «аудитор не гарантирует отсутствие ошибок» следует указать, что не исследовалось и как это влияет на вывод. Например, тест выполнен на одной языковой группе, доступ к производственным журналам отсутствовал либо версия облачной модели не могла быть независимо зафиксирована.

Отчёт должен позволять отличить отсутствие существенных выявленных недостатков от доказанного соответствия всем заявленным критериям. Эти выводы могут требовать разных процедур и объёма доказательств. Термины официальной схемы или стандарта применяются только при фактическом соблюдении соответствующей процедуры и наличии необходимых оснований.

Часть отчёта	Обязательное содержание предлагаемой формы
Предмет	Объект, версия, назначение и период
Основание	Критерии и их правовой или договорный статус
Работа	Методы, данные, доступ и существенные ограничения
Результат	Подтверждённое и выявленные расхождения
Последствия	Исправления, повторная проверка и условия использования

Краткое резюме должно сохранять смысл полного документа. Если в полном отчёте есть критическое ограничение, реклама не может сообщать, что система полностью проверена и безопасна. Условия допустимого цитирования заключения согласуются с этой логикой.

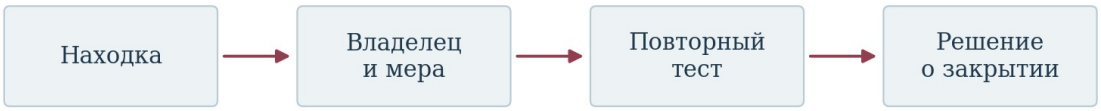


Рисунок 16.4. Находка и подтверждённое исправление. Каждая мера связывается с владельцем и повторной проверкой. Заявление об исправлении не заменяет испытание.

Практический разбор 16.20

Рекомендация: Связывать находку с исполнимой мерой

Условный пример: В десяти ответах ссылка ведёт на другой документ с похожим названием.

Разбор: Общее замечание о плохом качестве не показывает механизм и не задаёт проверку исправления.

Практические действия: Аудитор описывает примеры подмены, назначает проверку идентификатора документа и повторный тест на близких названиях.

Критерий результата: У находки есть владелец, срок и подтверждающий повторный результат.

21 Мониторинг и пересмотр результата

Аудит как оценка определённого состояния

Заключение описывает исследованное состояние системы. После его выдачи могут измениться модель, данные, назначение, пользователи или правовые требования. Поэтому в программе нужно определить события, при которых прежний вывод требует пересмотра. Календарная повторная проверка полезна, но не заменяет реагирование на существенное изменение между установленными датами.

Предлагается включить в условия актуальности идентификатор объекта, разрешённое назначение, существенные зависимости и контроль изменений. Например, переход от подготовки черновика к автоматической отправке существенно меняет последствия ошибки даже при неизменной модели. Такой переход не должен прикрываться прежним положительным отчётом.

Событие	Причина пересмотра
Новая модель или настройки	Возможно изменение поведения
Новый язык или тип документов	Прежняя выборка может быть непригодна
Расширение полномочий	Возрастает возможный вред
Изменение источников	Меняется информационная основа
Новое правовое требование	Нужно пересмотреть критерии
Существенная жалоба	Возможно наличие ранее неучтённого сценария

Решение о повторной проверке должно быть пропорциональным. Иногда достаточно ограниченного испытания затронутой функции, иногда требуется новая программа. Важно документировать основание, а не автоматически начинать весь аудит заново или считать любое изменение несущественным.

Для ARGA такая практика позволяет сохранять аналитическую ценность заключений: пользователь видит не только дату выдачи, но и условия, при которых результаты ещё можно применять.

Показатели наблюдения

Мониторинг должен отражать реальные пути ошибки. Общая доступность сервиса не показывает достоверность юридических выводов, а среднее время ответа — качество человеческой проверки. Поэтому показатели выбираются по установленным рискам и связываются с действиями при отклонении.

Для юридического помощника могут быть полезны доля неподтверждённых ссылок, случаи утраты существенного условия, число исправленных человеком ответов и обращения по разглашению данных. Эти показатели требуют определений и надёжного сбора. Рост числа зарегистрированных ошибок иногда отражает улучшение обнаружения, а не ухудшение системы.

Показатель	Что нужно дополнительно понимать
Исправления пользователем	Какие изменения существенны и почему
Отказы системы	Какие задачи и группы затронуты
Жалобы	Доступен ли канал и какова осведомлённость людей
Ошибки источников	Обнаруживаются ли они до внешнего использования
Изменения конфигурации	Проверены ли затронутые свойства

Предлагается сочетать количественные данные с разбором отдельных значимых случаев. Один серьёзный инцидент может раскрыть дефект, который теряется в большой общей статистике. В то же время единичный необычный ответ не должен автоматически объявляться массовым явлением без исследования масштаба.

У каждого сигнала должен быть ответственный и предусмотренное действие: проверка, ограничение применения, запрос поставщику или повторное испытание. Система показателей без такого маршрута создаёт отчётность, но не обеспечивает управление риском.

Инцидент и сохранение возможности расследования

При серьёзном событии необходимо сохранить достаточные сведения о версии, входных данных, действиях и последствиях. Непродуманная очистка журналов или немедленная замена конфигурации может затруднить установление причины. Одновременно сохранение доказательств не должно необоснованно продолжать опасное поведение системы.

Предлагается заранее определить порядок согласованной фиксации: кто принимает решение, какие материалы сохраняются, где ограничивается доступ и как отмечаются последующие изменения. Если техническая команда вынуждена срочно исправить систему, фиксируются доступные сведения до изменения и его точное содержание.

Вопрос первичной записи	Практическое значение
Что произошло	Отделить наблюдаемый факт от предположения
Какая версия работала	Связать событие с конкретным состоянием
Кто затронут	Оценить необходимые защитные действия
Какие материалы сохранены	Обеспечить последующую проверку
Что изменено после события	Избежать смешения исходной и исправленной системы

Обязанности уведомления зависят от применимого режима и характера события. Аудитор не должен автоматически переносить срок из одного закона на другой вид инцидента. Для конкретного случая юрист определяет основания, адресатов и сроки, а техническая команда предоставляет необходимые факты.

Результаты расследования используются для пересмотра программы аудита. Если причина показывает новый путь вреда, его включают в будущие сценарии и проверяют на других релевантных применениях. Простое закрытие одного обращения не всегда означает устранение системного риска.

Практический разбор 16.21

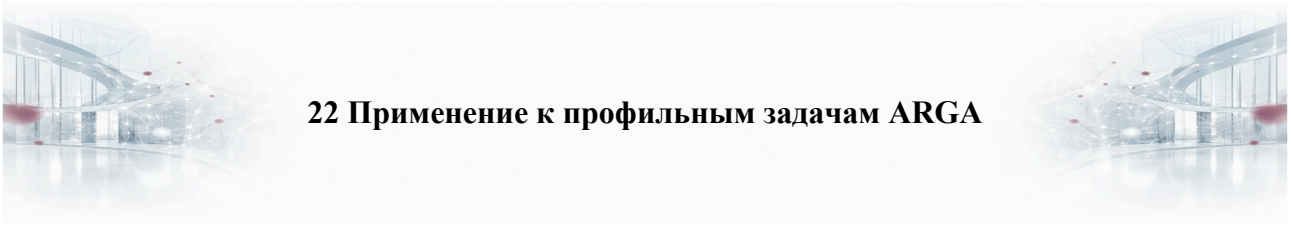
Рекомендация: Задавать триггеры повторной оценки

Условный пример: После аудита помощнику разрешили автоматически создавать исходящие письма.

Разбор: Область последствий расширилась; прежний тест чтения не покрывал отправку.

Практические действия: Владелец приостанавливает новое полномочие до проверки адресата, подтверждения человеком и журналирования отправки.

Критерий результата: Новое действие разрешено отдельным решением на основании отдельного теста.



22 Применение к профильным задачам ARGA

Поиск совпадений и работа с именами

При поиске совпадений по имени задача аудита состоит в проверке того, как сигнал превращается в вывод о конкретном лице. Совпадение строк не устанавливает тождество людей. Значение имеют дополнительные признаки, качество источника, варианты написания и возможность исправления ошибочной связи.

Условная программа включает однофамильцев, различные порядки имени и фамилии, транслитерацию, отсутствующую дату рождения и устаревшие сведения. Отдельно исследуются ложные совпадения и пропуски. Нельзя оценивать систему только на заранее известных положительных примерах, если в реальном потоке большинство лиц не относится к искомой категории.

Этап	Что проверяется
Получение исходных сведений	Официальность, актуальность и принадлежность записи
Сопоставление	Правила работы с неоднозначными признаками
Представление сигнала	Отделение возможного совпадения от установленного факта
Человеческая проверка	Доступ к дополнительным идентификаторам
Исправление	Удаление ошибочной связи и её последствий

Для ARGA полезен формат результата, в котором у каждого совпадения указаны подтверждающие и противоречащие признаки. Итоговая уверенность модели не должна заменять это объяснение. Если доступных идентификаторов недостаточно, правильным состоянием может быть неопределённость, а не принудительный положительный или отрицательный вывод.

Данный сценарий является методическим примером. Он не описывает фактическую архитектуру INTERPOL, банка или иной конкретной организации и не утверждает, что определённая система уже была проверена ARGA.

Финансовая аналитика и корпоративные связи

В анализе финансовых и корпоративных данных система может обнаруживать связи между лицами, организациями и операциями. Аудит должен отличать найденную связь от доказанного контроля, нарушения или преступного намерения. Общий адрес, совпадающее имя или последовательность переводов могут иметь несколько объяснений.

Предлагается проверять происхождение каждого ребра аналитической схемы: из какого документа оно получено, на какую дату относится и какую именно связь обозначает. Владение, управление, представительство и историческое участие нельзя обозначать одним символом без пояснения. Ошибка в типе связи способна изменить юридический вывод даже при правильном извлечении имени.

Вывод системы	**Необходимая проверка**
Лицо связано с компанией	Вид, дата и источник связи
Установлен контроль	Применимое понятие и достаточность фактов
Операция необычна	Контекст и критерий сравнения
Предполагается обход ограничения	Альтернативные объяснения и правовое основание
Риск признан высоким	Значение категории и последующее действие

Проверка должна учитывать изменения во времени. Реестр на текущую дату может не подтверждать структуру владения в момент операции. Если система объединяет разные периоды, аудитор исследует, не создаётся ли искусственная цепочка отношений, никогда не существовавшая одновременно.

Для ARGA практический результат состоит в проверяемой карте фактов и неопределённостей. Алгоритмический сигнал может направлять дальнейший поиск, но обвинительный вывод требует самостоятельного доказательственного и правового анализа.

Обобщение материалов и защита процессуальной позиции

В юридическом досье важно сохранить различие между показанием, утверждением стороны, содержанием документа и установленным обстоятельством. Генеративное обобщение может незаметно стереть эти границы. Поэтому аудит проверяет не только наличие фактов в исходном тексте, но и сохранение их доказательственного статуса.

Условная программа включает противоречивые объяснения, отрицания, оговорки о неполноте сведений и документы разных дат. Проверяется, не превращает ли система вероятность в достоверность, отсутствие подтверждения в опровержение или чужую позицию в собственный установленный вывод. Для таких ошибок обычный подсчёт совпадающих слов малоинформативен.

Исходное состояние	Недопустимое усиление в обобщении
Сторона утверждает факт	Факт установлен
Документ не представлен	Документ не существует
Связь требует проверки	Контроль доказан
Данные относятся к прошлому периоду	Состояние продолжается сейчас
Имеются противоречия	Противоречия исключены без основания

Для оценки используются правила смысловой разметки и профильный рецензент. Система может помогать подготовить структуру и извлечь фрагменты, но каждое значимое обобщение должно оставаться связанным с исходным материалом. Если источник недоступен, это ограничение переносится в вывод.

Предлагаемый первоначальный продукт ARGA — ограниченная проверка такого помощника на искусственных досье. Она позволит отработать процедуру без необоснованной передачи реальных материалов доверителей и даст основу для решения, нужна ли дальнейшая техническая экспертиза.

Практический разбор 16.22

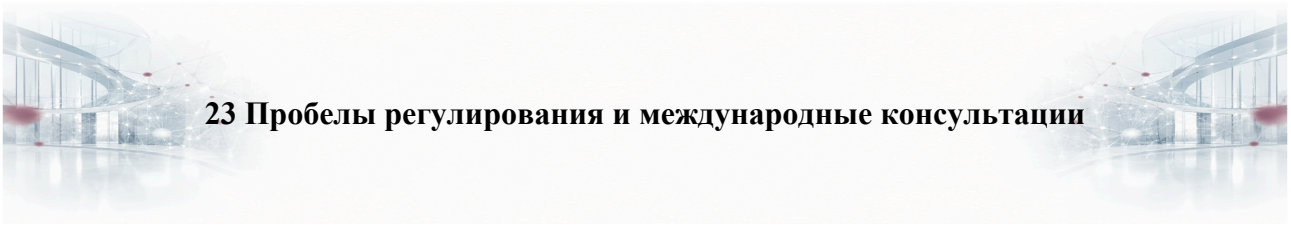
Рекомендация: Сохранять доказательственный статус сведений

Условный пример: Помощник превратил фразу клиента о предполагаемом давлении в установленный факт.

Разбор: Обобщение изменило силу утверждения и способно исказить обращение.

Практические действия: Юрист вводит метки факта, сообщения и гипотезы; проверяет хронологию по каждому источнику.

Критерий результата: В проекте обращения предположение не выдано за подтверждённое событие.



23 Пробелы регулирования и международные консультации

Где существующие механизмы оставляют неопределённость

Сравнение показывает несколько устойчивых проблем. Разные документы называют аудитом проверки с различным предметом. Заказчику трудно понять, относится ли результат к организации, модели, конкретному применению или узкой группе показателей. Общий знак качества способен создать более широкое впечатление, чем позволяет фактически выполненная работа.

Вторая проблема — ограниченный доступ к материалам поставщика. Независимый исполнитель может иметь достаточную квалификацию, но не получать сведения, без которых нельзя проверить значимое утверждение. Третья — зависимость от коммерческого заказчика и возможность контролировать публикацию. Четвёртая — изменение системы после проверки без ясного механизма пересмотра.

Пробел	**Предлагаемое направление решения**
Неясный предмет заключения	Обязательное краткое описание объекта и критериев
Непрозрачные ограничения	Указание недоступных данных рядом с выводом
Конфликт интересов	Раскрытие существенных связей и проверка защитных мер
Непереносимые результаты	Паспорт версии, среды и допустимого назначения
Разрыв после выдачи отчёта	Условия пересмотра и контроль исправлений

Предлагаемые меры не требуют объявлять любой аудит полноценной сертификацией. Их ценность состоит в минимальной сопоставимости информации: пользователь должен понимать, что проверено и почему результат заслуживает определённой степени доверия. Более строгие требования могут устанавливаться для конкретных сфер и последствий.

Риск чрезмерной формализации также существенен. Большой пакет документов может быть недоступен небольшим организациям и при этом мало говорить о реальном поведении системы. Поэтому предложения ARGA связывают каждое требование с доказательственной пользой и допускают пропорциональный объём работы без сокрытия ограничений.

Открытые процедуры EDPB

На 20 октября 2026 года официальные страницы EDPB указывают срок обсуждения проектов руководств 03/2026 о сборе веб-данных для генеративного ИИ и 02/2026 об анонимизации: до 30 октября 2026 года, 23:59 CET. Это две отдельные процедуры с разным предметом; подготовленные предложения должны быть адресными. [27–28]

Для обсуждения сбора данных ARGA может предложить минимальную структуру доказательств происхождения: источник, дата получения, назначение, применимое основание, существенные ограничения и история последующей обработки. Независимая проверка должна показывать, какие утверждения можно подтвердить по выборке, а какие требуют более полного доступа. Это предложение настоящего исследования, а не пересказ обязательного перечня проекта.

Для обсуждения анонимизации полезна позиция о воспроизводимости оценки: описание получателей, дополнительных данных, допустимых сценариев связывания и условий пересмотра. Публичный результат должен сохранять сведения о существенных ограничениях без раскрытия материалов, способных усилить риск идентификации. Проект руководства рассматривается именно как предмет обсуждения. [31]

Процедура	Предлагаемый вклад ARGA
EDPB 03/2026	Происхождение данных и проверяемость заявлений поставщика
EDPB 02/2026	Доказательства анонимности и условия повторной оценки
Общее методическое предложение	Независимость исполнителя и раскрытие пределов проверки

Перед подачей необходимо повторно проверить доступную форму, условия публикации и правила для представителя организации. Вклад согласуется в установленном внутреннем порядке. Подготовка доклада не подтверждает участия ARGA в процедурах; отправка и получение фиксируются отдельными документами. [36]

Другие адресаты и приоритеты позиции

Консультация Европейской комиссии о влиянии технологий на авторское право открыта до 3 ноября 2026 года. Для неё релевантны предложения о проверяемости происхождения обучающих материалов и разграничении публичного резюме с конфиденциальным доказательственным пакетом. Официальная публикация прямо называет неправительственные организации среди предполагаемых участников. [29]

Британский ICO проводит сбор материалов об агентном ИИ до 20 ноября 2026 года. Выводы настоящего доклада можно использовать для предложений об аудите полномочий инструментов, журналировании действий и эффективном человеческом подтверждении. Это национальная процедура зарубежного регулятора, которую следует отличать от наднациональной консультации ЕС. [26]

Адресат	Предмет предложения	Срок по официальной странице
Европейская комиссия	Проверяемость использования охраняемых материалов	3 ноября 2026 года
ICO	Ответственность и контроль действий агентного ИИ	20 ноября 2026 года

Предлагается подготовить короткие адресные позиции, а не направлять весь доклад одинаковым текстом каждому органу. Для каждой позиции выбираются конкретный вопрос, предлагаемое решение и объяснение доказательственной пользы. Дополнительные приложения включаются только тогда, когда они помогают адресату оценить предложение.

Приоритетом являются процедуры EDPB, прямо связанные с независимой проверкой данных и близким сроком. Это исследовательская рекомендация, а не утверждённое обязательство

ARGA подать материалы. Фактическая подача, подтверждение получения и возможная публикация учитываются отдельно; подтверждение участия не означает одобрения позиции регулятором.

Практический разбор 16.23

Рекомендация: Формулировать адресное предложение регулятору

Условный пример: В консультационный ответ включили просьбу повысить прозрачность без описания проверяемой записи.

Разбор: Адресат не получает конкретного изменения процедуры и не может оценить реализуемость.

Практические действия: Автор предлагает хранить версию, источник и основание существенного вывода; добавляет условный пример восстановления ошибки.

Критерий результата: Каждое предложение называет требуемую запись, её назначение и ограничение раскрытия.



24 Практическая программа ARGA и итоговые выводы

Ограниченный пилотный проект

Первый пилот предлагается посвятить обобщению юридического досье с обязательным указанием источников и сохранением доказательственного статуса сведений. Сравниваются две точно зафиксированные конфигурации, например исходная и версия с проверкой цитат. Задачи для разработки инструкции отделяются от итогового набора. Пилот проверяет пригодность классификатора ошибок, полноту журнала и воспроизводимость расчёта; выбор победившей модели не является его единственной целью.

До запуска утверждаются исследовательский вопрос, основной исход и практически значимое улучшение, допустимые новые дефекты, единица отбора и полномочия принимающего решение. После разработки проводится пробная независимая разметка; затем инструкция фиксируется и применяется к итоговому набору. Изменения после начала проверки заносятся в журнал отклонений. Финальное заключение сопровождается проверкой исходных количеств и мотивированным разбором всех существенных разногласий.

Этап	Конкретный результат
Постановка	Утверждённое задание и границы вывода
Подготовка	Паспорт объекта, данные и инструкция оценки
Испытание	Журналы, исходные ответы и расчёты
Рецензия	Проверенные находки и разрешённые замечания
Решение	Допустимый следующий шаг и условия пересмотра

Нельзя начинать с обещания подтвердить надёжность поставщика в целом. Полезный первоначальный результат может быть отрицательным: например, установлено, что система не сохраняет доказательственный статус утверждений. Такое заключение позволяет сузить применение или изменить процедуру и потому имеет самостоятельную ценность.

Реализация требует внутреннего решения о ресурсах, компетенциях и порядке использования результата. Доклад предлагает программу, но не утверждает, что ARGA уже выполнила соответствующие испытания или получила официальный статус. [36]

Минимальная система качества

Для устойчивой экспертной работы предлагается вести реестр заданий, доказательств, конфликтов интересов, находок и изменений методики. Эти записи должны быть связаны между собой. Не требуется создавать отдельную бюрократическую систему для каждого небольшого проекта, но существенный вывод должен оставлять проверяемый след.

Руководство ОЭСР по должной осмотрительности в сфере ответственного ИИ рассматривается как международный ориентир работы с неблагоприятными последствиями. Для ARGA практический смысл состоит в связи исследовательского результата с предотвращением и исправлением вреда, а не только с выпуском отчёта. Конкретная программа ниже является предложением доклада. [34]

Элемент качества	Минимальное подтверждение
Компетентность	Назначение специалистов по задачам
Независимость	Проверка связей и решение по угрозам
Прослеживаемость	Реестр материалов и процедур
Рецензия	Проверка существенных выводов другим специалистом
Исправление	Версии заключения и причины изменений
Обратная связь	Рассмотрение обоснованных возражений

Перед внешним использованием результата проверяются факты, ссылки, язык и соответствие реально выполненной работе. Внутреннее утверждение и международное согласование проводятся по действующему порядку ARGA; настоящий доклад не заменяет их. Любое заявление об аккредитации, полномочиях или признании требует отдельного документального основания. [36]

Критерием развития программы следует считать не число выданных заключений само по себе, а способность воспроизвести существенный вывод и показать его практическое последствие. Если рабочее дело этого не позволяет, увеличение объёма публикаций не исправляет методологический недостаток.

Итоговое заключение

Обоснованный аудит ИИ требует согласованности измерения, статистического вывода и решения о применении. Теория измерений определяет, что считается юридически существенным дефектом; дизайн испытания показывает, на какие задачи распространяется наблюдение; независимость процедуры позволяет сохранить неблагоприятный результат. Отсутствие любого из этих условий ограничивает итоговое мнение, даже если остальные части выполнены качественно.

Для предлагаемого пилота установлена последовательность допуска: сначала пригодность эталона и задания, затем воспроизводимость данных и расчётов, далее сравнение основного исхода и анализ новых дефектов, после этого — решение о границах использования. Положительная средняя динамика не компенсирует неустранённую критическую ошибку. Расширение на иной язык, тип досье или способ действия требует нового обоснования области применимости.

Главный риск рынка — разрыв между узким результатом и широким обещанием. Организационный сертификат, отдельный тест, аудит кадрового инструмента и правовая экспертиза имеют разные предметы. Их можно сочетать в одной программе, но нельзя представлять взаимозаменяемыми доказательствами общей безопасности или полной законности любой системы.

Практический результат для ARGA — воспроизводимый протокол ограниченного испытания, а также набор проверяемых требований к внешнему оценщику. Предлагается начать с одного назначения юридического помощника, подтвердить работоспособность разметки и повторного анализа, после чего решать вопрос о расширении. Методология допускает содержательно отрицательный итог: он показывает, какой механизм требует исправления и какое применение пока не обосновано.

Дальнейшая проверка должна установить, помогает ли этот протокол разным экспертам получать сопоставимые выводы по одному материалу и выявлять ошибки, существенные для правового решения. До получения таких данных методика остаётся аргументированным исследовательским предложением. Основанием внешнего заявления служит фактически выполненная процедура и её результат, а не название использованной модели или объём подготовленного текста.

Практический разбор 16.24

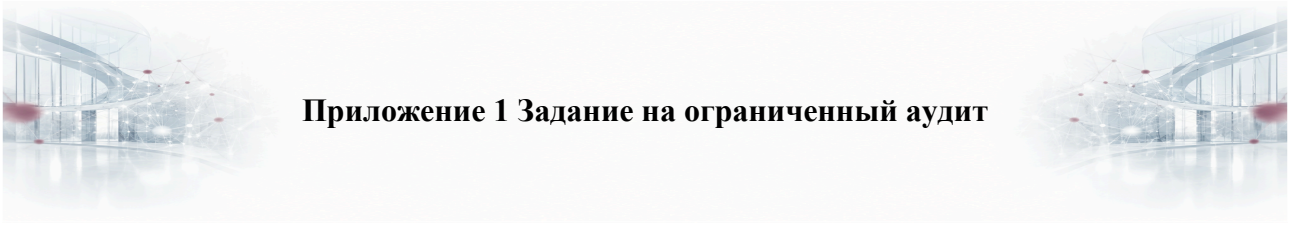
Рекомендация: Начинать с ограниченного пилота и решения по его итогам

Условный пример: Команда готова проверить десять досье, но планирует сразу объявить надёжность всех юридических функций.

Разбор: Малый пилот проверяет выполнимость протокола, а не универсальную эффективность. Без реестра ошибок и рецензии расширение преждевременно.

Практические действия: Куратор утверждает одно назначение, разметку, реестр находок и независимую проверку; перед публикацией сверяет формулировки с результатом.

Критерий результата: Выдано решение о продолжении, доработке либо остановке с основаниями и ответственными.



Приложение 1 Задание на ограниченный аудит

Заполненный пример предмета и критериев

Ниже приведён условный образец задания для системы J-1.2, которая готовит обобщения искусственных юридических досье. Это проект формы, а не запись выполненной ARGA проверки. Предполагаемый пользователь — юрист, который изучает источники и самостоятельно принимает решение о дальнейшем использовании текста.

Проверяемый набор D-16 содержит 100 искусственных досье с одним основным вопросом к каждому. Включены противоречащие документы, отрицания, неполные сведения и различающиеся даты. Набор специально сконструирован для проверки значимых ошибок; его результаты не объявляются оценкой частоты ошибок всех будущих обращений.

Поле задания	Условное заполнение
Назначение	Подготовка черновика обобщения с указанием источников
Объект	J-1.2 и зафиксированная конфигурация поиска
Данные	D-16, искусственные документы без реальных досье
Основные свойства	Достоверность, сохранение условий, разделение материалов
Исключено	Внешняя отправка и автоматическое принятие решений
Адресат	Руководитель пилотного проекта

Критерий достоверности означает, что существенное утверждение поддерживается указанным материалом и сохраняет его доказательственный статус. При недостатке сведений допустимо обозначить неопределённость. Уверенное заполнение пробела вымышленным фактом считается дефектом, даже если общий текст выглядит убедительно.

До начала работы стороны согласуют, кто размечает эталон, кто выполняет запуск и кто рецензирует вывод. Разработчик может объяснять устройство системы, но не должен единолично определять правильность каждого спорного ответа. Все изменения задания после просмотра результатов подлежат отдельной фиксации.

Доступ и условия завершения

Для рассматриваемого примера достаточно контролируемой тестовой среды, интерфейса пользователя, доступных журналов и копии конфигурации. Поставщик подтверждает состав

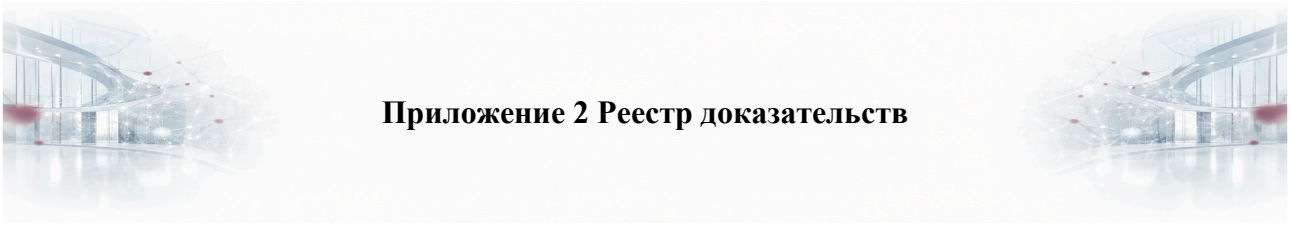
доступных функций. Если система обращается к внешним сервисам, эти обращения включаются в карту объекта до проведения тестов.

Условие	Предлагаемая запись в задании
Доступ	Именные учётные записи с ограниченными правами
Сохранение	Все исходные ответы, параметры и существенные события
Критическая находка	Остановка затронутой процедуры и уведомление ответственного
Возражения поставщика	Рассматриваются с приложением подтверждающих материалов
Рецензия	Проверка расчётов и существенных выводов другим специалистом
Итог	Заключение только о согласованном назначении и версии

Завершение не связывается с получением положительного результата. Работа считается выполненной, если проведены согласованные процедуры, рассмотрены существенные противоречия и оформлен обоснованный вывод. Если доступ оказался недостаточным, итог отражает ограничение, а не скрывает его под формальной отметкой об исполнении задания.

В договоре полезно предусмотреть отдельный этап проверки исправлений. Его объём определяется найденными дефектами. Если новая версия существенно отличается, нельзя автоматически считать дополнительный запуск продолжением прежнего аудита без пересмотра объекта и критериев.

Публичное описание результата допускается только в пределах фактического вывода. Формула «проверен ограниченный сценарий подготовки черновиков» соответствует предмету примера; формула «полностью сертифицированный юридический ИИ» не следует из выполненной программы.



Приложение 2 Реестр доказательств

Пример связанных записей

В условном рабочем деле каждой записи присваивается постоянный идентификатор. Он используется в протоколе и находках. Переименование файла не должно менять связь с доказательством. Если материал обновлён, сохраняется история, позволяющая установить, на какой версии основывался первоначальный вывод.

Код	Материал	Что подтверждает	Ограничение
E-001	Паспорт J-1.2	Согласованный объект	Основан на доступных сведениях о сервисе
E-002	Набор D-16	Состав заданий	Специально подготовленные примеры
E-003	Экспертная разметка	Критерии правильного ответа	Спорные случаи отмечены отдельно
E-004	Исходные ответы	Фактический результат запусков	Относится к установленной конфигурации
E-005	Расчёт показателей	Обработку результатов	Зависит от корректности E-003
E-006	Журнал проверки человеком	Действия проверяющего	Охватывает только наблюдавшиеся задания

В расширенной записи указываются дата получения, источник, способ выгрузки, доступ и контроль целостности. Конфиденциальные сведения о местонахождении материалов можно хранить отдельно от общего перечня. Рецензент получает доступ к необходимому содержанию по предусмотренной процедуре.

Реестр полезен только при содержательной связи с выводом. Сотня файлов без указания подтверждаемого утверждения не образует доказательственный пакет. Для каждой существенной находки аудитор должен объяснить, какие записи её поддерживают и какие материалы могли бы изменить оценку.

Если доказательство получено со слов сотрудника, это прямо указывается. Интервью не переименовывается в подтверждённый технический факт. При возможности объяснение сопоставляется с журналом, настройкой или наблюдением в среде.

Проверка полноты и изменений

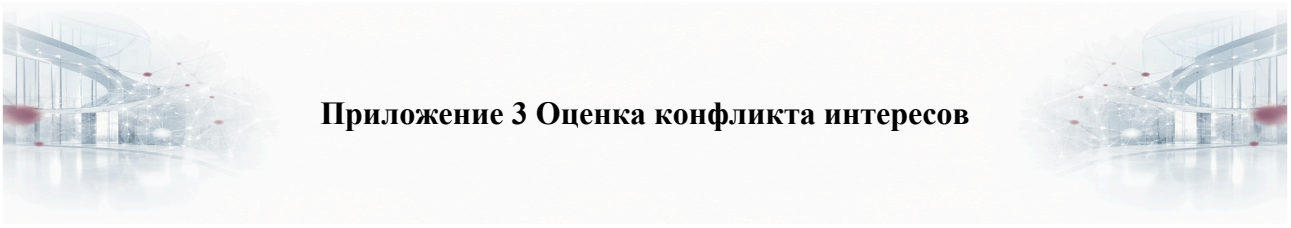
Предположим, после передачи материалов поставщик заменил паспорт конфигурации и сообщил, что прежняя версия содержала ошибочное название модели. Аудитор сохраняет оба документа, выясняет влияние изменения и проверяет, какой объект фактически использовался в запуске. Новая запись не стирает первоначальное противоречие.

Событие	Запись в рабочем деле
Получен новый паспорт	E-001-v2 с датой и источником
Выявлено расхождение	Описание изменённого поля
Запрошено объяснение	Вопрос о фактически использованной модели
Получено подтверждение	Журнал или иное достаточное свидетельство
Принято решение	Сохранить вывод, ограничить его либо повторить тест

Если объяснение не подтверждено, заключение не должно заявлять точную версию, которую аудитор не смог установить. Возможен ограниченный вывод о поведении доступного сервиса в указанную дату. Его доказательственная сила уже, чем вывод о полностью идентифицированной конфигурации.

Перед закрытием задания проверяются целостность ссылок, наличие ключевых материалов и доступность расчёта. Это не требует повторять всё исследование: задача состоит в том, чтобы другой уполномоченный специалист мог восстановить существенный путь от данных к выводу.

Срок хранения назначается на основании применимых обязанностей и согласованных целей. После его окончания действует предусмотренный порядок удаления или архивирования. Реестр не оправдывает бессрочное сохранение персональных данных, не нужных для защиты прав и подтверждения выполненной работы.



Приложение 3 Оценка конфликта интересов

Четыре условные ситуации

Предлагаемая форма используется для добровольного аудита и не заменяет специальные требования обязательных процедур. Решение принимается по реальным обстоятельствам. Наличие связи не всегда означает запрет любой работы, но может исключать независимое подтверждение определённого утверждения.

Ситуация	Существенная угроза	Предлагаемое решение
Исполнитель настраивал проверяемый алгоритм	Проверка собственной работы	Передать соответствующую оценку другой стороне

Ситуация	Существенная угроза	Предлагаемое решение
Премия зависит от положительного заключения	Финансовая заинтересованность	Изменить условие либо отказаться от задания
Рецензент связан с руководителем поставщика	Личная зависимость	Назначить другого рецензента
Заказчик оплатил фиксированный объём процедур	Сам факт оплаты не определяет вывод	Проверить остальные условия и раскрытие роли заказчика

Для первой ситуации недостаточно поменять название подразделения, если тот же специалист определяет тест и оценивает свой результат. Для второй недостаточно декларации объективности при сохранении зависимой оплаты. Защитная мера должна воздействовать на источник угрозы.

Если независимое подтверждение невозможно, стороны могут рассмотреть консультационную работу с точным описанием роли. Такой вариант не даёт права использовать слово «независимый» в публичном пересказе. Пользователь результата должен понимать, что исполнитель участвовал в создании или настройке проверяемого решения.

Решение по конфликту подписывает лицо, не заинтересованное в сохранении спорной роли. При отсутствии такого лица нужен внешний компетентный рецензент либо отказ от соответствующей части задания.

Запись решения и последующий контроль

Карточка конфликта должна быть достаточно подробной, чтобы рецензент мог понять логику. Формальное поле «конфликтов нет» допустимо только после предусмотренной проверки. Если обстоятельства неизвестны, они не должны автоматически считаться отсутствующими.

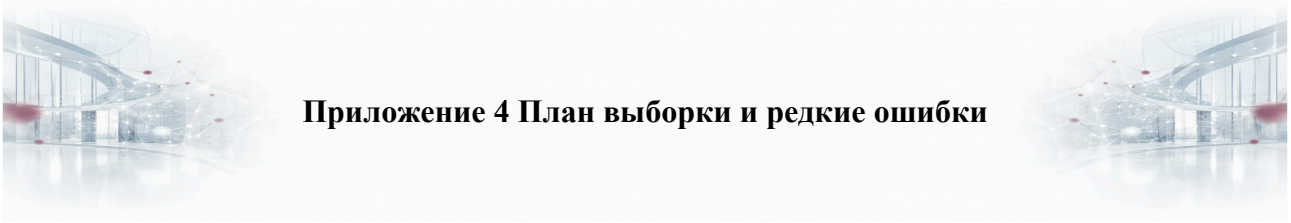
Поле	Условное заполнение
Обстоятельство	Специалист ранее разрабатывал шаблон запросов J-1.2
Затронутая задача	Оценка ошибок, связанных с этим шаблоном
Мера	Исключение специалиста из оценки и назначения критериев
Допустимая роль	Предоставление фактических пояснений с фиксацией
Дополнительная проверка	Независимый выбор сценариев и рецензия результатов

Поле	Условное заполнение
Остаточное ограничение	Общая коммерческая связь организации оценивается отдельно

Из карточки видно, что персональная замена не обязательно устраняет интерес всей организации. Поэтому проверка продолжается на уровне договора, управления и финансирования. Если эти связи делают неблагоприятный вывод практически невозможным, предложенная мера недостаточна.

Повторная оценка проводится при изменении команды, оплаты или характера помощи заказчику. Например, аудитор не должен незаметно перейти от выявления дефекта к разработке его исправления, а затем самостоятельно подтвердить успех без рассмотрения возникшего конфликта.

В итоговом документе раскрывается существенная информация в допустимом объёме. Конфиденциальные детали договора могут оставаться закрытыми, но это не должно скрывать обстоятельства, необходимые пользователю для оценки независимости заключения.



Приложение 4 План выборки и редкие ошибки

Нулевая ошибка в 300 испытаниях

Условная задача состоит в проверке вероятности конкретного отказа защитной меры. Предполагается 300 независимых испытаний с одинаковой вероятностью ошибки. Во всех попытках ошибка отсутствует. В рамках указанной биномиальной модели односторонняя верхняя 95-процентная граница равна 1 − 0,05^(1/300), то есть примерно 0,9936%.

Параметр	Значение
Число испытаний n	300
Наблюдаемые ошибки	0
Односторонний уровень	95%
Расчёт	1 − 0,05^(1/300)
Верхняя граница	Около 0,99%

Этот результат не означает, что вероятность ошибки точно равна нулю. Он также не означает, что система безопасна для любого назначения. Если ошибка может привести к тяжёлому и необратимому вреду, даже такая неопределённость может быть неприемлемой без дополнительных мер.

Для нулевых ошибок минимальное число независимых испытаний, позволяющее получить такую границу не выше 1%, определяется условием $0{,}99^n \le 0{,}05$. Отсюда n должно быть не меньше 299. Расчёт справедлив только для выбранной модели и заранее установленного плана; он не обосновывает универсальное требование к объёму аудита.

В реальном проекте дополнительно проверяются представительность сценариев, корректность регистрации ошибок и независимость попыток. Без этих условий красивый численный предел может описывать искусственную ситуацию, не связанную с предполагаемым применением.

Когда этот расчёт неприменим

Предположим, вместо 300 независимых случаев использованы 30 документов, каждый из которых отправлялся системе десять раз. Повторы полезны для исследования вариативности ответа, но ошибки внутри одного документа могут быть связаны. Применять приведённую формулу так, будто получены 300 независимых ситуаций, без дополнительного обоснования нельзя.

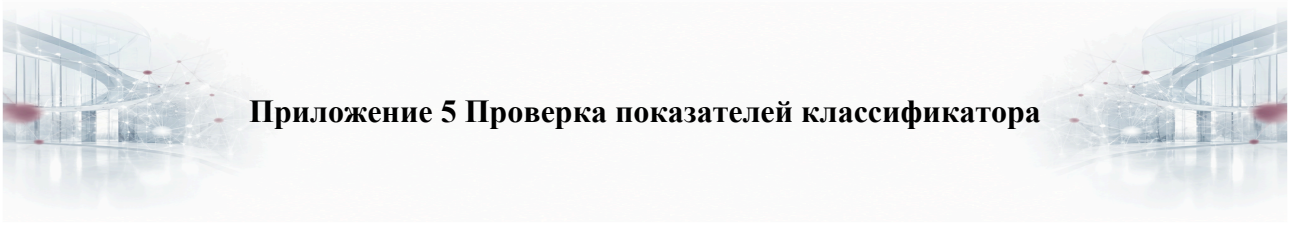
Дизайн	Что действительно исследуется
300 независимых подходящих случаев	Ошибка в определённой совокупности при заданных условиях
30 документов по 10 повторов	Вариативность на ограниченном наборе документов
300 специально сложных запросов	Устойчивость к выбранным трудным сценариям
300 заранее известных примеров	Поведение на знакомом наборе, возможна настройка под тест

Если проверка прекращалась после удачной серии, необходимо учитывать правило остановки. Необъявленный выбор наиболее благоприятного момента искажает смысл фиксированного доверительного расчёта. План остановки и порядок обработки промежуточных результатов устанавливаются до испытания.

Необходимо также проверить единицу ошибки. Если один запрос содержит десять независимых утверждений, доля ошибочных утверждений и доля ответов хотя бы с одной ошибкой различаются. Для юридического использования второй показатель может быть важнее, поскольку любой существенный дефект способен испортить весь документ.

Финальная запись должна описывать дизайн словами, а не только формулой. Пользователь должен понимать, какие ситуации охвачены и почему статистический вывод относится к

ним. Справочник NIST используется как основание выбора интервального подхода; приведённые числа рассчитаны для этого условного примера. [32]



Приложение 5 Проверка показателей классификатора

Расчёт по исходным количествам

Возвращаемся к условному набору из 10 000 случаев. Положительная категория присутствует в 100 случаях. Система выдаёт 90 верных положительных сигналов и 99 ложных, пропускает 10 положительных случаев и правильно отклоняет 9801 отрицательный. Все показатели ниже вычисляются непосредственно из этих четырёх чисел.

Показатель	Формула	Результат
Чувствительность	90 / (90 + 10)	90%
Специфичность	9801 / (9801 + 99)	99%
Доля верных среди сигналов	90 / (90 + 99)	47,62%
Общая правильность	(90 + 9801) / 10 000	98,91%
Частота ложных сигналов среди отрицательных	99 / 9900	1%

Аудитор проверяет не только арифметику, но и смысл положительной категории. Для поиска документов она может означать релевантность; для сопоставления имён — установленное эталоном тождество. Нельзя использовать одну и ту же формулу, меняя значение категории без пояснения.

Внешнему пользователю особенно полезно показать, сколько сигналов потребуют дополнительного разбора. В данном наборе ложных сигналов больше, чем верных. Это не обязательно делает инструмент бесполезным для предварительного поиска, но исключает автоматическое принятие сигнала за факт.

Расчёт не даёт интервала неопределённости и сам по себе не доказывает переносимость результата. Для реального вывода дополнительно нужны описание выборки, эталона и предполагаемой среды применения.

Изменение распространённости категории

Предположим, чувствительность остаётся 90%, а частота ложноположительных результатов среди отрицательных случаев — 1%. При снижении распространённости искомой категории доля верных сигналов уменьшается. Это свойство расчёта, а не обязательно ухудшение алгоритма.

Условная совокупность	Верные сигналы	Ложные сигналы	Доля верных
10 000 случаев и 1% положительных	90	99	47,62%
100 000 случаев и 0,1% положительных	90	999	8,26%

Во второй строке положительных случаев также 100, а отрицательных — 99 900. Один процент от последнего числа даёт 999 ложных сигналов. Поэтому число положительных ответов системы не может интерпретироваться без знания контекста, в котором они возникают.

Для ARGA это особенно важно при анализе редких совпадений. Поставщик может сообщить впечатляющую чувствительность, но пользователь сталкивается главным образом с необходимостью отвергать ложные сигналы. Задание на аудит должно учитывать фактический поток и требуемые ресурсы человеческой проверки.

Предпосылка неизменности чувствительности и частоты ложных сигналов здесь введена только для учебного расчёта. В новой среде они тоже могут измениться. Поэтому перенос показателей поставщика на другой рынок, язык или состав документов требует дополнительной проверки.

Приложение 6 Различия между группами

Наблюдаемое отношение результатов

В условном процессе группа А включает 200 заявлений, из которых 100 отобраны, группа Б — 100 заявлений, из которых отобраны 30. Доли положительного результата равны 0,50 и 0,30. Отношение второй доли к первой составляет 0,60, а абсолютная разница — 20 процентных пунктов.

Группа	Заявления	Отобраны	Доля отбора
А	200	100	50%
Б	100	30	30%

Группа	Заявления	Отобраны	Доля отбора
Сопоставление Б к А	—	—	0,30 / 0,50 = 0,60

Правильное описание: в исследованном наборе наблюдается различие результатов. Неправильное усиление: система доказанно совершает незаконную дискриминацию. Для правового вывода нужны применимые нормы, сопоставимость ситуаций, механизм решения и другие существенные обстоятельства.

Нельзя и игнорировать различие только потому, что оно ещё не квалифицировано юридически. Оно является основанием для дополнительной проверки. Аудитор выясняет, где возникает расхождение: в составе заявлений, доступности данных, правилах модели или последующих действиях человека.

Порог из одной зарубежной процедуры не переносится автоматически в другую юрисдикцию. Если организация добровольно использует определённый показатель, это фиксируется как методический выбор с обоснованием. Он не должен представляться всеобщей юридической границей допустимого поведения.

Дополнительные процедуры

Следующий этап — проверка объяснений. Например, группы могут различаться по типу документов или качеству распознавания, а система может чаще отклонять неполные записи. Это не устраняет вопрос о правах: нужно понять, оправданы ли критерии и доступны ли людям способы исправить недостаток сведений.

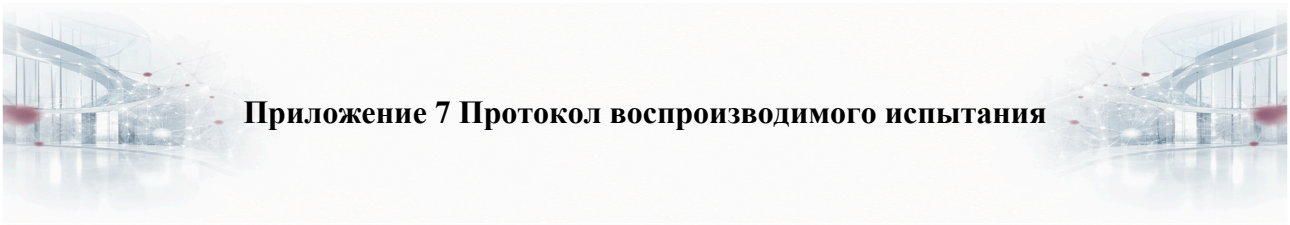
Возможная причина	Дополнительная проверка
Различное качество исходных файлов	Сопоставимые поднаборы и проверка распознавания
Пропуски необходимых сведений	Причины пропусков и возможность дополнения
Особенности модели	Ошибки при сопоставимых обстоятельствах
Действия сотрудников	Последовательность пересмотра и подтверждения
Недоступность обжалования	Проверка реального пользовательского маршрута

Предлагается проверить небольшие пересекающиеся группы, когда это законно и статистически осмысленно. Результаты с малым числом наблюдений сопровождаются соответствующей осторожностью. Нельзя скрывать такую группу лишь ради гладкой общей картины, но нельзя и делать уверенный количественный вывод по недостаточным данным.

После изменения системы сравнение повторяется на независимых данных или по обоснованному плану. Если улучшился только один показатель, исследуются возможные

новые ошибки. Коррекция должна соответствовать механизму проблемы и правовому контексту.

Итог приложения — образец направления исследования, а не готовый вывод о реальных группах людей. Все числа условны и не описывают кадровую, финансовую или иную практику ARGA.



Приложение 7 Протокол воспроизводимого испытания

Паспорт запуска

Протокол предназначен для сохранения условий, без которых повторение результата будет неоднозначным. В условном примере проверяется J-1.2 на наборе D-16. Каждый запуск получает идентификатор, который используется в реестре доказательств и при описании находки.

Поле	Условное заполнение
Идентификатор запуска	RUN-016-01
Объект	J-1.2, согласованная тестовая конфигурация
Набор	D-16, установленная версия
Инструкция	Зафиксированный текст задания и системных условий
Внешние источники	Только разрешённый набор документов
Инструменты	Поиск и чтение без внешней отправки
Оценка	Экспертная инструкция и независимый разбор спорных случаев
Результаты	Все ответы, отказы и ошибки выполнения

В рабочем протоколе дополнительно сохраняются фактическое время, доступные идентификаторы модели и параметры. Если некоторые настройки поставщик не раскрывает, поле не заполняется предположением. Указывается, какая информация недоступна и почему это ограничивает повторение.

Порядок заданий может влиять на результат, если сохраняется история взаимодействия. Поэтому необходимо описать, начинается ли каждое досье с нового контекста и какие сведения переносятся между запросами. Непредусмотренное смешение досье само может быть существенной находкой.

Протокол подписывается фактическим исполнителем в принятой организацией форме. Подпись руководителя не заменяет сведения о том, кто выполнил процедуру и какие отклонения от плана возникли.

Проверка повторного результата

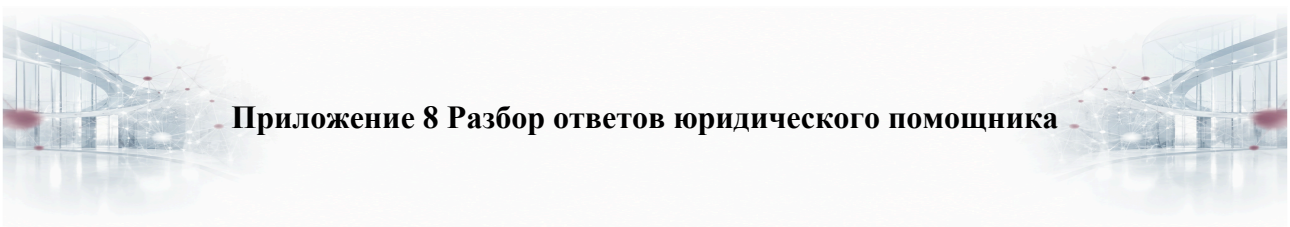
Предположим, второй специалист получил иной ответ на тот же вопрос. Сначала проверяются видимые различия условий: версия, параметры, история диалога, состав источников и состояние внешнего поиска. Только после этого можно обсуждать, свидетельствует ли расхождение о нестабильности системы.

Шаг сопоставления	Что фиксируется
Проверка входа	Совпадает ли полный материал, а не только вопрос
Проверка конфигурации	Совпадают ли доступные параметры и права
Проверка контекста	Нет ли скрытого влияния прежних сообщений
Сравнение ответа	Какие существенные утверждения различаются
Оценка последствий	Меняется ли допустимое действие пользователя

Различие формулировок не обязательно означает содержательный дефект. Два ответа могут одинаково сохранять смысл и ограничения. Напротив, небольшая замена слова «может» на «обязан» способна изменить правовой вывод. Сопоставление должно учитывать значимость, а не только буквальное совпадение.

Если повторное выполнение невозможно из-за обновления облачного сервиса, сохраняется проверка по материалам исходного запуска. Итоговая запись прямо различает повторный расчёт и повторный запуск. Нельзя объявлять поведение воспроизведённым, если проверена только арифметика итоговой таблицы.

Для последующих проектов полезно вести реестр причин невоспроизводимости. Он показывает, какие условия необходимо улучшить в договоре, журналировании или выборе поставщика, чтобы будущие заключения опирались на более устойчивую доказательственную основу.



Приложение 8 Разбор ответов юридического помощника

Шесть существенных ошибок в условном наборе

В наборе из 100 вопросов система дала 60 содержательных ответов, из которых 54 признаны правильными по принятому критерию. В шести обнаружены существенные ошибки: три

неподтверждённых вывода по источнику, два смешения дат и одна утрата отрицания. В остальных 40 случаях система обозначила недостаток сведений или отказалась отвечать.

Вид результата	Количество	Предлагаемое исследование
Правильный ответ	54	Контроль сохранения условий и источников
Неподтверждённый вывод	3	Сопоставление ссылки с реальным содержанием
Смешение дат	2	Проверка временной структуры документов
Потеря отрицания	1	Проверка распознавания и обобщения
Отказ или неопределённость	40	Анализ причин и доступного продолжения работы

Эти результаты нельзя объединить в одно заявление «качество 90%» без пояснения. Число 90% относится только к правильности среди 60 выданных ответов. Пользователю также нужно знать охват и то, что шесть ошибок могут иметь разные последствия.

Для каждого дефекта сохраняется исходный вопрос, материал, ответ и объяснение оценщика. Если разметчики расходятся, случай рассматривается отдельно. Правовая неоднозначность не должна автоматически превращаться в ошибку системы или исключаться из анализа без следа.

В реальном проекте такая таблица была бы результатом испытания. Здесь она представляет учебный пример применения методики; никакая реальная модель этими числами не оценивается.

Решение о применении и исправлении

Для условного инструмента не обоснован переход к автоматической подготовке окончательных внешних документов. Возможен ограниченный вспомогательный режим, если каждый существенный вывод проверяет компетентный человек по источнику, а процедура действительно предотвращает перенос ошибки. Само обещание проверки человеком без подтверждённого процесса недостаточно.

Дефект	Предлагаемая мера	Повторная проверка
Неподтверждённые выводы	Усилить связь утверждения с источником	Новые примеры с похожим, но нерелевантным текстом
Смешение дат	Отдельно представлять периоды действия фактов	Досье с изменением обстоятельств во времени

Дефект	Предлагаемая мера	Повторная проверка
Потеря отрицания	Проверить обработку и смысловую оценку	Новые документы с отрицаниями и оговорками
Избыточные отказы	Уточнить назначение и маршрут продолжения	Сопоставимый новый набор задач

После исправления необходимо повторить исходные проблемные случаи и проверить новые. Первый набор показывает устранение известных симптомов, второй — перенос улучшения. Если поставщик демонстрирует только заранее исправленные ответы, общий вывод об эффективности изменения остаётся ограниченным.

Решение о допустимом режиме принимает уполномоченное лицо с учётом рисков и применимых обязанностей. Аудитор предоставляет доказательства и пределы вывода. Положительное решение не должно превращать условный пилотный результат в обещание безошибочной юридической работы.

Если человеческая проверка не выявляет убедительно сформулированные ошибки, требуется изменить процесс или ограничить использование. Дополнительная инструкция сама по себе не доказывает, что проблема устранена.

Приложение 9 Образец ограниченного заключения

Основная часть заключения

Приведённый текст является образцом для условного проекта. Он не удостоверяет реальную проверку и не предназначен для использования как готовое подтверждение соответствия конкретного поставщика. Перед применением форма заполняется фактическими сведениями и проверяется по установленной процедуре.

Предмет проверки: система J-1.2 в конфигурации подготовки черновых обобщений искусственных юридических досье. Критерии: связь существенных утверждений с источниками, сохранение временных условий и отрицаний, обозначение недостатка сведений. Данные: специально подготовленный набор D-16 из 100 заданий. Из области исключены реальные клиентские досье и автоматическая внешняя отправка.

Выполненные в условном примере процедуры включают фиксацию доступной конфигурации, запуск заданий, экспертную оценку ответов и повторный расчёт показателей. Из 60 содержательных ответов 54 отвечают принятому критерию, шесть содержат существенные ошибки. Для 40 заданий получен отказ либо указание неопределённости. Количества не являются оценкой всей будущей эксплуатации.

Вывод	Основание и предел
Автоматическое внешнее использование не обосновано	Обнаружены существенные смысловые дефекты
Вспомогательный режим требует дополнительных условий	Необходима эффективная проверка каждого значимого утверждения
Нужна проверка исправлений	Изменения должны пройти исходные и новые сценарии

Пользователь результата должен видеть и удачные ответы, и ограничения. Формулировка заключения не превращает отсутствие исследованных данных в положительное подтверждение. Любое расширение назначения требует отдельной оценки.

Ограничения и использование документа

Образец относится к указанной версии и специально подготовленным данным. В нём не исследованы все языки, все виды юридических задач и внутренняя архитектура облачного поставщика. Не оценивались сертификационная схема, статус нотифицированного органа или соответствие полным текстам платных стандартов ISO.

В реальном заключении ограничения должны быть конкретизированы по фактическим обстоятельствам. Если доступ к журналам предоставлен, нельзя оставлять шаблонную оговорку об их отсутствии. Если часть данных была исключена, необходимо назвать причины и влияние на результат. Шаблон помогает структуре, но не заменяет добросовестное описание работы.

Элемент завершения	Что должно быть заполнено
Исполнители	Фактические роли и необходимые компетенции
Независимость	Существенные связи и принятые меры
Рецензия	Кто проверил существенные выводы
Версия заключения	Дата, изменения и действующая редакция
Использование	Адресаты и допустимый объём цитирования
Пересмотр	Изменения, при которых вывод требует обновления

Если после выдачи обнаружена существенная ошибка самого отчёта, выпускается исправленная версия с указанием затронутых выводов. Нельзя молча заменить документ так, чтобы получатели прежней редакции не узнали об изменении, важном для их решения.

Образец не является юридическим отказом от ответственности за небрежную работу. Его назначение — точно обозначить выполненное исследование, доказательства и пределы использования результата. Содержание должно соответствовать реальным действиям исполнителей.



Приложение 10 Пакет предложений для ARGA

Минимальные требования к качеству внешнего аудита

Для обсуждения с регуляторами и профессиональным сообществом предлагается набор минимальных требований к раскрытию результата. Он не задаёт единую международную сертификацию и не подменяет специальные правовые процедуры. Цель — сделать заключения разных исполнителей понятнее и уменьшить риск необоснованных заявлений о надёжности.

Предлагаемое требование	Доказательственная польза
Назвать объект, версию и назначение	Ограничить перенос результата
Указать критерии и их статус	Отделить закон от добровольной методики
Описать доступ и выборку	Показать фактическую основу проверки
Раскрыть существенные связи	Позволить оценить независимость
Сохранить протокол и расчёты	Обеспечить проверяемость вывода
Сообщить существенные ограничения	Предотвратить вводящий в заблуждение пересказ
Определить пересмотр	Учесть изменения системы после аудита

Пропорциональность достигается объёмом процедур и детализацией отчётности, но не сокрытием предмета. Небольшая документальная проверка может иметь короткий отчёт, если читатель видит, что технические испытания не проводились. Более широкое утверждение требует более сильных доказательств.

Для ARGA этот набор может стать основой внутреннего стандарта подготовки экспертных материалов. Перед принятием его необходимо адаптировать к ресурсам, компетенции и видам проектов. Настоящий доклад предлагает содержание такого решения, но не утверждает, что оно уже принято организацией.

Подготовка адресной консультационной позиции

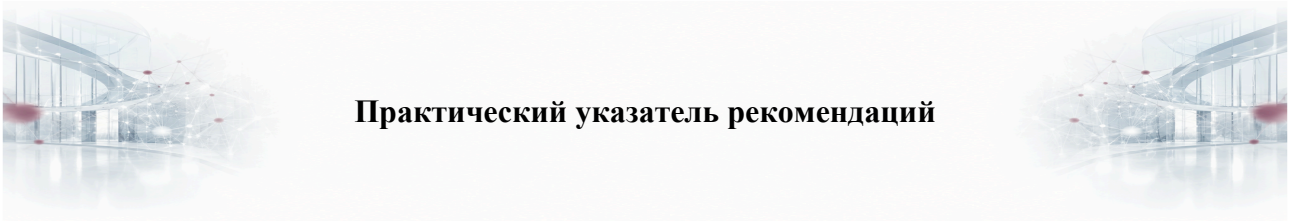
Для каждой внешней процедуры выбирается одна конкретная проблема и предложение, связанное с её предметом. Например, для обсуждения анонимизации — воспроизводимость оценки с учётом получателя данных; для сбора данных — прослеживаемость происхождения; для агентного ИИ — проверка полномочий и действий инструментов. Один общий текст не заменяет такую адаптацию.

Часть позиции	Предлагаемое содержание
Вопрос	Какой пробел создаёт практическую неопределённость
Предложение	Конкретное требование к доказательству или процедуре
Обоснование	Как это помогает проверке и защите прав
Пропорциональность	Как выполнить требование без избыточного раскрытия
Пример	Искусственный сценарий с ясным пределом вывода
Подтверждение подачи	Отдельная запись после фактического отправления

Сроки, подтверждённые на дату доклада: процедуры EDPB 02/2026 и 03/2026 — 30 октября 2026 года в 23:59 CET; консультация Европейской комиссии по технологиям и авторскому праву — 3 ноября; сбор материалов ICO об агентном ИИ — 20 ноября. Перед направлением условия проверяются повторно. [26–29]

Позиция проходит внутреннее утверждение по действующему порядку ARGA. Для инициатив, затрагивающих глобальное управление или институциональное сотрудничество, учитывается предусмотренное международное согласование. Данные о фактической отправке и получении сохраняются отдельно от самого исследования. [36]

Итоговая практическая последовательность: выбрать адресата, подготовить короткую позицию, проверить источники и условия, получить необходимое внутреннее утверждение и только затем подать материал. Наличие настоящего доклада означает готовность аналитической основы, а не состоявшееся участие в консультации.



Практический указатель рекомендаций

Выберите задачу по рекомендации и перейдите к разбору с соответствующим кодом. Код содержит номер доклада и номер основного раздела. В каждом разборе даны пример, его анализ, действия и критерий результата.

16.01 Фиксировать узкое проверяемое утверждение
Раздел 1. Результат: Заключение относится к сохранению событий и источников в исследованном наборе.

16.02 Составлять карту применимых требований
Раздел 2. Результат: Каждое требование связано с конкретным применением и источником.

16.03 Проверять полномочия внешнего оценщика
Раздел 3. Результат: В договоре указан допустимый вид заключения без расширения статуса исполнителя.

16.04 Разделять юрисдикции и назначения проверки
Раздел 4. Результат: Для каждого перенесённого вывода указано основание либо ограничение.

16.05 Выбирать критерий до испытания
Раздел 5. Результат: Итог содержит все заранее выбранные показатели, включая неблагоприятные.

16.06 Устранять конфликт интересов до назначения
Раздел 6. Результат: Автор механизма не утверждает вывод о его эффективности.

16.07 Назначать компетенции по рискам задания
Раздел 7. Результат: У каждой существенной части заключения есть компетентный проверяющий.

16.08 Закреплять разрешённый доступ и остановку
Раздел 8. Результат: Журнал подтверждает отсутствие неразрешённых действий и соблюдение границ доступа.

16.09 Описывать всю проверяемую конфигурацию
Раздел 9. Результат: Паспорт запуска позволяет отличить изменение модели от изменения поиска.

16.10 Сохранять оригинал и цепочку преобразований
Раздел 10. Результат: Любую спорную дату можно проследить до исходной страницы.

16.11 Проверять эталон на неоднозначных фактах
Раздел 11. Результат: Сохранены исходные метки, причина расхождения и версия инструкции.

16.12 Разделять выборочную и целевую проверку
Раздел 12. Результат: Доля ошибок случайной выборки не смешана с долей в стресс-наборе.

16.13 Показывать исходные количества и новые дефекты
Раздел 13. Результат: Решение учитывает новые существенные ошибки, а не только среднюю разницу.

16.14 Проверять различия групп на сопоставимых задачах
Раздел 14. Результат: В отчёте показаны состав групп и устойчивость различия после сопоставления.

16.15 Проверять возможность изменить результат
Раздел 15. Результат: Сотрудник мотивированно отменяет ошибочный флаг; решение отражено в журнале.

16.16 Проверять факты и полноту независимо
Раздел 16. Результат: Ответ сохраняет исключение и указывает соответствующий источник.

16.17 Подтверждать исправление повторным испытанием
Раздел 17. Результат: Дефект не воспроизводится в проверенном наборе, штатная задача выполняется.

16.18 Минимизировать данные без разрушения теста
Раздел 18. Результат: Тест сохраняет однофамильцев и различающиеся даты без раскрытия реальных лиц.

16.19 Передавать воспроизводимый пакет запуска
Раздел 19. Результат: Независимый пересчёт совпадает, расхождения объяснены в журнале.

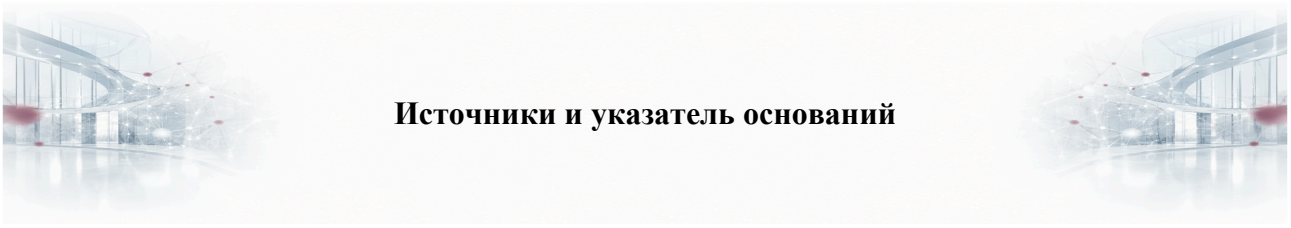
16.20 Связывать находку с исполнимой мерой
Раздел 20. Результат: У находки есть владелец, срок и подтверждающий повторный результат.

16.21 Задавать триггеры повторной оценки
Раздел 21. Результат: Новое действие разрешено отдельным решением на основании отдельного теста.

16.22 Сохранять доказательственный статус сведений
Раздел 22. Результат: В проекте обращения предположение не выдано за подтверждённое событие.

16.23 Формулировать адресное предложение регулятору
Раздел 23. Результат: Каждое предложение называет требуемую запись, её назначение и ограничение раскрытия.

16.24 Начинать с ограниченного пилота и решения по его итогам
Раздел 24. Результат: Выдано решение о продолжении, доработке либо остановке с основаниями и ответственными.



Источники и указатель оснований

Дата доступа — 20 октября 2026 года. Ссылки ведут к официальным публикациям; позиция 36 является внутренним документом. Русские названия передают предмет источника.

[1] Европейская комиссия. Регламент ЕС об ИИ, статья 113
Сроки применения. Портал отражает консолидацию на 27 июля 2026 года; специальные сроки нельзя переносить на все обязанности. Официальный источник

[2] Европейская комиссия. Регламент ЕС об ИИ, статья 31
Независимость и организационные требования к нотифицированным органам. Этот статус отличается от обычной договорной экспертной работы. Официальный источник

[3] Европейская комиссия. Регламент ЕС об ИИ, статья 43
Различные маршруты оценки соответствия. Для дела проверяются категория системы, секторное законодательство и применимые даты. Официальный источник

[4] Европейская комиссия. Регламент ЕС об ИИ, приложение VII
Оценка системы управления качеством и технической документации. Не рассматривается как универсальная процедура для любого внешнего аудитора. Официальный источник

[5] Европейский парламент и Совет ЕС. Регламент 2016/679 о защите данных
Официальная публикация EUR-Lex. Принципы обработки, роли, безопасность, оценка воздействия и передачи данных проверяются по конкретным обстоятельствам. Официальный источник

[6] Европейская комиссия. Вступление в силу правил независимого аудита по DSA
Официальное сообщение от 22 февраля 2024 года. Ежегодный независимый аудит особо крупных платформ и поисковых систем. Официальный источник

[7] Европейская комиссия. Делегированный регламент о независимых аудитах по DSA
Официальная публикация о подготовке аудиторских отчётов и отчётов об исполнении рекомендаций в специальном режиме цифровых услуг. Официальный источник

[8] Нью-Йорк, DCWP. Инструменты автоматизированных кадровых решений
Официальное описание Local Law 144: предмет, аудит на предвзятость, публикация сведений и уведомления. Официальный источник

[9] Нью-Йорк, DCWP. Окончательные правила об автоматизированных кадровых решениях
Разделы 5-300–5-304. Независимость аудитора, данные, показатели и публикация результатов в установленной сфере. Официальный источник

[10] Правительство Великобритании. Введение в подтверждение надёжности ИИ
Объяснение AI assurance и различий между видами оценки. Методический материал, не универсальная обязательная сертификация. Официальный источник

[11] DSIT. Дорожная карта независимой внешней оценки ИИ
Опубликована 3 сентября 2025 года. Политика развития качества услуг, компетенций и рынка внешней оценки. Официальный источник

[12] GAO. Рамка подотчётности искусственного интеллекта
GAO-21-519SP, 30 июня 2021 года. Управление, данные, результативность и мониторинг как направления проверочных вопросов. Официальный источник

[13] Федеральные банковские регуляторы США. Пересмотренное руководство по модельному риску
17 апреля 2026 года. Банковская сфера; генеративные и агентные модели прямо исключены из области руководства. Официальный источник

[14] NIST. Рамка управления рисками ИИ 1.0
NIST AI 100-1, 2023 год. Добровольная рамка; не задаёт универсальный приемлемый уровень риска или общий сертификат. Официальный источник

[15] NIST. Профиль рисков генеративного ИИ
NIST AI 600-1, 2024 год. Риски ошибочного содержания и проверка источников при оценке генеративных систем. Официальный источник

[16] ISO. ISO 19011:2026
Открытая карточка четвёртой редакции, опубликованной в мае 2026 года. Руководство по аудиту систем менеджмента; полный платный текст не исследовался. Официальный источник

[17] ISO. ISO/IEC 42001:2023
Открытое описание системы управления ИИ организации. Вывод о конкретном продукте не следует автоматически из сертификата такой системы. Официальный источник

[18] ISO. ISO/IEC 42006:2025
Открытое описание требований к органам аудита и сертификации систем управления ИИ; дополнение к ISO/IEC 17021-1. Официальный источник

[19] ISO. ISO/IEC 17029:2019
Открытое описание принципов и требований к органам валидации и верификации. Не используется для реконструкции полного текста стандарта. Официальный источник

[20] ISO. ISO/IEC 17025:2017
Открытое описание компетентности испытательных и калибровочных лабораторий. Для применения проверяется конкретная область компетенции. Официальный источник

[21] ISO. ISO/IEC 23894:2023
Открытая карточка руководства по управлению рисками ИИ. Полный перечень требований по карточке не восстанавливается. Официальный источник

[22] ISO. ISO/IEC 42005:2025
Открытое описание оценки воздействия систем ИИ. Использовано для разграничения предметов оценки. Официальный источник

[23] NIST. Подход к выявлению и управлению предвзятостью ИИ
NIST SP 1270, 2022 год. Человеческие, организационные и технические источники предвзятости. Официальный источник

[24] NIST. Таксономия атак на машинное обучение и мер противодействия
NIST AI 100-2e2025. Терминология и направления анализа угроз; конкретные испытания требуют разрешённой среды и отдельного задания. Официальный источник

[25] NIST. Практики безопасной разработки генеративного ИИ
NIST SP 800-218A, окончательная версия 2024 года. Профиль безопасной разработки, не самостоятельное подтверждение безопасности продукта. Официальный источник

[26] ICO. Сбор материалов об агентном ИИ
8 октября — 20 ноября 2026 года. Темы ответственности, прозрачности и защиты данных; условия ответа и публикации. Официальный источник

[27] EDPB. Консультация по проекту руководства 03/2026
Сбор веб-данных для генеративного ИИ. Официальный срок: 30 октября 2026 года, 23:59 CET. Официальный источник

[28] EDPB. Консультация по проекту руководства 02/2026
Анонимизация. Официальный срок: 30 октября 2026 года, 23:59 CET; процедура отличается от обсуждения веб-сбора. Официальный источник

[29] Европейская комиссия. Консультация о технологиях и авторском праве
Публикация 29 сентября 2026 года. Приём до 3 ноября; среди участников названы неправительственные организации. Официальный источник

[30] EDPB. Карточка проекта руководства 03/2026
Официальное подтверждение предмета и консультационного статуса документа. Предложения доклада не выдаются за окончательные требования проекта. Официальный источник

[31] EDPB. Проект руководства 02/2026 об анонимизации
Версия 1.0 от 7 июля 2026 года для публичного обсуждения. Контекст участника и разумно вероятные средства идентификации. Официальный источник

[32] NIST/SEMATECH. Доверительные интервалы для долей
Статистический справочник, раздел 7.2.4.1. Приведённые в докладе числовые примеры рассчитаны отдельно и являются условными. Официальный источник

[33] AI Security Institute. Документация Inspect по журналам испытаний
Официальное описание сохранения результатов оценивания. Программный инструмент не гарантирует корректность критерия или независимость оценки. Официальный источник

[34] ОЭСР. Руководство по должной осмотрительности для ответственного ИИ
Публикация 2026 года. Международный ориентир работы с неблагоприятными последствиями, используемый в пределах предмета исследования. Официальный источник

[35] Суд ЕС. Dun & Bradstreet Austria, C-203/22
Решение от 27 февраля 2025 года, ECLI:EU:C:2025:117. Содержательная информация о логике обработки в контексте права на доступ. Официальный источник

[36] ARGA Observatory. План аналитической и международной институциональной работы
Внутренний план от 8 октября 2026 года. Тема №16, критерии доклада и порядок согласования внешних позиций; не внешний нормативный источник.

Научные и методологические источники

[M1] Chouldechova A. et al. A Shared Standard for Valid Measurement of Generative AI Systems' Capabilities, Risks, and Impacts. 2024. Научный препринт. https://arxiv.org/abs/2412.01934

[M2] Raji I. D. et al. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. FAT* 2020, pp. 33–44. https://doi.org/10.1145/3351095.3372873

[M3] Keller A. et al. Expanding the AI Evaluation Toolbox with Statistical Models. NIST AI 800-3. 2026. https://doi.org/10.6028/NIST.AI.800-3

[M4] McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12, 1947, pp. 153–157. https://doi.org/10.1007/BF02295996